IJCSI

International Journal of Computer Science Issues

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

© IJCSI PUBLICATION
www.IJCSI.org

IJCSI proceedings are currently indexed by:

© IJCSI PUBLICATION 2012
www.IJCSI.org

IJCSI Publicity Board 2012

Dr. Borislav D Dimitrov
Department of General Practice, Royal College of Surgeons in Ireland
Dublin, Ireland

Dr. Vishal Goyal
Department of Computer Science, Punjabi University
Patiala, India

Mr. Nehinbe Joshua
University of Essex
Colchester, Essex, UK

Mr. Vassilis Papataxiarhis
Department of Informatics and Telecommunications
National and Kapodistrian University of Athens, Athens, Greece

IJCSI Editorial Board 2012

Dr Tristan Vanrullen
Chief Editor
LPL, Laboratoire Parole et Langage - CNRS - Aix en Provence, France
LABRI, Laboratoire Bordelais de Recherche en Informatique - INRIA - Bordeaux, France
LEEE, Laboratoire d'Esthétique et Expérimentations de l'Espace - Université d'Auvergne, France

Dr Constantino Malagôn
Associate Professor
Nebrija University
Spain

Dr Lamia Fourati Chaari
Associate Professor
Multimedia and Informatics Higher Institute in SFAX
Tunisia

Dr Mokhtar Beldjehem
Professor
Sainte-Anne University
Halifax, NS, Canada

Dr Pascal Chatonnay
Assistant Professor
MaÎtre de Conférences
Laboratoire d'Informatique de l'Université de Franche-Comté
Université de Franche-Comté
France

Dr Karim Mohammed Rezaul
Centre for Applied Internet Research (CAIR)
Glyndwr University
Wrexham, United Kingdom

Dr Yee-Ming Chen
Professor
Department of Industrial Engineering and Management
Yuan Ze University
Taiwan

Dr Gitesh K. Raikundalia
School of Engineering and Science,
Victoria University
Melbourne, Australia

Dr Vishal Goyal
Assistant Professor
Department of Computer Science
Punjabi University
Patiala, India

Dr Dalbir Singh
Faculty of Information Science And Technology
National University of Malaysia
Malaysia

Dr Natarajan Meghanathan
Assistant Professor
REU Program Director
Department of Computer Science
Jackson State University
Jackson, USA

Dr Deepak Laxmi Narasimha
Department of Software Engineering,
Faculty of Computer Science and Information Technology,
University of Malaya,
Kuala Lumpur, Malaysia

Dr. Prabhat K. Mahanti
Professor
Computer Science Department,
University of New Brunswick
Saint John, N.B., E2L 4L5, Canada

Dr Navneet Agrawal
Assistant Professor
Department of ECE,
College of Technology & Engineering,
MPUAT, Udaipur 313001 Rajasthan, India

Dr Panagiotis Michailidis
Division of Computer Science and Mathematics,
University of Western Macedonia,
53100 Florina, Greece

Dr T. V. Prasad
Professor
Department of Computer Science and Engineering,
Lingaya's University
Faridabad, Haryana, India

Dr Saqib Rasool Chaudhry
Wireless Networks and Communication Centre
261 Michael Sterling Building
Brunel University West London, UK, UB8 3PH

Dr Shishir Kumar
Department of Computer Science and Engineering,
Jaypee University of Engineering & Technology
Raghogarh, MP, India

Dr P. K. Suri
Professor
Department of Computer Science & Applications,
Kurukshetra University,
Kurukshetra, India

Dr Paramjeet Singh
Associate Professor
GZS College of Engineering & Technology,
India

Dr Shaveta Rani
Associate Professor
GZS College of Engineering & Technology,
India

Dr. Seema Verma
Associate Professor,
Department Of Electronics,
Banasthali University,
Rajasthan - 304022, India

Dr G. Ganesan
Professor
Department of Mathematics,
Adikavi Nannaya University,
Rajahmundry, A.P, India

Dr A. V. Senthil Kumar
Department of MCA,
Hindusthan College of Arts and Science,
Coimbatore, Tamilnadu, India

Dr Mashiur Rahman
Department of Life and Coordination-Complex Molecular Science,
Institute For Molecular Science, National Institute of Natural Sciences,
Miyodaiji, Okazaki, Japan

Dr Jyoteesh Malhotra
ECE Department,
Guru Nanak Dev University,
Jalandhar, Punjab, India

Dr R. Ponnusamy
Professor
Department of Computer Science & Engineering,
Aarupadai Veedu Institute of Technology,
Vinayaga Missions University, Chennai, Tamilnadu, India

Dr Nittaya Kerdprasop
Associate Professor
School of Computer Engineering,
Suranaree University of Technology, Thailand

Dr Manish Kumar Jindal
Department of Computer Science and Applications,
Panjab University Regional Centre, Muktsar, Punjab, India

Dr Deepak Garg
Computer Science and Engineering Department,
Thapar University, India

Dr P. V. S. Srinivas
Professor
Department of Computer Science and Engineering,
Geethanjali College of Engineering and Technology
Hyderabad, Andhra Pradesh, India

Dr Sara Moein
CMSSP Lab, Block A, 2nd Floor, Faculty of Engineering,
MultiMedia University, Malaysia

Dr Rajender Singh Chhillar
Professor
Department of Computer Science & Applications,
M. D. University, Haryana, India

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 1, January 2012 (IJCSI Vol. 9, Issue 1, No 1). The acceptance rate for this issue is 30.1%.

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

1. Development Model for Knowledge Management System (KMS) to Improve University Performance(Case Studies in Indonesia University of Education) — 1-6
Munir and Dedi Rohendi

2. A Clustering Method of Highly Dimensional Patent Data Using Bayesian Approach — 7-11
Sunghae Jun

3. A hybrid algorithm of Harmony Search and Bees Algorithm for a University Course Timetabling Problem — 12-17
Khang Nguyen, Phuc Nguyen and Nuong Tran

4. Applying Genetic Algorithms for Inventory Lot-Sizing Problem with Supplier Selection under Storage Capacity Constraints — 18-23
Chirawat Woarawichai, Kritsana Kuruvit and Paitoon Vashirawongpinyo

5. An Agile Method for E-Service Composition — 24-29
Pouya Fatehi and Seyyed Mohsen Hashemi

6. A Survey on Steganography Techniques in Real Time Audio Signals and Evaluation — 30-37
Abdulaleem Z. Al-Othmani, Azizah Abdul Manaf and Akram M. Zeki

7. The Adaptability of Conventional Data Mining Algorithms through Intelligent Mobile Agents in Modern Distributed Systems — 38-46
Dost Muhammad Khan and Nawaz Mohamudally

8. Principal Objects Detection Using Graph-Based Segmentation and Normalized Histogram — 47-49
Pham The Bao and Bui Ngoc Nam

9. Correlated Node Behavior Model based on Semi Markov Process for MANETS — 50-59
A.H Azni, Rabiah Ahmad, Zul Azri Muhamad Noh, Abd Samad Hasan Basari and Burairah Hussin

10. An Application of Genetic Algorithm with Iterative Chromosomes for Image Clustering Problems — 60-67
R. Balakrishnan and U. Karthick Kumar

11. Automatic System to Remove Unpleasant Images Detected by Pupil-Size Changes — 68-73
Koji Kashihara

12. A Discussion on Elliptic Curve Cryptography and Its Applications — 74-77
Sharad Kumar Verma and D. B. Ojha

13. A Hand Finding and Segmenting in a Partially Cluttered Scene — 78-83
R. M. Farouk, Gamal F. Elhadi and Abdalhakeem. T. Issa

14. A Novel Seven Input Majority Gate in Quantum-dot Cellular Automata — 84-89
Keivan Navi, Amir Mokhtar Chabi and Samira Sayedsalehi

15. A Priority Based Bandwidth Reservation Protocol for Wireless Mesh Networks — 90-95
K. Valarmathi and N. Malmurugan

16. Analysis of the Similarities and Differences between MPLS Label Distribution Protocols RSVP and CR-LDP — 96-103
Muhammad Asif, Zahid Farid, Muhammad Lal and Junaid Qayyum

17. Annotating Speech Corpus for Prosody Modeling in Indian Language Text to Speech Systems 104-108
Kiruthiga S and Krishnamoorthy K

18. A Novel Algorithm for Manets using Ant Colony 109-113
Javad Pashaei Barbin, Majid Taghipoor and Vahid Hosseini

19. An efficient algorithm for generating AoA networks 114-120
Nasser Eddine Mouhoub and Abdelhamid Benhocine

20. Modify the uCS-51 Architecture to SIMD, VLIW and Superscalar uC 121-128
Abdelmoneim M. Fouda, Assem Badr and Abdelsamie Kodb

21. A Strategic Framework for Designing E-Learning System with Focus on University Entrepreneurship 129-138
Masoud Asgarimehr, Babak Shirazi, Meisam Jafari Eskandari and Salman Rostami

22. EQNM2L: Towards a Metamodel for Extended Queuing Networks 139-149
Abdelhabib Bourouis and Brahim Belattar

23. Evaluation of coordination contracts for a two stage Supply Chain under price dependent demand 151-159
Noureddine Labiad, Zitouni Beidouri and Othmane Bouksour

24. Steady State and Transient Study of The Electron Transport In N-Type GaSb Using Three Valley Monte Carlo Model 160-164
Belhadji Youcef and Bouazza Benyounes

25. Comparison of Phase Only Correlation and Neural Network for Iris Recognition 165-171
Yogendra Kumar Jain and Manoj Kumar Verma

26. A modified Kernelized Fuzzy C-Means (KFCM) algorithm for noisy images segmentation: Application to MRI images 172-176
Abdenour Mekhmoukh, Karim Mokrani and Mohamed Cheriet

27. Gossip Based Routing Protocol Design for Ad Hoc Networks 177-181
Toqeer Mahmood, Tabbassam Nawaz, Rehan Ashraf and Syed M. Adnan Shah

28. Design and Implementation of Mobile IPv6 Data Communication in Dual Networks 182-190
Chaman Singh, Sanjeev Kumar, Sanjeev Kumar and K.L.Bansal

29. A Data Transfer in Wireless Sensor Networks Using AODV Protocol 191-200
B.Manimekala and M.Kayalvizhi

30. An Advanced Survey on Cloud Computing and State-of-the-art Research Issues 201-207
Mohiuddin Ahmed, Abu Sina Md. Raju Chowdhury, Mustaq Ahmed and Md. Mahmudul Hasan Rafee

31. An Empirical Study of Underwater Acoustic Sensor Network and Terrestrial Wireless Network 209-213
Mohd. Ehmer Khan and Farmeena Khan

32. Application of Decision Tree in Analysis of Intra College Festival Data set 214-218
Sumit Srivastava, Megha Yadav and Sandeep Devra

33. String Matching and its Applications in Diversified Fields 219-226
Vidya Saikrishna, Akhtar Rasool and Nilay Khare

34. A study on the behavior of a neural network for grouping the data 228-234
Suneetha Chittineni and Raveendra Babu Bhogapathi

35. Performance Evaluation of Bluetooth andZigbee Using Monte Carlo Simulation 235-237
Shoaib Ur Rehman, Obaid Ur Rehman, Zeshan Akbar and Javed Iqbal

36. Cloud Load Balancing Techniques: A Step Towards Green Computing 238-246
Nidhi Jain Kansal and Inderveer Chana

37. Wind Turbine Micrositing: Comparison of Finite Difference Method and Computational Fluid Dynamics 247-252
Samina Rajper and Imran Amin

38. Comparative Study between Various Aspects of the Academic and Nonacademic Institutions Websites 253-265
Khalid Samy Mohamed Husain

39. Image Processing and Analysis for DTMRI 266-272
Kondapalli Srinivasa Vara Prasad, Manike Chiranjeevi, Kundan Kumar Mishra and Kdbs Tanuja

40. Sensitivity Analysis of EADARP Multicast Protocol 273-287
Dina Darwish

41. Fault Tolerance-Challenges, Techniques and Implementation in Cloud Computing 288-293
Anju Bala and Inderveer Chana

42. A Study of Edge Detection Method for different Climate Condition Images using Digital Image Processing 294-298
P. Balakrishnan, S. Muruganand and K. Sriram

43. Survey of Tolerant Network in Mobile Communication 299-309
M.Chithik Raja, N. Mohame Yasin and M.S.Durai Rajan

44. An analytical model for Energy Consumption in Y-MAC Protocol 310-316
Ramchand V and D. K. Lobiyal

45. Performance Based Novel Techniques for Semantic Web Mining 317-326
Mahendra Thakur and Geetika S. Pandey

46. Optimized Energy-Aware Algorithm for Collaboration in Visual Sensor Networks 327-335
Harish H.Kenchannavar and Umakant P.Kulkarni

47. Comment on-Quantum Multiplexer Designing and Optimization applying Genetic Algorithm 336-338
Subhasis Pal, Arijit Roy and Dibyendu Chatterjee

48. Software Size Estimation Using Fuzzy Backpropagation Network Method 339-348
B V A N S S Prabhakar Rao and P Seetha Ramaiah

49. Improving Web Performance by a Differencing and Merging System 349-355
Aleksandar Jevremovic, Ranko Popovic, Dejan Zivkovic, Mladen Veinovic and Goran Shimic

50. Verdict Accuracy of Quick Reduct Algorithm using Clustering and Classification Techniques for Gene Expression Data 357-363
T.Chandrasekhar, K.Thangavel and E.N.Sathishkumar

51. A Quantitative Study and Comparison of AODV, OLSR and TORA Routing Protocols in MANET **364-369**
Tamilarasan Santhamurthy

52. Proposal of a New Block Cipher reasonably Non-Vulnerable against Cryptanalytic Attacks **370-373**
Abhijit Chowdhury, Angshu Kumar Sinha and Saurabh Dutta

53. Hybridizing PSM and RSM Operator for Solving NP-Complete Problems: Application to Traveling Salesman **374-378**
Chakir Tajani, Otman Abdoun and Jaafar Abouchabaka

54. Toward an example-based machine translation from written text to ASL using virtual agent animation **379-388**
Mehrez Boulares and Mohamed Jemni

55. Mercator as web crawler **389-395**
Priyanka Saxena

56. Noncooperative Iris Segmentation **396-401**
Elsayed Mostafa, Maher Mansour and Heba Saad

57. IPTV-Internet Protocol Televisions Key Broadcast Building Blocks **402-410**
Muhammad Asif, Zahid Farid, Muhammad Lal, Junaid Qayyum and Fayyaz Gul

58. A Review of QoS in Mobility Phillic Communication **411-421**
Sumit Joshi and Rajeev Kumar

59. Conceptualizing of Social Networking Sites **422-428**
J. S. Sodhi and Shilpi Sharma

60. Email Mining: A Survey **429-434**
Pranjal S. Bogawar and Kishor K. Bhoyar

61. Data optimization from a multiple species network using modified SFLA(Shuffled Frog Leaping Algorithm) **435-445**
N. Kannaiya Raja, N. Kannaiya Raja, K. Arulanandam, K. Arulanandam, and S. K. Sugunedham

62. Fuzzy ID3 Decision Tree Approach for Network Reliability Estimation **446-450**
Ak.Ashakumar Singh and Momtaz Thingujam

63. Implementation and Throughput Analysis of Perfect Difference Network (PDN) in Wired Environment **451-457**
Yogesh Golhar, R.K.Krishna and Mahendra A. Gaikwad

64. The application of methods of nonlinear dynamics for ECG in Normal Sinus Rhythm **458-467**
Nisar Hundewale

65. Performance evaluation of Statistical Approaches for Automatic Text-Independent Speaker Recognition using Robust Features **468-476**
R. Rajeswara Rao, A. Prasad and Ch. Kedari Rao

66. Secured Distributed Cognitive MAC and Complexity Reduction in Channel Estimation for the Cross Layer based Cognitive Radio Networks **477-486**
Niraj Shakhakarmi

Development Model for Knowledge Management System (KMS) to Improve University's Performance (Case Studies in Indonesia University of Education)

Munir[1] and Dedi Rohendi[2]

[1] Study Program of Computer Science, Faculty of Mathematics and Sciences Education, Indonesia University of Education, Bandung, Indonesia,

[2] Mechanical Engineering Education Department, Faculty of Technology and Vocational Education, Indonesia University of Education, Bandung, Indonesia

Abstract

Knowledge Management System (KMS) is one way to organize and documented the knowledge of institution. The purpose of this study is to develop a prototype KMS in organizing and documenting the knowledge in the university and to do a document aggregation based on the total number, subjects and writers (lecturers). The development method of KMS prototype is using sharepoints which are able to collect, store and publish all digital data available at the university which can be accessed online. Aggregation is using the percentage between the number of documents, subjects, and writers (lecturers). The aggregation's result between the number of digital files compared to the number of courses and the number of lecturers are each below 10%. In the future, the university still needs to encourage all lecturers or academic community to increase the collection of digital files that can indirectly improve the quality and educational services.

Keywords: Knowledge Management System (KMS, knowledge based, competitive advantage, resources based, digital files, online.

1. Introduction

The advance development of technology and science changes the paradigm of the university which initially relies on resources based into knowledge based, so the knowledge exists in university should be managed into an asset of the university. Problem arose in the university is the number of knowledge of individuals are not documented and well organized, and even the university itself does not realize that the individuals have knowledge that can enhance competitive advantage. Management of Science-based system is known by the term Knowledge Management System (KMS). In general, the implementation of KMS aims to increase the competitive advantage of university and can be one factor to improve the performance of the university.

The impact of the implementation of KMS can cause changes into university, therefore before applying KMS, it is necessary to do some series of activities first aimed to assess the readiness of the university and do an inventory of knowledge so that university know the knowledge had by individuals. In general, some universities are implementing KMS which focused more on the development of KM system (which emphasizes on information technology) only without noticing aspects to human-related such as socio-cultural factors and leadership. This is emphasized by Karl Wiig that the understanding and implementation of KMS should focus more on human beings (people-focused) since knowledge and empowerment are closely related to humans [4].

The success and the failure of the implementation of KMS in a university depends on organizational culture formation process generated by social interaction, effective communication and mutual trust among individuals within the university; in other words, a desire from individuals within the university to share knowledge. According to Charles T. Townley, KMS most effectively applied to learning management. Learning management is the management which is able to create, add and forward the knowledge or information and modify itself to reflect knowledge and new insights [9].

From the explanation above, it can be concluded that the application of KMS in a university that has a category of public service has to pay attention and make a socio-cultural factors into factors that influence the success of KMS. The differences of socio-cultural factors in each university create a unique planning and implementation of KMS. For example, the implementation of KMS for the public and private sector has little differences. David E. Mc.Nabb mentioned that this is due to the management between the two sectors are different [5].

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

David E. McNabb also described the characteristics of the public sector, which are (i) having limitations in the selection procedure, (ii) conducting activities within the mandate of the organization supported by political organizations, (iii) dealing with external controls in more and more specific actions, (iv) dealing with external influences on what and how they do things, (v) requiring the approval of many and diverse stakeholders, (vi) having variety goals which sometimes contradictory, (vii) having limited autonomy and control in decision making and human resources, (viii) having a limitation in increasing the incentives for staff, and (ix) being known and exposed for the mistakes made by either large or small faults to the public [5].

The implementation of KMS on the college level in Indonesia tends to be seen from technology aspects. Yet according to Ahza, the implementation of KMS is not only the completion of the KMS but also the role of the human involved in the organization. According to Sirajuddin bin Suhaimee et al, based on the study in 2005 the implementation of KMS in the Public Institution of Higher Education (PIHE) in Malaysia was still in the Intermediate level, and recorded only 47.1%; which also means only eight universities that have implemented or started to implement KMS. It is likewise in Thailand, the implementation of KM in higher education institutions has not been going well. This is because the paradigm of "Knowledge is Power" is still regarded as one of the factors that make individuals feel "safe" in the university [7].

Indonesia University of Education is a public university that is in the state of going into BHMN and soon became the Public Service Board. It allows KMS since the process of academic independence is continued still. By implementing KMS, UPI is expected to increase competitive advantage and to improve the performance of the university. The implementation should be prepared based on demands. Yet, UPI should first do the supported activities so that the prepared planning is in accordance with the conditions of organization and the applied strategy of KMS which can facilitate knowledge transfer within the University. The development of KMS is very wide and complex and yet the development of KMS in this study is focused on data gathering in the form of digital files that can be accessed online and aggregates are made based on the number of courses and lecturers in UPI.

2. Research Questions

The research conducted is designed to answer the following questions:

1. How is the model of knowledge management system (KMS) being applied in improving the performance of Indonesia University of Education?
2. How far is the achievement of the implementation of knowledge management system (KMS) in improving the performance of Indonesia University of Education?

3. Aim of The Research

This research includes two major areas, namely developing a prototype of Knowledge Management System (KMS) at the university and studying towards the achievement of Knowledge Management System (KMS) to improve the performance of the university.

4. Theoritical Foundation

Knowledge Management System (KMS) is a computerized system of Knowledge Management (KM), especially which connected to data, information and knowledge. Nilmini Wickramasinghe & Dag von Lubitz (2006) stated that the importance of understanding the basic concepts of KM has a relevance to the KM initiative. According to Awad & Ghaziri, data is a collection of facts about a discretly event that has not been organized and processed; while information is a set of data that has been processed to facilitate in making decision; whereas knowledge is a human understanding of the certain areas that have been learned through education and experiences [1].

The term of KM was first introduced in 1986 in the European management conference (American Productivity and Quality Center, 1996). The concept of KM then expanded rapidly and attracted the attention of many parties, and also reaped a lot of criticism [3]. Nilmini Wickramasinghe & Dag Von Lubitz explained that the activity of knowledge centered on individual psychology processes such as experience, perception and emotion. In organizations there are four main activities of knowledge, which are (1) knowledge acquisition, which refers to the internalization of information, (2) knowledge creation, which are related to the creation of new knowledge (3) knowledge distribution, which is the phase distribution of knowledge within the organization, and (4) knowledge application, which refers to the use of knowledge in the practical [6].

The same thing was suggested by Awad & Ghizari (2003), which stated that within the knowledge life cycle, there are several processes undertaken by the organization. Phases of the KM life cycle process can be divided into creation, capturing, organizing, refining, and transfers. The phase of creating is a new phase of knowledge creation from both a research as well as the phenomenon of certain events. The phase of capturing is the phase of collecting and

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

comprehending knowledge which is documented or not. The next phase is the phase of organizing, which is the phase of organizing knowledge so that can be searched easily and reused by individuals within the organization. Method of organizing knowledge can be done with indexing, clustering, cataloging, filtering, codifying, ontology, etc. The phase of refining is an activity in which the organized knowledge is refined for example by mining. The final phase is the phase of transfering/ disseminating, which is a phase that exchanges the knowledge which either can be procedure, tutorial or guide.

Knowledge organization is a place where individuals exchange knowledge of various functional areas within an organization by using technology and determined processes. The next step of Knowledge then is internalized and adopted into organizational culture. By referring to Grover & Davenport, Gottschalk stated that generally categorization of knowledge is divided into two which are explicit and tacit knowledge. The explicit knowledge can be in the form of words, numbers, data forms, scientific formulas, specifications, manuals, etc. [2]. This type of knowledge can be categorized as knowledge which can directly share with others in a systematic or formal way. Tacit knowledge is the opposite of explicit knowledge, which has a very personal nature and difficult to be formalized, so that will be more difficult to communicate with each other. Tacit knowledge is stored in the human mind and hard to be defined as it is relating to experiences, activities, and the value experienced by individuals. The same opinion stated by Tiwana who mentioned that knowledge category can be classified into two categories: tacit and explicit [8].

In the environment of university/higher education, there are two categories of knowledge namely the Academic Knowledge and Organizational Knowledge. The Academic Knowledge is the main aim for higher education, while the Organizational Knowledge includes knowledge which refers to the whole business process of the institution of education in the term of lack or excellence. Huang explained that the Academic Knowledge Framework is suggested to go through four main processes to form a culture of knowledge sharing and collaboration, namely (1) making knowledge visible (2) increasing knowledge intensity (3) building knowledge infrastructure (4) developing knowledge culture.

Referring to Galbereath, from the perspective of academic knowledge, learning commnity must begin at the individual stages, building sections which manage knowledge in organizations, creating an interesting domain across departments, creating a knowledge network of organizations and networks with other organizations. Thus, in the context of academic knowledge, Galbereath emphasized three strategies that built ie the individual strategy, institutional strategy and network strategy. From the above explanation, the collective knowledge built through sharing within the community, from individual through the team and the group into the sharing level of the organization. Individual strategies aimed at competence development of teachers into professional teachers. The role of KMS in this case is to develop the skills, the ability of teachers, and experiences through e-learning, and doing research. After the individual knowledge is captured, then the teams, groups and processes must be able to make the individual knowledge into organizational knowledge. After becoming an organizational knowledge, the role of KMS should increase at the level of inter-organization, namely the organizational strategy.

At the level of organizational strategy, knowledge sharing is expected to occur between organizations. This can be applied in various activities such as teacher's community, seminars, and culture sharings. It should be understood that at this stage, knowledge sharing is not limited to the organization only. The American Productivity and Quality Center (APQC) and Arthur Andersen, stated that in Organizational Knowledge Framework there are four aspects of the KM strategy namely culture, leadership, technology and measurement. Coukos-Semmel provides an example in the strategy of culture at the university, including staff development, training, community of practice, and development of organizational learning. KM strategy in leadership at the university includes KM strategic plan, in line with the vision and mission, employs a knowledgeable staff and asseses employer in his/her contribution to the knowledge. University is responsible for the provision of infrastructure and systems (intranet, webpages, electronic repositories, databases, etc.). Measurement strategy taken are comparisons with other universities in the field of KM, the allocation of resources, improvement of the knowledge base, connection and assessment of the impact of KM on the term of strategic plan.

5. Research Methode

This research is intended to improve the performance of Indonesia University of Education by using knowledge management system (KMS). The method used are Research and Development. The steps taken in the process of this research leads to a cycle based on the findings of research studies and then develop a product. Product development based on preliminary findings of this research was tested in a situation and made revisions to the test results until finally obtained a product. The product itself is knowledge management system (KMS) in improving the performance of Indonesia University of Education. The procedure of this study uses the techniques of research and development with the following steps: (i)

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

development of models, which are: preliminary study, planning, field testing and revision of the model. (ii) analysis of the KMS model impact on the performance of the university.

6. Research Findings

6.1 KMS Model of Indonesia University of Education

Based on the analysis needs, KMS should be able to fulfill the needs of KM process that has been running. Besides, KMS should be able to facilitate various components of knowledge within the UPI. Therefore, the KMS was developed by considering the following matters:

a) Content Management; the source content can be from the academic community or from outside community and digital libraries. The three sources needs to be managed, filtrated and stored in the database which can be gained from: External Source, Internal source, and Domain knowledge and information. Those sources of Content Management should be comprehended, filtrated and stored in the Knowledge Base (storage media). With a specific platform, this knowledge can be reused for the testing instrument.
b) Experience Management; the organization members as knowledge worker should participate in knowledge management. This knowledge worker is the academic community of UPI whom needed as the source of KM along with the external knowledge sources that are expected to fill the KMS to be used in the process of sharing knowledge within the organization.
c) Process Management; are the source of knowledge which can be gained from sharing results, access of the knowledge base and the sources from both internal and external which will be a cycle, and a best practice or a reference for users.

The three components above would be a contributor to the formation of knowledge management with the following cycle: (i) Knowledge Acquisition; knowledge within the academic community of UPI should be collected and categorized according to the specified category, (ii) Record Knowledge: Knowledge must be recorded and stored on storage media by using standard technology platforms and applications that are used in the organization, (iii) Knowledge Storage; new knowledge which has been digitalised will be stored on a centralized database system, (iv) Knowledge Sharing; is the process that allows the sharing of knowledge and valuable information stored on the system through the use of internet or intranet.

From the description of the needs, then KMS should be able to accommodate the needs of KM processes, namely: (i) The development of KMS should be qualified in accordance with the development flatform of system information of UPI, (ii) KMS must have the features that can represent the process of KM namely the collection, recording, storage and sharing knowledge, (iii) KMS should be designed in such a way that users do not experience access confusions.

The developed KMS has features such as (i) web-based system so it can be accessed anywhere, (ii) user authentication, (iii) upload facility, (iv) an adjustment facility for the user, (v) connection with other document storage systems owned by UPI, (vi) and a searching facility.

The system architecture developed includes 7 layers which are the interface layer, access and authentication, collaborative intelligence and filtering, application, middleware, transport and repository, as can be seen in the Fig 1.

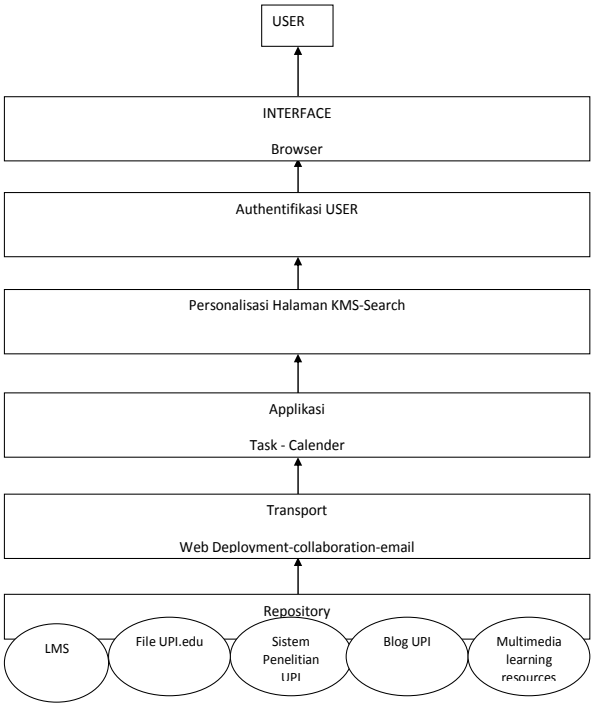


Fig. 1 The system architecture developed.

Descriptions:
Interface: This layer transfers information to and from the KMS. It is the layer that connects users who use the KMS. The interface of KMS UPI is using a web-browser. This is needed so the built system is in accordance with the policy of system development of UPI and can be integrated with other systems in UPI.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

User Authentication: This layer authenticates the validity of user and access in the KMS. The security and restricted access is part of this layer. Authentication is done by using the username and password owned by the academic community to access internet.
Personalized page of KMS: The goal of personalized pages is the effectiveness of the user; it is meant to allow users more efective in accessing KMS and avoid overloaded information.
Aplication: the existed aplication in KMS means to accommodate the process of KMS.
Transport: this layer is based on the infrastructure of intranet which provides a basic functionality to synchonized and asynchronized communication, data sharing and document and as a tool to organize electric asset in general and spesifically web content.
Repository: Repository is a place to storage and categorize knowledge data; those should be synchronized with KMS.

Based on the system architecture and features needed by KMS of UPI, the development of KMS is using Microsoft Sharepoint software. Microsoft SharePoint is one of the software released by Microsoft that has the facilities needed by the KMS such as: Knowledge Network, Search and Indexing, Content Management, Web content, Records Management, Document Management, and Collaboration. The selection of sharepoint is based on the share point architecture that allows developing the capabilities and services of SharePoint itself without disrupting the core of the platform, and in terms of scalability SharePoint has the capability to add applications that are needed without having to design and to develop from scratch. In addition, SharePoint is an application that is responsible for providing document repository services, document workflow services, intranet sites, search indexing, and collaboration service. The data storage in Microsoft SharePoint uses Microsoft SQL database server to provide data storage, retrieval, modification, and destruction.

6.2 Knowledge Achievement on The Implementaion of Knowledge Management System Model (KMS) in Improving The Perpformance of Indonesia University of Education

KMS has been running now and can be used by the academic community of UPI to share knowledge. The early stages of the implementation focused on gathering knowledges which are obtained from research results, articles or teaching materials. It is more clearly illustrated in the Table 1:

Table 1: Number of File Accessed by UPI Peoples in KMS

System	Number of File
Document storage system of KMS	2153
file.upi.edu	74162
System of research	2425
Learning management system	2020

All files have the character as the type of portable document file (pdf) of Microsoft Word, and other types of files. The files categorize based on the institution where the sources are, for example the Faculty of Science Education, the Faculty of Matematics and Science Education and so forth. The categorization is used to search files easily for general users and have not considering the prevailing taxonomy of knowledge.

7. Conclusions

The development of KMS is expected to be the starting point of inventoring the knowledge within UPI, so it is hoped that the knowledge of the academic community of UPI can be seen and used. The development was done by using a waterfall approach method and supported by Microsoft SharePoint software, which was considered to represent and have the characteristic of the required features. KMS of UPI can be accessed at http://kms.upi.edu, and have these features namely:

1. The web-based system development which can be accesed from anywhere
2. User authentication
3. Upload facility
4. Adjustment fasility for user to make some adjustment according to their needs
5. Connected to other storage document system owned by UPI
6. Searching facility

KMS has been running now and can be used either, but the success of KMS is not only determined by technological factors. There are other factors that must be considered and prepared such as the factor of people and process. Both factors mentioned above in the study were not influenced factors. Further research is needed to make the factor of people and process in KM can be run as expected. The successful implementation of KM should be supported also by the leadership policy of the university so that the academic community has a desire and willingness to share knowledge, and given support through policy and reward for the academic community who share the knowledge. And for the results of data analysis on the number of digital files in the database of UPI compared to the number of courses and lecturers at UPI have not shown encouraging results; it is still below 10%.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

References

[1] Awad, M. Elias & Ghaziri, M. Hassan, (2003) Knowledge Management. New York: Prentice Hall.

[2] Gottschalk, P. (2005). Strategic knowledge management technology. hershey PA: Idea Group Publishing (an imprint of Idea Group Inc.).

[3] Jan Hidajat, D. C. (2006). Knowledge Management dalam Konteks Organisasi Pembelajar. Bandung: SBM-ITB.

[4] Karl Wiig (2004), How effective Decision Making Leads to Cooperate Success. USA: Butterworth-Heinemann.

[5] McNabb, David E. (2007), Knowledge Management in the Public Sector: A Blueprint for Innovation in Government; Accessed 14 March 2009: http://books.google.com/books?id= WSc102GMZ9cC&dq= knowledge+management+in+public+sector&source=gbs_summary_s&cad=0.

[6] Nilmini Wickramasinghe, D. v. (2006). Knowledge-Based Enterprise:Theories and Fundamentals. London: Idea Group publishing.

[7] Sirajuddin bin Suhaimee,Ahmad Zaki Abu Bakar, Rose Alinda Alias. (2005). "Knowledge Sharing Culture in Malaysian Public Institution of Higher education : an overview. Proceedings of the Postgraduate Annual Research Seminar, (pp. 47-51).

[8] Tiwana, A. (1999). Knowledge Management Toolkit. NJ: Prentice Hall.

[9] Townley, T, Charles (2012). Knowledge Management and Academic Libraries. USA: American Library Association.

Munir, is currently a lecturer of the Study Program of Computer Science, Faculty of Mathematics and Sciences Education, Indonesia University of Education. He received a Ph.D degree in Information Science from Universiti Kebangsaan Malaysia in 2001 on “Multimedia in Education for Literacy”. Dr. Munir’s research interests are in Multimedia for Education and e-learning. He was Director of ICT Directorate, Indonesia University of Education

Dedi Rohendi, is a lecturer of Mechanical Engineering Education Department, Faculty of Technology and Vocational Education and the Study Program of Computer Science, Faculty of Mathematics and Sciences Education, Indonesia University of Education. He received Master degree in Computer and Informatics Systems from Gadjah Mada University and Doctor degree in Mathematics Education from Indonesia University of Education. Dr. Dedi Rohendi’s research interest are in implementation e-learning in Mathematics Education.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

A Clustering Method of Highly Dimensional Patent Data Using Bayesian Approach

Sunghae Jun

Department of Statistics, Cheongju University
Cheongju, Chungbuk 360764, Korea

Abstract

Patent data have diversely technological information of any technology field. So, many companies have managed the patent data to build their R&D policy. Patent analysis is an approach to the patent management. Also, patent analysis is an important tool for technology forecasting. Patent clustering is one of the works for patent analysis. In this paper, we propose an efficient clustering method of patent documents. Generally, patent data are consisted of text document. The patent documents have a characteristic of highly dimensional structure. It is difficult to cluster the document data because of their dimensional problem. Therefore, we consider Bayesian approach to solve the problem of high dimensionality. Traditional clustering algorithms were based on similarity or distance measures, but Bayesian clustering used the probability distribution of the data. This idea of Bayesian clustering becomes a solution for the problem in this research. To verify the performance of this study, we will make experiments using retrieved patent documents from the United States Patent and Trademark Office.

Keywords: *Patent Clustering, Bayesian Clustering, Highly Dimensional Problem, Probability Distribution, Bayesian Learning.*

1. Introduction

The applied patent documents have been increased rapidly according to the progress of intellectual property systems. Most governments and companies have applied their results of the developed technologies to the patents in the world. Also, they analyze the patent data for efficient patent management because the patent data have diversely technological information of any technology field. Many companies have managed the patent data to build their R&D policy. Patent analysis is an approach to the patent management. Also, patent analysis is an important tool for technology forecasting. The classification and clustering of patent documents are popular methods of patent analysis. In this paper, we propose a clustering method of patent documents. The patent document data are sparse and have high dimensional data structure. So, it is difficult to cluster the patent documents by traditional clustering methods such as hierarchical and K-means clustering algorithms. To solve the dimensional problem of the patent data, we consider Bayesian clustering in this research. The idea of Bayesian analysis is to combine previous and current information of given domain data [1]. There were so many researches of Bayesian approach [2-3]. Efficient clustering was a popular issue of Bayesian data analysis [4-5]. Most results of Bayesian clustering were model based clustering [6]. This clustering method provided the determination of the number of clusters optimally [7]. So, we will use this approach to select the number of clusters in this paper. Bayesian clustering provides the number of clusters by the posterior values using Bayes' rule. The clustering is to assign multivariate data into a number of clusters by similarity or distance measures [8]. Another clustering approach was introduced by Symons [9]. This method used a probabilistic model by mixtures of multivariate normal distribution. Bayesian hierarchical clustering was one of Bayesian clustering approaches [10]. In this paper, we construct a dendrogram as a clustering result using posterior probability as a distance value. This is one of the results of Bayesian hierarchical clustering. We will also build a log posterior plot for determining the number of clusters. We make experiment for verifying the improved performance of this research using retrieved patent documents from United States Patent and Trademark Office (USPTO) [11]. The topic of retrieved patent documents is social network service (SNS). In our experiment, we will cluster the searched patent data using Bayesian clustering.

2. Background Research

2.1 Clustering Patent Data with High Dimension

Patent document clustering is similar to general document clustering. But, the data type of patent document is different to general document in one instance. That is, patent document has issued date, citation, and family patent information. These are not in general document. So, we need to consider the characteristics of patent document. One of them is high dimensionality problem. We have to transform the patent document data into structured data for quantitative clustering methods. In this process, the high dimensionality problem is occurred. The structured data

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

are consisted of large matrix with documents and terms. The row and column represent document and term respectively. Each value of the matrix is the occurred frequency of a term in each document. Also, most values are 0 because the dimension of the matrix is so high. So, it is difficult to analyze the matrix. To settle this difficulty, we consider Bayesian analysis in this paper. Using the probabilistic approach of Bayesian clustering, we will cluster the patent documents.

2.2 Bayesian Clustering

A general method of Bayesian clustering was based on finite mixture model of probability distribution [12]. This probability model partitions a set of elements to some clusters using predictive distribution [13]. That is, Bayesian clustering is performed by the conditional probability of comparing to the existing clusters. Sometimes, the clustering makes a new cluster by the probability distribution. The elements and clusters are represented by $D=\{d_1, d_2, \ldots, d_n\}$ and $C=\{c_1, c_2, \ldots, c_k\}$ respectively. We can define the joint probability distribution of the elements as follow.

$$p(d_1, d_2, \ldots, d_n) = \prod_{i=2}^{n} p(d_i \mid d_{i-1}, \ldots, d_2, d_1) \qquad (1)$$

Where, the probability of $p(d_1, d_2, \ldots, d_n)$ can be defined as the conditional probability of $p(d_i|d_{i-1}, \ldots, d_2, d_1)$. Using this predictive distribution, we can assign the elements to existing clusters or new clusters. Markov Chain Monte Carlo (MCMC) method is more advanced Bayesian approach [3]. But, this demands larger computing resource and time cost for analyzing the given data. This research does not consider MCMC method because we have the highly dimensional problem of the document data.

3. A Proposed Method

Bayesian analysis has a powerful method to combine current data information with previous analyzer's knowledge. The combined result produces updated information for optimal decision. This is based on Bayes' rule [14]. The parameter θ of Bayesian model is a random variable [14]. This has a probability distribution. That is, the prior distribution of the parameter $p(\theta)$ is defined subjectively and initially without the information of given data. The other way, likelihood distribution is based on given data x. This is a conditional distribution, $l(x|\theta)$. By multiplying the prior distribution and likelihood function, Bayes' rule computes the posterior distribution $p(\theta|x)$ of the parameter of Bayesian model as follow [15].

$$p(\theta \mid x) = \frac{l(x \mid \theta) p(\theta)}{P(x)} \qquad (2)$$

This is the analyzer's updated belief of the parameter of Bayesian model for optimal decision (clustering). To solve the highly dimensional problem of patent document clustering, we use this Bayesian updating approach. Bayes' theorem was updated as the following formula [14].

$$Posterior \propto Prior \times Likelihood \qquad (3)$$

The prior has the previous knowledge of given data. Also, the current information of the data is included in the likelihood function. So, we can compute posterior probability by multiplying prior distribution and likelihood function. In this paper, we construct a clustering method for the patent document clustering using Bayesian updating. Bayesian clustering is a probabilistic model on data partition for efficient document clustering. In this paper, we use the hierarchical model of Bayesian clustering for suitable clustering of the patent documents with the problem of high dimensionality. This research constructs a hierarchical dendrogram by posterior probability as a similarity measure. We are able to partition the patent documents to the defined clusters using posterior result of Bayesian updating. Also, we will verify the efficiency of this research by efficient clustering of the patent documents. We have to select the parameter of Bayesian distribution family for the document clustering. The retrieved patent documents are transformed to matrix data by the preprocessing of text mining. The column of this matrix is the extracted term (cluster variable). Each document (observation) is represented by a row of the matrix. This matrix is used for input data of the hierarchical Bayesian clustering in this paper. The distribution family of Bayesian model has Gaussian and Laplace in this model. The proposed method has two steps of input and output as follow.

Input:
(1) Given data, $X=\{x_1, x_2, \ldots, x_n\}$
(2) Prior distribution, $p(\theta)$
(3) Likelihood function, $l(x|\theta)$

Output:
(1) Posterior distribution, $p(\theta|x)$
(2) Updated parameters of hierarchical Bayesian model
(3) Dendrogram of Bayesian clustering

(Step1) Initialization
(1-1) Setting parameters $(\theta_1, \theta_2, \ldots, \theta_p)$ of prior distribution

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

(1-2) Computing likelihood function $l(X|\theta_1, \theta_2, \ldots, \theta_p)$
(1-3) Updating posterior distribution $p(\theta_1, \theta_2, \ldots, \theta_p|X)$
(1-4) Selecting the number of clusters, k
(1-5) Setting $X=\{x_1, x_2, \ldots, x_n\}$

(Step2) Repetition (Clustering), k>1
(2-1) Searching x_i and x_j with input and output with the highest probability
(2-2) Assigning x_i and x_j into x_k

We initialize the parameters and the prior distribution types of Bayesian model in step1. The number of clusters and data are built in this step. That is, all parameters of Bayesian model such as prior distribution are decided in step1. In the step2, we repeat the Bayesian updating for finding proper parameters of final Bayesian model for patent documents assignment. Therefore, we perform the efficient clustering by combining the results of step1 and step2. We also use the result of the dendrogram of the hierarchical Bayesian clustering.

4. Experimental Result

We used the retrieved patent documents from USPTO to verify our improved performance. The patent data of 'social network service (SNS)' were searched for this experiment. We retrieved all applied patent documents in the U.S. up to now. Total 71 patent documents were searched in this research. The following figure shows the number (frequency) of issued patent documents until December 1, 2011. Also, we use R-project as the analytical tool for Bayesian clustering in this experiment [16-17].

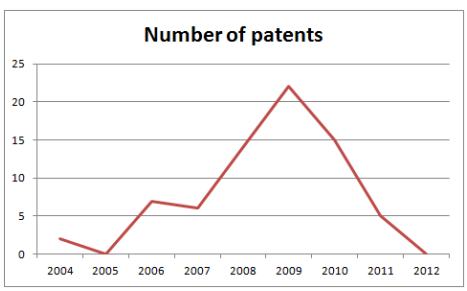


Fig. 1 Number of applied patents by year.

The first patent of the SNS was applied in 2004. We knew the SNS patent documents were increased 2008 through 2009 rapidly. But, the number of applied patents of SNS was decreased after that. So, we concluded that most needed technologies of SNS were developed until 2009. This is considerable for the R&D planning of SNS technologies. Next, we transformed the retrieved patent documents into matrix data. In this paper, we had two steps to construct the matrix data for Bayesian clustering. We constructed the first matrix consisting of retrieved documents and extracted terms. Next, we made second matrix from the first matrix. This matrix had the documents and predicted principal component (PC) scores. All PC scores were computed by principal component analysis (PCA) using the correlation structure of the terms [8]. Next table shows the first (term based matrix) and second (PC score based matrix) data.

Table 1: First and second matrix data

Matrix type	Row	Column
First matrix	71 documents	1559 terms
Second matrix	71 documents	71 PC scores

In this paper, we used the second matrix for input data of Bayesian clustering. The second data had smaller dimension than first data. We identified the result of dendrogram of the hierarchical Bayesian clustering. This dendrogram is shown in next figure.

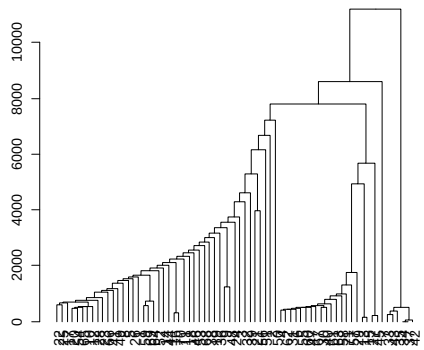


Fig. 2 Dendrogram of hierarchical Bayesian clustering.

The X and Y axes represent patent documents and log posterior values respectively. We could the hierarchical structure of the SNS patent data using the dendrogram. In this experiment, we found the number of clusters for SNS patents clustering from the following figure.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

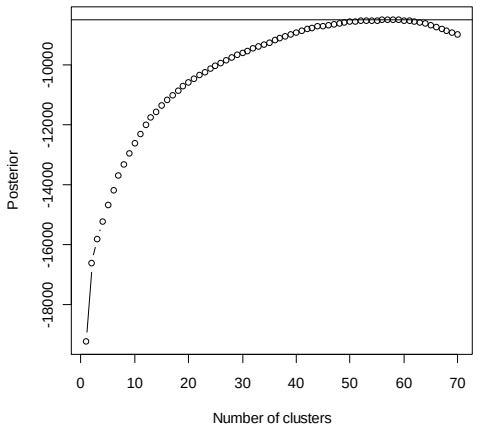


Fig. 3 Number of clusters by posterior.

The result of the hierarchical Bayesian clustering for SNS patent (technology) data gives about 55 clusters. This meant that the developed technologies of SNS were in diverse and wide areas. Next figure represents the importance of clusters variables. In this experiment, we used 71 cluster variables.

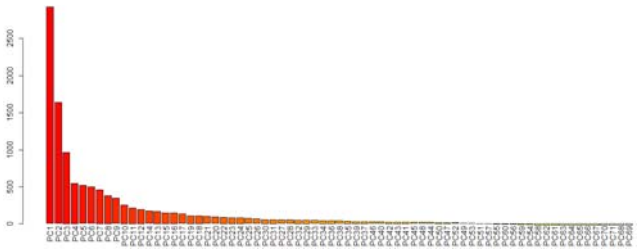


Fig. 4 Importance of cluster variables.

We knew that a few variables affected the clustering result among 71 clustering variables. The largest and smallest important variables were PC1 and PC69 respectively.

5. Conclusions and Future Works

In this paper, we proposed a clustering method of patent documents using the hierarchical Bayesian clustering. General clustering methods such as K-means clustering had some problems for patent documents clustering because the patent data are sparse and have high dimensionality. So, in this research, we considered Bayesian clustering as an efficient method for of the patent document clustering.

In this experimental result, we verified the performance of this study. We determined the optimal number of clusters by the log posterior distribution. Also, we assigned the retrieved patent documents using the result of Bayesian clustering. Also, we found the technological trend of SNS technology.

Our future work is to develop more advanced Bayesian clustering model by the MCMC method and Bayesian mixture mode for the patent document clustering.

References

[1] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, Bayesian Data Analysis, Chapman & Hall, 1995.

[2] M. D. Escobar, and M. West, “Bayesian density estimation and inference using mixtures”, Journal of the American Statistical Association, Vol. 90, 1995, pp. 577-588.

[3] G. J. McLachlan, and K. E. Basford, Mixture models: Inference and applications to clustering, Marcel Dekker, 1988.

[4] J. D. Banfield, and A. E. Raftery, “Model-based Gaussian and non-Gaussian clustering”, Biometrics, Vol. 49, 1993, pp. 803-821.

[5] F. A. Quintana, and P. L. Iglesias, “Bayesian Clustering and Product Partition Models”, Journal of the Royal Statistical Society Series B, Vol. 65, 2003, pp. 557-574.

[6] C. Fraley, and A. E. Raftery, “Model-Based Clustering, Discriminant Analysis, and Density Estimation”, Journal of the American Statistical Association, Vol. 97, 2002, pp. 611-631.

[7] C. Fraley, and A. Raftery, “How many clusters? Which clustering methods? Answers via model-based cluster analysis”, Computer Journal, Vol. 41, 1998, pp. 578-588.

[8] J. Han and M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann, 2006.

[9] M. J. Symons, “Clustering criteria and multivariate normal mixtures”, Biometrics, Vol. 37, 1981, pp. 35-43.

[10] K. A. Heller, and Z. Ghahramani, “Bayesian hierarchical clustering”, Proceedings of the 22nd International Conference on Machine Learning, 2005.

[11] United States Patent and Trademark Office (USPTO), www.uspto.gov

[12] J. M. Bernardo, and J. Giron, “A Bayesian approach to cluster analysis”, Questiio, Vol. 12, No. 1, 1988, pp. 97-112.

[13] F. A. Quintana, “A Predictive View of Bayesian Clustering”, Journal of Statistical Planning and Inference, Vol. 136, Iss. 8, 2006, pp. 2407-2429.

[14] S. J. Press, Bayesian Statistics: Principles, Models, and Applicaions, Wiley, 1989.

[15] P. Giudici, Applied Data Mining, Statistical Methods for Business and Industry, Wiley, 2003.

[16] V. P. Nia, R Package ‘bclust’, R-Project, 2011.

[17] R Development Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org. 2010.

Sunghae Jun is associate professor in the department of statistics, Cheongju University, Korea. He received the B.S., M.S., and Ph.D. degrees in department of statistics from Inha University, Korea, in

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

1993, 1996. Also he took the doctor's degree in computer science and engineering from Sogang University, Korea, 2007. He worked in NCR as a data mining consultant from 2000 to 2001. He was a visiting scholar in department of statistics, Oklahoma State University, OK, USA from 2009 to 2010. His research fields are machine learning, evolutionary computing, and management of technology.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

A hybrid algorithm of Harmony Search and Bees Algorithm for a University Course Timetabling Problem

Khang Nguyen[1], Phuc Nguyen[2] and Nuong Tran[3]

[1] HCMC University of Science, Vietnam

[2] Cao Thang Technical College, Vietnam

[3] HCMC University of Science, Vietnam

Abstract

This paper is concerned with the development of a new hybrid metaheuristic approach for solving a practical university course timetabling problem in Vietnam. Our hybrid method is a combination of Harmony Search (HS) algorithm and the Bees algorithm. The proposed method has been tested on 14 real-world data instances and compared with some other metaheurisitic approaches, which are Variable Neighborhood Search, Tabu Search and Bee Algorithms. Numerical results indicate the effectiveness of the hybrid HS – Bees algorithm over the others for the specific problem.

Keywords: *Bees algorithm, harmony search, university course timetabling.*

1. Introduction

The univeristy timetabling problems in general are known to be NP-complete problems [4]. It often involves in assignment of a set of courses to a given number of periods and rooms subject to some hard and soft constraints. Hard constraints must be satisfied to produce a feasible solution. Soft constraints should be satisfied as much as possible. In the timetabling literature, there are several meta-heuristic approaches have been proposed and efficiently applied for solving these problems, such as Tabu Search [3,9], Simulated Annealing [12], Variable Neighborhood Search (VNS) [10], Memetic Algorithm [11], Genetic Agorithm [13], Harmony Search [14,18], Great Deluge[15], etc. Many approaches have been introducted for solving specific curiculum-based course timetabling problem in Vietnam [2,3,4,5,15,16]. In this paper, we proposed the hybrid of Harmony Search (HS) [17] and Bees algorithm [1] for this problem. Our approach is tested over 14 real-world data which are taken from the Faculty of Information Technology, HCMC University of Science in Vietnam and compared with previous works [2,3,4]. The results show that our approach is capable of finding high quality solutions for this problem.

This paper is organized as follow: section 2 describes the details of the considered course university timetabling problem, section 3 explains details of the proposed algorithm, section 4 presents the experimental results and the comparison between proposed approach and the previous works, and section 5 shows some conclusions.

2. Problem Description

The general aim of the educational timetabling problem is to find an appropriate assignment of a set of courses (events) into a limited number of periods and rooms in such a way that satisfying a number of pre-defined constraints. The problem considered in this paper is a real-world curriculum – based course timetabling problem. This problem is highly constrained with 19 constraints, including 8 hard constraints, which must be satisfied, and 12 soft constraints, which should be satisfied as much as possible. Each soft constraint is associated with a weight factor to represent its level of importance.

- Time concepts: including the concepts of period, lecture and session.
  + Period: each period often last about 45 minutes.
  + Session: is a group of consecutive periods that forms a distinct part of the day, e.g., morning session includes all the periods in the morning.
  + Lecture: is a part that lasts one period of a course
- Resource concepts: including the concepts of lecturer, class, room and building
  + Class: is a group of students that will attend the same courses.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

+ Building of the faculty: is a set of near rooms. Each building is often distanced far from the others, so there should be some constraints to eliminate the movement of lecturers and students between these buildings in the same day.

- Major concepts: including the concepts of course and curriculum
  + Course: is a group of lectures that have the same lecturer, the same attending classes and subject. The main information of a course includes the lecturer teaching this course, one or some attending classes and the number of lectures in a week that belonging to this course, i.e., the number of periods that this course must hold in a week. Note that some courses may be pre-assigned, i.e., their periods and rooms are pre-determined by the staff, these courses will not be re-scheduled again but their pre-assigned information (e.g., the assignment of their lecturer, classes) will be considered during the timetabling process of the other courses.
  + Curriculum: is a group of courses that should not be overlapped due to the special requirements of the university. In our considered problem, a course is allowed to belong to more than one curriculum.

Next, we present the constraints of the problems. As mentioned above, each soft constraint has a relevant weight and this value is shown just right before the description of the constraint.

- Eight hard constraints (denoted by letter H and their index):

[H1] Lecturers, classes and rooms could not be assigned to periods at which they're not available (the availability of those resources are pre-defined)
[H2] All the lectures of all courses must be assigned.
[H3] The pre-assignment of any course must be respected.
[H4] The capacity of a room that is assigned to a course must be equal or greater than the planned number of students attending that course.
[H5] Courses that have the same lecturer or class must not overlap.
[H6] No room could be assigned to two simultaneous courses.
[H7] The lectures of each course in a day must belong to the same session
[H8] All lectures that belong to the same course must be assigned to the same room

- Twelve soft constraints (including one high weighted soft constraint, denoted by HS and 11 low weighted soft constraints, denoted by letter S, followed by their weight values):

[HS1] (60) Courses that belong to the same curriculum should not be overlapped.
[LS1] (10) The movement of a lecturer between two buildings in the same day should be avoided.
[LS2] (5) The movement of a class between two buildings in the same day should be avoided.
[LS3] (10) Courses should be assigned to periods that they're preferred.
[LS4] (5) The idle time between consecutive courses in the same session for a lecturer or a class should be avoided.
[LS5] (5) The number of periods that a class is assigned in a day should be equal or less than maxStudyingPeriods.
[LS6] (5)The idle time between consecutive courses in the same day for a lecturer or a class should be avoided.
[LS7] (10) The number of sessions that a lecturer was assigned to should be as small as possible.
[LS8] (10) The number of periods that a lecturer is assigned in a day should be equal or less than maxTeachingPeriods.
[LS9] (5) The number of days that a class was assigned to should be as small as possible.
[LS10] (10) In the timetable of a class in session, if there are three periods in which this class is idle, these should be consecutive periods (this is a special constraint requested by the timetabling staff because they need use these idle periods for another purpose).
[LS11] (10) All courses that belong to the same curriculum which are scheduled to the same session should be assigned to the same building at that session.

3. The Algorithm

3.1. Hybrid HS-Bees Algorithm

In this subsection, we briefly review the general ideas of Harmony Search algorithm and Bees algorithm, besides that, the hybrid approach combining those two algorithms, namely the Hybrid HS-Bees algorithm is presented.

3.1.1. Harmony Search Algorithm

The harmony search (HS) algorithm is a new population-based metaheuristic algorithm proposed by Geem et al. [17] in 2001. The algorithm mimics the characteristics of a musician. The mimicked characteristics of a musician include the random selection, memory consideration and pitch adjustment. The goal of the process is to reach a perfect state of harmony. HS's parameters include the Harmony Memory Size (HMS), which is the number of harmonies (solutions) in the harmony memory; Harmony Memory Considering Rate (HMCR); Pitch Adjusting Rate (PAR) and the number of improvisations (NI). The harmony memory (HM) is a memory location (history) where all the harmonies (solutions) are stored. The four basic steps of HS algorithm are shown in Figure 1, and details of the improvising step are described in Figure 2.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

1. Initialize a Harmony Memory (HM).
2. Improvise a new harmony from HM.
3. If the new harmony is better than the worst harmony in HM, put the new harmony in HM, and exclude the worst harmony from HM.
4. If stopping criteria are not satisfied, go to 2.

Fig. 1. The basic steps of the Harmony Search

Define a new harmony (solution) $x_{new}=(x_1, x_2,\ldots, x_n)$
1. While (i <= the number of components (n))
2. If (rand < HMCR)
3. Choose a random value from HM for the component x_i.
4. If (rand < PAR), Adjust the value of component x_i
5. End If
6. Else
7. Choose a random value for component x_i.
8. End If.
9. End while
10. Return x_{new}

Fig. 2. Improvise a new harmony from HM

3.1.2. The Bees Algorithm

The bees algorithm is a population-based search algorithm proposed by Pham et. al. [1]. It's inspired by the foraging behaviour of honey bees and performs a kind of neighbourhood search combined with random search to find the global optimum. The algorithm requires a number of parameters to be set, namely the number of scout bees (n), the number of sites selected for neighborhood search (out of n visited sites) (m), the number of elite bees among m selected sites (e), the number of bees recruited for patches visited by e elite bees (nep), the number of bees recruited for the other (m-e) selected patches (nsp) and size of patches (ngh).

1. Initialise the population with random solutions
2. Evaluate the fitness of the population
3. while (stopping criterion is not met)
//Forming new population
4. Select sites for neighbourhood search
5. Recruit bees for selected sites (the number of bees for a sites is proportional to the quality of that site) and evaluate the fitness.
6. Select the fittest bee from each patch
7. Assign remaining bees to search randomly and evaluate their fitness.
8. end while.

Fig. 3. The basic steps of the Bees Algorithm

The algorithm starts with the n scout bees being placed randomly in the search space. The fitness values of the sites that scout bees visited are evaluated in step 2. The bees having the highest fitness values are selected as "elite bees" in step 4 and sites visited by them are chosen for neighbourhood search in steps 5-6. The algorithm recruits more bees to follow the elite bees (e) than the other selected bees (m-e) in order to search around the neighbourhood of sites visited by the elite bees. In step 6, only one bee with the highest fitness value will be selected for each site to generate the next bee population. Besides, the remaining bees (n-m) in the bee population are randomly assigned around the search space in order to explore new potential solutions in step 7. These steps are repeated until the stopping criterion is met.

3.1.3. The Hybrid HS-Bees Algorithm

The Hybrid HS – Bees algorithm combines the optimization capabilities of harmony search algorithm and bees algorithm. The disadvantages of the basic bees algorithm are premature convergence and not obtaining experiences between solutions in a population. The Harmony search algorithm handles intensification and diversification. The diversification is controlled by pitch adjustment and random selection, which would help good local solutions to be retained. Random selection explores global search space more widely and efficiently while the pitch adjustment makes the new solution good enough and near the existing good solutions. The intensification in Harmony search algorithm is controlled by memory consideration, leading the searching process toward the searching space of good solutions. The hybrid HS-Bees algorithm involves two phases of optimization: (a) The bees algorithm using neighbourhood search and random search, (b) The harmony search using memory consideration, random selection and pitch adjustment. The general process of the algorithm is shown in Figure 4.

1. Initialise population with random solutions
2. Evaluate fitness of the population
3. While (stopping criterion not met)
// Bees algorithm
4. Select sites for neighbourhood search
5. Recruit bees for selected sites (more bees for best e sites) and evaluate fitnesses.
6. Select the fittest bee from each patch
7. Assign remaining bees to search randomly and evaluate their fitnesses.
// Harmony Search Algorithm
8. Initialize Harmony Memory (HM).
9. Improvise a new harmony from HM.
10. Improve the new harmony using Hill Climbing

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

11. If the new harmony is better than the worst harmony in HM, include the new harmony in HM, and exclude the worst harmony from HM.
12. End While.

Fig. 4. The proposed hybrid HS - Bees Algorithm

3.2. The Hybrid HS-Bees Algorithm for the considered university course timetabling problem

3.2.1. Objective function

As described in section 2, the considered problem has 8 hard constraints and 12 soft constraints. Each soft constraint has a weighted value to represent its importance and the quality of a solution will be evaluated by the objective function $f(x)$, which is a linear weighted combination of the violation of soft constraints. In this approach, hard constraints H5 and H6 are relaxed. There're two reasons for this relaxation: first, those constraints is not easily to satisfied during the search because its relevant feasible assignment are dynamic, second, relaxing them will help to avoid the case of stucking in some local region by the lack of connectivity in the searching space. Those relaxed constraints will be treated as soft constraints during the searching process, with dominantly high weighted value compared to the original soft constraints (here we use the weighted value of 1000), helping the search not go too far from the real feasible region.

3.2.2. Creating random initial solutions

In those random initial solutions, every course is randomly assigned to periods and rooms in such a way that the non-relaxed hard constraints are not violated.

3.2.3. Moves used in the neighbourhood search

In Bees algorithm's phase, neighbour solution is created from current solution by randomly choosing a kind of move in the set of two kinds of moves, including: Single moves and Swap moves. These moves are used for searching in the neighbourhood of the selected sites with *ngh* parameter is 10 percent of a random available set of courses of current solution.

- Single moves: moves that changing either the period assignment or the room assignment of a course. Single move includes 3 components: a Course k, the new first period p and the new room r. We consider a set of courses (neighbourhood) (called N_{cur}) in current solution. Firstly, for each course k in N_{cur}, examine all available first period p for this course k. Secondly, find the best appropriate room r which is available at period p and its capacity must be equal or greater than the planned number of attending students. Finally, we choose the best single move in the set of single moves (k, p, r) which are created above.

- Swap moves: moves that exchanging either the period assignment or the room assignment of two courses. Each swap move includes 2 components: Course k_i and Course k_j. We consider a set of courses (neighbourhood) (called N_{cur}) in current solution. Firstly, for each course k_i in N_{cur}, examine all courses k_j in N_{cur} which is able to swap first period and room with course k_i. Secondly, we choose the best swap move in a set of swap moves (k_i, k_j) which are created above.

Create_A_Neighbour_Solution(X)
1. Randomly Select $\alpha \in \{1 \ldots ngh\}$
2. Set X' = X
3. Randomly choose a kind of move in two kinds of moves (Single move, Swap move)
4. Take a best solution X' in the neighbourhood of X created by the chosen kind of move with α percent of available of set of courses of X.
5. return X'

Fig. 5. The method to create a neighbour solution of X

3.2.4. Details of the application of Harmony Search Algorithm in proposed approach

Initialize Harmony Memory: In this step, all solutions in bees' population is sorted in ascending order by objective function values. Then, we initialize HM based on the the HMS best solutions in the population.

Improvise a new harmony: In Harmony search algorithm, three Harmony search operations are applied during each improvement of a new harmony, namely Memory consideration, random selection and pitch adjustment. Each course in a solution is a variable in the solution's relevant harmony.

a. *Memory consideration*: Memory consideration operators would decide the first period and room of every course of the new harmony based on solutions stored in HM with HMCR rate.

b. *Random selection*: In this operator, we choose randomly the new first period and new room for courses of the new harmony with (1-HMCR) rate.

c. *Pitch adjustment*: Pitch adjustment operator is applied with PARxHMCR rate, the first period and room of the chosen course of the new harmony are taken from the best harmony in HM.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

d. *Additional Hill Climbing algorithm*: Hill Climbing is an iterated algorithm for optimization problems. In this problem, it attempts to improve a newly generated harmony as a new operator of Harmony search algorithm. A neighbour solution is created by using single move and swap move in the neighbourhood search above.

3.2.5. Stop criteria

The hybrid algorithm will stop when the stop criteria (the maximum number of allowance iterations) is met.

4. Experimental results

The hybrid HS-Bees algorithm is implemented in C++. The computer used to test has the configuration of Core2 Duo CPU, 4GB RAM and Windows Vista OS. Fourteen real- world data instances taken from the Faculty of Information Technology, HCMC University of Science in Vietnam are used to evaluate the hybrid algorithm.
The hybrid algorithm was run on each instance for ten times and the best and average results are reported in table I. In this implementation, the weights of soft constraints are: w(HS1) = 60, w(S1) = 15, w(S2) = 10, w(S3) = 10, w(S4) = 5, w(S5) = 2, w(S6) = 5, w(S7) = 10, w(S8) = 10, w(S9) = 5, w(S10) = 10, w(S11) = 10. In Bees algorithm's phase, we choose n (scout bees) in the set of {20, 30}, m (selected sites) in set of {4, 5}, e (number of elite bees) in set of {3, 4}, nep (number of bees recruited for patches visited by e elite bees) in set of {4, 5}, nsp (number of bees recruited for the other (m-e) selected patches) in set of {2, 3}, ngh (patch of size) in set of {5, 10}. Otherwise, In HS phase, we choose HMCR rate in the set of {0.7,0.8,0.95}, PAR rate in the set of {0.3, 0.4, 0.5}. We test all possible pairs of these parameters' values and find out that the pair {n=30, m=5, e=4, nep=4, nsp=2, ngh=10, HMCR=0.95, PAR=0.4} gives the best result.

Table 1. Details of 14 data instances

# instance	# courses	# teachers	# classes	# rooms	# curricula
Data 1	82	44	13	40	11
Data 2	80	44	10	46	10
Data 3	95	44	14	45	11
Data 4	84	45	11	40	10
Data 5	89	48	12	51	11
Data 6	51	31	19	45	0
Data 7	96	50	14	47	9
Data 8	88	44	11	47	5
Data 9	107	43	15	46	5
Data 10	78	47	12	48	3
Data 11	83	45	13	53	7
Data 12	87	52	21	53	5
Data 13	84	51	20	53	6
Data 14	102	55	26	53	9

Table 2. Best results on 14 data instances

#instance	Hybrid HS-Bees Algorithm	VNS	The Bees Algorithm	Tabu Search
Data 1	**172**	2647	435	272
Data 2	**194**	1209	382	1141
Data 3	**305**	767	569	577
Data 4	**188**	867	405	1657
Data 5	**142**	693	444	572
Data 6	**155**	1595	1876	1450
Data 7	213	1343	498	142
Data 8	1378	358	1420	154
Data 9	**269**	564	480	282
Data 10	1100	312	343	1167
Data 11	592	354	343	147
Data 12	1447	319	315	1125
Data 13	497	405	466	231
Data 14	**185**	490	297	210

5. Conclusion

In this paper, the hyrid HS - Bees algorithm is presented to solve a real-world university timetabling problem. The results show that the proposed algorithm is capable of efficiently solving this problem and has been confirmed by comparing with known results: VNS, The Bees Algorithm, Tabu Search. For future research, the hybrid HS - Bees algorithm can be applied other timetabling problems like high school timetabling, examination timetabling.

References

[1] D. T. Pham, Ghanbarzadeh A., Koc E., Otri S., Rahim S., and M.Zaidi. "The Bees Algorithm - A Novel Tool for Complex Optimisation Problems", Proceedings of IPROMS 2006 Conference, pp.454-461, 2006.

[2] NTTM Khang, NB Phuc, TTH Nuong, "The Bees Algorithm for a Practical University Timetabling Problem in Vietnam", Computer Science and Automation Engineering (CSAE), 2011 IEEE International Conference, pp. 42-47, Shanghai, 2011.

[3] NTTM Khang, DTT Nguyen, TT Khon, TTH Nuong, "Automating a Real-World University Timetabling Problem with Tabu Search Algorithm", Computing and Communication Technologies, Research, Innovation, and Vision for the Future (RIVF), pp. 1-6, Hanoi, 2010.

[4] NTTM Khang, NT Quang, TTT Hien, NB Phuc, TTH Nuong, "Variable Neighborhood Search for a Real-world Curriculum-based University Timetabling Problem", The Third International Conference on Knowledge and Systems Engineering (KSE 2011), Hanoi, 2011.

[5] NTTM Khang, NQ Nam, NQ Vi, NB Phuc, TTH Nuong, "Applications of particle swarm optimization algorithm to a practical university timetabling problem", 2011 International

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Conference on Modeling and Optimization (ICMO 2011), Cairo, Egypt, 2011 (to appear).

[6] C Lara, J Flores, F Calderón, "Solving a School Timetabling Problem Using a Bee Algorithm", Proceedings of the 7th Mexican International Conference on Artificial Intelligence: Advances in Artificial Intelligence, pp. 664 – 674, Springer-Verlag Berlin, 2008.

[7] B. McCollum, P. McMullan, B. Paechter, R. Lewis, A. Schaerf, L. Di Gaspero, A. J. Parkes, R. Qu, E. Burke: Setting the Research Agenda in Automated Timetabling: The Second International Timetabling Competition. Technical Report, Internaltional Timetabling Competition 07-08, 2007.

[8] TB Cooper and JH Kingston, "The Complexity of Timetable Construction Problems" in the Practice and Theory of Automated Timetabling, ed. EK Burke and P Ross, pp. 283-295, Springer-Verlag (Lecture Notes in Computer Science), 1996. Basser Department of Computer Science, University of Sydney, Australia, 1996.

[9] LÄu Z, Hao JK, "Adaptive tabu search for course timetabling". European Journal of Operational Research doi: 10.1016/j.ejor.2008.12.007, 2009.

[10] Abdullah S.,Burke E.K. and McCollum B.:"An Investigation of a Variable Neighborhood Search Approach for Course Timetabling". Proceedings of the 2nd Multidisciplinary Conferenceon Scheduling: Theory and Applications, 18-21July, NY, USA, 413-427, July, 2005.

[11] S. N. Jat and S. Yang, "A memetic algorithm for the university course timetabling problem," Proc. 20th IEEE Int. Conf. Tools Artif. Intell. 2008, vol. 1, pp. 427–433, 2008.

[12] D. Abramson. "Constructing school timetables using simulated annealing: sequential and parallel algorithms". In Management Science, vol.37, pp. 98-113, 1991.

[13] Burke E.K., Elliman D.G. and Weare R.F. "A Genetic Algorithm Based University Timetabling System" East-West Conference on Computer Technologies in Education, Crimea, Ukraine pp. 35-40, 1994.

[14] M. A. Al-Betar, A.T. Khader, T.A. Gani, "A harmony search algorithm for university course timetabling", PATAT 2008.

[15] McMullan, "An Extended Implementation of the Great Deluge Algorithm for Course Timetabling", Lecture Notes in Computer Science, Springer, Vol 4487, 538-545, 2007.

[16] DT Anh, VH Tam, NQV Hung, "Combined Interactive and Automatic Course Timetabling", Proceedings of International Workshop on Advanced Computing and Applications, ACOMP'2007, HoChiMinh City, March 14-16, 2007, pp.2-7, 2007.

[17] Geem, Z.W., Kim, J.H., Loganathan, G.V.: A New Heuristic Optimization Algorithm: Harmony Search. Simulation. 76(2), pp.60–68, 2001.

[18] M. A. Al-Betar, A.T. Khader, I.Y. Liao. "A Harmony Search with Multi-pitch Adjusting Rate for the University Course Timetabling", Recent Advances in Harmony Search Algorithm, SCI 270, pp. 147-161, Springer-Verlag Berlin Heidelberg, 2010.

Khang Nguyen received B.S degree and M.S degree in Computer Science from HCMC University of Science, Vietnam in 1996 and 2001. Currently, he is a Ph.D candidate in Computer Science in HCMC University of Science, Vietnam. Since 1996, he has been a lecturer at Faculty of Information Technology, HCMC University of Science, Vietnam. His research interests include timetabling problem, optimization problem. Khang Nguyen is co-author over ten papers in international conferences and national conferences.

Phuc Nguyen reveived B.S degree and M.S degree in Computer Science from HCMC University of Science, Vietnam in 2007 and 2011. Since 2008, he has been a lecturer at Faculty of Electronic and Informatics, Cao Thang Technical College, Vietnam. His research interests in timetabling problem.

Prof. Dr. Nuong Tran received Ph.D degree in Mathematics from University of Warsaw, Poland in 1989. Currently, she is a lecturer at Faculty of Mathematics, HCMC University of Science, Vietnam. Her research interests include timetabling problem, optimization problems, optimization global. Nuong Tran is co-author over thirty papers in international journals and international conferences.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Applying Genetic Algorithms for Inventory Lot-Sizing Problem with Supplier Selection under Storage Capacity Constraints

Chirawat Woarawichai[1], Kritsana Kuruvit[1], Paitoon Vashirawongpinyo[2]

[1] Department of Industrial Engineering, Faculty of Engineering, Rajamangala University of Technology Lanna Tak, Thailand

[2] Department of Industrial Engineering, Faculty of Engineering, Siam University Bangkae Bangkok, Thailand

Abstract

The objective of this research is to calculate the optimal inventory lot-sizing for each supplier and minimize the total inventory cost which includes joint purchase cost of the products, transaction cost for the suppliers, and holding cost for remaining inventory. Genetic algorithms are applied to the multi-product and multi-period inventory lot-sizing problems with supplier selection under storage capacity. Also a maximum storage capacity for the decision maker in each period is considered. The decision maker needs to determine what products to order in what quantities with which suppliers in which periods. It is assumed that demand of multiple products is known over a planning horizon. The problem is formulated as a mixed integer programming and is solved with the Genetic algorithms. The detailed computation results are presented.

Keywords: *Genetic Algorithms, Supplier selection, Storage capacity.*

1. Introduction

Lot-sizing problems are production planning problems with the objective of determining the periods where production should take place and the quantities to be produced in order to satisfy demand while minimizing production and inventory costs [1]. Since lot-sizing decisions are critical to the efficiency of production and inventory systems, it is very important to determine the right lot- sizes in order to minimize the overall cost.

Lot-sizing problems have attracted the attention of researchers. The multi-period inventory lot-sizing scenario with a single product was introduced by Wagner and Whitin [2], where a dynamic programming solution algorithm was proposed to obtain feasible solutions to the problem.

Soon afterwards, Basnet and Leung [3] developed the multi-period inventory lot-sizing scenario which involves multiple products and multiple suppliers. The model used in these former research works is formed by a single-level unconstrained resources indicating the type, amount, suppliers and purchasing time of the product. This model is not able to consider the capacity limitations. One of the important modifications we consider in this paper is that of introducing storage capacity. One of the important modifications we consider in this paper is that of introducing storage capacity constraints. With the advent of supply chain management, much attention is now devoted to supplier selection. Rosenthal et al. [4] study a purchasing problem where suppliers offer discounts when a “bundle” of products is bought from them, and one needs to select suppliers for multiple products. Then a mixed integer programming formulation is presented. Jayaraman et al. [5] proposed a supplier selection model that considers quality, production capacity, lead-time, and storage capacity limits.

In this paper based on Basnet and Leung [3] genetic algorithms (GAs) are applied to the multi-product and multi-period inventory lot-sizing problem with supplier selection under storage space. Also a maximum storage space for the decision maker in each period is considered. The decision maker needs to determine what products to order in what quantities with which suppliers in which periods. The objective of this research is to calculate the optimal inventory lot-sizing for each supplier and minimize the total inventory cost.

This paper is organized as follows: Section 2 the genetic algorithm approach is applied to problem. Section 3 we describe our model. In Section 4 presents a numerical example of the model. Finally, computation results and conclusions are presented in Section 5 and 6.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

2. Methodology

The genetic algorithms (GAs) approach is developed to find optimal (or near – optimal) solution. Detail discussion on GAs can be found in Holland [6], Michalewicz [7], and Gen and Cheng [8] [9]. In this section, we explain GAs procedure is illustrated in Fig. 1 to start the search GAs are initialized with a population of individuals. The individuals are encoded as chromosomes in the search space. GAs use mainly two operators namely, crossover and mutation to direct the population to the global optimum. Crossover allows exchanging information between different solutions (chromosomes) and mutation increases the variety in the population. After the selection and evaluation of the initial population, chromosomes are selected on which the crossover and mutation operators are applied. Next the new population is formed. This process is continued until a termination criterion is met [1].

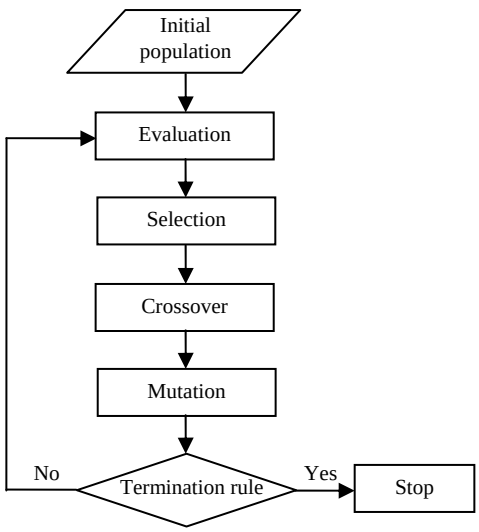


Fig. 1 The genetic algorithm procedure

3. Formulation

We also make the following assumptions and mathematical for the model:

3.1 Assumptions

- Demand of products in period is known over a planning horizon.
- All requirements must be fulfilled in the period in which they occur: shortage or backordering is not allowed.
- Transaction cost is supplier dependent, but does not depend on the variety and quantity of products involved.
- Holding cost of product per period is product-dependent.
- Initial inventory of the first period and the inventory at the end of the last period are assumed to be zero.
- Product needs a storage space and available total storage space is limited.

Base on the above assumption of model, Fig 2 shows the behavior of the model considering the scenario of multi-period inventory lot-sizing problem with supplier selection under storage space and budget constraints.

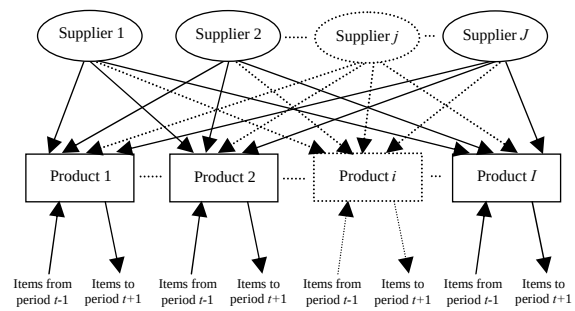


Fig. 2 Behavior of the model in period t .

3.2 Mathematical modeling

This paper is built upon Basnet and Leung [3] model. We formulate the multi-product and multi-period inventory lot-sizing problem with supplier selection under storage space and budget constraints using the following notation:

Indices:

i = $1,\ldots,I$ index of products
j = $1,\ldots,J$ index of suppliers
t = $1,\ldots,T$ index of time periods

Parameters:

D_{it} = demand of product i in period t
P_{ij} = purchase price of product i from supplier j
H_i = holding cost of product i per period
O_j = transaction cost for supplier j
w_i = storage space product i
S = total storage capacity

Decision variables:

X_{ijt} = number of product i ordered from supplier j in period t
Y_{jt} = 1 if an order is placed on supplier j in time period t , 0 otherwise

Intermediate variable:

R_{it} = Inventory of product i , carried over from period t to period $t+1$

Regarding the above notation, the mixed integer programming is formulated as follows:

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

$$Minimize\ (TC) = \sum_i \sum_j \sum_t P_{ij} X_{ijt} + \sum_j \sum_t O_j Y_{jt} + \sum_i \sum_t H_i \left(\sum_{k=1}^{t} \sum_j X_{ijk} - \sum_{k=1}^{t} D_{ik} \right) \quad (1)$$

Subject to:

$$R_{it} = \sum_{k=1}^{t} \sum_j X_{ijk} - \sum_{k=1}^{t} D_{ik} \geq 0 \quad \text{for all } i \text{ and } t, \quad (2)$$

$$\left(\sum_{k=t}^{T} D_{ik} \right) Y_{jt} - X_{ijt} \geq 0 \quad \text{for all } i,\ j, \text{ and } t, \quad (3)$$

$$\sum_i w_i \left(\sum_{k=1}^{t} \sum_j X_{ijk} - \sum_{k=1}^{t} D_{ik} \right) \leq S \quad \text{for all } t, \quad (4)$$

$$Y_{jt} = 0 \text{ or } 1 \quad \text{for all } j \text{ and } t, \quad (5)$$

$$X_{ijt} \geq 0 \quad \text{for all } i,\ j, \text{ and } t, \quad (6)$$

The objective function as shown in (1) consists of three parts: the total cost (TC) of 1) purchase cost of the products, 2) transaction cost for the suppliers, and 3) holding cost for remaining inventory in each period in $t+1$. Constraint in (2) all requirements must be filled in the period in which they occur: shortage or backordering is not allowed. Constraint in (3) there is not an order without charging an appropriate transaction cost. Constraint in (4) each products have limited capacity. Constraint in (5) is binary variable 0 or 1 and Constraint in (6) is non-negativity restrictions on the decision variable. According to a large optimal problem, a GAs approach is applied to solve this problem.

4. A numerical example

In this section we solved a numerical example of the model using the LINGO. We consider a scenario with three products over a planning horizon of five periods whose requirements are as follows: demands of three products over a planning horizon of five periods are given in Table 1. There are three suppliers and their prices and transaction cost, holding cost and storage space are show in Table 2 and Table 3, respectively.

Table 1: Demands of three products over a planning horizon of five periods (D_{it}) and Budget of them (B_t).

	Planning Horizon (Five Periods)				
Products	1	2	3	4	5
A	12	15	17	20	13
B	20	21	22	23	24
C	20	19	18	17	16

Table 2: Price of three products by each of three suppliers X, Y, Z (P_{ij}) and transaction cost of them (O_j).

	Price		
Products	X	Y	Z
A	30	33	32
B	32	35	30
C	45	43	45
Transaction Cost	110	80	102

Table 3: Holding cost of three products A, B, C (H_i) and storage space of them (w_i).

	Products		
	A	B	C
Holding Cost	1	2	3
Storage Space	10	40	50

The total storage capacity (S) is equal to 200.

The results of applying the proposed method are shown in Table 4. The solution of this problem (I = 3, J = 3, and T = 5) is to place the following orders.
All other X_{ijt} = 0:

Table 4: Order of three products over a planning horizon of five periods (X_{ijt}).

	Planning Horizon (Five Periods)				
Products	1	2	3	4	5
A	X_{111} = 12	X_{132} = 15	X_{113} = 37	-	X_{135} = 13
B	X_{231} = 20	X_{232} = 21	X_{213} = 22	X_{234} = 23	X_{235} = 24
C	X_{321} = 20	X_{332} = 19	X_{313} = 18	X_{334} = 17	X_{335} = 16

Cost calculation for this solution:
Purchase cost for product 1 from supplier 1, 3
= (37×30) + (12+15+13) × 32 = 2,390.
Purchase cost for product 2 from supplier 1, 3
= (22×32) + (20+21+23+24) × 30 = 3,344.
Purchase cost for product 3 from supplier 1, 3
= (18×45) + (20+19+17+16) × 45 = 4,050.
Transaction cost from supplier 1, 3
= (1×110) + (4×102) = 518.
Holding cost for product 1
$R_{13} = X_{113} - D_{13} = 37 - 17 = 20.$
$= H_1 \sum R_{1t} = 1 \times (0 + 0 + 20 + 0 + 0) = 20.$
Thus, the total cost for this solution
= 2,390 + 3,344 + 4,050 + 518 + 20
= 10,322.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

5. Computation results

In this section the comparison of the two methods solved problem size is using a commercially available optimization package like LINGO and GAs code is developed in MATLAB. Experiments are conducted on a personal computer equipped with an Intel Core 2 duo 2.00 GHz, CPU speeds, and 1 GB of RAM. The transaction costs are generated from *int* [50; 200], a uniform integer distribution including 50 and 200. The prices are from *int* [20; 50], the holding costs from *int* [1; 5], the storage space from *int* [10; 50], and the demands are from *int* [10; 200].

The result in Table. 5 shows the GAs comparing with LINGO for the nine problem sizes. A problem size of *I; J; T* indicates number of suppliers = *I*, number of products = *J*, and number of periods = *T*. Computation time limit is set at 120 minutes. For comparison, the percentage error is calculated by (7) and (8).

Percentage error of LINGO

$$= \left[\frac{\text{Upper bound - Lower bound}}{\text{Upper bound}}\right] \times 100 \qquad (7)$$

Percentage error of GAs

$$= \left[\frac{\text{Upper bound LINGO - GAs}}{\text{Upper bound LINGO}}\right] \times 100 \qquad (8)$$

The solution time of LINGO to optimal is a short time as the small problem size (with the problem sizes 3 x 3 x 5; 3 x 3 x 10; 3 x 3 x 15; and 4 x 4 x 10). For large problems sizes LINGO cannot obtain optimal solutions within limit time due to as the larger problem size (with the problem sizes 4 x 4 x 15; 5 x 5 x 20; 10 x 10 x 50; 10 x 10 x 80; and 15 x 15 x 50). The GAs can optimally solve when the problem size is small (with the problem sizes 3 x 3 x 5; 3 x 3 x 10; 3 x 3 x 15; 4 x 4 x 10; 4 x 4 x 15; 5 x 5 x 20; and 10 x 10 x 50). There are two problems which GAs cannot obtain optimal solutions (with the problem sizes 10 x 10 x 80; and 15 x 15 x 50). Next, we study differences in the problem sizes between solutions from the optimization with LINGO and the GAs. The results are show in Fig. 4, a plot of the problem size versus solution time. LINGO uses longer computation time more than GAs with seven problem sizes, but uses equal time with two problem sizes. As show in Fig. 5 a plot of the problem size versus % error when the problem size is very large, LINGO used a maximum % error from the optimal solutions is found to be 4.41% (at the problem size 10 x 10 x 80) which has more % error than GAs. The GAs can solve small % error of two problem sizes (at the problem size 10 x 10 x 80; and 15 x 15 x 50). Fig. 6 and Fig. 7 show compares result between LINGO and GAs in problem size 3 x 3 x 5. Thus, it is evident that GAs is an effective means for solving the problem. GAs solution is optimal when the problem size is small. For larger problems GAs can find feasible solution within time limit for which LINGO fails to find the optimum.

Table 5: Comparative results of the two methods

Problem size	Optimization approach with LINGO			Genetic Algorithms (GAs)		
	Total cost	Solution time (minute)	% Error	Total cost	Solution time (minute)	% Error
3 x 3 x 5	10,322	0.01	0	10,322	0.02	0
3 x 3 x 10	20,644	0.14	0	20,644	0.21	0
3 x 3 x 15	30,966	14.35	0	30,966	1.45	0
4 x 4 x 10	25,436	6.34	0	25,436	0.51	0
4 x 4 x 15	38,154[a] , 37,828[b]	120	0.85	38,154	2.47	0
5 x 5 x 20	60,218[a] , 59,527[b]	120	1.14	60,200	3.36	0.03
10 x 10 x 50	285,344[a] , 274,758[b]	120	3.70	284,940	108.50	0.14
10 x 10 x 80	456,494[a] , 436,317[b]	120	4.41	455,904	120	0.12
15 x 15 x 50	417,800[a] , 405,155[b]	120	2.66	416,473	120	0.31

[a]LINGO12 = Upper bound, [b]LINGO12 = Lower bound.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

However, the GAs provides superior solutions to those from LINGO that are close to optimum in a very short time, and thus appears quite suitable for realistically sized problems.
Additionally, the computation time when using GAs is also short, making it a very practical means for solving the multiple products and multi-period inventory lot-sizing problem with supplier selection under storage capacity.

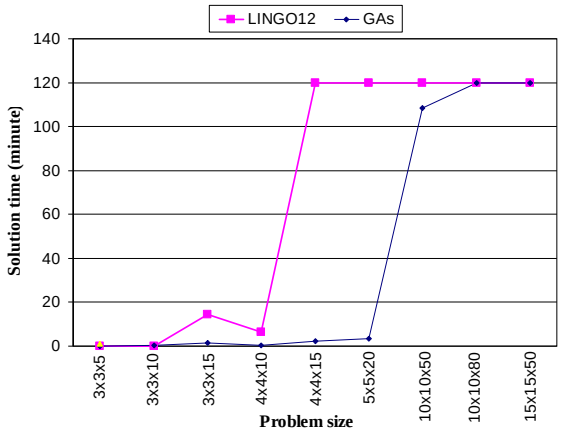


Fig. 4 Plot of the problem size vs. solution time (minute)

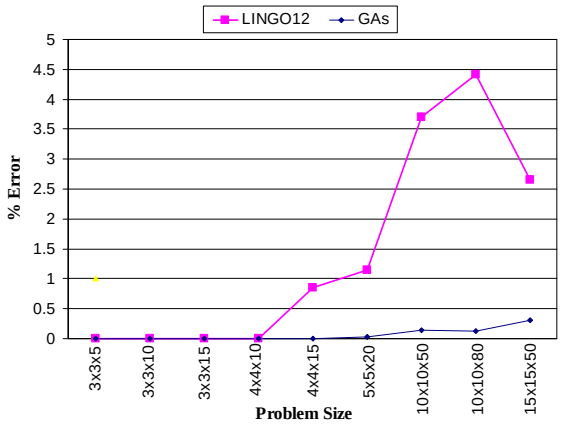


Fig. 5 Plot of the problem size vs. % error

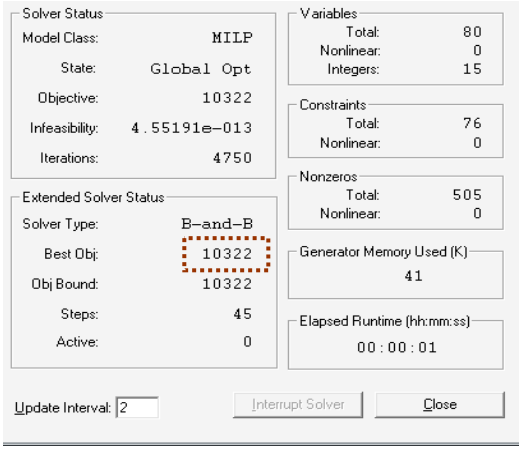


Fig. 6 The best objective of LINGO

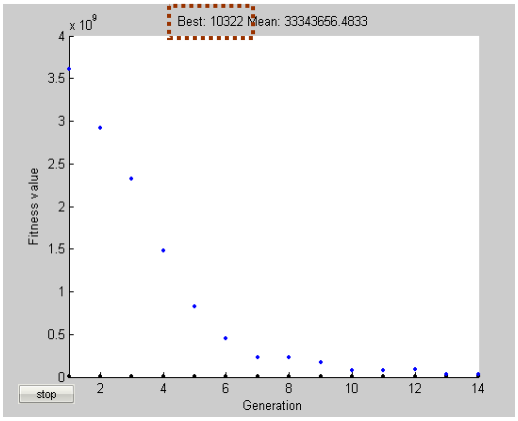


Fig. 7 The fitness value of GAs

6. Conclusions

In this paper, we present genetic algorithms (GAs) applied to the multi-product and multi-period inventory lot-sizing problem with supplier selection under storage capacity. Also a maximum storage space for the decision maker in each period is considered. The decision maker needs to determine what products to order in what quantities with which suppliers in which periods. The mathematical model is give and the use of the model is illustrated though a numerical example. The problem is formulated as a mixed integer programming and is solved with LINGO and the GAs. As compared to the solution of optimization package like LINGO, the GAs solutions are superior.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

References

[1] G.H. Goren, S. Tunali, and R. Jans, "A review of applications of genetic algorithms in lot sizing", Journal of Intelligent Manufacturing. Springer Netherlands. 2008.

[2] H. M. Wagner, and T. M. Whitin, "Dynamic version of the economic lot-size model", Management Science. 5, pp, 89–96, 1958.

[3] C. Basnet and J. M. Y. Leung, "Inventory lot-sizing with supplier selection", Computers and Operations Research. 32, pp, 1–14, 2005.

[4] E.C. Rosenthal, J.L. Zydiak, and S.S. Cauchy, "Vendor selection with bundling", Decision Sciences. 26, 1995, 35–48.

[5] V. Jayaraman, R. Srivastava, and W.C. Benton, Supplier selection and order quantity allocation: a comprehensive model", The Journal of Supply Chain Management. 35, 1999, 50–58.

[6] J. H. Holland, Adaptation in Natural and Artificial Systems. The University of Michigan Press, Ann Arbor. 1975.

[7] Z. Michalewicz, Genetic Algorithms + Data Structures = Evolution Programs. *AI Series*. Springer-Verlag, New York. 1944.

[8] M. Gen, and R. Cheng, Genetic Algorithms and Engineering Design. Wiley, New York. 1977.

[9] M. Gen, and R. Cheng, Genetic Algorithms and Engineering Optimization. Wiley, New York. 2000.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

An Agile Method for E-Service Composition

Pouya Fatehi[1], Seyyed Mohsen Hashemi[2]

[1] Department of Computer Software , Science and Research Branch, Islamic Azad university Tehran, Iran

[2] Department of Computer Engineering, Science and Research Branch, Islamic Azad university Tehran, Iran

Abstract

Nowadays, application of Service Oriented Architecture is increasing rapidly; especially since introduction of distributed electronic services on the web. SOA software has a modular manner and works as a collaboration of independent software components. As a result, e-service approach is sufficient for software with independent components, each of which may be developed by a different company. Such software components and their cooperation form a composite service. Agile methodologies are the best candidate for developing small software components. Composite services and its building blocks are small pieces of software, making agile methodology a perfect fit for their development.

In this paper, we introduce an agile method for service composition, inspired by agile patterns and practices. Therefore, across the agile manifesto, we can develop low cost, high quality composite services quickly using this method.

Keywords: *Agile model, E-Service, Service Composition, software as Service*

1. Introduction

rends to inter enterprise communication and creation of on demand-services, made e-services more popular technology. Another reason of e-service popularity is Software as a Service (SaaS) model which is a rapidly growing model of software licensing. *"SaaS is seen as a possible replacement to traditional software where the buyer obtains a perpetual license and installs and maintains all necessary hardware, software and other technical infrastructure. Under SaaS, the software publisher (seller) runs and maintains all necessary hardware and software and buyers obtain access using the Internet"*.[1]

A *Service* is a discoverable software resource which has an advertised service description. The service description is available in a repository, called a registry that can be searched by a potential consumer of the service. Given the service description, a service consumer can bind to and use the service. In real world e-services are represented by web service which has different types (grid services, distributed services, etc).

Making advanced features and goals, companies uses a combination of several services to develop composite service. The composability feature of services leads to low-cost development of high quality services. "*The process of integrating existing Web services to achieve higher-level business tasks that cannot be fulfilled by any individual service alone is often referred to as Web service composition*"[2]

Based on SOA modeling on IBM [3] service composition is 4th step of SOA modeling. Service Identification, Service Specification, Service Realization are earlier steps and last one is Service Implementation. Whereas services are small software components, different software methodologies could be used for developing a composed service.

RUP and similar guides, like the PMBOK, Software Engineering Institute's (SEI's) Capability Maturity Model Integrated (CMMI), or the UK's IT Infrastructure Library (ITIL) standards impose unnecessary process overhead for smaller projects. In contrast agile methodologies allow for fast and tight increments or phases; cut down overhead; and ensure a close relationship between developer and customer. Comparing agile methodologies with other methodologies, demonstrates that agile methodologies is more useful for developing services. So in this paper we are trying to represent a new agile methodology for services; especially composed Services.

To explore which agile methods are the most effective for composite service development, we first investigate if any combinative use of agile methods (or, at least, employing more than one agile method) can have an effect on the four outcomes. Based on research that has been done in [4], using two agile methods is better in productivity rather than using only one method, and there appeared to be no significant further advantage in increasing the number of methods used beyond two. Also [4] expressed that in terms of both quality and productivity, there was a significant difference between the eXtreme Programming/Scrum combination and all the other pairs of methods. However there was no significant difference in either cost or satisfaction. This clearly tells us that the eXtreme Programming/Scrum combination is a good pairing of methods to adopt.

This result can be seen to make some sense in that, eXtreme Programming (XP) is very much oriented towards technology based practices and programmer activity. In contrast, Scrum is more focused on agile project management aspects (Abrahamsson et al.2002). In addition, Scrum is explicitly intended as a wrapper around other engineering approaches. Therefore XP and Scrum can be seen to be complementary from a practical point of view, supporting the claims made by Mar and Schwaber (2002).

Accordingly we use a combination of these pair methods to bring agility in our service composition model.Since changing development process can not implemented in a single shot, we defined two implementation levels. By defining separate levels in development process, preparation of implementation process would be simple like what

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

CMM[1] done. Also IBM defined a capability maturity model for services (SIMM) [5]. This model consists of 7 steps and companies should implement dynamic service composition to achieve 7th step. First level of AM4SC also is useful for describing our model strategy.

2. Agile Model for Service Composition (AM4SC)

AM4SC described by 4 activities which derived from Scrum methodology and 6 practices which comes from XP method. Mainly activities are management process and practices increases agility and development quality. The 4 activities are described as below:

1- Planning process: In this process functionality domain described and with TDD rules an abstract model for composed service would be designed. As a result expected inputs and outputs are represented. Obviously according to agile manifesto, proposed model would not be a complete model but it is started with a simple model and will be finalized with iteration and incremental in later evolves.

2-Arrangement process: After planning process, features list that should be included in next iteration prepared. This process will be done at a periodical session and as a output of services and features that should be included in next iteration (something like backlog in scrum) reported. Some of the services listed above, should be refactor to fulfill the composed service rules.

3-Modeling process: As in the previous steps, expected input, expected output and requirement list determined, modeling could be done with any available service modeling tools; In the simplest case, by using pen and paper. The test process can be started from here. Across to the input, output and designed model in first step, tests can be introduced before implementing composed service. Each service will be considered as a block and their outputs are determined in each specific input.

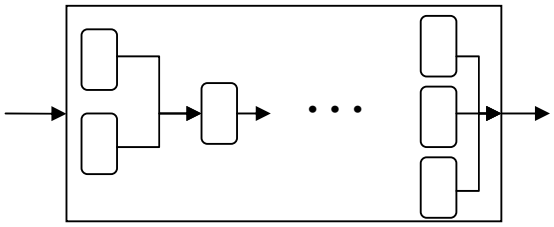


Fig. 1 *composite service modeling*

4- Implementation Process: At this step the platform of independent model for composite service that offered in previous step mapped to platform of specific model. The result is a composite service that provides the desired functionality. In the last, this result released and delivered as iteration to customers.

This life cycle will be continued till getting customers feedback and their new requirements, then composite service incrementally completed in next iterations.

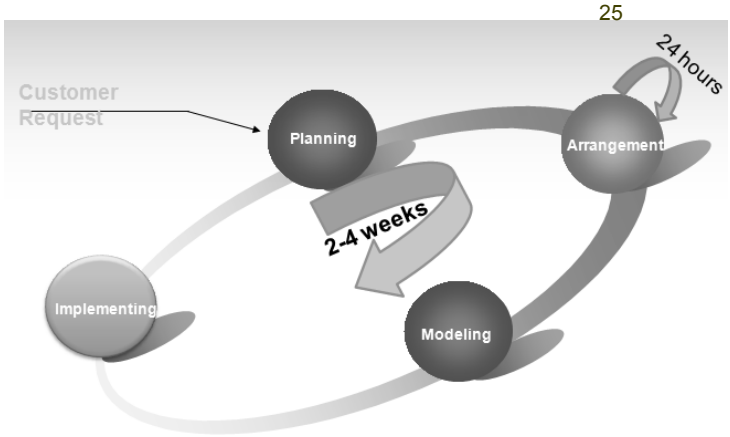


Fig 2 AM4SC Activities

Practices bring more agility to AM4SC. These practices are based on XP practices and make our model more adaptive to agile manifesto as bellow:

Collaborative programming: Developed services may implement in different platforms and with different technologies by several development teams. Although service description is available through UDDI and with Web Service Description Language (WSDL), but developers of services can share their experiences for modeling and implementing composite service in low cost at easiest way.

Iterative and Incremental development: Requirement should be categorized. So main features developed in first iteration then with refactoring other features would be inserted to composite service incrementally.

Online customers: composite service customers also called service requester. Their feedbacks will be used to increase quality and acceptance of developed service. Also distributing beta versions in web, delivers more feedbacks.

Continues integration: In each iteration, whole service model should be integrated to implement composite service, this makes our model more reliable.

Reusability through design: This is a fundamental agile practice that involves actively ensuring that a piece of functionality is implemented in a single place. We use these techniques to make sure that services which used in composite service, implement functionality once. Also same techniques could be used to brings more functionality and cover more requirements with less services. Design reviews are used to remove duplicate logic across existing services. In order to increase the potential of reusing the services, ensure that they do not have channel, technology platform or transport specific logic which is coupled with the service's domain logic.[5]

Service model refactoring: After each periodical release, service model should be changed to achieve new features and requirements while across these changes, logics of model sections must stay unchanged. This practice is useful for decreasing cost and time to apply changes that consumer expects. Therefore based on customer request which may received in WSBPEL or other format, a

[1] Capability Maturity Model

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

translator interfere it to a known service model and then model will be refactored to satisfy requested features.

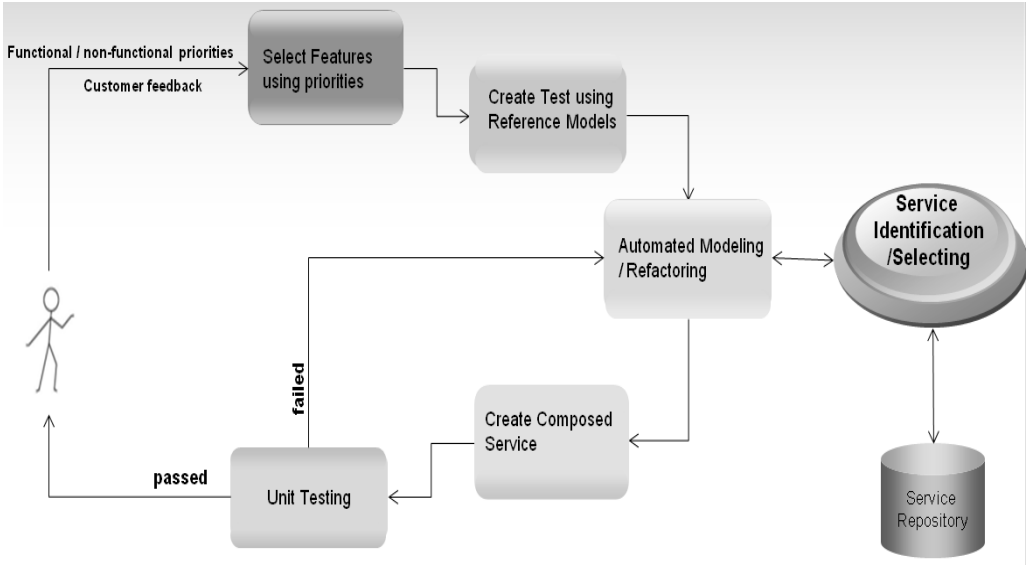


Fig3- collaboration of AM4SC agents

3. Second level of AM4SC

Today, service developers use SaaS for distributing their service in the web. To use this method through service development, the AM4SC method should use a dynamic method for developing composed services. For reaching this goal, we defined an automated model for AM4SC that not only have a same agility feature of first level but also brings capability of developing composed service in run time and on demand.

The execution of AM4SC starts with pair element of request and request priorities defines by service requester. After modeling and developing compose service, despite of unit test acceptance, a new version of service delivers to the customer. In AM4SC second level, six agents collaborate together to achieve model goals as bellow:

At first, an agent called feature selector, select features by pairs that service requester sent as an input. Additionally based on meet in the middle architecture, customer's requests should be categorized across the developer's policies and their priorities which provided by requester. Therefore in this agent, customer's request or feedback priorities with company policy variables would be used for selecting features that will be develop at the next iteration to decrease risk of development. Output of this agent, also is useful tool for accessing progress of development process.

Second agent which named unit test maker, searches among reference models for defined problem domain that described by service requester requirements. This agent acts as a TDD practice to increase quality of developed service. Unit tests, generated by this agent would be used as artifacts to describe requirements of the problem.

Then automated refactor/modeler agent, creates requests for binding required service based on selected features and artifacts that defined by unit test maker agent. The request should first translated to WSDL format then requests would be sent to service broker agent that search UDDI registry for requested services, delivers binding information of services to automated modeler agent. Then automated modeler use an automated modeling method for model composed service using bind information. The model simply can be converted to BPL or BPEL format and then creation process of service will be started y next agent.

Another task that would be done by this agent is model refactoring when unit tests fails or in next iteration of service development, cycles to add more features in composed service.

Refactoring is one of main principle of DRY practice that is responsible for locate opportunities for eliminating duplication among services. Therefore refactoring agent also can be used to increase reusability of composed service.

At the next step, an agent named Service Creator binds services and creates needed relationship designed in service model between input and outputs of services that described by BPEL.

Tester agent, Tests composed service with created unit tests and surveys the results. If Tests pass, the composed service delivers to service requester, otherwise if one or more tests failed, model with failed unit tests delivered to automated modeler for refactoring and changing model to satisfy all unit tests. As for unit tests are artifacts that describes features, if unit tests pass we can ensure that all planned features have been covered.

At the end of iteration, after designed unit tests passed, an integrated composed service would be delivered to service requester. Customer's feedback and recommendations could be used to improve productivity and usability of composed service in next iteration of developing composed service.

4. Comparison of AM4SC with other Service Composition Methods

In this part we are considering to compare AM4SC to other service composition methods from agility and main service composition features perspective.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Table 1 Comparison of AM4SC with other composition methods

Composition Types	Features / Methods	1- Service composition support	2- Meet in the middle support	3-Participation of stake holders and customers	4- Repetitive production	5-Categorizing requirements	6-Managing creation process	7-Fast release	8-Optimal usage of available resource	9- Reusability	10- Refactoring	11- Continues integration	12- Simple Design	13- flexibility	14- Periodical release	15- Dynamic Composition	16 - Availability of Techniques	17- Modeling capability	18- Quality Assurance
SOA Methodology	SOMA[6]	✓	✓																
	OASIS Methodology[7]			✓(2)	✓(1)	✓	✓												
	Microsoft Motion Methodology[8]		✓	✓	✓			✓	✓										
Agile methods in SOA	MSMUA(5)[5]	✓	✓		✓	✓(4)		✓		✓	✓	✓(3)	✓	✓					
	SRI-DM[9]	✓			✓						✓		✓	✓	✓				
Service Composition Methods	SCDC(6)[10]	✓														✓			
	Multi Agent model[11]	✓														✓	✓		
	Model Based[12]	✓															✓	✓	
	Ant colony based model[13]	✓														✓	✓		
Agile Methodologies	XP Scrum FDD AMDD DSDM ,…[14]		✓	✓	✓		✓	✓	✓		✓	✓	✓	✓	✓		✓		✓
Offered Method	AM4SC	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

[1] In first level once a year and in second level after fourth iteration
[2] In collaborative working practice
[3] Done-Done practice
[4] With user stories and their updates
[5] Modern SOA Methodology and SOA Adoption Using Agile Practices
[6] A Service Compatibility Model to Support Dynamic Cooperation of Cross-Enterprise Services

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Some of these methods (like SOMA, OASIS, etc) are not specially designed for service composition but covers service composition as a part of SOA modeling.

Table1 represents these comparisons. Some features (4, 6, 11, 9, 10, 12, and 19) cover productivity and reliability of composed service while some others are for decreasing risk of production. In next section we are describing that how these features are supported by AM4SC.

1- **Service composition:** In our comparison we take look at other SOA methodology to sure that we are supporting their features. Although some of these methods don't include composition process, but supports all other service development features.
2- **Meet in the middle methodology**: features that would be generated at next iteration should be selected by features selector agent based on user priorities and company policies to decrease risks.
3- **Participation of customers**: Customers feedbacks and priorities cause to low risk reliable software. AM4SC infrastructure is on demand development method that composes services to satisfy customer requirements. Customers also, send his request priorities to service development process that would be used for selecting next iteration features which should be developed.
4- **Repetitive production:** by periodical and repetitive production that described above, service stability and quality would be guaranteed.
5- **Categorizing requirement**: development would be started from simple model and completed in several iterations. So we will sure about main featured that should be developed and then consider about other features.
6- **Managing creation process**: As for AM4SC is derived from scrum methodology process will be managed in a good manner.
7, 12-**Fast release/simple design**: Development of composite service started with some base features, and final product would be completed incrementally. This kind of development caused to fast product delivery.
8- **Optimal usage of available resource:** One of the agile manifesto rules is optimal usage of human resources for developing software in low cost. These 6 agents are responsible for 5 activities that described above which are minimum required activities that should be used to cover both scrum and XP methodology.
10- **Reusability/Refactoring**: Refractor agent is responsible for redesigning of composite service model without logical change in service features as described before. As composed service, may use as a component of another composed service; so reusability of service through design should be considered. Based on [5], user stories and refactoring are two main practices that would be caused to increase reusability of service.
11- **Continues integration**: In each iteration, whole composite service integrated for delivering a fresh version to customer. Several periodical iteration, increase quality of service and reliability by decreasing side effects of new inserted features.
13- **Flexibility**: New feature should be added based on customer's feedback and requests in each iteration. So composed service is so flexible that could take a look at all features based described priorities.
14- **Periodical release:** in each period we have a fresh version of composed service and consumer can choose best one that fit his/her requirements.
15- **Dynamic composition support :** using agents that automatically perform service composition tasks, enables on demand composition and as a result, composed service can be described in run time.
16- **Availability of techniques:** Since all agents duties and practices derived from agile methods, these techniques are available and tested before.
17- **Modeling capability:** Modeling that will be done by automated modeler, increase readability for both customer and developers.
18- **Quality assurance:** Test driven development that would be done by unit test maker and unit tester agent is best practice that satisfies quality assurance of development

5. Conclusions

In this paper an agile method based XP and Scrum methodology used to brings more agility for service composition process. So service production process become faster and more flexible, additionally all other agile production features will be achieved.

References

[1] G Vidyanand Choudhary, "Software as a Service: Implications for Investment in Software Development" Proceedings of the 40th Annual Hawaii International Conference on System Sciences (HICSS'07) © 2007 IEEE
[2] W. Liu et al.: E-Service Composition Tools from a Lifecycle Perspective, Studies in Computational Intelligence (SCI) 37,125- 146, © Springer-Verlag Berlin Heidelberg 2007
[3] J.Amsden, "Modeling SOA: Part 4. Service composition" in http://www.ibm.com/developerworks/rational/library/07/1023_amsden/index.html

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

[4] Parsons, D., Ryu, H., and Lai, R., "The Impact of Methods and Techniques on Outcomes from Agile Software Development Projects" 2007, in IFIP International Federation for Information Processing, Volume 235, Organizational Dynamics of Technology-Based Innovation: Diversifying the Research Agenda, eds.McMaster, T., Wastell, D., Femeley, E., and DeGross, J. (Boston: Springer), pp. 235-249.
[5] V .Narayanan, "Modern SOA Methodology and SOA Adoption Using Agile Practices" Published: September 15, 2010 (SOA Magazine Issue XLIII: September 2010)
[6] A.Arsanjani, Service-oriented modeling and architecture How to identify, specify, and realize services for your SOA , IBM developer works, 2004
[7] U.Homann, J,Tobey, From Capabilities to Services: Moving from a Business Architecture to an IT Implementation, MICROSOFT CORPORATION, APRIL 2006
[8] S.Jones, M.Morris, A METHODOLOGY FOR SERVICE ARCHITECTURES, OASIS Draft, October 2005
[9] David E. Millard • Y.Howard • N.Abbas • Hugh C. Davis • Lester Gilbert •Gary B. Wills • Robert J. Walters, Pragmatic web service design: An agile approach with the service responsibility and interaction design method, © Springer-Verlag 2009
[10] Jing-fan Tang, 2006, A Service Composibality Model to Support Dynamic Cooperation of Cross-Enterprise Services, in IFIP International Federation for Information Processing, Volume 220, Information Technology for Balanced Manufactimng Systems, ed. Shen, W., (Boston; Springer)
[11] Minhong Wang , William K. Cheung2, Jiming Liu , Xiaofeng Xie and Zongwei Luo , S. Dustdar, J.L. Fiadeiro, and A. Sheth (Eds.), E-Service/Process Composition Through Multi-agent Constraint Management,© Springer-Verlag Berlin Heidelberg 2006
[12] V.Torres ,V.Pelechano, R.Quintero, Model Centric Approach of Web Services Composition, © 2007 Birkh¨auser Verlag Basel/Switzerland
[13] T.Ghafarian, M.Kahani, Semantic Web Service Composition Based on Ant Colony Optimization Method, 978-1-4244-4615-5/09, ©2009 IEEE
[14] D.Rosenberg, M.Stephens, M.Collins-Cope, Agile Development with ICONIX Process:People, Process, and Pragmatism, ISBN 1-59059-464-9, 2005

Pouya Fatehi is software designer in Iran Rayaneh Company and is responsible for service architecture of BPM project which names FilerPlus for 5 years.
He has a Masters degree in Software Engineering from Science and research Beranch of Islamic Azad university of Tehran. He is specialist in software as service paradigm and has a several research in the field of agile methodologies and SOA.

Seyyed Mohsen Hashemi received the M.S. degree in Computer Science from Amirkabir University of Technology (Tehran Polytechnic University) in 2003, and the PhD degree in Computer Science from the Azad University in 2009. Moreover, he is currently a faculty member at Science and Research Branch, Azad University, Tehran. His current research interests include Software Intensive Systems, E-X systems (E-Commerce, E-Government, E-Business, and so on), Global Village Services, Grid Computing, IBM SSME, Business Modeling, Agile Enterprise Architecting through ISRUP, and Globalization Governance through IT/IS Services..

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

A Survey on Steganography Techniques in Real Time Audio Signals and Evaluation

Abdulaleem Z. Al-Othmani[1], Azizah Abdul Manaf[2] and Akram M. Zeki[3]

[1] Advanced Informatics School (AIS), Universiti Teknologi Malaysia (UTM)
Kuala Lumpur, Malaysia

[2] Advanced Informatics School (AIS), Universiti Teknologi Malaysia (UTM)
Kuala Lumpur, Malaysia

[3] Kulliyyah Information and Communication Technology, International Islamic University Malaysia (IIUM)
Kuala Lumpur, Malaysia

Abstract

Steganography has proven to be one of the practical ways of securing data. It is a new kind of secret communication used mainly to hide secret data inside other innocent digital mediums. Most of existing steganographic techniques use digital multimedia files as cover mediums to hide secret data. Audio files and signals make appropriate mediums for steganography due to the high data transmission rate and the high level of redundancy. Hiding data in real time communication audio signals is not a simple mission. Steganography requirements as well as real time communication requirements are supposed to be met in order to construct a useful and useful data hiding application. In this paper we will survey the general principles of hiding secret information using audio technology, and provide an overview of current functions and techniques. These techniques will be evaluated across both, steganography and real time communication requirements.

Keywords: *Real time communication, data hiding, LSB, audio steganography, signal processing.*

1. Introduction

Steganography has a long history of been used as a way to protect security and privacy of valuable information. While cryptography focuses on protecting the secret message by jumbling its content, steganography concerns on protecting the secret message by concealing its mere existence. The concealment of secret messages is achieved by embedding them into other seemingly-innocent host mediums [1], [2], [3]. Figure 1 below illustrates the basic idea of any steganography process.

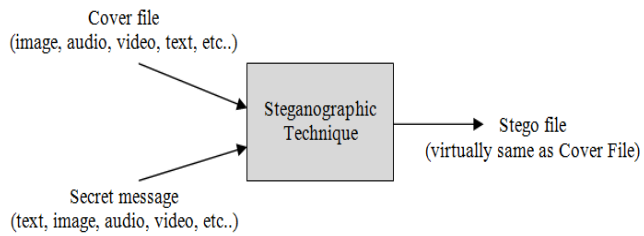


Fig. 1 Fundamental scheme of steganography process.

Steganographic application can hide different types of data within a cover medium. The resulting stego message contains the hidden information, even though it is seemingly identical to the cover medium. Steganography basically exploits human awareness and observation because human senses are not qualified to seek files that have information hidden inside them, while there are many third party programs can do what is called Steganalysis, which an art of inverse steganography aims to analyze and break a specific steganographic system. Normally, steganography is required where cryptography techniques are ineffective [4].

Generally, all digital mediums, signals, or files can be used in steganography process as cover media, but some formats are more suitable than others depending on the level of redundancy [5]. For instance, text steganography is believed to be the hardest type of steganography because of the low degree of redundancy in text as compared to image, audio or video. Redundancy can be described as the bits of a media, signal or file that offer accuracy more than needed for the object's use [6]. The redundant bits of an object may also be defined as those bits that can be easily altered without this change being

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

noticed easily [7]. Image, video and audio files in particular fulfill this requirement, while studies have also revealed other file formats that are suitable to be used for information hiding. Figure 2 illustrates the main categories of file formats or signals that can be efficiently used for steganography.

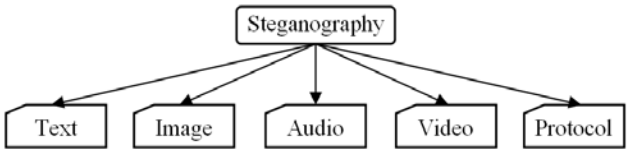


Fig. 2 Categories of steganographic cover mediums.

2. Structure of Steganography

Given the increased general attention over steganography technique and practices, some common terminology that most of the applications have in common have been discussed and determined [8]. The items agreed on are:

- **Emb (*m*):** Some information data or signal to be hidden, in other media.
- **Stego (*s*):** The output of the steganography process which is the signal, file or data that has the embedded message hidden in it.
- **Cover (*c*):** The input to the information hiding process which represents the innocent carrier signal or file.
- **Stegokey** or simply **key (*k*):** This is additional un-imbedded secret data which may be needed in the information hiding process. In particular, this key (or a related one) is typically needed to extract the embedded message again in the final destination.

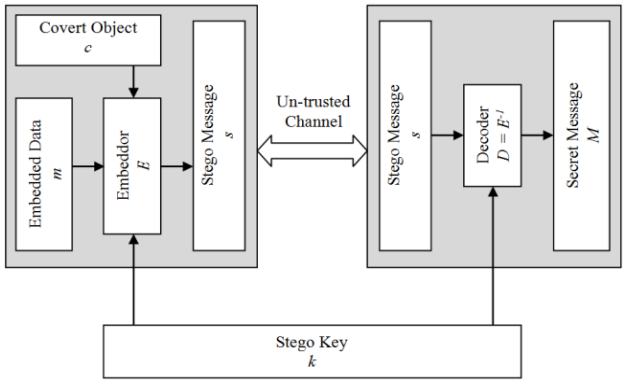


Fig. 3 Steganography terminology.

In general, steganographic system consists of embedding or encoding phase and extracting or decoding phase. As illustrated in Figure 3, the embedding process is accomplished by encoding or embedding the secret message into a covert innocent message using a stego key.

The result of this process is the stego message which contains both cover and secret messages combined according to the stego key. On the other hand, the decoding phase requires having the same stego key in order to be able to extract the embedded secret data from the stego message. In most steganography techniques, failing to have the stego key will make the process of extracting the secret message almost impossible.

3. Overview of Audio Steganography

Audio steganography is focused in hiding secret information in an innocent cover audio file or signal securely and robustly. Communication security and robustness are vital for transmitting important information to authorized entities while denying access to not permitted ones. By embedding secret information using an audio signal as a cover medium, the very existence of secret information is hidden away during communication. This is a serious and vital issue in some applications such as battlefield communications and banking transactions [9].

The secret message is concealed into the audio media by slightly changing the binary sequence of the audio file. Hiding secret information into digital audio media is generally more complicated than hiding secret information into other media, such as digital images. In order to hide secret information successfully, a range of techniques for inserting information into digital audio have been introduced. These techniques vary from simple ones that embed information as signal noises to more powerful ones that take advantage of complicated signal processing techniques to embed the secret message

3.1 Digital Audio Signal

Digital audio signals are different from other traditional analogue sounds in the fact that they are discrete signal rather than continuous ones. Discrete signals are produced by sampling continuous analogue signals at specific rates. For instance, the typical sampling rate for CD digital audio is 44 kHz. Figure 4 below, shows a continuous analog audio signal wave being sampled to create digital audio signal wave.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

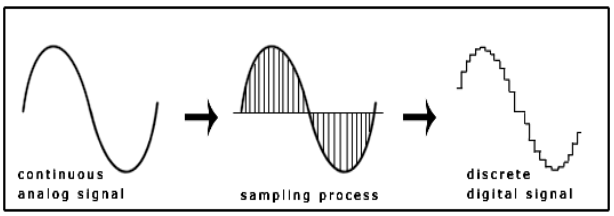


Fig. 4 Sampling of audio signal.

Figure 4 above emphasizes the discrete nature of digital audio signals. Nonetheless, typical sampling rate is generally set at a level in which the produced discrete signal is not imperceptibly distinguishable from the original continuous signal.

Digital audio files are stored in computers as a series of 0's and 1's. With a correct tool, it is possible to change the bits that structure a digital audio file. Such accurate controls permit changes to be performed to the binary bits that are not perceptible to the human sense

3.2 Methods of Audio Steganography

There are many steganographic techniques for hiding secret data or messages in audio in a way that the modifications made to the audio file are perceptually indiscernible. Several recent methods necessitate previous familiarity with signal processing techniques, Fourier transform, and other high level mathematics areas. Common techniques include [10] – [16]:

A) Least Significant Bit (LSB) Coding

Least significant bit (LSB) coding is the easiest and simplest method to hide secret data in a digital audio media. By replacing the least significant bit of each sample words with a bit of the secret data, LSB coding permits a big size of secret data to be embedded. LSB audio steganography techniques have the same previously discussed advantages and disadvantages of common LSB steganography techniques on other cover media.

In computing, the least significant bit (LSB) is the bit in the right most position of a binary number, which also determines whether the number is even or odd. It is equivalent to the least significant digit of a decimal number, which is the digit in the ones (right-most) position.

1	0	1	1	0	1	0	1

Fig. 5 Binary representation of decimal 181.

Figure 6 illustrates how the message “Hi” is encoded in a 16-bit quality audio sample using the LSB method. Here the secret information is “Hi” and the cover file is an audio file. “Hi” is to be embedded inside the audio file. First the secret information “Hi” and the audio file are converted into bit stream. The least significant column of the audio file is replaced by the bit stream of secret information “Hi”. The resulting file after embedding secret information “Hi” is called Stego-file.

Audio stream sample (16-bits)	“Hi” in binary	Stego audio Stream (w embedded message)
1 1 0 1 1 1 0 1 1 1 0 0 1 0 0 1	0	1 1 0 1 1 1 0 1 1 1 0 0 1 0 0 0
0 0 0 1 1 0 0 0 0 1 1 0 0 1 1 0	1	0 0 0 1 1 0 0 0 0 1 1 0 0 1 1 1
1 1 1 0 0 1 0 1 1 1 0 1 1 0 1 0	0	1 1 1 0 0 1 0 1 1 1 0 1 1 0 1 0
0 0 0 1 1 0 0 0 0 1 1 0 0 0 0 0	0	0 0 0 1 1 0 0 0 0 1 1 0 0 0 0 0
1 1 1 0 0 0 0 1 1 1 0 1 0 1 1 0	1	1 1 1 0 0 0 0 1 1 1 0 1 0 1 1 1
0 0 0 0 1 0 1 1 0 0 1 0 0 0 0 0	0	0 0 0 0 1 0 1 1 0 0 1 0 0 0 0 0
1 1 1 1 1 0 0 0 1 1 0 0 0 1 1 1	0	1 1 1 1 1 0 0 0 1 1 0 0 0 1 1 0
0 1 0 0 1 1 1 1 0 1 0 1 1 0 1 0	0	0 1 0 0 1 1 1 1 0 1 0 1 1 0 1 0
0 1 0 0 0 0 0 0 0 1 1 0 0 0 1 1	0	0 1 0 0 0 0 0 0 0 1 1 0 0 0 1 0
0 0 1 1 1 0 1 1 0 1 0 0 1 1 1 0	1	0 0 1 1 1 0 1 1 0 1 0 0 1 1 1 1
0 1 1 0 0 0 0 0 0 0 1 1 0 0 1 0	1	0 1 1 0 0 0 0 0 0 0 1 1 0 0 1 1
1 0 0 0 1 1 0 1 0 1 0 1 1 1 0 0	0	1 0 0 0 1 1 0 1 0 1 0 1 1 1 0 0
0 1 1 0 0 0 1 0 1 0 1 0 0 0 1 0	1	0 1 1 0 0 0 1 0 1 0 1 0 0 0 1 1
1 1 0 0 1 0 0 0 0 1 0 0 0 0 0 0	0	1 1 0 0 1 0 0 0 0 1 0 0 0 0 0 0
0 0 0 0 0 0 1 0 1 1 1 1 1 0 1 1	0	0 0 0 0 0 0 1 0 1 1 1 1 1 0 1 0
1 1 0 1 1 1 0 0 1 1 0 0 0 1 0 1	1	1 1 0 1 1 1 0 0 1 1 0 0 0 1 0 1

Fig. 6 LSB audio coding example.

In LSB coding, the idyllic secret data communication rate is 1 kbps for 1 kHz. In some variations of LSB coding, however, the two least significant bits of a sample word are use to hide two bits secret message data. This increases the capacity of embedded data but also increases the nose and therefore increases risk of being perceptible and eventually breakable.

Using LSB is possible, as modifications will typically not create perceptible changes to the sounds. Another method involves taking advantage of human sound system limitations. It is possible to encode messages using frequencies that are inaudible to the human ear. Using any frequencies above 20.000 Hz, messages can be hidden inside sound files and will not be detected by human checks [14].

B) Parity Coding

In parity coding, audio signal is broken down into separate areas of samples and hide the secret message in the parity bit of each sample area. If the parity bit of a sample area does not match the secret message bit to be embedded, the LSB of one of the samples in the area is inverted. Therefore, this will give a wider range of choices on where to hide the secret bit, and will keep the change in the signal more unobservable [17].

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

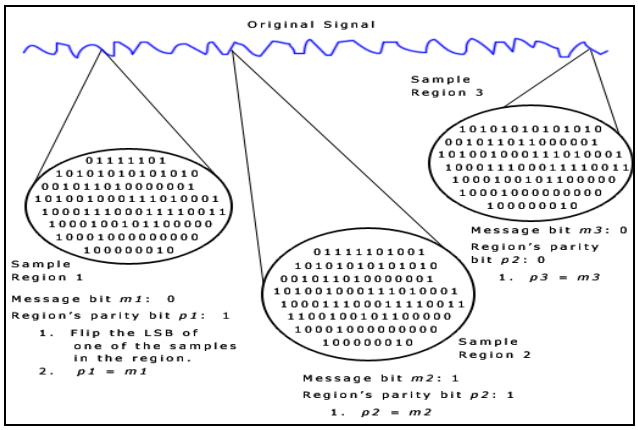


Fig. 7 Parity coding procedure.

C) Phase Coding

Phase coding deals with the weaknesses of the previously discussed audio steganography techniques which induce noises to the medium. Phase coding is based on the reality that, unlike noises, audio phase components are imperceptible to the human ear. Rather than adding noises, this technique encodes the secret data bits to phase shifts in the phase spectrum of the audio signal, attaining inaudible encodings in terms of signal-to-noise ratio [14].

$$phase_{new} = \begin{Bmatrix} \frac{\pi}{2} & if\ message & bit = 0 \\ -\frac{\pi}{2} & if\ message & bit = 1 \end{Bmatrix} \quad (1)$$

In phase coding, the phase of an initial audio segment is substituted with a reference phase that represents the data. Following segments phase is modified back to maintain the relative phase between segments. Phase coding, when applicable, is one of the most efficient audio steganographic methods in terms of the signal to noise ratio (SNR). When the phase relation between each frequency component is dramatically changed, noticeable phase dispersion will occur. On the other hand, on condition that the alteration of the phase is small enough, an inaudible steganography can be accomplished [14], [15].

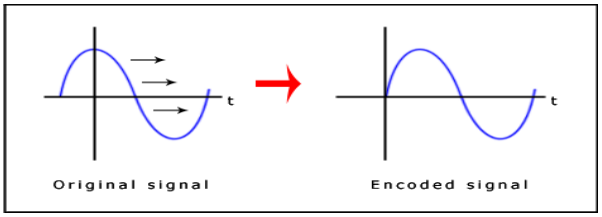


Fig. 8 The signals before and after Phase coding procedure.

D) Spread Spectrum

In the field of audio steganography, fundamental spread spectrum (SS) techniques attempts to distribute secret data throughout the frequency spectrum of the audio signal to the maximum possible level. This is equivalent to implementing LSB coding by spreading the secret data bits over the entire audio signal. However, different from LSB coding, the SS techniques spread the secret bits over the frequency spectrum of the audio media by using a code that is not reliant on the genuine signal. Consequently, the resultant signal will utilize a bandwidth wider than what is essentially needed for communication [15].

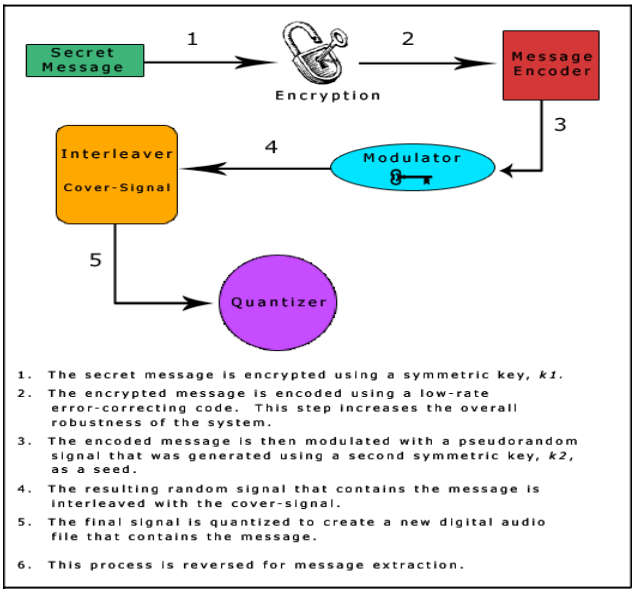


Fig. 9 Spread spectrum.

Two types of spread spectrum are utilizable in audio steganography: the direct sequence and frequency hopping schemes. In direct sequence spread spectrum, the secret data is distributed using a constant named the chip rate then adapted with a pseudorandom signal and then interleave with the cover signal. In frequency hopping spread spectrum, the frequency spectrum of the audio medium is changed so that it hops quickly among frequencies.

E) Echo Hiding

In echo hiding techniques, secret data is inserted into an audio medium by introducing an echo into the discrete signal. Similar to SS technique, it also offers benefits as it allows high data communication rates and offers greater robustness compared to the earlier noise-inducing techniques.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

In order to hide secret message effectively, three echo-related factors are involved and changed: amplitude, decay rate, and offset (delay time) from the genuine audio signal. All of those factors should be set lower than the human hearing threshold in order to keep the echo imperceptible. Additionally, offset values are changed corresponding to the binary secret data targeted. A specific offset value represents a binary one, and another offset value represents a binary zero. Figure 10 illustrates the echo hiding process

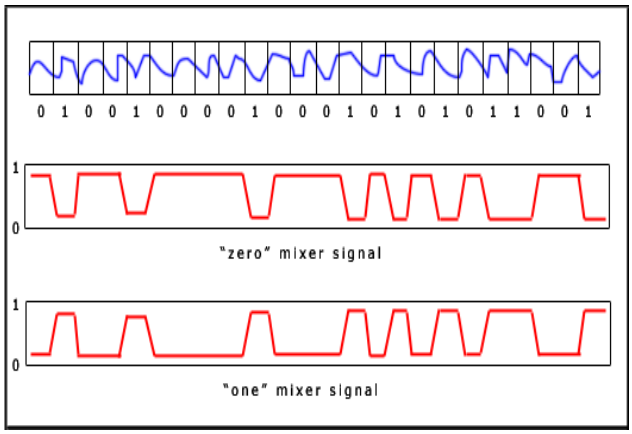


Fig. 10 Echo hiding.

3.3 Steganography in Real Time Audio Signals

Real-time communications (RTC) is any form of telecommunications in which information can be exchanged among the users instantly or with insignificant latency. RTC may occur in half-duplex or full-duplex approaches. Data can be transmitted in both directions on a single carrier simultaneously in full-duplex mode, while in half-duplex RTC data can be transmitted in both directions on a single carrier but not in the same time. Any type of communication, especially digital real-time one, needs to have one or more protocols to control and standardize the communication process among the clients. Generally, theses protocols used to rule the segmentation, framing, sampling, transmission, traffic controlling, receiving, and other main protocol tasks. The protocols used to send the sampled data using packets. Packet payloads are basically encoded multimedia data and they may contain any type of multimedia data.

Usually, protocol packets have some extra unused bits mainly for future use purposes as well as for some special situations. It is very useful and practical to hide secret information into these redundant bits allocated in the structure of every packet sent. This provides the ability to modify these bits to hide data without any perceptible change in the encoded multimedia contents. The sample word size, used by different multimedia encoding formats is an important aspect in finding out the maximum amount of available space in the cover medium for hiding the secret information. In general, least significant bits of each word value are probably usable to be modified with no perceptible change in the quality of the multimedia content. Thus, as an example, only half the amount of available space in a16-bit encoded audio cover medium will be available in comparison with a cover medium with an 8-bit word size.

The throughput of the real time communication system is another important factor used to define the performance of any proposed RTC steganography technique. As an example, utilizing the LSB of every sample in some kind of data compression protocol which has the packet size of 160 bytes, a suggested total of 20 bytes of secret data can be successfully embedded. If the throughput of this system is around 50 packets per second unidirectional, this results in approximately 1,000 bytes of full-duplex throughput of secret data within this covert communication channel

4. Test and Evaluation

All steganography techniques have to fulfill a few specific and essential requirements. A set of criteria has been proposed to further describe the quality of a steganography algorithm as illustrated in Figure 11 and discussed below.

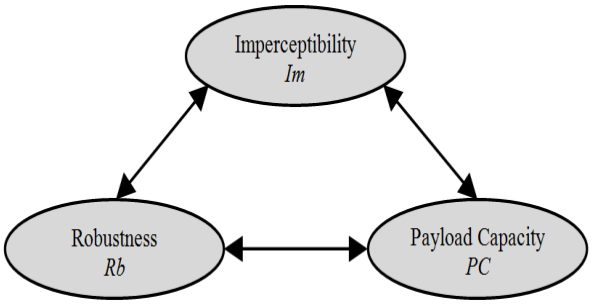


Fig. 11 Steganography requirements.

In this evaluation, since there are a number of sub layers, we think that it is better to measure each requirement separately.

4.1 Imperceptibility (*Im*)

The imperceptibility is the most important requirement of a steganography system, as the strength of steganography system depends on its ability to be unnoticed by the human senses (visually or acoustically). If it is noticed that

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

the cover medium has been altered, the steganography technique or system is not practical anymore [19]. The steganography system is said to be accurately imperceptible only if it is impossible to distinguish the covert data from the hidden secret information. It is sometimes enough that the alterations in the stego message be unobserved, on condition that the embedded message is not compared with the original covert message [20], [21].

4.2 Robustness (*Rb*)

Robustness defines how strong the used steganographic technique against changes. It measures the capability of the embedded secret data to endure different types of intentional and unintentional modifications. The hidden embedded message should be very hard to eliminate or modify without altering the quality of the covert medium [22], [23].

It is important for steganography techniques to be robust against both intentional and unintentional modifications to the message [19]. Many steganography techniques leave some kind of 'signature' when embedding the secret information. Typically, these signature data can be easily identified using statistical analysis. It is necessary that a steganography algorithm does not leave such a mark in the covert medium in order to be able to avoid and pass by such statistically analysis without being detected. In the communication of a stego message by reliable systems, the message may go through manipulations in an attempt to remove potential hidden information. Data manipulation, such as compression or rotating, might be applied to the message prior reaching its final destination. Depending on the manner in which the message is embedded, these manipulations may or may not destroy the hidden message depending on the method or technique used to hide the secret message

4.3 Payload Capacity (*PC*)

Payload capacity is the size of embedded data that can be hidden into a particular innocent cover medium relative to the size of this medium. The real challenge is how to hide as much secret data as possible while keeping the quality of the medium untouched and without infringe the imperceptibility requirement. Generally, increasing the embedding capacity makes the secret hidden information more conspicuous in viewing. To calculate the embedding capacity of a particular steganography system, the size of the embedded secret message is divided by the total size of the cover medium as shown in Equation 2.

$$Payload\ Capacity = \frac{Total\ number\ of\ bits\ of\ hidden\ data}{Total\ number\ of\ bits\ of\ cover\ file} \quad (2)$$

4.4 Real Time Suitability (*RTS*)

Steganography in real time audio signals involves additional requirements such as system complexity, throughput, bandwidth, delay, absence of duplications, failure recovery, and service setup time. These requirements directly affect the real time communication process, and hence, may have influences on the real time steganography processes [24], [25].

5. Discussion

Table 1 tabularizes the comparison of audio signal steganographic techniques based on the proposed evaluation criteria. The levels of technique fulfillment to the requirements have been defined as high, medium or low. A high indicates that the technique totally meets the requirement, while a low level indicates that the technique has some limitations or weaknesses relating to this requirement. A medium level indicates that the requirement is dependent on external factors [7], [19].

Table 1: Evaluation of steganography techniques in real time audio signals

Technique	***PC***	***Im***	***Rb***	***RTS***
LSB	High	Medium	Low	High
Parity Coding	Medium	Medium	Low	Low: Delay
Phase Coding	Low	High	High	Medium
Spread Spectrum	High	Low	High	Low: Bandwidth
Echo Hiding	High	Low	Medium	Medium

There are two major drawbacks related to the use of methods like parity coding. The human ear is sensitive and can often detect even the slightest bit of noise introduced into a sound signal, although the parity coding method almost makes the generated noise impossible to hear. Another drawback of using parity coding is the low robustness. Phase coding has the disadvantage of a low data transmission rate because the secret message is embedded in the first signal segment only. That is what makes phase coding method practical only when a small quantity of secret data needs to be transmitted. 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.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Regarding real time communication requirements, only LSB technique shows a high level of suitability in terms of delay, throughput, complexity, and reliability requirements of RTC. Parity coding technique requires the signal to be broken down into segments to apply the parity hiding on these segments and then transmits these segments to the destination. This introduces delays on the transmission process which severely affects the real time communication process. Spread spectrum coding expands the frequency bandwidth of the audio signal which in turn might not be available by the communication channel provider. Phase coding and echo hiding techniques have shown a high level of compatibility with real time requirements except some negative aspects related to system complexity, throughputs, and service setup time.

6. Conclusion

In this paper, several techniques are discussed as potential methods for embedding data in real time audio signals. While a degree of success has been achieved, each one of the proposed methods has its limitations. The ultimate goal of attaining protection of large amounts of secret data against deliberate attempts at removal may be still far from being obtained. The five techniques discussed above offer numerous choices and make this data hiding technology more obtainable and accessible. Prioritizing the importance of communication and security characteristics such as data rate, bandwidth, robustness, and noise audibility, must be done before choosing the steganographic technique which should completely fits the nature, environment and requirements of the application.

Although some data hiding techniques have been proposed by various researchers, the specific requirements of each data hiding technique vary from one application to another; with each of these techniques have some advantages and disadvantages. The flexible nature of audio formats, signals and files, is what makes them good and practical medium for steganography. Another aspect of audio steganography that makes it so attractive and promising is the ability to combine steganography techniques with existing cryptography technologies. We do not have to depend on one technique only. Secret data not only can be encrypted, they can be hidden and encrypted at the same time

References

[1] Mazurczyk, W. and K. Szczypiorski, "Covert Channels in SIP for VoIP Signalling", in Global E-Security, H. Jahankhani, K. Revett, and D. Palmer-Brown, Editors. 2008, Springer Berlin Heidelberg. p. 65-72.

[2] Hui, T., et al. "An M-Sequence Based Steganography Model for Voice over IP". in Communications, 2009. ICC '09. IEEE International Conference on. 2009.

[3] Tian, H., et al., "A Covert Communication Model Based on Least Significant Bits Steganography in Voice over IP", in Proceedings of the 2008 The 9th International Conference for Young Computer Scientists. 2008, IEEE Computer Society. p. 647-652.

[4] Nutzinger, M., C. Fabian, and M. Marschalek. Secure Hybrid Spread Spectrum System for Steganography in Auditive Media. in Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2010 Sixth International Conference on. 2010.

[5] Petitcolas, F.A.P., R.J. Anderson, and M.G. Kuhn, Information hiding-a survey. Proceedings of the IEEE, 1999. 87(7): p. 1062-1078.

[6] Currie, D.L. and C.E. Irvine, 1996. Surmounting the effects of lossy compression on steganography. Proceedings of the 19th National Information Systems Security Conference, Oct. 22-25, Baltimore, Maryland, pp: 194-201

[7] Anderson, R.J. and F.A.P. Petitcolas, On the limits of steganography. Selected Areas in Communications, IEEE Journal on, 1998. 16(4): p. 474-481.

[8] Pfitzmann, B. (1996) (collected by): Information Hiding Terminology – Results of an informal plenary meeting and additional proposals; Information Hiding, LNCS 1174, Springer-Verlag, Berlin 1996, 347-350

[9] Gopalan, K. and S. Wenndt, Audio Steganography for Covert Data Transmission by imperceptible Tone Insertion. in Proc. the IASTED International Conference on Communication Systems and Application (CSA 2004), Banff, Canada, July 8-10, 2004, track 422-025.

[10] Shahreza, S.S. and M.T.M. Shalmani. High capacity error free wavelet Domain Speech Steganography. in Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on. 2008.

[11] Johnson, N.F. and S. Jajodia, Exploring steganography: Seeing the unseen. Computer, 1998. 31(2): p. 26-34.

[12] Bhattacharyya, D., et al., Hiding Data in Audio Signal. Advanced Communication and Networking, C.-C. Chang, et al., Editors. 2010, Springer Berlin Heidelberg. p. 23-29.

[13] Chungyi, W. and W. Quincy. "Information Hiding in Real-Time VoIP Streams". in Multimedia, 2007. ISM 2007. Ninth IEEE International Symposium on. 2007.

[14] Kumar S. B., D. Bhattacharyya, P. Das, D. Ganguly and S. Mukherjee, “A tutorial review on Steganography”, International Conference on Contemporary Computing (IC3-2008), Noida, India, August 7-9, 2008, pp. 105-114.

[15] Bender, W., W. Butera, D. Gruhl, R. Hwang, F. J. Paiz, S. Pogreb, “Techniques for data hiding”, IBM Systems Journal, Volume 39, Issue 3-4, July 2000, pp. 547 – 568.

[16] Vapnik, V.N. “Statistical Learning Theory”. John Wiley and Sons, New York, USA, 1998.

[17] Dutta, P. Bhattacharyya, D. and Kim, T. (2009). Data Hiding in Audio Signal: A Review. International Journal of Database Theory and Application Vol. 2, No. 2, June 2009

[18] Kahn, D. “The History of Steganography”, Proc. of First Int. Workshop on Information Hiding, Cambridge,UK, May30-June1 1996, Lecture notes in Computer Science, Vol.1174, Ross Anderson (Ed.), pp.1-7

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

[19] Morkel, T., J.H.P. Eloff, and M.S. Olivier. An Overview of Image Steganography. In Proceedings of the ISSA 2005 New Knowledge Today Conference, Information Security South Africa (ISSA), pp 1-11 (CD).
[20] Setyawan, I. "Geometric Distortion in Image and Video Watermarking, Robustness and Perceptual Quality Impact". Ph.D. Thesis. University of Technology, Department of Electrical Engineering, Indonesia, 2004.
[21] Langelaar, G.C., I. Setyawan, and R.L. Lagendijk, Watermarking digital image and video data. A state-of-the-art overview. Signal Processing Magazine, IEEE, 2000. 17(5): p. 20-46.
[22] Chen, W.Y., A Comparative Study of Information Hiding Scheme Using Amplitude, Frequency and phase Embedding, PhD thesis. National Cheng Kung University, Tainan, Taiwan, 2003.
[23] Voyatzis, G. and I. Pitas, Protecting digital image copyrights: a framework. Computer Graphics and Applications, IEEE, 1999. 19(1): p. 18-24.
[24] Ferrari, D., Client requirements for real-time communication services. Communications Magazine, IEEE, 1990. 28(11): p. 65-72.
[25] Gaitonde, S.S., D. Jacobson, and A.V. Pohm, Bounding delay on a multifarious token ring network. Commun. ACM, 1990. 33(1): p. 20-28.

Abdulaleem Z. Al-Othmani is a Ph.D. student in Computer Science at University Technology Malaysia (UTM), KL. He received his Bachelor of Science in Computer Engineering in 2002 at Baghdad University, Iraq and a Master of Science in Computer Science in 2010 from the University of Technology Malaysia, Malaysia. In 2010, he started his Ph.D. in Computer Science at the Faculty of Computer Science and Information Systems, UTM. His research interests include audio signal processing, data hiding, cryptography, mobile steganography, and real time communication.

Azizah Abdul Manaf is a professor of image processing and Pattern Recognition from University Technology Malaysia (UTM). She graduated with B.Eng (Electrical) 1980, MSc. Computer Science (1985) and PhD (Image Processing) in 1995 from UTM. Her current areas of interest and research are image processing, watermarking, steganography and computer forensics and she has postgraduate students at the Master and PhD level to assist her in these research areas. She has written numerous articles in journals and presented an extensive amount of papers at national and international conferences on her research areas. Prof. Dr. Azizah has also held management positions at the university and faculty levels such as Head of Department, Deputy Dean, Deputy Director, and Academic Director. She is currently the Deputy Dean – Academic of Advance Informatics School (AIS) at UTM Second Author biography appears here. Degrees achieved followed by current employment are listed, plus any major academic achievements. Do not specify email address here.

Akram M. Zeki has obtained B.Sc. from University of Jordan, Amman, Jordan. And Master in Computer Graphics from Faculty of Computer Science and Information Technology at University Putra Malaysia. His PhD was from Faculty of Computer Science and Information System at University Technology Malaysia. Recently he is Assistant Professor at Kulliyyah of Information and Communication Technology, International Islamic University Malaysia. His research interest including: Watermarking, Steganography, Information Security and Image Processing.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

The Adaptability of Conventional Data Mining Algorithms through Intelligent Mobile Agents in Modern Distributed Systems

Dost Muhammad Khan[1], Nawaz Mohamudally[2]
[1]*Assistant Professor, Department of Computer Science & IT, The Islamia University of Bahawalpur, Pakistan & PhD student, School of Innovative Technologies & Engineering, University of Technology, Mauritius*

[2]*Associate Professor, & Consultancy & Technology Transfer Centre, Manager, University of Technology, Mauritius (UTM)*

Abstract

Intelligent mobile agents are today accepted as powerful tools for data mining in a distributed environment. The use of data mining algorithms further beefs up the intelligence in software agents. Knowledge discovery and data mining algorithms are applied to discover hidden patterns and relations in complex datasets using intelligent agents. The distributed computing provides remote communication, fault tolerance, high availability and remote information access. The uses of intelligent mobile agents using data mining algorithms make distributing computing more popular and acceptable. This paper is about the use of conventional data mining algorithms through intelligent mobile agents in distributed systems.

Keywords: *Data Mining, Distributed Systems, Multiagent System (MAS)*

1. Introduction

The ever growing amount of data that are stored in distributed form over networks of heterogeneous and autonomous sources poses several problems such as network bandwidth, communication, autonomy preservation, scalability, data buffering and privacy protection. The techniques of mining information and knowledge from huge data sources such as weather databases, financial databases or emerging diseases information systems were originally developed for handling the distributed cases. The increasing demand to extend data mining technology to datasets inherently distributed among a large number of autonomous and heterogeonous sources over a network with limited bandwidth has motivated the development of new approaches in distributed systems and knowledge discovery. A few of these approaches make use of intelligent mobile agents. Due to the adaptive and deliberative reasoning features of intelligent agents, they are well suited to cope up with the problems of distributed systems. An intelligent, learning and autonomous agent is capable of capturing and applying domain specific knowledge, learning, information and reasoning, to take actions in pursuit of a goal. The data is available every where now. The issue is to produce knowledge from this mountain of data. This is an era of 'knowledge' and 'knowledge' can be obtained with the combination of 'data' and 'information' by using data mining techniques. Data mining is successful due to it comes up with the precise formulation of problem solving and using the right data [17][18][19][20][21]. This paper is about the adaptability of commonly used data mining algorithms through multiagent system in distributed system.

The rest of the paper is organized as: Section 2 is the related literature, section 3 discusses the methodology, section 4 is about results and discussion and finally the conclusion is drawn in section 5.

2. Related Literature

In this section we discuss the most commonly used data mining algorithms and modern distributed computing systems, their issues and applications.

2.1. Data Mining Algorithms

The users may think that only by loading the computer software; data mining will happen. Before moving forward with data mining many issues must be considered such as data pre-processing. There is often a misconception that data mining is a data warehousing, SQL queries and reporting, software agents and online analytical processing (OLAP). The answer is; these are not data mining. Data mining in fact increases computing power, improves data collection and management and it has statistical and learning algorithms. It is clear that decisions are not made by data mining; the people have to decide with their knowledge and experience. The main properties of data mining algorithms are robustness, scalability and efficiency. Data mining is now used in bioinformatics, genetics, medicine, education & electrical power. The major objectives of data mining are to verify the hypothesis prepared by the user and to discover or uncover new patterns. Classification, Clustering, Regression model and Association rule Learning are the main areas of data mining [15][16]. Figure 1 shows the data mining algorithms.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

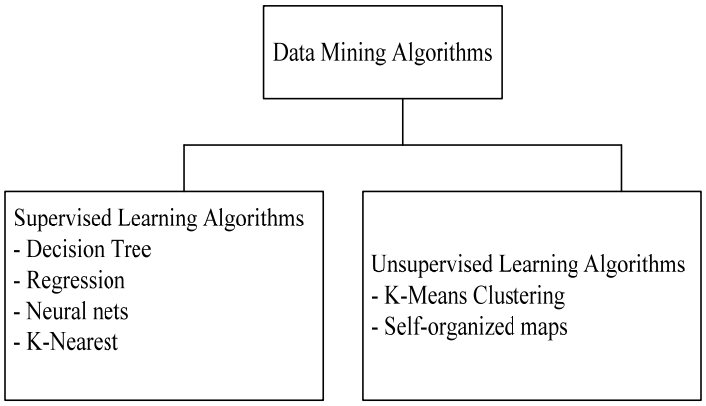


Figure1. Data Mining Algorithms

2.1.1. Neural Networks

Neural networks are used in system performing image and signal processing, pattern recognition, robotics, automatic navigation, prediction and forecasting and simulations. Figure 2 depicts Neural Networks with a hidden layer.

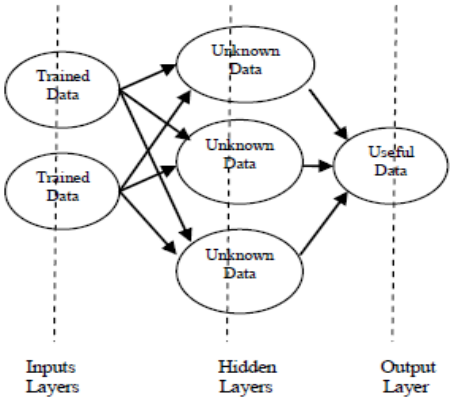


Figure 2. Neural Network with one hidden layer

The issues in NNs are: The process it goes through is considered to be hidden and therefore left unexplained. This lack of explicitness may lead to less confidence in the results and a lack of willingness to apply those results from data mining, since there is no understanging of how the results came about. It is abovious, as the number of variables of a dataset increases, it will become more difficult to understand how the NNs come to it conclusion.The algorithm is better suited to learning on small to medium sized datasets as it becomes too time inefficient on large sized datasets [1][2][3].

2.1.2. C4.5

C4.5 (Decision tree) is used to produce classifiers, which will predict with the values of its available input attributes and the class for an entity. The classification is the process of dividing the samples into pre-defined groups. It is used for decision rules as an output. In order to do mining with the decision trees, the attributes should have continuous discrete values, the target attribute values must be provided in advance and the data must be sufficient so that the prediction of the results will be possible. Decision trees are faster to use, easier to generate understandable rules and simpler to explain. They also help to form an accurate, balanced picture of the risks and rewards that can result from a particular choice. The decision rules are obtained in the form of "if-then-else", which can be used for the decision support systems, classification and prediction. Figure 3 illustrates how decision rules are obtained from decision tree algorithm.

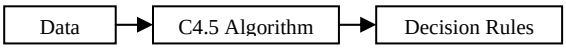


Figure 3. Function of C4.5 Algorithm

The issues in C.5 are: It is good for small problems but quickly becomes cumbersome and hard to read for intermediate-sized problems. Special software is required to draw that tree. If there is a noise in the learning set, it will fail to find a tree. The data must be interval or categorical. Any data not in this format will have to be recorded to this format. This process could hide relationships. Over fitting, large set of possible hypotheses, pruning of the tree is required. C4.5 generally represents a finite number of classes or possibilities. It is difficult for decision makers to quantify a finite amount of variables. This sometimes affects the accuracy of the output, hence misleading answer. If the list of variables increases the if-then statements created can become more complex. This method is not useful for all types of data mining, such as time series [1][2][3].

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

2.1.3. K-means Clustering

Unsupervised Learning depends on input data only and makes no demands on knowing the solution. It is used to recognize the similarities between inputs or to identify the features in the input data. It is used for finding the similar patterns due to its simplicity and fast execution. It starts with a random, initial partition and keeps re-assigning the samples to clusters, based on the similarity between samples and clusters, until a convergence criterion is met. The basic working of all the clustering algorithms is represented in figure 4.

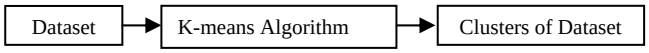


Figure 4. Function of the K-means Algorithm

The issues in K-means are: The algorithm is only applicable to datasets where the notion of the mean is defined. Thus, it is difficult to apply to categorical data sets. There is, however, a variation of the *k*-means algorithm called *k*-modes, which clusters categorical data. The algorithm uses the mode instead of the mean as the centroid. The user needs to specify the number of clusters *k* in advance. In practice, several *k* values are tried and the one that gives the most desirable result is selected. The algorithm is sensitive to outliers. Outliers are data points that are very far away from other data points. Outliers could be errors in the data recording or some special data points with very different values. The algorithm is sensitive to initial seeds, which are the initially selected centroids. Different initial seeds may result in different clusters. Thus, if the sum of squared error is used as the stopping criterion, the algorithm only achieves local optimal. The global optimal is computationally infeasible for large data sets. The algorithm is not suitable for discovering clusters that are not hyper-ellipsoids or hyper-spheres [1][2][3].

2.1.4. Genetic Algorithms (GAs)

They are used in optimization problems, selection, crossover and mutation. They are called natural selection and evolution of the problem. The issue in GAs is: It is not used in the large scale problems. It requires a significant computational effort with respect to other methods with parallel processing is not employed [1][2][3].

2.1.5. Fuzzy Logic (FL)

In contras to binary, multi valued logic to deal with imprecise or vague data. It is a new field not very widely used. FL is a fad. There is no guarantee that it will work under all circumstances. The issues in FL are: The language barrier is the major problem. In Japanese term it is ‘clever’ but in American it is ‘fuzzy’. This technology is still somewhat under developed in United States. It seems as if many American researchers have shunned it. It is not yet popular in data mining so it may not get the status due to its name [1][2][3].

2.1.6. Data Visualization

This method provides the user better understanding of data. Graphics and visualization tools better illustrate the relationship among data and their importance in data analysis cannot be overemphasized. The distributions of values can be displayed by using histograms or box plots. 2D or 3D scattered graphs can also be used. [3]Visualization works because it provides the broader information as opposed to text or numbers. The missing and exceptional values from data, the relationships and patterns within the data are easier to identify when graphically displayed. It allows the user to easily focus and see the patterns and trends amongst data. The major issues in data visualization are: As the volume of the data increases it becomes difficult to distinguish patterns from data sets. Another problem in using visualization is displaying multi-dimensional or multi-variable models because only two-dimensions can be shown on a computer or paper [1][2][3].

2.1.7. K-Nearest Neighbor (K-NN)

It is a classification technique that considers the solutions of similar problems that have been solved previously. It decides where to place a new case “k” after examining the most similar cases of “k” or its neighbors. For example “N” is a new case; it is assigned to the class “x”, because the algorithm assigns new case on the bases of most similar cases of “N” or its neighbor. It is illustrated in figure 5 [1][2][3].

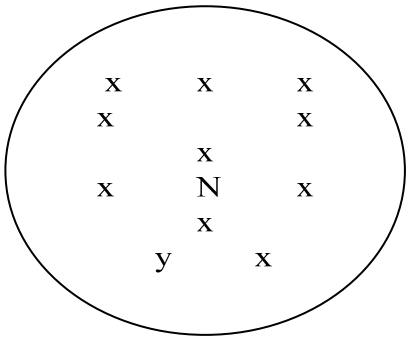


Figure 5. K-Nearest Neighbor

K-NN models can be used in modeling of non-standard data type i.e. text and these models are suitable for few predictor variables because the output is easy to understand. The first thing is to calculate the distance between attributes in the data. Once the distances between cases is calculated, then select the set of already classified cases then decide the range of neighborhood to do comparison also count the neighbors themselves. The major issues in K-NN are: It requires large computational load on the computer therefore all the data is kept in memory. This enhances the speed of K-NN. It is also known as K-NN Memory-based reasoning. It processes new cases rapidly as compared to d. tree or neural networks. It requires new calculation for new cases. Numeric data can easily be handled by this algorithm, categorical variables need special handling [2][3][7].

2.1.8. Bayesian Classification

As its name implies, Bayesian classification attempts to assign a sample x to one of the given classes using a probability model defined according to the Bayesian theorem. The latter calculates the posterior probability of an event, conditional on some other event. Basic prerequisites for the application of Bayesian classification are: Knowledge of the prior probability for each class. Knowledge of the conditional probability density function for each class. It is then possible to calculate the posterior probability using the Bayesian formula

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

$$q(ci \mid x) = (p(x \mid ci)p(ci)) / p(x) \quad (1)$$

where p(x) is the prior probability of x. Each new data tuple is classified in the class with the highest posterior probability. The issue in Bayesian Classification is: Major drawbacks of Bayesian classification are the high computational complexity and the need for complete knowledge of prior and conditional probabilities [4].

2.2. Distributed Computing Systems

The goal of distributed computing is to solve large computational problems. These are popular due to two main reasons: First, the nature of the problem requires using communication network that connects several computers and second the use of distributed system is beneficial for practical reasons. The distributed computing provides remote communication, fault tolerance and remote information access. Figure 6 below shows different forms of distributed computing:

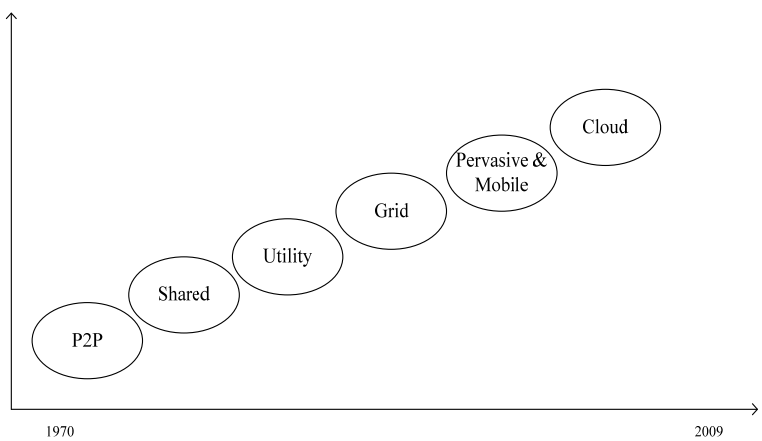


Figure 6. Forms of Distributed Computing

2.2.1. Cloud Computing

It is a new type of distributed computing still emerging field in computer science. The remote machines owned by other company which will run everything for the user is called cloud computing. It will change entire computer industry. The only thing the user has to run interface software of cloud computing system. There is a significant workload shift i.e. the user's computer will not run the applications. It will decrease the demand of hardware and software. There is no limit of its applications. Everything can be done through cloud computing. The major advantage of this is the client can access his data any where at any time. It will reduce the need for advanced hardware which will bring the cost of hardware down. The client can take the advantage of network processing power if the cloud computing is using a grid at its back end. This is a step backward to early computers having only keyboard and terminal. The major issues in cloud computing are: Data governance: Enterprises have sensitive data that requires proper monitoring and protection, moving data into cloud enterprise will lose their governance on own data. Manageability, Monitoring, Compliance, Cross-country Data migration, Reliability, availability and recovery and Security and privacy: are major concerns and issues in cloud computing [7].

2.2.2. Shared Computing

It is a network of computers that work together to complete a task. It is sharing of processing power and other resources. A user can access the processing power of entire network. It is used only for the complex problems not for others. Its administration and design is complicated. The main issues are: The safety and privacy is issue in shared computing. It needs a plan when a system goes offline or unavailable. Power consumption in shared computing is high which produces the heat. The major concern about shared computing is that they are not comprehensive. They uses only processing power not the other resources like storage. The grid computing is more applicable then shared because of its resource sharing [7].

2.2.3. Utility Computing

It is a business model in which a company outsources its computer support to other company. This support can be in the form of processing power, storage, hardware and software applications. The major advantage of utility computing is convenience because client has not to buy all hardware and licensed software for his business. He has to rely on another party to provide these services. The main issues are: This type of computing model is suitable for medium or large scale enterprise, not suitable for small business. Another main disadvantage of utility computing is reliability i.e. clients may hesitate to hand over duties to a smaller company where they feel the lost of data. It is an easy target of hackers. The major challenge in utility computing is that the consumers are not educated about its service. Its awareness is not very widespread [5].

2.2.4. Grid Computing

It is a type of distributed computing where every computer can access the resources such as processing power, memory and data storage of other computer on the network, turning it into a powerful supercomputer. It is a high performance computing. The grid computing can provide an effective computational support from applications for knowledge discovery. The basic services of grid computing are communication, authentication, information and resource management. This is not a new concept but not yet perfect. People are still working on creating, establishing and implementing standards and protocols. The applications of grid computing are limitless. The main challenges are: Resource sharing & coordinated problem, Coordinated problem solving in dynamic, Multi-institutional virtual organizations, Data protection, No clear standard, Better understanding as simple as possible, Difficult to develop, Lack of grid-enabled software, Centralized management, and the limited number of users are allowed the full access of network otherwise the control node will be flooded with processing requests which can create deadlock situation [10][13].

2.2.5. Pervasive Computing

It is a share of small, inexpensive, robust networked processing devices distributed at all scales in everyday life. The main challenges in pervasive computing are: System design and engineering, System modeling and human computer interaction models [9].

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

2.2.6. Mobile Computing

It is an ability to use technology while moving. It is a building block for pervasive computing. It is a tomorrow's network technology. It will revolutionize the way computers are used. Wireless communication, mobility and portability are salient features of mobile computing. It is a paradigm shift in distributing computing. The limitations of mobile computing are: There are no such standards of security of data. The time of power supply is limited. There may be signal problem due to transmission interferences. The excess use of mobile devices may cause potential health hazards. Small and limited user interface for human on the device is available. Small storage capacity, Risks to data, Mobility, moving from one coverage area to another and insufficient bandwidth [7].

3. Methodology

The traditional centralized data analyzing is not suitable for distributed applications. In distributed environment analyzing the distributed data is a non-trivial problem because of many constraints such as limited bandwidth, privacy-sensitive data and distributed nodes. The distributed problems solving environment fit well with the multiagent system (MAS) since the solution requires autonomous behavior, collabartion and reasioning. The agents perform the underlying data analysis task very efficiently in distributed manner. The MAS offer an architecture for distributed problem solving, deal with complex applications that require distributed problem solving. The combination of data mining algorithms and MAS for data analyzing will further enhance the processing power of the application.

Two medical datasets 'Diabetes' and 'Breast Cancer' are selected for this paper. The data is pre-processed, called data standardization. The interval scaled data is properly cleansed by applying the *range method*. The attributes of the dataset/testbed 'Diabetes' are: Number of Times Pregnant (NTP)(min. age = 21, max. age = 81), Plasma Glucose Concentration a 2 hours in an oral glucose tolerance test (PGC), Diastolic Blood Pressure (mm Hg) (DBP), Triceps Skin Fold Thickness (mm) (TSFT), 2-Hour Serum Insulin (m U/ml) (2HSHI), Body Mass Index (weight in kg/(height in m)^2) (BMI), Diabetes Pedigree Function (DPF), Age, Class (whether diabetes is cat 1 or cat 2) [22][23][24][25]. The attributes of dataset 'Breast Cancer' are: Clump Thickness (CT), Uniformity of Cell Size (UCS), Uniformity of Cell Shape (UCSh), Marginal Adhesion (Mad), Single Epithelial Cell Size (SECS), Bare Nuclei (BNu), Bland Chromatin (BCh), Normal Nucleoli (NNu), Mitoses , Class (benign, malignant). The vertical partitions of the chosen datasets are created. The architecture of MAS is shown in figure 7.

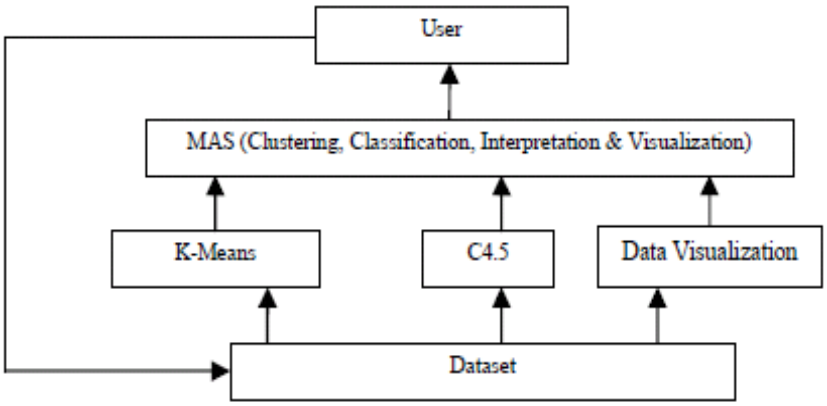


Figure 7. The Architecture of MAS

The partitioned datasets are used as inputs, three intelligent agents for conventional data mining algorithms such as k-means, c4.5 and data visualization, take these datasets and produce the outputs clusters, decision rules and 2D graphs which are helpful for users to take decisions. The MAS is deployed on a client-server based local distributed system in the form of a 'grid computing' shown in figure 8.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

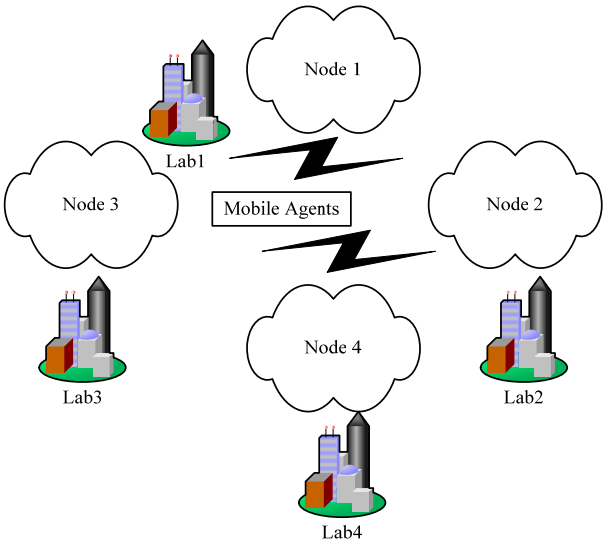


Figure 8. A distributed network using MAS

These agents can roam from one node to other node freely and can be stored at any node in the distributed network. The results of these agents can be stored at any where in the network. The user has the access to all the outputs [11][12][14].

4. Results and Discussion

The results and discussion part of this paper has two parts. In the first part, we draw a comparison of commonly used data mining algorithms in table 1. In order to build a predictive model, understanding of the data is the must. The data is in the forms of continuous and categorical (ordinal or nominal).

Table 1. Comparison of Data Mining Algorithms

Properties	**NNs**	**DTs**	**K-means**	**FL**	**GAs**	**Data visualization**	**K-NN**	**Bayesian Classification**
Training/Learning of a data set	High	Low	Low	None	High when no parallel processing	None	High	Low
Explicitness	Yes	No	No	None	No	No	No	No
Lack of Knowledge	Yes	No	Yes	It is a fad	No	No	No	No
Lackness of problem solving	Yes	Yes	Yes	None	No	Yes	Yes	No
Type of data in data sets	All types of data sets	Interval or categorical	Categorical and Numeric	None	It is suitable for strings	All types of data used in graphs	Numeric	Numeric
Estimation or prediction	Yes	No estimation but prediction is possible	Yes	Not popular in data mining	General-purpose search algorithm	Yes	Yes	Yes
Resource consumption	High	Medium	Medium	None	Low	Low	High	High
Size of data sets	Small to Intermediate	Small	Small to intermediate	None	Small	Small to Intermediate	Small to Intermediate	Small to Intermediate
Complexity	O(m)	O(m)	O(nkl) for time and O(k+n) for space	O(m)	O(m)	O(m)	O(m)	O(m)

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

The main properties of data mining algorithms are robustness, scalability and efficiency. It is clear from this comparison that all the data mining algorithms are suitable and produce useful results only on small to medium scale datasets. They are not appropriate for large datasets. This is a problem of scalability. There is another overhead; the learning and training of the dataset is required in all algorithms. If there is any noise in datasets, the extracted knowledge may be misleading. The choice of the algorithm depends on the intended use of extracted knowledge. The data can be used either to predict future behavior or describe patterns in an understandable form within discover process.

We also discuss different distributed systems and their issues. The induction of cloud computing has change the paradigm; it is a step backward to early computer having only keyboard and terminal. The distributed systems provide remote access to information, when resources are shared in any type of distributed computing; the users have to compromise on privacy and security of their data. These are two major issues in all form of distributed computing. There are no standards in design, implementation and management of networks in any form of distributed computing. The migration of data across the country is another issue. A mountain of data is there and to fetch the correct data on the request of a user from the clouds of data is another issue.

In the second part, the results obtained from the agents of the MAS are discussed. The output of these agents is in the form of 2D graphs of the clusters, 2D graphs of whole partitioned clustered dataset and decision rules in the form of if-then-else. We discuss some of the graphs and decision rules in this paper. The cluster by cluster graphs and the whole partitioned clustered graphs have their own significance and the user can use either of the graphs or both for the future prediction of the data. The purpose of 2D graph is to identify the type of relationship if any, between the attributes of the given dataset. The graph is used when a variable exists which is being tested and in this case the attribute or variable 'class' is a test attribute of both selected medical datasets.

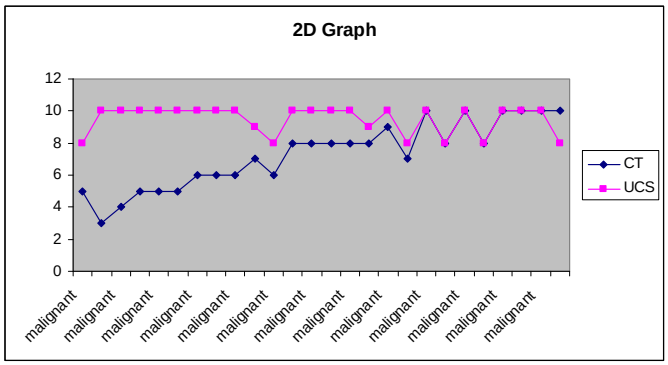


Figure 9. 2D graph between attributes 'clump thickness' & 'uniformity of cell size' of dataset 'Breastcancer'

The graph in figure 9 shows that the value of the attributes 'clump thickness' and 'uniformity of cell size' is almost constant at the beginning and at the end of the graph both attributes have equal value. The outcome of this graph is that if the value of these attributes is either constant or equal then the patient has 'malignant' class of breast cancer.

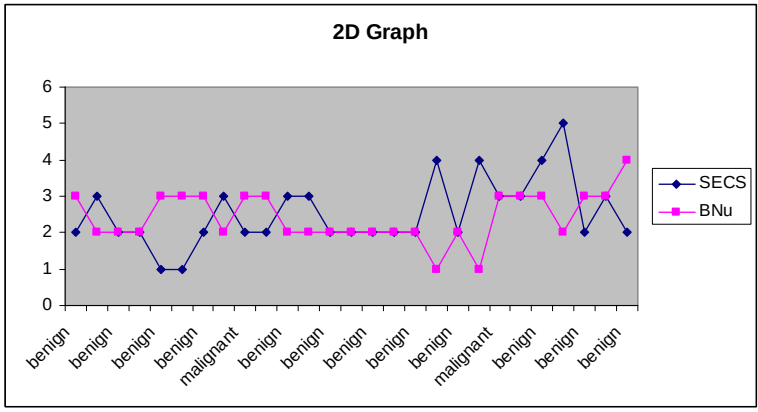


Figure 10. 2D graph between attributes 'single epithelial size'
& 'bare nuclei' of dataset 'Breastcancer'

The structure of graph in figure 10 is almost similar in figure 9. This graph has three sections, the value of attributes 'single epithelial size' and 'bare nuclei' is constant in the beginning, then the value of these attributes is variable, then in some region of the graph the value is equal and then again it is constant. The conclusion of this graph is that if the value of these attributes is either constant or equal then the patient has 'benign' class of breast cancer and the variable values of these attributes give the 'malignant' class of breast cancer.

Figure 11 shows the decision rules of a cluster of dataset 'Breast Cancer'.

1. Rule: 1 if Mitoses = 1 then
2. Rule: 2 if BCh = 7 then
3. Class = malignant , benign else
4. Rule: 3 if BCh = 1 then
5. Class = benign else
6. Rule: 4 if BCh = 2 then
7. Class = benign , malignant else
8. Rule: 5 if BCh = 3 then
9. Class = benign , malignant else
10. Rule: 6 if BCh = 4 then
11. Class = malignant , benign else
12. Rule: 7 if BCh = 8 then
13. Class = malignant

Figure 11. The Decision Rules of a cluster of dataset 'Breast Cancer'

The decision rules in the form of 'if – then – else' are shown. It is an easy way to take the decision. The result is if the value of the attribute 'mitoses' is 1 and the value of attribute 'bland chromatin' is either 1, 2 or 3 then the patient has 'benign' class of breast cancer otherwise the 'malignant' class of breast cancer. These rules can be used in a simple query to further validate the results.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

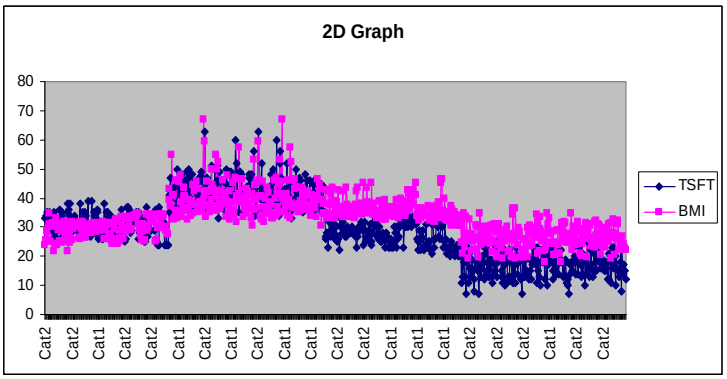


Figure 12. 2D graph between attributes 'triceps skin fold thickness' &
'body mass index' of dataset 'Diabetes'

The graph of the whole partitioned clustered dataset 'Diabetes' is shown in figure 12. The structure of the graph is complex as compared to the graphs of 'Breast Cancer' dataset. The graph has two regions, either the value of attributes 'triceps skin fold thickness' and 'body mass index' is equal or constant. There is no variation in values of the attributes in this graph. It is difficult to take the decision either the patient has 'cat1' or 'cat2 type of diabetes.

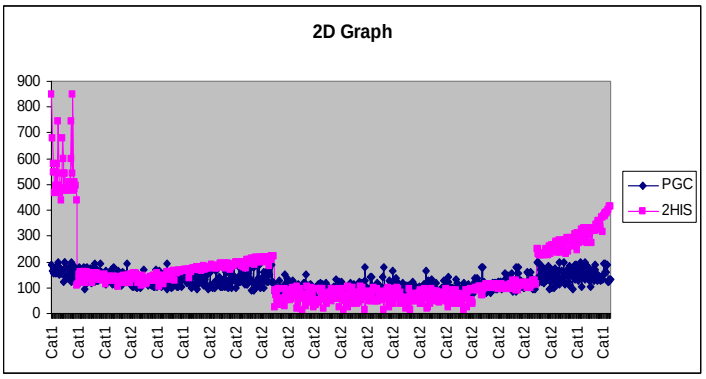


Figure 13. 2D graph between attributes 'plasma glucose concentration' &
'2-hour insulin serum' of dataset 'Diabetes'

The graph of the whole partitioned clustered dataset 'Diabetes' is shown in figure 13. The graph has two main regions, in the beginning and at the end the value of attributes 'plasma glucose concentration' and '2-hour insulin serum' is variable, which shows the 'cat1' type of diabetes in patient and in the middle of the graph the value of these attributes is either constant or equal, which gives 'cat2' type of diabetes in patient.

Figure 14 shows the decision rules of whole partitioned clustered dataset 'Diabetes'.

1. Rule: 1 if PGC = 165
   then
2. Class = Cat2
   else
3. Rule: 2 if PGC = 153
   then
4. Class = Cat2
   else
5. Rule: 3 if PGC = 157
   then
6. Class = Cat2
   else
7. Rule: 4 if PGC = 139
   then
8. Class = Cat2
   else
9. Rule: 5 if HIS = 545
   then
10. Class = Cat2
    else
11. Rule: 6 if HIS = 744
    then
12. Class = Cat2
    else
13. Class = Cat1

Figure 14. The Decision Rules of whole partitioned clustered dataset 'Diabetes'

The result for this whole partitioned clustered dataset 'Diabetes' is if the value of attribute 'plasma glucose concentration' is more than 130 and the value of attribute '2-hour insulin serum' is more than 500, as the rules show, then the patient has diabetes of cat2 otherwise cat1. The decision rules make it easy and simpler for the user to interpret and predict this partitioned dataset of diabetes. These rules can be used in a simple query to further validate the results.

Similarly, the 2D graphs and the decision rules can also be drawn for the other clusters and partitioned clustered datasets. These results show that the conventional data mining algorithms can be deployed using the MAS on the modern distributed systems.

5. Conclusion

The goals of data mining are pattern and feature extraction, visualization of data and evaluation of results. These goals can be achieved by using different data mining algorithms. In this paper we discuss the most commonly used data mining algorithms, their issues and applications and a comparison of these algorithms is drawn. If the user is clear about business goals, type of prediction and model then the selection of algorithms is very easy. Three data mining 'k-means clustering', 'C4.5' and 'data visualization' algorithms are selected, for clustering, classification, interpretation and visualization. A MAS approach comprising three intelligent agents, one for each algorithm is used. Two medical datasets 'Breastcancer' and 'Diabetes' are selected to test the MAS over a client-server based local distributed system in the form of a 'grid computing'. The vertical partitions of the given datasets, based on the similar values of the attributes are created. The results obtained are encouraging, acceptable, satisfactory and consistent. We conclude our paper, due to wide spread use of distributed systems nowadays, in order to access the data and produce 'knowledge', apply intelligent mobile agents in data mining algorithms over distributed systems.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Future Work

The study could be extended to large scale distributed systems with the other data mining algorithms so as to validate the effectiveness of the proposed methodology. For further investigation in this direction, one will undoubtedly has to take into account the parameters such as data caching and the validity of the agent framework.

Acknowledgment

The authors are thankful to 'The Islamia University of Bahawalpur, Pakistan' for providing financial assistance to carry out this research activity under HEC, Pakistan project 6467/F – II.

References

[1] Wang, John., "*Data Mining Opportunities and Challenges"*, Idea Group Publishing ISBN: 1-59140-051-1, chapter IX page 235 and chapter XVI page 381

[2] Liu, Bing., "*Web Data Mining Exploring Hyperlinks, Contents, and Usage Data"*, ISBN: 13 978-3-540-37881-5, Springer Berlin Heidelberg New York, chapter 3 and chapter 4.

[3] "*Introduction to Data Mining and Knowledge Discovery"*, ISBN: 1-892095-02-5, Third Edition by Two Crows Corporation, page numbers: 11,12,13,15.

[4] Symeonids, Andreas. Pericles, Mitkas., "*AGENT INTELLIGENCE THROUGH DATA MINING"*, ISBN 0-387-24352-6, chapter 1, 2, 3.

[5] Y. Chee, B. Rajkumar, D. Marcos, Y. Jia, S. Anthony, V. Srikumar, P. Martin, "*Utility Computing on Global Grids"*, ISBN: (978-0-471-78461-6, John Wil ey & Sons, New York, USA, 2007.

[6] L. Jiangchuan, Rao. Sanjay, Li. Bo, Z. Hui, "Opportunities and Challenges of Peer-to-Peer Internet Video Broadcast", *Proceedings of the IEEE* Volume 96, Issue 1, Jan. 2008 Page(s):11 - 24

[7] Satyanarayanan. M., "Fundamental Challenges in Mobile Computing", School of Computer Science, Carnegie Mellon University, *Proceedings of the fifteenth annual ACM symposium on Principles of distributed computing* Philadelphia, Pennsylvania, United States, Pages: 1 – 7, ISBN:0-89791-800-2

[8] Jimenez. Raul, Eriksson. Lars-Erik, Knutsson, Bj¨orn., "P2P-Next: Technical and Legal Challenges", *6th Swedish National Computer Networking Workshop (SNCNW'09) and 9th Scandinavian Workshop on Wireless Adhoc Networks (Adhoc'09) Sponsored by IEEE VT/COM Sweden*

[9] Satyanarayanan. M., "Pervasive Computing: Vision and Challenges", *School of Computer Science, Carnegie Mellon University, Personal Communications, IEEE* Aug 2001, Volume: 8, Issue: 4 On page(s): 10-17, ISSN: 1070-9916

[10] C. Mario., T. Domenico., T. Paolo., "Distributed data mining on the grid", *ICAR-CNR*, Via P. Bucci, Cubo 41-C, 87036, Rende (CS), Italy, August 2002.

[11] Voskob. Max., Howey. Rob., Panin. Nick., "Data mining and Privacy in Public Sector using Intelligent Agents", *Discussion paper*, November 2003, Wellington, New Zealand.

[12] Grossman. Robert, Kasif. Simon, Moore. Reagan, Rocke. David, Ullman. Jeff, "Data Mining Research: Opportunities and Challenges", *A Report of three NSF Workshops on Mining Large, Massive, and Distributed Data*, 1998.

[13] Dutta. Haimonti., "Empowering Scientific Discovery by Distributed Data Mining on the Grid Infrastructure"

[14] da Silva. C. Josenildo., et al, "Distributed Data Mining and Agents"

[15] Berkhin, Pavel., "*Survey of Clustering Data Mining Techniques"*, Accrue Software, Inc.

[16] Freitas, Alex A., "*A Survey of Evolutionary Algorithms for Data Mining and Knowledge Discovery"*, Pontificia Universidade Catolica do Parana Rua Imaculada Conceicao, 1155. Curitiba - PR. 80215-901. Brazil.

[17] Mento, Barbara. and Rapple, Brendan., "*Data Mining and Data Warehousing"*, Virtual Data Center Manager Boston College, July 2003

[18] Mangina, Eleni., "Intelligent Agent-Based Monitoring Platform for Applications in Engineering", *International Journal of Computer Science & Applications* Vol. 2, No. 1, pp. 38 – 48, Technomathematics Research Foundation, 2005.

[19] Chen, Ming-Syan, Han, Jiawei and Yu, Philip S., "Data Mining: An Overview from Database Perspective", *IBM T.J. Watson* Res. Ctr. P.O.Box 704, Yorktown, NY 10598, U.S.A.

[20] Maindonald, John., "Data Mining Methodological Weaknesses and Suggested Fixes", *Proc. Fifth Australasian Data Mining Conference (AusDM2006)*, 2006.

[21] Schommer, Christoph., "An Unified Definition of Data Mining", *arXiv:0809.2696v1 [cs.SC]* 16 Sep 2008.

[22]US Census Bureau. Iris, Diabetes, Vote and Breast datasets at URL: www.sgi.com/tech/mlc/db visited 2009.

[23]National Institute of Diabetes and Digestive and Kidney Diseases, Pima Indians Diabetes Dataset, http://archive.ics.uci.edu/ml/datasets/Diabetes, visited 2009

[24]Skrypnik, Irina., Terziyan, Vagan., Puuronen, Seppo., and Tsymbal, Alexey, "Learning Feature Selection for Medical Databases", *CBMS 1999*.

[25] Irene M. Mullins et al., "Data mining and clinical data repositories: Insights from a 667,000 patient dataset", *Computers in Biology and Medicines* 36 p. 1351 – 1377, 2006.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Principal Objects Detection Using Graph-Based Segmentation and Normalized Histogram

Bui Ngoc Nam and Pham The Bao

Faculty Mathematics and Computer Science, University of Science
Ho Chi Minh City, Vietnam

Abstract

In this paper, we introduce a new method to distinguish the principal objects in image datasets using graph-based segmentation and normalized histogram (PODSH). Unlike the usual object detection systems which require the input objects, we propose a new approach to recognize objects one might focus on when taking images. Motivated by the habit of taking picture, we suppose that the position of a main object is located near the image centre and this object always holds a large area. The normalized histogram is added to increase the effect of our system. In the experiment, we used images which consist of objects to test the precision of PODSH. Our system is implemented by Matlab.

Keywords: *objects detection, graph-based segmentation, normalized histogram, principal objects.*

1. Introduction

Locating principal objects plays an important role for many learning systems. In recent years, researchers has explored in this section. However, most projects focus on the object recognition problem in which the interested objects is initially given. There are many algorithms have got a significant result such as SIFT [1], SURF [2], PCBR [3] and Discriminatively Trained Deformable Part Models [4]. Nevertheless, consider a problem that one wants to find the set of considerable objects in complex image dataset. These methods have just listed cannot solve this problem because there are no initial required objects. Therefore, we conduct a new object measurement in order to identify the important objects.

Motivated by the high performance of the graph based-segmentation algorithm developed by Pedro F. Felzenszwalb and Daniel P. Huttenlocher [5] as mentioned in [6], we proposed the new quantity to characterize the 'importance' of an object. Based on the target when taking a picture, we realize that the main objects always located near the centre of image and occupy a large square on an image. But using just two quantities to describe the 'importance' of image causes many drawbacks will be discussed later. Therefore, we imposed new rules on the normalized histogram of each object to correct these mistakes. In addition, due to the requirement of graph based-segmentation algorithm [5], we use neural network model to train parameters of this method.

2. Determine parameters in graph-based segmentation

The use of graph based algorithm proposed by Pedro F. Felzenszwalb and Daniel P. Huttenlocher to the segment problem due to its best result and low time performance. The segments or objects in image are nonlinear data due to the fact that their area can be large or small and cannot be predicted. Therefore, this algorithm entails three input parameters to extract fine segmentations. However, these parameters are also nonlinear which is not proper to locate objects automatically. Due to that reason, we conduct a method that find out the suitable nonlinear parameters for each image by using neural network training. This model is processed by using Levenberg-Marquardt training algorithm with 20 neurons in hidden layer [7]. We used the Matlab neural network toolbox to train our model.

Pedro and Daniel used three parameters which are the variance of Gaussian function, the variable threshold function and the minimum square of component. These factors are symbolized by σ, k and min [5] respectively. Given an input image, our purpose is characterizing the (σ, k, min) vector which is also the output vector of the neural network model. The input vector is the features of an input image. Due to experiences, we recognize the relationship between an image and these aspects below:

- The size of images.
- The average value in each Red, Green, Blue and Gray color space.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

- The standard deviation value in each Red, Green, Blue and Gray color space.

One might wonder what target to evaluate the fitness of an output vector. The answer is in the requirement of our problem; the most important objects need to be extracted. After testing 500 colour images, we proposed the proper number objects are from 5 to 15. Fortunately, this number is also the quantity to estimate the fitness of (σ, k, min) vector. That means the (σ, k, min) vector must satisfy the condition; the number of components must be between 5 and 15 after segmenting.

The 500 colour images are used again to test our model. This 70% images have the number of components in range 5 and 15.

3. Determine measurement for detecting principal objects

The main idea for extracting principal objects automatically by using segment method is each image component can be considered as a real life object. However, we should constraint many rules to locate the objects that one concentrates on. The first two factors come from the scenario taking picture. Because the main object is targeted, they will be situated near the image centre and take a large areas. Based on these factors, the two elements of importance vector are the square and position of an object. However, this two issues are not sufficient because in some cases the background focal point is close the image midpoint than that of the main object. Moreover, background always occupies a large square in image. This problem is illustrated by the figure 1.

Because of those disadvantages, we propose another constraint relating to the normalized histogram. In order to make this idea clearly, we use the car image above to illustrate our view. We first compare the different between the two normalized histogram respects to the car and background object in figure 2.

If we limit bins have small frequencies, the energy range of car is over 200 wider than that of background (under 100). In this paper, we propose the energy range measurement to distinguish the main objects and background. This supposition based on our experiment result.

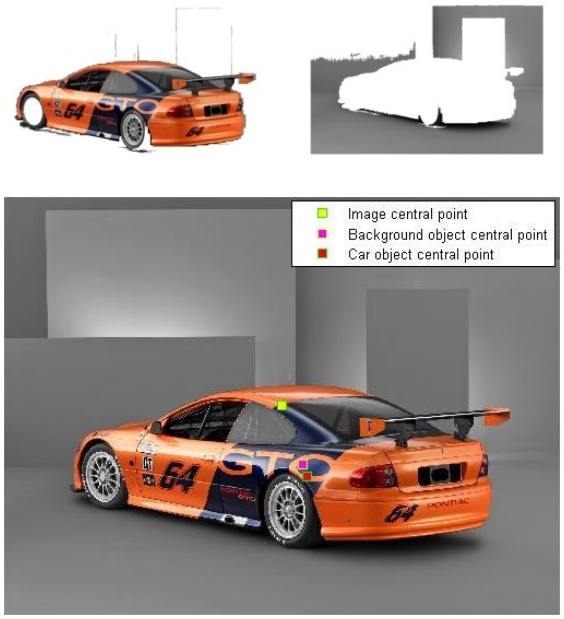


Fig. 1 A comparision between the car object and the backgroud object in terms of square and focal point. The top left image is the car object which has 603878 pixels and the focal point is the red spot on the bottom image. The top right image is the backgroud object which is compoed by 1160178 pixels and its central point is the magenta spot. The yellow spot is the image central point.

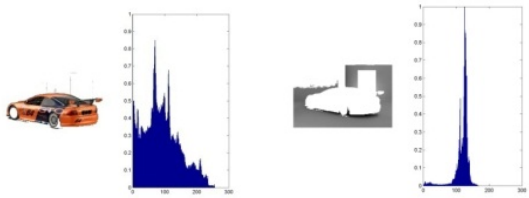


Fig. 2 The normalized histogram of car and background object.

The proven of this hypothesis relies on the probability theory as we did in [6]. Given 326 images, we extracted 447 principal objects by using PODSH. There are 347 precise main objects. Based on the central limit theorem, we gain the following result:

$$\frac{f-p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1) \qquad (1)$$

where n is 447, f is 347/447 and p is the probability this method isolates main objects.
With the 90% confidence interval in the Eq. (1), we get the formula:

$$-1.6449 \leq \frac{f-p}{\sqrt{\frac{p(1-p)}{n}}} \leq 1.6449 \qquad (2)$$

Solving the Eq. (2), we have p is in between 0.7422 and 0.807. That means the probability that PODSH produced

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

the main objects is from 74% to 80% with 90% confidence interval.

After contributing the principal object measurement, the objects are filtered based on this quantity. However, the role of each value in this vector is different. Particularly, the object area is the most important factor because the small object can cause noisy in both focal point and normalized. Therefore, we should limit the small object first. After that, the sum of two left issues is taken and their average value is used to be the threshold. Objects has the summation larger than the threshold are considered as the principal objects.

4. Experiments

The image dataset consists of 2258 objects extracted from 1648 images in two categories: vehicles and train-planes. Despite of clustered background, PODSH only preserves main objects, particularly trains, planes and vehicles. The 2258 objects extracted from this dataset are used to test the precision of PODSH. The precision is the percentage of main objects account for in 2258 objects. The average result and time consuming is given by table 1. Figure 3 gives some example results of this system.

Table 1: Average precision and time consuming of principal object detection method

Average precision	Average time consuming
79.6722%	6.2453s

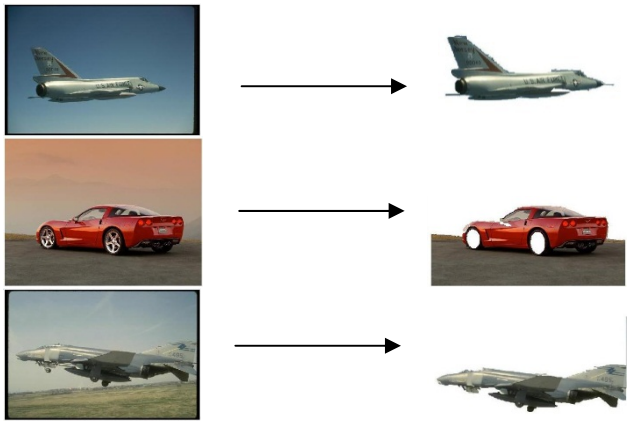


Fig. 3 The planes and cars are extracted from images with clusterd background using PODSH. The images on the left side is the original image as the input parameters of PODSH and on the other side is the principal objects extracted by using PODSH.

5. Conclusions

Beside the encouraging result, this approach still has many shortcomings due to the lack of information of edge and corner when using histogram. The results can growth if the edge, corner and frequency factors are considered. In many cases, the focal point of background always located in the image center. Therefore, more rules should be added to improve our system.

Acknowledgments

We would like to thank DOLSOFT Inc for providing us a large-scale image database with complex background.

References

[1] D.G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision, Springer Netherlands, Vol. 60, No. 2, 2004, pp. 91-110.
[2] H. Bay, A. Ess, T. Tuytelaars, L. Van Gool, "SURF: Speeded Up Robust Features", Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, 2008, pp. 346-359.
[3] H. Deng, W. Zhang, E. Mortensen, T. Dietterich, and L. Shapiro, "Principal Curvature-based Region Detector for Object Recognition", in IEEE Conference on Computer Vision and Pattern Recognition, 2007.
[4] P. Felzenszwalb, D. McAllester, D. Ramaman, "A Discriminatively Trained, Multiscale, Deformable Part Model", Proceedings of the IEEE CVPR, 2008.
[5] P.F. Felzenszwalb, D.P. Huttenlocher, "Efficient Graph-Based Image Segmentation", International Journal of Computer Vision, Springer Netherlands, Vol. 59, No. 2, 2004, pp. 167-181.
[6] N.D. Anh, P.T. Bao, B.N. Nam, and N.H. Hoang, "A New CBIR System Using SIFT Combined with Neural Network and Graph-Based Segmentation", in Proc. ACIIDS (1), 2010, pp. 294-301.
[7] M.T. Hagan and M. Menhaj, "Training feedforward networks with the Marquardt algorithm", IEEE Trans. Neural Networks, Vol. 5, No. 6, 1994, pp. 989–993.

Bui Ngoc Nam holds MSc degree in Computer Science from the University of Nottingham in 2011. He received the BSc degree in Department of Computer Science from University of Science, Ho Chi Minh City, Vietnam in 2009. He is currently a lecturer in Faculty of Computer Science of University of Science Ho Chi Minh City, Vietnam. His research interest includes computer vision, image processing and parallel computing.

Pham The Bao has obtained his Master degree and PhD in Computer Science at University of Science, Ho Chi Minh City, Vietnam. He is currently a Vice-Dean in Faculty of Computer Science and Head of Computer Science Department at University of Science Ho Chi Minh City, Vietnam. His interested research field includes computer vision, image process, parallel computing, soft computing and recognition.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Correlated Node Behavior Model based on Semi Markov Process for MANETS

A.H Azni[1], Rabiah Ahmad[2], Zul Azri Muhamad Noh[3], Abd Samad Hasan Basari[4], Burairah Hussin[5]

[1,2,3,4,5]Universiti Teknikal Malaysia Melaka
Faculty of Information and Communication Technology,
76100 Durian Tunggal, Melaka, Malaysia

Abstract

This paper introduces a new model for node behavior namely Correlated Node Behavior Model which is an extension of Node Behavior Model. The model adopts semi Markov process in continuous time which clusters the node that has correlation. The key parameter of the process is determined by five probabilistic parameters based on the Markovian model. Computed from the transition probabilities of the semi-Markov process, the node correlation impact on network survivability and resilience can be measure quantitatively. From the result, the quantitative analysis of correlated node behavior on the survivability is obtained through mathematical description, and the effectiveness and rationality of the proposed model are verified through numerical analysis. The analytical results show that the effect from correlated failure nodes on network survivability is much severer than other misbehaviors.

Keywords: *network resilience, survivability, node behavior model, traffic profiling, MANETS.*

1. Introduction

The node behavior pattern plays an important role in performance analysis of mobile and wireless networks. In mobile networks, for example, a node may change its behavior from normal to misbehave node which directly affects the connectivity and availability of the network [1]. The impact of misbehave node is quite challenging due to multiple failures caused by node mobility, energy depletion and Denial of Services (DoS) attacks. In addition, node behavior has major effect on route discovery, packets forwarding, and network control message [2][3][4]. However, nodes in the network not only affect individual node, but it may affect multiple nodes which has a direct or indirect correlation between nodes.

The modeling of correlated node behavior is thus an essential part in analyzing and designing survivability framework in wireless network. In this paper, Markov process is represented to model and analyze the stochastic properties of the node's behavior. We also design a correlated node behavior model that could dynamically affect network resilience in MANETS. The rest of the paper is organized as follows Section 2 gives related works of node behavior theory. Section 3 defines node behavior in MANETS. Sections 4 presents a proposed model defining four-sate transition of correlated node behaviors and describe probabilistic theory to model correlated nodes behavior. Section 5 discusses the parameters use as a performance evaluation. Finally we conclude our paper.

2. Related Works

The studies of node behavior have helped many researchers to understand many research problems such as the design of fault tolerant routing protocol and analysis of network performance, thoroughly. Currently, node behavior modeled by [5][6] profile the normal behavior of wireless network. It used genetic algorithm to learn current behavior based on past history. This approach has been used widely in intrusion detection system to detect anomalies in the network. These model measures network performance based on single or individual node. As mention by [7], individual node may not accurately measure real life application. It needs to take into account an impact of correlated node instead of individual node. Work in [8][9][10] deals with modeling of selfish node behavior using game theory. Game theory is a branch of economics that deals with strategic and rational behavior [11]. It appears to be a natural tool for both designing and analyzing the interaction among players/nodes. The theory seems suitable to model selfish behavior where energy conservation is an issue to selfish node. It tends to limit their support to other nodes as this will costs energy. Thus, to stimulate cooperation, the nodes need to be rewarded every time it cooperates. The other theory used to model behavior is Markov chain theory [12][13][14]. Markov chain is a mathematical system that undergoes transition from one state to another in a chain manner [15] . Markov chain theory suits for nodes in ad hoc network as it is general enough to capture the major characteristics and yet

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

simple enough to formulate nodes long-run behavior [16][17]. To the authors knowledge, neither one of the research in this area involve correlated nodes behavior in their studies.

Work done in this paper applied Semi Markov to determine node behavior transition. Semi Markov is a probabilistic system that made its transitions according to transition probability matrix of Markov process, but whose time between transitions could be an arbitrary random variable which depends on the transition. This work is an extension of a research done by [7] which model network survivability based on individual node behavior. He assumed that node behavior has no correlation with another node's behavior in wireless network. Thus, his model is tractable, however, this assumption may not always hold in real environment. In real environment node has a correlation with other nodes in such a way that if a node has more and more neighbors failed, it may need to load more traffic originally forwarded by those failed neighbors, and thus might become failed faster due to excessive energy consumption. Similarly, it is also possible that the more malicious neighbors a node has, the more likely the node will be compromised by its malicious neighbors. Therefore, the paper examines how these correlated behaviors will affect the network resilience.

3. Node Behavior Definitions

In MANETS, the nodes are dynamically and arbitrarily change its behavior from cooperative to misbehave node. Based on [9] node behaviors in MANETS are classified into four types of behaviors. They can be classified as:

- ***C****ooperative Nodes* are active in route discovery and packet forwarding, but not in launching attacks
- *Failed Nodes* are not active in route discovery
- *Malicious Nodes* are active both in route discovery and launching attacks
- *Selfish Nodes* are active in route discovery, but not in packet forwarding. They tend to drop data packets of others to save their energy so that they could transmit more of their own packets and also to reduce the latency of their packets.

Whenever node joints the network, it is assume as normal or cooperative. It may change its behavior to misbehave node due to various reasons. In this model, we specify rules for a node transition. Figure 1 shows the transition of node behavior in MANETS.

a) Cooperative node (C) may change its state either to selfish, malicious or failure node. At cooperative state, it is exposed to become failed due to energy exhaustion, misconfiguration, and so on.
b) It is also possible to convert a selfish (*S*) node to be cooperative again by means of proper configurations. A selfish or cooperative node can become malicious due to being compromised or failed due to power depletion.
c) A malicious (*M*) node can become a failed node (*F)*, but it will not be considered to be cooperative or selfish any more even if its disruptive behaviors are intermittent only.
d) A failed node (*F)* can become cooperative again if it is recovered and responds to routing operations.

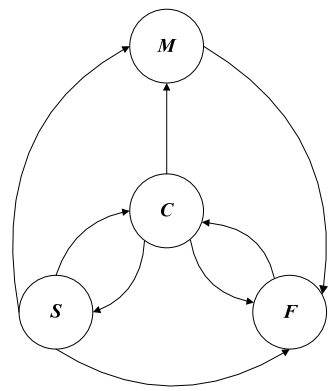


Fig. 1: Node behavior transition

4. Correlated Node Behavior Model

In this section, we use a Semi-Markov process to model node behavior transitions, analyzed the stochastic properties of node behavior, formulating the transition matrix and model state transition for correlated node behavior.

4.1 Semi Markov and Stochastic Properties Node Behavior Model

Due to the fact that node in MANETS is more inclined to be failed over time, we find that, the probability changes its behavior dependent on time. Therefore, node transition cannot simply described by Markov chain because of its time-dependent property. The semi-Markov process denoted by:

$$Z(t)= X_n, \quad \forall t_n \leq t < \forall t_{n+1} \tag{1}$$

with a state space S = *{C (cooperative), S (selfish), M (malicious), F (failed)}*. X_n denotes the *embedded* Markov chain of Z(*t*), which has a finite state space S**,** and the *n*th state visited [18].Thus, X_n is *irreducible* and *ergodic*. By

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Collolary 9 -11 (pp 325) in [7] we know that *Z(t)* is irreducible and *Z(t)* is the state of process at its most recent transition. The transition probability from state *i* to state *j* is defined as follows:

$$P_{ij} = \lim_{t \to \infty} \Pr(X_{n+1)} = j, t_{n+1} - t_n \le t \mid X_n = i)$$
$$= \Pr(X_{n+1} = j \mid X_n = i) \quad (2)$$

Then a matrix $\overline{P} = (P_{ij})$ is the transition probability matrix (TPM) of $\{X_n\}$. The construction of $\overline{P}$ can be determined by the observation of empirical results. The state transition diagram of semi Markov node behavior model is shown in Fig.2, which is determined by characteristics of node behaviors.

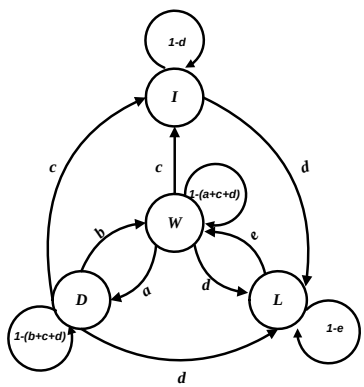


Fig. 2: Semi-Markov Process for Node Behavior

A node in proposed model is viewed as having in four states namely forward (*W*), drop (*D*), inject (*I*) and loss (*L*) which adequately describe the behavior of the node during forwarding, dropping, injecting and loss state. Let the probability of dropping packets due to selfishness and the probability of forwarding packets due to altruistic nature shown by the node *a* and *b* respectively and they are independent of each other. Let *c* be the probability of injecting packets due to malicious activity and *d* is the probability of loss packets due to exhausted battery power, out of transmission range or malfunction. Probability of recovery shown by *e* define the recovery of node from failure state to cooperative again after it has been recovered, recharged or repaired. To determine the current behavioral status of neighbor node *m* at instant time *t* is formally given as:

$$N^{(m)}_{(t)} \begin{cases} \text{W, if the level of dropping, a < threshold and level of forwarding b > threshold; and e=1} \\ \text{D, if level of dropping a > threshold;} \\ \text{I, if the injecting level, c > threshold and level of forwarding b<threshold} \\ \text{L, if level of loss d=1;} \end{cases} \quad (3)$$

4.2 Formulation of Correlated Node Behavior Transition Probability Matrix (TPM).

Based on node behavior above, transition probability matrix (TPM) of $\{X_n\}$ is given below:

$$P = \begin{array}{c} \\ W \\ D \\ I \\ L \end{array} \begin{array}{c} \begin{array}{cccc} W & D & I & L \end{array} \\ \begin{bmatrix} 1-(a+c+d) & a & c & d \\ b & 1-(b+c+d) & c & d \\ 0 & 0 & 1-d & d \\ e & 0 & 0 & 1-e \end{bmatrix} \end{array} \quad (4)$$

The "0" in the matrix means that it is not possible to make transition between the two states based on the rules specify in section 3.0. Since it is a stochastic matrix, the summation of transition probabilities to a state must be equal to 1. $\{Z(t)\}$ is also associated with the time distributions between two successive transitions. Let T_{ij} denote the time spent in state *i* given the next state *j*. Then $Z_{ij}(t)$ is a commonly used notation for cumulative distribution function (CDF) of T_{ij}, defined by :

$$Z_{ij}(t) = \Pr(T_{ij} \le t)$$
$$= \Pr(t_{n+1} - t_n \le \mid X_n = i, X_{n+1} = j) \quad (5)$$

where $i,j \in S$.

Proof: Let P_{ij} be the parameter for a, b, c d, and e define above for the node N^m at time instant t, we can specify:

$a = P[N^{(m)}_{(t-1)} = C \mid N^{(m)}_{(t)} = S]$
$b = P[N^{(m)}_{(t-1)} = S \mid N^{(m)}_{(t)} = C]$
$c = P[N^{(m)}_{(t-1)} = C \mid N^{(m)}_{(t)} = M]$ *or*

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

$$P[N^{(m)}_{(t-1)} = S \mid N^{(m)}_{(t)} = M]$$
$$d = P[N^{(m)}_{(t-1)} = C \mid N^{(m)}_{(t)} = F] \text{ or } P[N^{(m)}_{(t-1)} = S \mid N^{(m)}_{(t)} = F] \text{ or } P[N^{(m)}_{(t-1)} = M \mid N^{(m)}_{(t)} = F]$$
$$e = P[N^{(m)}_{(t-1)} = F \mid N^{(m)}_{(t)} = C] \quad (6)$$

By knowing the transition probability P_{ij} and transition time distribution $Z_{ij}(t)$, which are defined above, the transient distribution P_{ij} converge to a limiting P_i can be calculated by

$$P_i = \lim_{t\to\infty, \forall j \in S} \Pr(Z(t) = i \mid Z(0) = j) = \frac{\pi_i E[T_i]}{\sum_j \pi_j E[T_j]} \quad (7)$$

where π_i is the stationary probability of state i of X_n. To obtain the steady state probabilities, we need to solve the equation

$$\prod . \mathbf{P} = \prod$$
$$\sum_{i \in s} \pi_I = 1, \ \pi_i \geq 0, \text{ and } E[T_i] = \sum_{j \in s} P_{ij} E[T_{ij}] \quad (8)$$

Let $\prod$ is the probability vector in steady state and **P** is the matrix representing the transition probability distribution from Fig.3. Thus, to calculate P_i, we only need to estimate P_{ij} and $E[T_{ij}]$, which are normally easier to obtain from statistic, then we can use equation in (8) to obtain π_i and $E[T_i]$. After π_i and $E[T_i]$ are obtained, we can use equation (7) to derive P_i.

*Proof: First, for the given embedded Markov Chain of $\{X_n\}$ associated with the state space S and TPM **P** defined by (4), it is trivial to prove that $\{X_n\}$ is irreducible and positive recurrent. Thus, the SMP $\{Z(t)\}$ is irreducible. Second, based on our assumptions in Section 3.0, a node works in any behavior state for a finite time, which implies $E[T_i] < \infty$ and $\sum_{i \in S} E[T_i] < \infty$. Thus, the SMP $\{Z(t)\}$ is positive recurrent as well. Finally, by Theorem 9-3* [19] *the limiting probability exists and can be given by (7).*

In order to formulating TPM for correlated node behavior, let N^1 and N^2 are two nodes connected in a network. The corresponding TPM for N^1 and N^2 are given below:

Table 1a: Node N^1

	W	D	I	L
W	$1-(a+c+d)^1$	a^1	c^1	d^1
D	b^1	$1-(b+c+d)^1$	c^1	d^1
I	0	0	$1-d^1$	d^1
L	e^1	0	0	$1-e^1$

Table 1b: Node N^2

	W	D	I	L
W	$1-(a+c+d)^2$	a^2	c^2	d^2
D	b^2	$1-(b+c+d)^2$	c^2	d^2
I	0	0	$1-d^2$	d^2
L	e^2	0	0	$1-e^2$

With two nodes, the state space occupies sixteen finite states such as { WW,WD, WI, WL, DW,DD, DI, DL, IW, ID, II, IL, LW, LD, LI, LL} at any point of time. It requires $O(4^m)$ computation complexity, where m refer to number of nodes in the network. However, for correlated nodes behavior, we proposed a clustering method which groups the nodes according its current status and behavior and thus, reduce the computational complexity. Fig 3 illustrated the clustering of correlated transition probability distribution between different correlated state and correlated PTM matrix is also given in table 2 below. The four correlated states behavior are grouped by $\{S_0, S_1, S_2, S_3\}$ mapped against {WW, DD, II, LL}.

Let {u, v, w, x} be a function mapping states into correlated state behaviors. Denote π_i the probability of being in steady state. If the state space is finite, then the equation in (8) can be solve to obtain π_i .The status of a node at current t time is given by (3). Thus, the cluster of correlated node behavior involving two nodes N^1 and N^2 are:

$$N^{(1,2)}_{(t)} \begin{cases} S_0 = C, \text{ if } [(N^{(1)}_{(t)} = W) \text{ and } (N^{(2)}_{(t)} = W)] \\ S_1 = S, \text{ if } [(N^{(1)}_{(t)} = D) \text{ and } (N^{(2)}_{(t)}) = D)] \\ S_2 = M, \text{ if } [(N^{(1)}_{(t)} = I) \text{ and } (N^{(2)}_{(t)} = I)] \\ S_3 = F, \text{ if } [(N^{(1)}_{(t)} = L) \text{ and } (N^{(2)}_{(t)} = L) \end{cases} \quad (9)$$

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

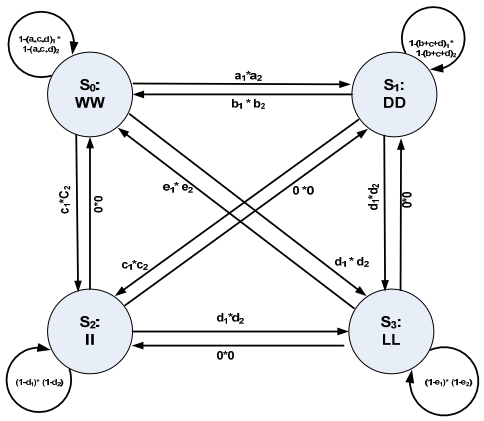


Fig.3: Correlated State Transition representing the node behavior

Table 2: Correlated PTM for four state Markov chain

	S_0	S_1	S_2	S_3
S_0	$1-(a+c+d)^1$ * $1-(a+c+d)^2$	a^1*a^2	c^1*c^2	d^1*d^2
S_1	b^1*b^2	$1-(b+c+d)^1$ * $1-b+c+d)^2$	c^1*c^2	d^1*d^2
S_2	0*0	0*0	$(1-d)^1*(1-d)^2$	d^1*d^2
S_3	0* 0	0*0	0*0	$(1-e)^1*(1-e)^2$

Table 2 is constructed to map PTM against correlated state transition S_0, S_1, S_2, and S_3 in Fig.3. We propose a correlated function based on conditional probabilities:

$$u = P[N^{(1\ldots m)}_{(t-1)} = C \mid N^{(1\ldots m)}_{(t)} = S]$$

$$v = P[N^{(1\ldots m)}_{(t-1)} = S \mid N^{(1\ldots m)}_{(t)} = C] \text{ or}$$

$$P[N^{(1\ldots m)}_{(t-1)} = F \mid N^{(1\ldots m)}_{(t)} = C]$$

$$w = P[N^{(1\ldots m)}_{(t-1)} = C \mid N^{(1\ldots m)}_{(t)} = M] \text{ or}$$

$$P[N^{(1\ldots m)}_{(t-1)} = S \mid N^{(1\ldots m)}_{(t)} = M]$$

$$x = P[N^{(1\ldots m)}_{(t-1)} = C \mid N^{(1\ldots m)}_{(t)} = F] \text{ or}$$

$$P[N^{(1\ldots m)}_{(t-1)} = S \mid N^{(1\ldots m)}_{(t)} = F] \text{ or}$$

$$P[N^{(1\ldots m)}_{(t-1)} = M \mid N^{(1\ldots m)}_{(t)} = F] \qquad (10)$$

To show how correlated function works, let N^1 and N^2 be the connected node in the network. Resultant from equation (6) and (10), we get:

$$u^1 = P[N^{(m)}_{(t-1)} = C \mid N^{(m)}_{(t)} = S]$$

$$v^1 = P[N^{(m)}_{(t-1)} = S \mid N^{(m)}_{(t)} = C] \text{ or}$$

$$P[N^{(m)}_{(t-1)} = F \mid N^{(m)}_{(t)} = C]$$

$$w^1 = P[N^{(m)}_{(t-1)} = C \mid N^{(m)}_{(t)} = M] \text{ or}$$

$$P[N^{(m)}_{(t-1)} = S \mid N^{(m)}_{(t)} = M]$$

$$x^1 = P[N^{(m)}_{(t-1)} = C \mid N^{(m)}_{(t)} = F] \text{ or}$$

$$P[N^{(m)}_{(t-1)} = S \mid N^{(m)}_{(t)} = F] \text{ or } P[N^{(m)}_{(t-1)} = M \mid N^{(m)}_{(t)} = F]$$

Then, using equation (8), we can obtain u^1 as:

$$u^{(1)} = \frac{P[N^{(1)}_{(t-1)} = C \mid N^{(1)}_{(t)} = S]}{P[N^{(1)}_{(t-1)} = C]} * P[N^{(1)}_{(t)} = S]$$

$$= \frac{a^1}{1-(\pi_1+\pi_2+\pi_3)} * \pi_1$$

where π_1, π_2, π_3 are obtained from solving equation (8) for correlated state model shown in Fig.3. Rest of the function can be calculated in similar ways. The equation only represents a single pair of neighbor nodes only. As part of correlated state transition, its functions ($u^{(1,2)}$, $v^{(1,2)}$, $w^{(1,2)}$, $x^{(1,2)}$), can be represented as per equation (10). The next step, it is proposed to have an iterative approach to model the correlated nodes to logically present nodes as correlated clusters. Let $m \geq 2$ be the total number of neighbor nodes. In the first iteration, neighbor nodes N^1 and N^2 into equivalent node. Then the resulting equivalent node is combined with N^3, and so on, until all m neighbor nodes are combined together to form a function of correlated node. This iteration is repeated till all nodes are clustered according to their correlated state. Using the resultant result from correlated state model in Fig.3 and equation set (9), the correlated function can be rewritten can as:

$$u^{(1\ldots m)} = \frac{P\left[N^{(1\ldots m)}_{(t-1)} = C \mid N^{(1\ldots m)}_{(t)} = S\right]}{P\left[N^{(1\ldots m)}_{(t-1)} = C\right]} * P\left[N^{(1\ldots m)}_{(t)} = S\right]$$

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

$$= \frac{a^{1\ldots m} * a^{1\ldots m}}{1-(\pi_1+\pi_2+\pi_3)} * \pi_1$$

$$v^{(1..m)} = \frac{P\left[N_{(t-1)}^{(1..m)} = S \mid N_{(t)}^{(1..m)} = C\right]}{P\left[N_{(t-1)}^{(1..m)} = S\right]} * P\left[N_{(t)}^{(1..m)} = C\right] +$$

$$\frac{P\left[N_{(t-1)}^{(1..m)} = F \mid N_{(t)}^{(1..m)} = C\right]}{P\left[N_{(t-1)}^{(1..m)} = F\right]} * P\left[N_{(t)}^{(1..m)} = C\right]$$

$$= \frac{b^{1..m} * b^{1..m}}{1-(\pi_0+\pi_2+\pi_3)} * \pi_0 + \frac{e^{1..m} * e^{1..m}}{1-(\pi_0+\pi_1+\pi_2)} * \pi_0$$

$$w^{(1\ldots m)} = \frac{P\left[N_{(t-1)}^{(1..m)} = C \mid N_{(t)}^{(1..m)} = M\right]}{P\left[N_{(t-1)}^{(1..m)} = C\right]} * P\left[N_{(t)}^{(1..m)} = M\right] +$$

$$\frac{P\left[N_{(t-1)}^{(1..m)} = S \mid N_{(t)}^{(1..m)} = M\right]}{P\left[N_{(t-1)}^{(1..m)} = S\right]} * P\left[N_{(t)}^{(1..m)} = M\right]$$

$$= \frac{c^{1..m} * c^{1..m}}{1-(\pi_1+\pi_2+\pi_3)} * \pi_2 + \frac{c^{1..m} * c^{1..m}}{1-(\pi_0+\pi_2+\pi_3)} * \pi_2$$

$$x^{(1\ldots m)} = \frac{P\left[N_{(t-1)}^{(1..m)} = C \mid N_{(t)}^{(1..m)} = F\right]}{P\left[N_{(t-1)}^{(1..m)} = C\right]} * P\left[N_{(t)}^{(1..m)} = F\right] +$$

$$\frac{P\left[N_{(t-1)}^{(1..m)} = S \mid N_{(t)}^{(1..m)} = F\right]}{P\left[N_{(t-1)}^{(1..m)} = s\right]} * P\left[N_{(t)}^{(1..m)} = F\right] +$$

$$\frac{P\left[N_{(t-1)}^{(1..m)} = M \mid N_{(t)}^{(1..m)} = F\right]}{P\left[N_{(t-1)}^{(1..m)} = M\right]} * P\left[N_{(t)}^{(1..m)} = F\right]$$

$$= \frac{d^{1..m} * d^{1..m}}{1-(\pi_1+\pi_2+\pi_3)} * \pi_3 + \frac{d^{1..m} * d^{1..m}}{1-(\pi_0+\pi_2+\pi_3)} * \pi_3 +$$

$$\frac{d^{1..m} * d^{1..m}}{1-(\pi_0+\pi_1+\pi_3)} * \pi_3 \qquad (11)$$

5. Model Parameters

This section describes how to obtain transition probabilities which are described as *a, b, c, d* and *e* above. To determine P_{ij} (i,j ∈ {S}), we need the appropriate input parameters from routing packets to determine its behavior and network resilience. Whenever node initiates the link, it will periodically broadcast a message to update its link information with its neighbors [20]. The message carries routing information such as number of packets, number of bytes sent and received, and its residual energy. Initially, every node has the same initial energy and may turn off packet forwarding functionality once its residual energy below a 1/ η of its initial energy. The nodes tend to drop all the packets forwarded and become selfish at a time $T_{selfish}$ a s given below:

$$T_{selfish} = (1 - 1/\eta)\ \overline{L} \qquad (12)$$

where η is the selfish threshold parameter and $\overline{L}$ is the average life time of a node derived from:

$$\overline{L} = \frac{\text{Remaining power}}{\text{Power consumption rate}} \qquad (13)$$

Thus, the probability of selfish node a=1/$T_{selfish}$ and is given as:

$$a = \frac{\eta}{(\eta-1)} * \frac{1}{\overline{L}} \qquad (14)$$

The node is considered cooperative if 'a' is less than threshold and the average life time of the node will be longer. Similarly, the probability 'b' can be derived from monitoring neighboring nodes in terms of forwarding each other packets. Probability of 'b' can be derived as:

$$b = \frac{\text{Number of packets forwarded by the neighbors}}{\text{Number of packet received by the neighbors}} \qquad (15)$$

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

The nodes may also receive more packets that it supposed to. This scenario considers a misbehavior activity by nodes. It is considered malicious nodes which goal is to depleting the resources of nodes forwarding packets. Probability 'c' can be derived from:

$$c = \frac{\text{Number of packet received by the neighbors}}{\text{Number of packets forwarded by the neighbors}} \tag{16}$$

The nodes consider malicious when 'c' is greater than threshold. Injecting an overwhelming amount of traffic also can easily cause network congestion and decrease the node lifetime. Thus, the node tend to loss it packets when the energy power is exhausted. Therefore, probability of loss'd' is given as below:

$$d = \frac{b}{(\text{Num of packets forwarded}+\text{Num of packets receives})} \tag{17}$$

In the case when the node failed, it can be back in the network after the recovery time. The node recovery time can be derived using formula:

$$e = \frac{1}{\text{Average Recovery}} \tag{18}$$

6. Simulation and Analysis

In this paper we focus on the effect nodes correlation towards survivability. This section will verify the correctness and rationality of the equation discuss above . We use MATLABv7.10 to perform the simulation based on parameters discussed above. The survivability analysis in MANETS is considered from the following cases:

6.1 The effect on survivability from probability of cooperation (b)

As explain in section 3.0, node cooperation is represented by probability of forwarding *b*. The higher the *b* implies that the node is cooperative. To observe the effect of probability of cooperation clearly, $c + d + e = 0$ is set so that *b* varies only due to node selfishness attack. Figure 5 shows the analytic results of survivability under different nodes range 5, 15, 25, 50 nodes respectively. It is observed that cooperation level incline steady line with fewer nodes. This is due to the misbehavior node effect are less. Thus, the effect of node behavior is tractable with fewer nodes. This result proofs the concept from [7]. Cooperative nodes are affected by packet forwarding rate to obtain a higher survivability. Thus it is necessary to have a higher packet forwarding rate in order for network to survive. When drop packets are higher, the nodes become less cooperative and network survivability is impossible to achieve.

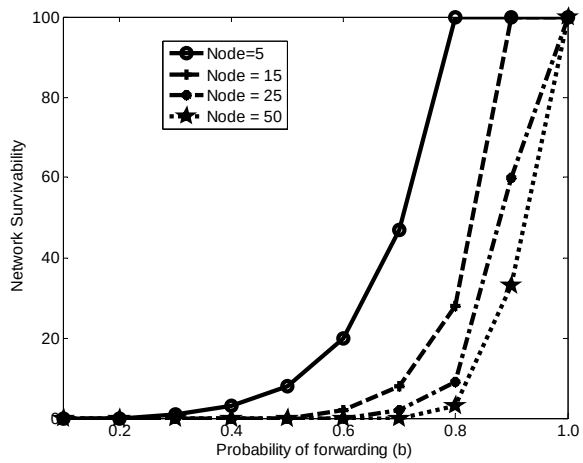


Fig.5: Effect on survivability of node cooperation based on probability of forwarding

6.2 The effect of survivability from the probability of selfishness (a)

In the same way, $c+d+e=0$, and the result can be obtained from selfish nodes in Fig.6. Similar to that in Fig 5, the plot in Fig 6 shows that the survivability decreases as probability of dropping (*a*) increases. The survivability does not change significantly at the beginning especially if network scalability is less. In contrast, survivability for fewer nodes starts to decline faster compared to networks with large nodes. Network also becomes unstable when the dropping level reaches 0.4 to 0.5.

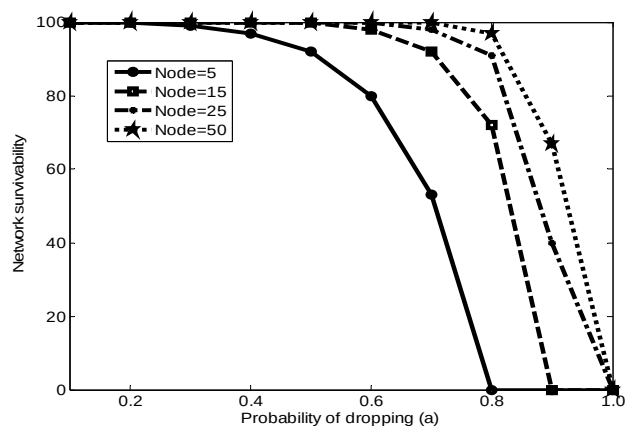


Fig 6. : Effect on survivability of node cooperation based on probability of dropping (a)

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

6.3 The effect on survivability from probability of malicious nodes (c).

To explain the effect of malicious node, a+d=0 is set to eliminate the impact of selfish and failure node. From Fig.7 the network survivability decreases very fast as probability of injection increase. It can be seen that network with more nodes could not sustain it survivability when network under attacks. When more packets were injected in the network, it will create a Denial of services (DoS) attack. DoS attack spread to other neighbors faster and will decrease network survivability faster too.

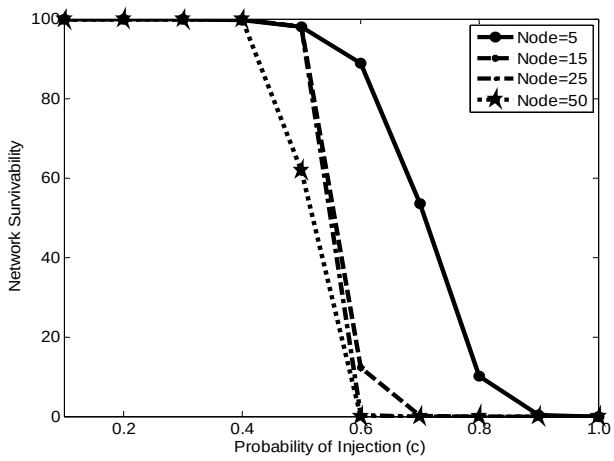


Fig 7. : Effect on survivability of node cooperation based on probability of injection (c)

6.4 The effect on survivability from probability of failure (d).

To explain the effect of node failures on network survivability, it is set that a+c=0 to eliminate the effect of misbehaving nodes in simulations. From Fig 8, survivability decreases very fast as probability of loss increase. For highly survival network, the effect of loss packet is more significant, e.g., the survivability drop to 0 when probability of packet loss (d) = 0.3. Compare to malicious nodes, failed nodes shown severe effect on network survivability. The severer impact of node failures is due to the fact that node failures are also isolated from the network, which reduces the density of active nodes [ ]. Therefore, the probability of network failure cannot be ignored especially for a large scale network.

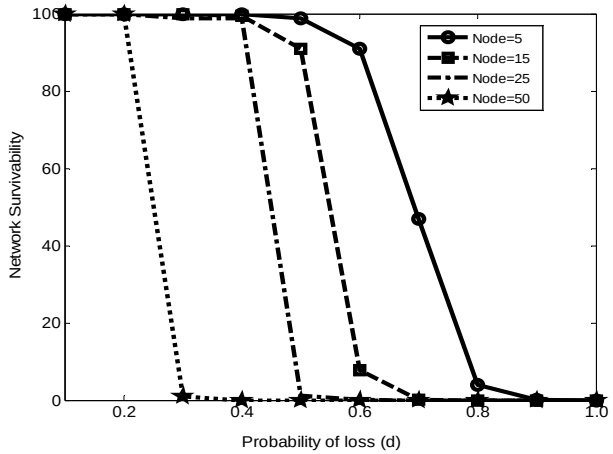


Fig 8. : Effect on survivability of node cooperation based on probability of loss (d)

5. Conclusions

In this paper, we focused on modeling and formulation correlated node behavior for dynamic topology network such as in MANETS. The nodes are classified into four types: cooperative, selfish, malicious and failed. Then, a correlated node behavior model was proposed by employing semi Markov process. The nodes were correlated according to their clusters describe as forwarding, dropping, injecting and loss. In our model, nodes changed their behavior according to correlated transition probability matrix and transition time distribution matrix. Further, the equilibrium states of correlated node being in each behavior state were obtained. The model showed accuracy in portraying the behavior of correlated node and thus can be used to study various performance metrics for network survivability and resilience. For future work, there is a need to solve the complexity of equation involving correlated transition probability matrix of $O(4^m)$ to simplify the computational processing and increase quality of services.

Acknowledgments

A.H.Azni would like to thank Universiti Sains Islam Malaysia (USIM) and Ministry of Higher Education (MOHE) for financial support throughout her studies in Universiti Teknikal Melaka Malaysia (UTEM), Melaka, Malaysia.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

References

[1] F. Xing and W. Wang, "Understanding Dynamic Denial of Service Attack in Mobile Ad hoc Networks," in *IEEE Military communication conference (MILCOM)*, 2006, pp. 1-7.

[2] P. Rai, "A Review of ' MANET ' s Security Aspects and Challenges '," *System*, pp. 162-166, 2010.

[3] F. Xing and W. Wang, "Modeling and analysis of connectivity in mobile ad hoc networks with misbehaving nodes," in *IEEE International Conference on Communications, 2006*, 2006, vol. 4, no. c, pp. 1879–1884.

[4] J. P. G. Sterbenz et al., "Resilience and survivability in communication networks: Strategies, principles, and survey of disciplines," *Computer Networks*, vol. 54, no. 8, pp. 1245–1265, Jun. 2010.

[5] M. M. Sivajothi and E. Naganathan, "Analysis of Reference Point Group Mobility Model in Mobile Ad hoc Networks with an Ant Based Colony Protocol," *Proceedings of the International MultiConference of Engineers and Computer Scientists*, vol. 1, 2009.

[6] V. R. V. D. G. Mohankumar, "Feature Analysis for Intrusion Detection in Mobile Ad-hoc Networks," *IJCSNS*, vol. 10, no. 9, p. 215, 2010.

[7] F. Xing, "Modeling, Design, and Analysis on the Resilience of Large-scale Wireless Multi-hop Networks," University of North Carolina, 2009.

[8] Y. Liu, C. Comaniciu, and H. Man, "Modelling misbehaviour in ad hoc networks: a game theoretic approach for intrusion detection," *International Journal of Security and Networks*, vol. 1, no. 3/4, p. 243, 2006.

[9] T. Sundararajan and A. Shanmugam, "Modeling the Behavior of Selfish Forwarding Nodes to Stimulate Cooperation in MANET," *International Journal*, vol. 2, no. 2, pp. 147-160, Apr. 2010.

[10] S. Buchegger, "Coping with Misbehavior in Mobile Ad-hoc Networks," Citeseer, 2004.

[11] R. Mahajan, M. Rodrig, D. Wetherall, and J. Zahorjan, "Experiences applying game theory to system design," *Proceedings of the ACM SIGCOMM workshop on Practice and theory of incentives in networked systems - PINS '04*, p. 183, 2004.

[12] S. K. Hwang and D. S. Kim, "Markov model of link connectivity in mobile ad hoc networks," *Telecommunication Systems*, vol. 34, no. 1-2, pp. 51-58, Dec. 2006.

[13] P. Pileggi and P. Kritzinger, "Semi-Markov Process Formulation of IEEE 802 . 11 DCF Analytic Performance Models," *Computer*.

[14] M. Zhao and W. Wang, "A novel semi-markov smooth mobility model for mobile ad hoc networks," in *Proc. of IEEE GLOBECOM*, 2006, pp. 1–5.

[15] M. Fygenson, "SEMI-MARKOV PROCESSES," *Science*, vol. 32, pp. 151-160, 1989.

[16] V. G. Chaganti, L. W. Hanlen, and T. A. Lamahewa, "Semi-Markov modeling for body area networks," in *Communications (ICC), 2011 IEEE International Conference on*, 2011, pp. 1–5.

[17] D. N. Alparslan and K. Sohraby, "A Generalized Random Mobility Model for Wireless Ad Hoc Networks and Its Analysis: One-Dimensional Case," *IEEE/ACM Transactions on Networking*, vol. 15, no. 3, pp. 602-615, Jun. 2007.

[18] S. Wang and J. T. Park, "Modeling and analysis of multi-type failures in wireless body area networks with semi-Markov model," *Communications Letters, IEEE*, vol. 14, no. 1, pp. 6–8, Jan. 2010.

[19] Daniel Heyman and Mattlew Sobel., *Stochastic Models in Operations Research.* McGraw-Hill, 1982.

[20] K. Komathy and P. Narayanasamy, "A Probabilistic Behavioral Model for Selfish Neighbors in a Wireless Ad Hoc Network," *IJCSNS*, vol. 7, no. 7, p. 77, 2007.

A.H. Azni is a Ph.D student at Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, with financial support from Ministry of Higher Education Malaysia and Universiti Sains Islam Malaysia. Azni received her M.sc in Digital Communication at Monash University, Clayton Australia in 2002 and B.sc in Computer Information System at Bradley University, Illinois, USA in 1998. She has work as a lecturer in Information and Security Assurance for many years, her interest domain is security, wireless and mobile network, survivability and resilience.

Rabiah Ahmad is an Associate Professor at theFaculty of Information Technology and Communication, University Technical Malaysia Melaka (UTeM), Malaysia. She received her PhD in Information Studies (health informatics) from University of Sheffield, UK, and MSc. (information security) from Royal Holloway University of London, UK. Her research interests include healthcare system security and information security architecture. She delivered papers at various health informatics and information security conferences on national as well as international level. She has also published papers in accredited national/international

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

journals. Besides that, she also serves as a reviewer for various conferences and journals.

Zul Azri Muhamad Noh is a lecturer at Faculty of Information and Communcation Technology, Universiti Teknikal Malaysia Melaka, Malaysia. He received his D.Eng. (Computer Science and Engineering), M.Eng. (Computer Science and Engineering), and B.Eng. (Electrical and Computer Engineering) at Nagoya Institute of Technology, Nagoya Japan in 2005, 2007, and 2010, respectively. His current research interests include wireless LAN, packet scheduling, and multimedia QoS.

Abd Samad Hasan Basari is a lecturer at Faculty of Information and Communcation Technology, Universiti Teknikal Malaysia Melaka, Malaysia. He received his B.Sc. (Mathematics) at National University of Malaysia in 1998, M.Sc in IT Education at Universiti Teknologi Malaysia in 2002 and Ph.D in Modeling at Universiti Teknikal Melaka Malysia in 2004y. His current research interests include Decision Support Technology, Modelling, Intelligent System

Burairah Hussin is an Associate Professor at the Faculty of Information Technology and Communication, University Technical Malaysia Melaka (UTeM), Malaysia. He received his PhD in Management Science from University of Salford, UK, and MSc. (Numerical Analysis and Programming) from University of Dundee, UK. His research interests include operational problem within industry, maintenance management, statistical analysis, creative media and software development. He has also published papers in accredited national/international journals. Besides that, he also serves as a reviewer for various conferences and journals.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

An Application of Genetic Algorithm with Iterative Chromosomes for Image Clustering Problems

Mr.R.Balakrishnan[1], Mr.U.Karthick Kumar[2]

[1]Assistant Professor & Head
Department of MCA & Software Systems
VLB Janaki Ammal College of Arts and Science
Coimbatore, Tamil Nadu, India -641042

[2]Assistant Professor
Department of MCA & Software Systems
VLB Janaki Ammal College of Arts and Science
Coimbatore, Tamil Nadu,India -641042

Abstract

Many heuristic algorithms have been applied to the clustering problem, which is known to be NP Hard. This paper represents a Genetic Algorithm for clustering on image data. Genetic algorithms have been used in a wide variety of fields to perform clustering, however, the technique normally has a long running time in terms of input set size. This paper proposes an efficient genetic algorithm for clustering on very large data sets, especially on image data sets. In this study, a heuristic method based on Genetic Algorithms (GA) is adopted to automatically determine the number of cluster centroids during unsupervised classification. Efficient time techniques are used as a performance measure for clustering on image data. This paper compares Genetic algorithm with K-Means algorithm for clustering on image data.

KeyWords: *Data Mining, Clustering, K-means algorithm, Genetic algorithm, Image data, Heuristic method.*

1. Introduction

Data mining techniques are the result of a long process of research and product development [20]. This evolution began when business data was first stored on computers, continued with improvements in data access, and more recently, generated technologies that allow users to navigate through their data in real time. Data mining takes this evolutionary process beyond respective data access and navigation to prospective and proactive information delivery. Data mining is ready for application in the business community because it is supported by three technologies such as Massive data collection, Powerful multiprocessor computers and Data mining algorithms [11]. Data mining deals with large databases that impose on clustering analysis additional severe computational requirements. Cluster analysis is mainly conducted on computers to deal with very large scale and complex datasets. With the development of computer based techniques, clustering has been widely used in data mining, ranging from web mining, image processing, machine learning, artificial intelligence, pattern recognition, social network analysis, bioinformatics, geography, geology, biology, psychology, sociology, customers behavior analysis, marketing to e-business and other fields[14].

The task of grouping data points into clusters of "similar" items are a form of unsupervised learning that has applications in many fields. For instance, current techniques used for machine vision require processing of digital information obtained from pixels [2]. A very important step in this digital information processing is to group the data in some fashion so that patterns can be recognized. Clustering can be used for this task. In the medical field, clustering of data can be used to determine if a drug provides greater benefits to a certain group of patients. Grouping of information is used in the engineering field to determine what factors lead to the failure of a component in a system [14]. And in marketing, data clustering can give a clearer picture of how to focus an advertising campaign to the proper audience. The concept of clustering has been around for a long time. It has several applications, particularly in the context of information retrieval and in organizing web resources. The main purpose of clustering is to locate information and in the present day context, to locate most relevant electronic resources [3]. The research in clustering eventually led to automatic indexing as well as to retrieve electronic records. Clustering method in which we make group of objects that are somehow similar in characteristics.

The ultimate aim of the clustering is to provide a grouping of similar records. A clustering is essentially a set of such clusters, usually containing all objects in the data set. Additionally, it may specify the relationship of the clusters to each other, for example a hierarchy of clusters embedded in each other. Clustering can be roughly distinguished by hard clustering in which each object belongs to a cluster or not and soft clustering in which each

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

object belongs to each cluster to a certain degree e.g. a likelihood of belonging to the cluster. This paper will discuss the use of Genetic Algorithms (GA) for the task of clustering data. The running time for most clustering becomes quite large as the size of the input data set increases. Here, the application of GA algorithm for clustering on very large data sets, such as image data sets, will be addressed. The paper is organized as follows. In section II we explain related work, in section III, we review the clustering techniques. In Section IV-we describe clustering problems and in section V detail Genetic algorithms for clustering on very large data sets. In section VI describe performance analysis and conclusions describes in section VII.

2. Related work

Clustering has become a widely studied problem in a variety of application domains, such as in data mining and knowledge discovery [1], [2] statistical data analysis [3], [4] data classification and compression [6], medical image processing [5] and bioinformatics [6]. Several algorithms have been proposed in the literature for clustering [7], [8]. A. L. Abul is explained about Cluster Validity Analysis Using Sub sampling [10]. The objective of all clustering algorithms is to divide a set of data points into subsets so that the objects within a subset are similar to each other and objects that are in different subsets have diverse qualities [11], [12], [13]. Bradley and Fayyad have proposed an algorithm for refining the initial cluster centers. Not only are the true clusters found more often, but the clustering algorithm also iterates fewer times [15]. Some clustering methods improve performance by reducing the distance calculations. For example, Judd *et al*. proposed a parallel clustering algorithm P-CLUSTER [16] which uses three pruning techniques.

The K-Means algorithm [7] is well known for its efficiency in clustering large data sets. Fuzzy versions of the k-means algorithm have been reported in Ruspini [9] and Bezdek, where each pattern is allowed to have membership functions to all clusters rather than having a distinct membership to exactly one cluster. Kanungo *et al*. [17] proposed a filtering algorithm which begins by storing the data points in a *k-d* tree. For each node of the tree maintaining a set of candidate centers, they will be pruned, or filtered as they are propagated to the node's children. Kanungo *et al*.'s implementation of the filtering algorithm is more robust than Alsabti's, because Alsabti's method relies on a less effective pruning mechanism based on computing the minimum and maximum distances to each cell.

Ming-Chuan Hung, explained about an Efficient *k*-Means Clustering Algorithm Using Simple Partitioning [20]. The Genetic algorithm described in [18] uses a multi step procedure. The authors refer to this procedure as a semi supervised form of learning. In [19] a GA is used to solve the clustering problem for a data set of geographical data. Similarly, Yan-He Chen [22] describes about Genetic algorithm for Aerial image clustering. Bandyopadhyay, S., and Maulik describes Genetic algorithm for clustering using application of image classification by automatic evolution [24]. Rothlauf. F explained uses of genetic algorithm and evolutionary algorithms[23]. Yang, G., Reinstein define new genetic algorithm for optimisation problem[26].

3. Clustering Analysis

Cluster Analysis groups data objects based on information found in the data that describes objects and its relationship. The main goal of clustering is that a object within a group be similar to one another and different from the objects in other groups. The similarity within a group and the difference between the groups defines the clustering.

3.1 Types of Clustering

Cluster and clustering have different meaning. Cluster is a collection of objects and collection of clusters is referred as clustering and it has various types which have been mentioned below:

Hierarchical Cluster: Set of nested cluster which is organised as a tree. Each node in the tree refers clusters while parent is a main cluster and its children are subclusters and the root is a cluster containing all the objects.

Partitional Clustering: Set of data objects which is unnested non – overlapping subsets called clusters such that each data object have exactly one subset.

Exclussive Clustering: It assign each object to a single cluster.

Overlapping Cluster: Object in a data set may belongs to more than one group. It also called non-exclussive clustering which is used an object is between two or more clusters and assigned to any of these clusters.

Fuzzy or Soft Clustering: Every objects belongs to every cluster and hold value in between o or 1. Sum of value of object is 1 then it is absolutely belongs to that group and sum of value of object is 0 then it is absolutely not belongs to that group.

Complete Clustering: Assign all objects to cluster.

Partial Clustering: Not assign all objects to cluster when some objects in data sets may not belongs to group.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

3.2 Types of Cluster

Cluster groups define valid set of groups. There are several different cluster group are in practice such as

Well Seperated Cluster: A cluster is a set of objects which each object is very similar to other objects within a cluster and not similar to objects in other clusters.

Prototype Based Cluster: Prototype of a object is called centroid. A cluster is a set of objects in which each object is closer to prototype than to the prototype of other cluster.

Graph Based Cluster: If data is represented as a graph then group of object are connected to one another but no connection to outside group objects.
Density Based Cluster: If a cluster is a dense region of objects then it is seperated from other and it surrounded by a region of low density.

Conceptual Cluster: A cluster as a set of objects that share some property example: it share some closest centroid value.

3.2 Cluster Techniques

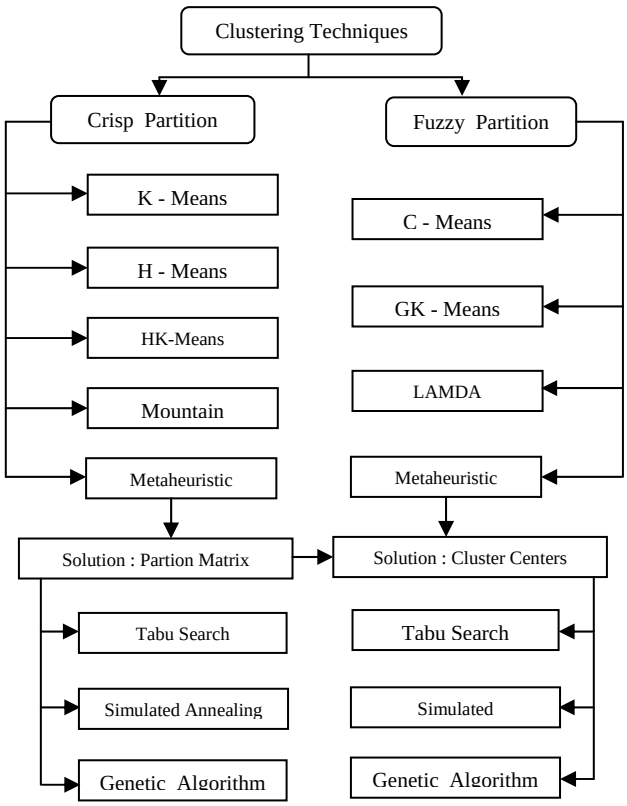


Figure. 1 Clustering Techniques

Bustos and Sellier categorise a number of lustering techniques **(Figure 1**) and show the K-means and the C-means to be amongst the most popular [25]. Cluster analysis has become an important technique in exploratory data analysis, pattern recognition, machine learning, neural computing, and other engineering. The clustering aims at identifying and extracting significant groups in underlying data. The four main classes of clustering algorithms are partitioning methods, hierarchical methods, density based clustering and grid-based clustering.

In the field of clustering, K-means algorithm is the most popularly used algorithm to find a partition that minimizes mean square error (MSE) measure. Although K-means is an extensively useful clustering algorithm, it suffers from several drawbacks. The objective function of the K-means is not convex and hence it may contain local minima. Consequently, while minimizing the objective function, there is possibility of getting stuck at local minima (also at local maxima and saddle point)[25]. The performance of the K-means algorithm depends on the initial choice of the cluster centers. Besides, the Euclidean norm is sensitive to noise or outliers. Hence K-means algorithm should be affected by noise and outliers. In addition to the K-means algorithm, several algorithms, such as Genetic Algorithm (GA) have been used for clustering.

4. Clustering Problem

The diversity of applications for clustering has lead to many problem definitions. The objective of all clustering algorithms is to divide a set of data points into subsets so that the objects within a subset are similar to each other and objects that are in different subsets have diverse qualities [11], [13], [20]. The fact that there are many different methods used to quantify the similarity and diversity of data points leads to the many different variations of the problem. For our comparison, we defined the clustering problem as the task of dividing an input data set into a desired number of subgroups so that the Euclidean distance between each data point and its corresponding cluster center is minimized. This is a very common method of defining the clustering problem. Total distances of each point to its cluster center is known as the total distance measurement of the clustering and is calculated as shown in (1).

$$S = \sum_{k=1}^{n} \sum_{x \in Ck} \sum_{a=1}^{A} (x_a - mk_a)^2 \qquad (1)$$

In this formula n is the number of clusters, x represents a data point, Ck represents cluster k, mk represents the mean of the cluster k, and A is the total number of attributes for a data point. This formula simply calculates the Euclidean distance of each point in the input data set to its cluster center. It minimizes the total distance measurement of a clustering leads to an optimal cluster solution. This definition, like all clustering definitions, requires finding

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

an optimal collection of subsets for a group of data points. This class of problem is known to be NP-Hard. Work has been done to develop both approximate and exact solution algorithms for solving various clustering problems [23] but the solutions appear to be impractical, as either the number of data points in the input set or the number of clusters desired becomes large. As a result, there have been a wide variety of heuristic algorithms developed for this clustering problem. These algorithms do not guarantee any quality in the solutions they find but they do run in polynomial time with respect to the number of objects in the input data set and the number of desired clusters.

5. GENETIC ALGORITHMS

Genetic Algorithms (GA), first proposed by John Holland in the 1975s, are a category of EC that use concepts derived from evolution. Proper application of a GA finds a balance between exploration and exploitation of a given optimization problem's search space. First, a population of chromosomes is created an initialized. These chromosomes each contain a collection of genes and each gene has a value. A single chromosome is an encoded version of a solution to the problem that the GA is attempting to optimize. The GA performs exploration or exploitation of the problems search space by evolving the population of chromosomes through a series of generations. During each generation of the GA, parent chromosomes are selected from the population. These parent chromosomes are combined to form children chromosomes and then the child chromosomes are mutated. In a generational type GA, an entirely new population for each generation is formed by creating multiple child chromosomes. For a steady state GA, the child chromosomes are used to replace members of the current population but a new population is not formed during each generation.

A very important step in the GA is the selection of parents for the next generation of chromosomes. In order to provide a guided search, which is appropriate for the given optimization problem, the selection of parents needs to be based on the quality of the solution that their chromosomes represent. A property called fitness is used to quantify the quality of a given solution and a fitness function is used to calculate the fitness value of each chromosome in a given population before parent selection is made. A variety of different selection methods are used by GA but they all use the principle that higher fit chromosomes are more likely to be chosen as parents. This fitness selection provides the GA direction for the search of an optimization problems search space.

5.1 A Genetic Algorithm On Image Data

GA has been successfully implemented for various clustering problems using different chromosome encoding schemes and fitness functions. A GA performs clustering on an input set of data objects so that supervised learning can be applied to predict class labels in the second step. The input for the GA is a set of data objects that have both numeric and label attributes and a desired number of clusters. The goal of the GA is to produce clusters of data objects that minimize cluster dispersion and are as pure as possible in relation to the label attributes. The GA uses a two component fitness function where the first component measures within cluster variance using a distance metric and the second component measures the similarity of the labeled attributes of the data objects.

A very large input data set can be preprocessed to make a representative set that can be used by the algorithm for better time and space efficiency. In GA implement two alternate preprocessing methods for clustering algorithm such as

- The first Preprocessing method used random sampling to obtain a data set with fewer points. This reduced data set was then used in evaluating the fitness of the chromosomes.
- The second preprocessing method used summarization of the input data set and is based on the work presented in reference [21].

For this method, a grid is first constructed and then the input data set is applied to this grid. A single point location and corresponding weight is calculated for each region defined by the grid. The location of the representative point is chosen as the mean value of all the points in the region and the weight of the representative point is equal to the number of points that it replaces.

5.2 Pseudocode for Genetic Algorithm

Here, we take n number of chromosomes for n number of centroids,m number of Samples S_j and S_j^i is j^{th} sample assigned to i^{th} chromosome. We take maximum iteration to found best fitness solution(F). In search space at each iteration it simultaneously found objects on label as L_a and input attribute as I_a. First chromosome is consider as a parent chromosome and in each iteration it build child chromosomes. Below are GA rule to find cluster construction.

Genetic Algorithm is more efficient because final solution is evaluated by frequensy of fitness value and samples matching value.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Step 1: Initialize n number of chromosomes to n number of centroids.
Step 2: Iteration(I)<=MaxIter
Step 3: Parent Chromosome P_i is selected randomly.
Step 4: Method1<=n
Step 5: Randomly select one sample in multi dimensional space.
Step 6: Sample j is assigned to chromosome i by

$$S_i^i = 1/n_i\{\log_a(L_a)*\log_a(I_a)\}.$$

Step 7: S_j^i value is either 0 or 1

- If S_j^i =1, Sample is matched to i^{th} chromosomes.
- If S_j^i = 0, Sample j belongs to some other chromosomes.

Step 8: If Chromosomes matched then fitness value of sample j to input attribute is calculated by

$$F_i(I_a) = 1/S_i^i(I_a/n_i).$$

Step 9: Fitness value of sample j for Label attribute is

$$F_i(L_a) = 1/S_i^i(L_a/n_i).$$

Step 10: Step 4 to Step 9 is repeated until Method1 reaches n samples in search space.
Step 11: Fitness value for chromosome i is calculated by

$$F_i^i = 1/n_i\{F_i(I_a)*F_i(L_a)\}.$$

Step 12: Repeat step 2 to step 10, until reach MaxIter.
Step 13: Finally, find minimum fitness value from the solution of each iteration by using frequency as

$$F=\mathrm{Min}\{1/n(f|F_i i-S_i^{i}|^2,0\}.$$

Figure. 2 Pseudocode for Genetic Algorithm

6. Performance Analysis

Clustering is an important task with applications in many fields. Heuristic algorithms are used for this task in an attempt to provide acceptable results, both in terms of solution quality and running time, because all of the non-trivial clustering problem variations are NP-Hard. In this paper comparison are done by using preprocessing techniques and performance measure for these algorithms are done on the basis of time. In K-means algorithm at different runs it produces poor results when the initial centroids are choosing randomly. It is important to realize that the choice of the initial centroid has a huge effect on the final result. K-means algorithm for multiple runs on large data sets is not work and it take more time to complete. For clustering on very large data sets, such as image data sets, the size of the associated databases makes it necessary to modify traditional GA because of their slow running times and combinations of input and label attributes.

Here, comparisons is done on two data sets first one is Artificial data sets created manually and it contains groups of color points. Genetic and K – means algorithm are applied to this data sets and clustering group of data by different colors. Second real group of image data sets is taken from google image. Two data sets contain six numerical attributes with values between 0 and 255. Each data set contains n points with the points centered around k cluster centers or chromosomes. The k cluster centers are first allocated by randomly and each of the six attributes values from a range of 0 to 255 is uniformly allocated in k- means algorithm and randomly allocated in Genetic algorithm. To calculate the minimum distance(D) between two cluster points is defined by D/r, where r is a variable used to define cluster.This process is repeated until it reach data points and clusters.Below are some sample images are done by experiments.

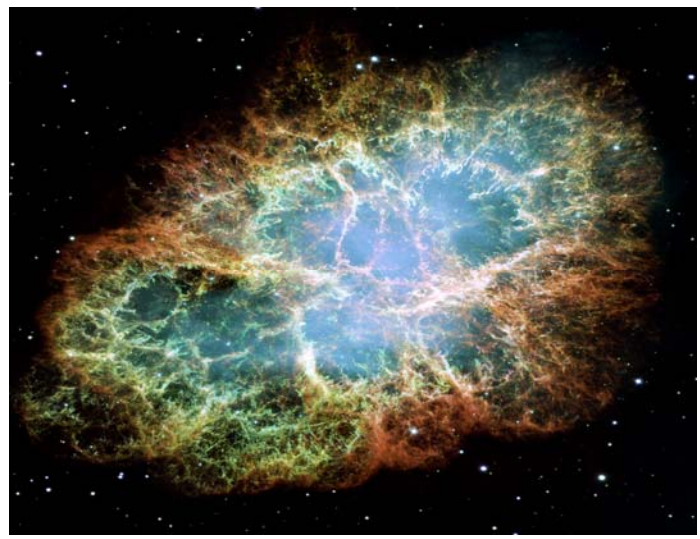


Figure. 3 Original Image taken from Google.

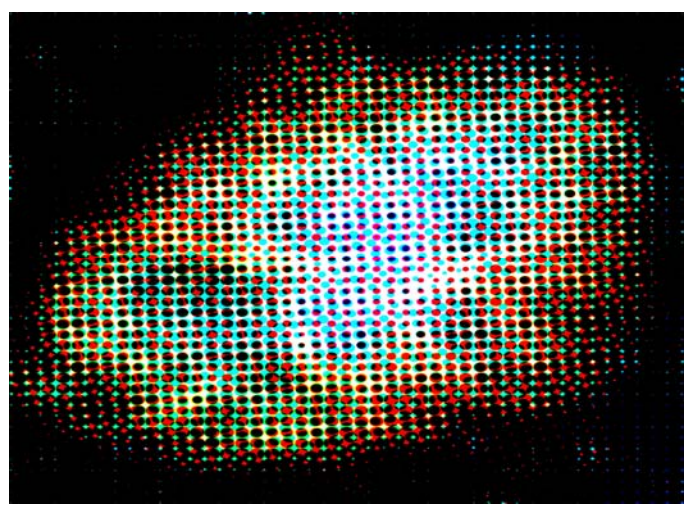


Figure. 4 Using GA to group red space.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

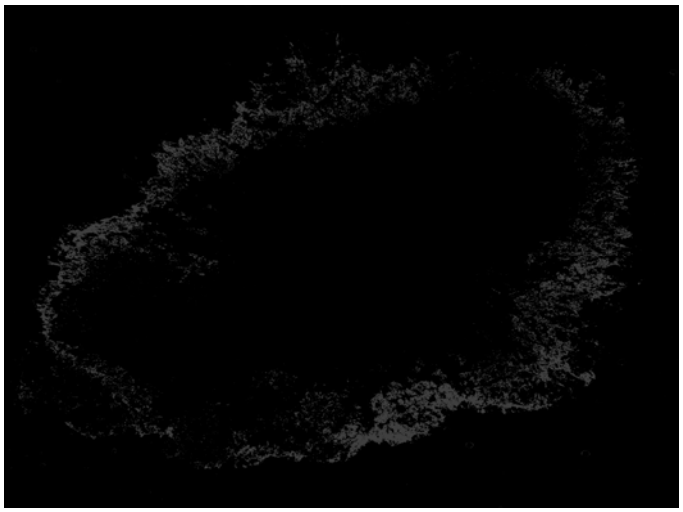


Figure. 5 Group red space using GA view in black and white.

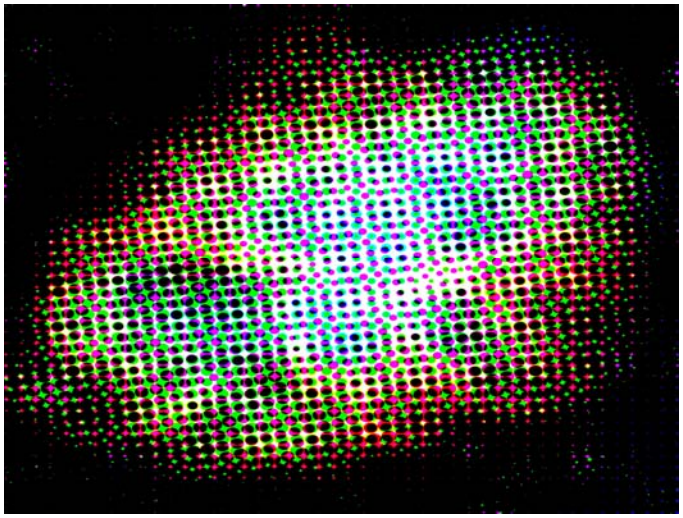


Figure. 6 Using GA to group green space.

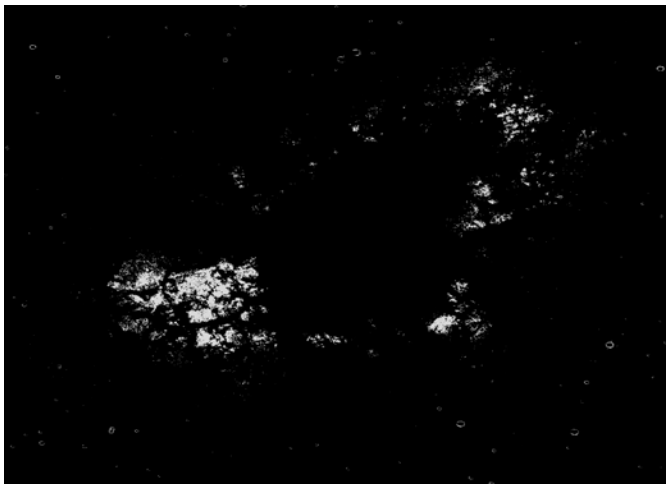


Figure. 7 Group green space using GA view in black and white.

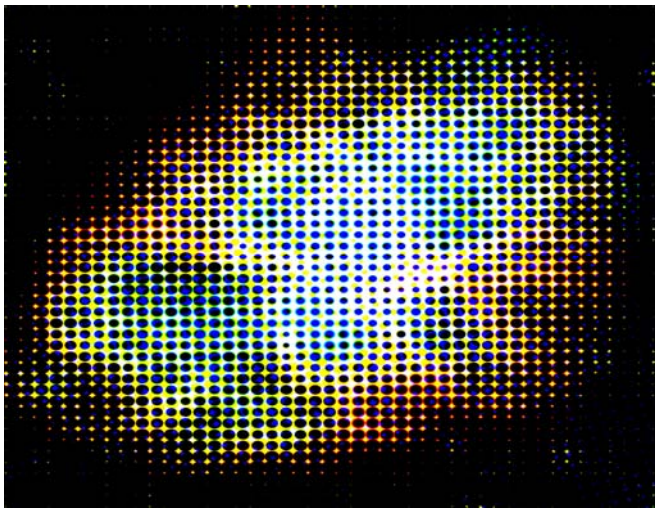


Figure. 8 Using GA to group blue space.

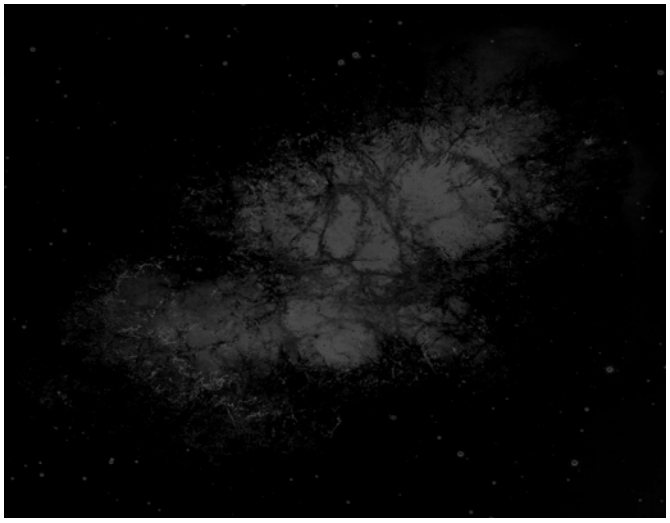


Figure. 9 Group blue space using GA view in black and white.

Table 1: Results On Artificial Data Set

Input set	***Algorithm***	***Running Time(sec)***	***Distance Measurement***
3 Centroids 1000 points	GA	230	5.23×10^{4}
	K-means	219	7.89×10^{4}
5 Centroids 10000 points	GA	457	4.56×10^{5}
	K-means	512	2.34×10^{6}
7 centroids 20000 points	GA	890	8.27×10^{8}
	K-means	2023	$1.10 \text{ X } 10^{10}$

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Table 2: Results On Real Image Data Set

Input set	*Algorithm*	*Running Time(sec)*	*Distance Measurement*
3 Centroids 500 images	GA	153	3.87×10^5
	K-means	127	7.56×10^5
5 Centroids 1000 images	GA	296	8.90×10^5
	K-means	313	2.23×10^6
7 centroids 2000 images	GA	789	9.27×10^7
	K-means	1567	3.43×10^9

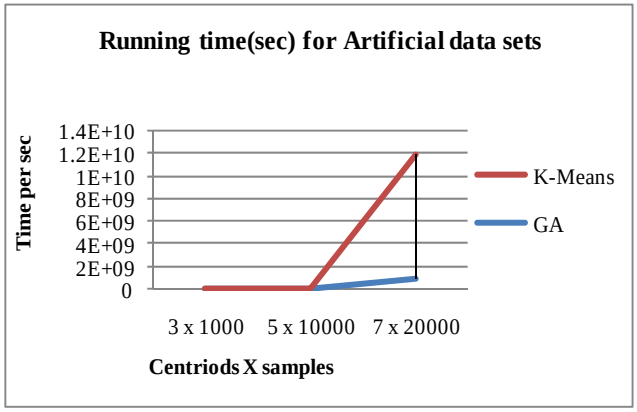


Figure. 10 Performance Analysis for running time using Artificial Data Sets

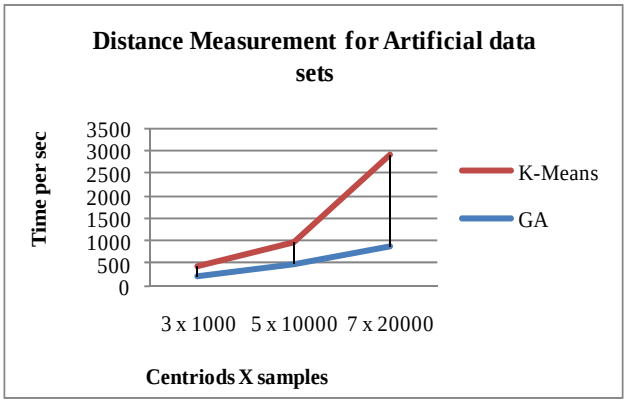


Figure. 11 Performance Analysis for distance measurement using Artificial Data Sets

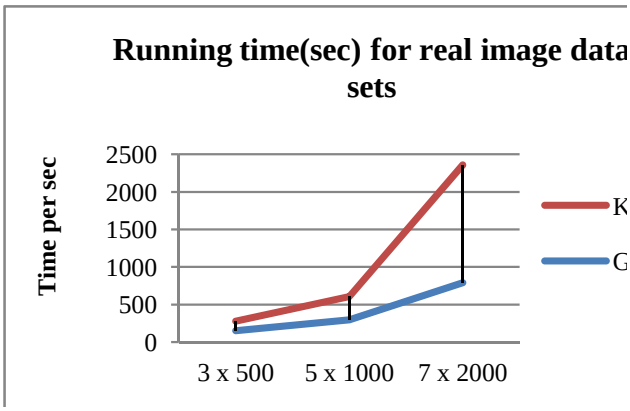


Figure. 12 Performance Analysis for running time usingArtificial Data Sets

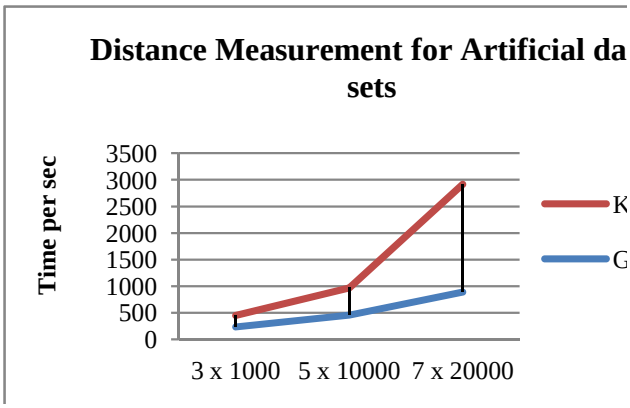


Figure. 13 Performance Analysis for distance measurement using Artificial Data Sets

7. Conclusion

The clustering performance of the two algorithms has been evaluated using real world and artificial data sets. The satisfactory results have demonstrated the effectiveness of the two algorithms in discovering structures in data. The scalability tests have shown that the two algorithms are efficient when clustering. But for large complex data sets in terms of number of records and the number of clusters GA shows better result as compared to k-means algorithm. The summary preprocessing method that prevent the creation of representative points for regions that contained less than a certain minimum threshold of points. This refinement removes the negative effect that outlier points have on the clustering quality. It makes the GA run faster than k-means algorithm because there would be fewer points in the processed data set.

Reference

[1]. Z. Huang, "Extensions to the k-means algorithm for clustering large data sets with categorical values," Data

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

[2]. Mining and Knowledge Discovery, Vol. 2, 1998, pp. 283-304.

[3]. J. R. Wen, J. Y. Nie, and H. J. Zhang, "Query clustering using user logs," ACM Transactions on Information Systems, Vol. 20, 2002, pp. 59-81.

[4]. J. Banfield and A. Raftery, "Model-based gaussian and non-gaussian clustering," Biometrics, Vol. 49, 1993, pp. 15-34.

[5]. J. L. Bentley, "Multidimensional binary search trees used for associative searching," Communications of the ACM, Vol. 18, 1975, pp. 509-517.

[6]. D.A. Clausi, "K-means iterative fisher unsupervised clustering algorithm applied to image texture segmentation," Pattern Recognition, Vol. 35, 2002, pp. 1959-1972.

[7]. F. X. Wu, W. J. Zhang, and A. L. Kusalik, "Determination of the minimum samples size in micro array experiments to cluster genes using K-means clustering," in Proceedings of 3rd IEEE Symposium on Bioinformatics and Bioengineering, 2003, pp. 401-406.

[8]. K. Alsabti, S. Ranka, and V. Singh, "An efficient k-means clustering algorithm," in Proceedings of 1st Workshop on High performance Data Mining, 1998.

[9]. R. C. Dubes and A. K. Jain,"Algorithms for Clustering Data", Prentice Hall, 1988.

[10]. E. R. Ruspini, "A new approach to clustering," Inform. Contr., vol. 19, pp. 22–32, 1969.

[11]. L. Abul, R. Alhajj, F. Polat and K. Barker "Cluster Validity Analysis Using Sub sampling," in proceedings of IEEE International Conference on Systems, Man, and Cybernetics, Washington DC, Oct. 2003 Volume 2: pp. 1435-1440.

[12]. J. Grabmeier and A. Rudolph, "Techniques of cluster algorithms in data mining," Data Mining and Knowledge Discover, 6, 2002, pp. 303-360.

[13]. L. O Hall, I. B. Ozyurt, J. C. Bezdek, "Clustering with a genetically optimized approach," IEEE Transactions on Evolutionary Computation,3(2), 1999, pp. 103-112.

[14]. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," ACM Computing Surveys, 31(3), 1999, pp. 264-323.

[15]. P. Berkhin, "A Survey of Clustering Data Mining Techniques" Kogan, Jacob; Nicholas, Charles; Teboulle, Marc (Eds.) Grouping Multidimensional Data, Springer Press (2006) 25-72.

[16]. P. S. Bradley and U. M. Fayyad, "Refining initial points for k-means clustering," in Proceedings of 15th International Conference on Machine Learning, 1998, pp. 91-99.

[17]. D. Judd, P. McKinley, and A. Jain, "Large-scale parallel data clustering," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, 1998, pp. 871-876.

[18]. T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithm: analysis and implementation," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, 2002, pp. 881-892.

[19]. Demiriz, K. P. Bennett, and M. J. Embrechts, "Semi-supervised clustering using genetic algorithms," R.P.I. Math Report No. 9901,Rensselaer Polytechnic Institute, 1999.

[20]. M. Painho and F. Bação, "Using genetic algorithms in clustering problems," in Proceedings of GeoComputation Conference, 2000.

[21]. J. Han and M. Kamber, "Data Mining: Concepts and Techniques," Morgan Kaufmann Publishers, 2000.

[22]. W. DuMouchel, C. Volinsky, T. Johnson, C. Cortes, and D. Pregibon, "Squashing flat files flatter," in Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 1999, pp. 6-15.

[23]. Ya-Wei Ho; Chih-Hung Wu; Chih-Chin Lai," Aerial image clustering using genetic algorithm,"IEEE transactions on Pattern Analysis and Machine Intelligence, Vol. 24, 2009.

[24]. P. K. Agarwal and C. M. Procopiuc, "Exact and approximation algorithms for clustering," in Proceedings of the ninth annual ACMSIAM symposium on Discrete algorithms, 1998, pp. 658-667.

[25]. Venkatesh Katari,Suresh Chandra Satapathy, JVR Murthy,PVGD Prasad Reddy, "Hybridized Improved Genetic Algorithm with Variable Length Chromosome for Image Clustering", IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.11, November 2007.

[26]. Yang, G., Reinstein, L.E., Pai, S., Xu, Z., Carroll, and D.L., "A new genetic algorithm technique in optimization of prostate implants". Medical Physics, 35(5), pp.104-112.

R.Balakrishnan MSc., M.Phil., Phd.,
He is a Phd Research Scholar in Manonmaniam Sundaranar University, Tirunelveli, Tamilnadu, India. He is working as a Assistant Professor and Head of the department in VLB Janaki Ammal Arts and Science College, Coimbatore. He has 12 years of experience in teaching line 6 years of experience in research. He conducted International, National Conference and he presented paper in International, National Conference and Journals His Interest areas are Data Mining, Image Processing, Current research project Genetic Algorithm using , Image Processing.

U.Karthick Kumar MSc., DCA., MCA., M.Phil.,
He is a Post Graduate with M.Phil from Bharathiar University, Coimbatore, Tamilnadu, India. He is working as a Assistant Professor in VLB Janaki Ammal Arts and Science College, Coimbatore. He has three years of experience in research. He presented paper in International, National Conference and Journals. His Interest areas are Grid Computing, Data Mining, Image Processing, Mobile Computing and Data Structures. Current research project Ant colony, PSO, Priority Based Pheromone Algorithm, Fair Scheduling, Sensor Network, Genetic Algorithm.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Automatic System to Remove Unpleasant Images Detected by Pupil-Size Changes

Koji Kashihara[1], Momoyo Ito[1], and Minoru Fukumi[1]

[1] Institute of Technology and Science, The University of Tokushima
2-1 Minamijyousanjima, Tokushima 770-8506, Japan

Abstract

An automatic filtering system to classify individual unpleasant emotions represented by pupil-size changes was proposed. The support vector machines classifier was applied to single-trial data of pupil size and indicated the possibility of the correct judgment of individually unpleasant states immediately after looking at emotional pictures. The framework was then constructed to automatically filter the similar unpleasant information from a picture database, applying the bag of features scheme to search for similar images.

Keywords: *Pupil Size, Emotion, Database Search, Support Vector Machines*

1. Introduction

For patients with amyotrophic lateral sclerosis (ALS) who cannot freely move their muscles [1], it may be desirable to construct an automatic system such as a brain-machine (BMI) or -computer interface (BCI) to express their emotions and to realize the smooth communication with others. Healthy people also receive various emotional changes to multimedia contents. However, the interactive method considering individually different emotional states has been insufficiently investigated. Individually designed devices to detect the underlying state of human mind would be effective as an interactive communication tool.

A human-machine or -computer interface considering individual preference would be realized by the camera image processing and analysis for facial expressions [2]. However, the individual impression of emotional pictures varies from person to person; there are some difficult cases to speculate or judge the emotional state from the camera image or facial information alone. The idea of an automatic system to abstract individually favorite data or to exclude unpleasant ones would increase the usability for the websites and databases with various kinds of images.

Individual emotions can be estimated by some physiological parameters. For example, the autonomic nervous system (heart rate, pupil size, etc.) as well as brain activity is known as an index associated with emotional states. Pupil size is particularly modulated by emotional changes such as pleasure and displeasure [3]. Because the pupil size can be noninvasively recorded with an eye camera, it is easier to work toward practical use of various devices.

There are a few examples as the application of eye-movement signals such as eye tracking and saccades [4]. However, the analytical method for a classifier to detect an individual's emotion based on pupil-size changes has never been established. Accordingly, this study aimed to develop a framework for an automatic system (1) to abstract individually unpleasant images from pupil-size changes, (2) to search for similar pictures from a database without textual information, and finally (3) to filter them from image materials such as websites.

The support vector machines (SVM) classifier with the high rate of classification in electroencephalography (EEG) signals was applied for this study. The noninvasive measurement of EEG signals requires the extraction of meaningful data from low spatial resolution with artificial noises. An SVM classifier can actually solve such a difficulty and has indicated highly accurate performance to detect important information from single-trial signals [5]. Therefore, we investigated the optimal features and parameters of SVM to classify the pupil-size change reflecting emotion. The automatic filtering system was then constructed to remove individually unpleasant information from a picture database, using the bag of features (BoF) scheme.

2. System Design

An automatic system to extract an individually unpleasant image, to search for similar images, and to filter them in multimedia contents was developed. In accordance with the following steps, the proposed system was designed.

(1) Measurement of Pupil Size. The pupil-size changes to emotional pictures were measured by an eye-mark

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

recorder. The picture stimuli were rated to evaluate the level of subjective emotion.

(2) Detection of Emotional State. The three features for a classifier were extracted after the preprocessing of the raw data on pupil size. The responses to emotional pictures were classified by the SVM scheme. The emotional state was subdivided into two categories: the unpleasant or neutral state.

(3) Validation of Classification. After the SVM classifier was firstly trained using single-trial data of pupil size, the accuracy of the trained classifiers with several features and parameters were assessed by the unknown data.

(4) Searching for Emotional Pictures. Referencing the unpleasant picture individually detected by the pupil-size change, the similar ones were explored from a picture database. Histograms of visual words were used for this method.

(5) Filtering System. On the emotional pictures detected by above procedures, we proposed an automatic system to remove individually unpleasant materials in websites.

3. Analysis of Pupil Size

3.1 Dataset

Participants in the dataset [6] were eight healthy volunteers with normal or corrected-to-normal vision (4 males and 4 females, age: 28.9 ± 4.9 years). Emotional pictures were selected from the International Affective Picture System (IAPS) [7] and included unpleasant and neutral categories. All pictures were converted to 8-bit grayscale bitmaps. The mean and distribution of the luminance of grayscale pictures were adjusted not to differ among the pictures sets. The picture size was 360 × 270 pixels (12.6 × 9.5º of visual angle).

Binocular pupil size was recorded using an EyeLink CL 1000 Desktop (SR Research Co.) with a software for EyeLink (Ver. 4.4) on a PC (DELL Co., Precision T3400), at a sampling rate of 500 Hz. A standard five-point camera calibration was performed; the absolute error of the gaze position was set within 0.5º of visual angle.

Visual stimuli were presented on a screen of a 19-inch monitor (a resolution of 1024 × 768 pixels) at the same height as the eyes. The distance from the stimuli was set at 60 cm. A trial started with a fixation cross (size of 0.76º), which was presented between 1 and 1.5 s. An unpleasant or neutral picture selected from the dataset was randomly presented for 100 ms at the center on a screen. The inter-trial interval was set at 8 s.

The emotional pictures for the visual stimuli were rated, using a 1-7 scale (the range from 1, extremely negative, to 7, extremely positive). In the rating scores, the paired t-test revealed that there was a significant difference between the picture conditions ($p < 0.01$): 4.4 ± 0.9 vs. 2.5 ± 0.5 in the neutral and unpleasant pictures.

3.2 Detection of Pupil-size Change

(1) Preprocessing: The trial data with eye blink artifacts were removed from the analysis. The responses of binocular pupil sizes were averaged at every sampling point. The signals were segmented into epochs ranging from 200 ms before to 1500 ms after stimulus onset. For the baseline correction of pupil sizes, the values of each sampling point were divided by the mean value of the 200-ms prestimulus interval in each block. A great difference between emotional conditions appeared in the average of pupil size between 400 and 1500 ms after stimulus onset.

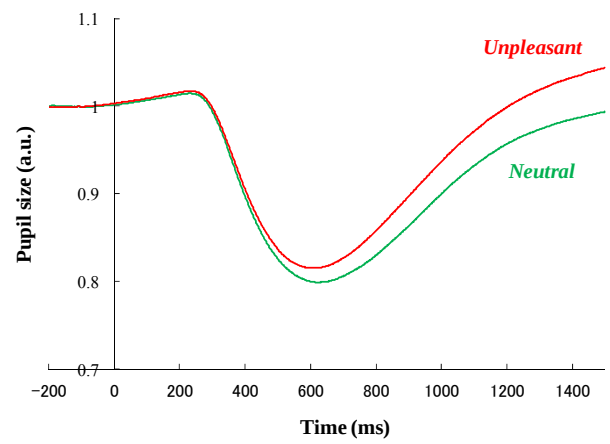


Fig. 1. Average responses of pupil size to emotional pictures.

(2) Feature Extraction: The following features were extracted from the raw data of pupil size and they were utilized for classifiers.

(a) Peak amplitude with latency. The peak value with latency was extracted from each single-trial data because of the remarkably negative responses to visual stimuli (Fig. 1).

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

(b) Time-course raw data. The time-course raw data between 400 and 1500 ms were extracted, focusing on the great change of responses to visual stimuli. Five hundred and fifty data points were used as the feature of each trial.

(c) Time-frequency data. The wavelet transform as an effective tool for the remove of external noises was applied to each trial. The signals were convoluted by the complex Morlet wavelet [8]:

$$w(t, f_0) = \exp\left(\frac{-t^2}{2\sigma_t^2}\right) \cdot \exp(2\pi f_0 it) / \sqrt{\sigma_t \sqrt{\pi}} \,. \qquad (1)$$

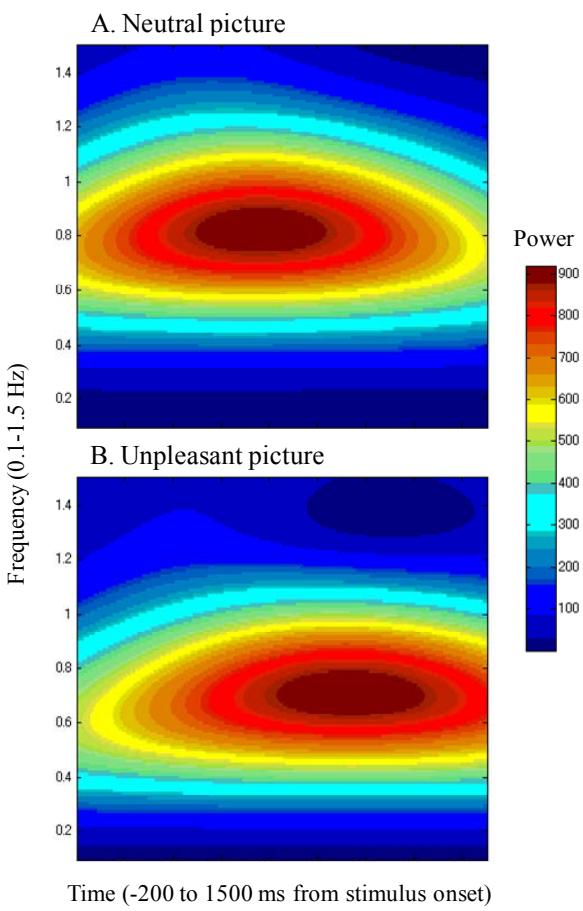


Fig. 2. Examples of the wavelet transform in pupil-size responses to emotional pictures within a subject.

The standard deviation (σ_t) of the time domain is inversely proportional to the standard deviation (σ_f) of the frequency domain [$\sigma_f = (2\pi\sigma_t)^{-1}$]. The f_0/σ_f determining the effective number of oscillation cycles comprised in the wavelet was set at 6 with f_0 ranging from 0.1 to 1.5 Hz in increments of 0.01 Hz.

After a linear trend was subtracted, the continuous wavelet transform of a time series [$u(t)$] was calculated as the convolution of a complex wavelet with $u(t)$:

$$\tilde{u}(t, f_0) = w(t, f_0) * u(t) \qquad (2)$$

The squared norm of the wavelet transform was calculated in a frequency band at around f_0.

(3) The characteristics of the wavelet transform remarkably appeared at frequencies between 0.2 and 1.2 Hz in both conditions (Fig. 2). Hence, the time-frequency powers at those frequencies were determined as the SVM feature every single trial. The latency of the peak power in the unpleasant picture was usually later than that in the neutral picture.

(4) Classification. SVM as a supervised learning method was selected for the classification of features; it could construct a hyperplane that optimally separates samples from two classes with maximum margin (see Appendix for detail), compared to other learning methods [9],[10]. The nonlinear kernel based on the radial basis function was chosen in this analysis.

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \qquad (3)$$

where γ is the kernel parameter to regulate the variance of the function.

4. Validation of Pupil-size Detection

The first three trials in each emotional category were removed for this validation owing to the possibility of initial unstable responses. The next 20 trials in each picture category were used as the training (the first 10 trials) and test (the last 10 trials) data for the ν-SVM analysis, adjusting the optimal parameter C (see Eq. (4) in Appendix) within a subject.

The nonlinear kernel parameter γ of Eq. (3) was tuned for accurate prediction of unknown samples: γ = 0.0001, 0.01, 1, and 1 / (the number of features). The maximum and minimum values of feature vectors of training dataset were recorded to normalize the data within the range between –1 and 1; they were used for the test dataset. For unknown feature vectors of the test dataset, the trained SVM outputs the predicted class labels.

The correct tuning of the parameter γ and the validation of a classifier were performed within each subject,

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

considering the individual differences in emotional responses. The value of the tolerance of termination criterion was set at e^{-12}. The average of the accuracy in each condition was calculated among all subjects. The accuracy was defined by the following equation: (the number of correct classification in the two categories) × 100 / (the total number of trials in the two categories) [%]. The paired *t*-test was applied for the evaluation of the significant differences among the features in each trial. *P* values less than 0.05 were considered statistically significant.

Table 1: Validation of a classifier for pupil size change.

Kernel Features	**Average rates of correct classification (%)**					
	Number of trials					
	1	***2***	***3***	***4***	***5***	***10***
(a) Peak Amplitude	56.3 (41.7)	62.5 (23.1)	66.7 (12.6)	65.6 (8.8)	65.0 (12.0)	62.5 (7.1)
(b) Time-course raw data	87.5 (23.1)*	75.0 (26.7)	72.9 (19.8)	70.3 (13.3)	71.3 (12.5)	65.0 (8.9)
(c) Wavelet transform	87.5 (23.1)*	75.0 (18.9)	68.8 (18.8)	65.6 (17.4)	65.0 (13.1)	61.3 (8.8)

All data are shown as mean ± (SD); *$p < 0.05$ vs. (a).

Table 2: Selected values in the nonlinear kernel parameter.

Subjects No.	***Values in the kernel parameter γ***		
	(a) Peak amplitude	***(b) Time-course raw data***	***(c) Wavelet transform***
1	1, *default*	1	1
2	0.0001	1	0.01
3	1	0.0001	1
4	0.01	0.0001	0.01
5	0.0001	0.0001	0.0001, *default*
6	0.0001	1	1
7	0.01	0.01	0.0001, 1, *default*
8	0.01	*default*	1

Default means 1 / (the number of features).

The statistical result of the accuracy for the SVM classifier with three kernel features is summarized in Table 1. The optimal value of the parameter γ indicating the highest accuracy at the final trial was selected within a subject. On the first trial, the accuracies in the features of raw data and the wavelet transform were significantly higher than that in the peak amplitude with latency. However, those values were gradually attenuated with the increase of trial numbers. Table 2 shows the optimal values selected as the nonlinear kernel parameter; their values were distributed among subjects.

5. Searching for Emotional Pictures

Searing for a picture dataset can be automatically started when one is looking at an unpleasant picture, which is detected by the pupil size change, in accordance with the above methods. The picture dataset (IAPS) including two emotional categories of 60 neutral and 60 unpleasant pictures [7] was set to assess this system.

5.1 Searching Method

The BoF based on the idea of the bag of words scheme for categorization of textural data was selected to search for similar emotional pictures. The BoF shows a flexible and adaptable framework owing to the independent techniques in each step. The following steps were performed to classify emotional images [11].

(1) Detection of image features: Speeded-up robust features (SURF) was utilized for the detection of the features in emotional pictures. SURF is a robust feature descriptor showing the distribution of the pixel intensities within a scale-dependent neighborhood of each interest point; the Haar wavelets to increase robustness and to decrease the computation time are adopted as simple filters.

(2) Codebook generation: Codebooks were made by assigning patch descriptors to a set of predetermined clusters with a vector quantization algorithm. Visual words were created by the k-means algorithm to cluster the feature vectors of SURF and to make a visual vocabulary. Here, the k-means clustering indicates the simplest square-error partitioning method. This algorithm proceeds by iterated assignments of points to their closest cluster centers and recomputation of the cluster centers [12].

(3) Construction of the bag of features: The BoF was constructed by counting the number of patches assigned to each cluster. Each picture was finally represented by a histogram of visual words. The generated histograms make it possible to compute the similarity among emotional images.

(4) Similarity calculation among pictures: Using the histograms constructed as the feature vector, the kd-tree method was applied for the similarity calculation between the target and database pictures. The searched similar pictures were automatically filtered from web materials.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

5.2 Triggered Filtering System

As a possible method for searching individually emotional pictures, similar pictures were explored from the database. The kd-tree method [13], which is a fast approximate nearest neighbor search algorithm, was applied for the selection of similar images and the similarity between images was calculated.

(1) Searching for similar pictures: Referencing the target picture detected by the pupil-size change, the top three similar images were searched from a picture database. The average of accuracy to correctly classify the unpleasant category was 89.4 % (161 of 180 images); that of a similarity between the target and searched images was 79.0 %.

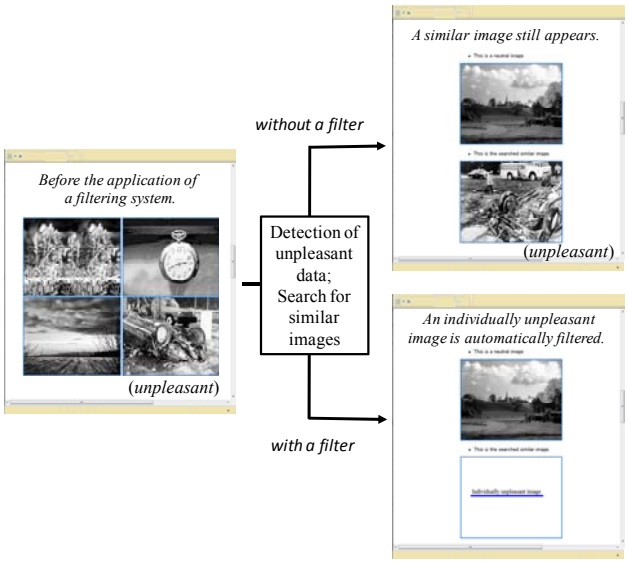


Fig. 3. An example of the applied filtering system for individually unpleasant information.

(2) Application to a database: Figure 3 shows an example of final output in the proposed system after the detection of a pupil-size change. The individually unpleasant pictures were automatically filtered immediately after the detection of an emotional change. The unpleasant image was firstly replaced with the link text; the hidden picture was able to be shown anytime by clicking it.

6. Discussion

The SVM classifier with the features of the raw data and the wavelet data resulted in the high accuracies for emotional judgment (Table 1). However, those accuracies were gradually decreased with the increase of trial numbers, suggesting that the nearest training data are effective for single-trial analysis in actual situations such as BMI and BCI. Therefore, it is desirable to recalculate the optimal parameters for SVM, renewing the training data recursively. The auto-tuning method for the optimal parameters might be also required to correctly classify individual preferences.

The classifier's accuracy in the peak response with latency showed the smallest of all features, which will be due to the shortage of significant features in time-course data. The meaningful changes in time-course data should be efficiently extracted for the selection of the classifier's features. Whereas the raw data include all features of time-course changes, there are some problems such as background noise and calculation time. In the single-trial data of pupil size, the features acquired from the wavelet transform will be effective for noise reduction [8].

Specific methodology would be required to search for emotional images because most of them are constructed by the combination of some objects and/or situations. For example, we probably feel different emotions when looking at a cute puppy or a barking dog, regardless of the same animal category and the very similar image features. A possible method for the improvement is the analysis of the text-based information on emotional images; however, it is difficult to apply it to all images in websites and databases. Considering the increment of a classifier's rate for generic object recognition, the algorithms for image processing to be able to judge and classify emotional situations will be needed in future studies.

7. Conclusions

An automatic system to identify individual emotion based on a pupil-size change was proposed. The SVM classifier applied to single-trial data of pupil size indicated the possibility of the correct judgment of the individually different unpleasant states when looking at emotional pictures. We also constructed the framework to automatically filter the individually unpleasant information from a picture database. The BoF scheme was able to efficiently search for similar ones detected from a pupil-size change. Further studies will be needed to apply the proposed system to tools for medical welfare and practical devices.

Appendix

The SVM algorithm is applied for a training dataset consisting of N samples. The principle of SVM for the

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

two-class problem is to optimally determine a separating hyperplane with largest margin between the classes.

The non-linear SVM can map the feature vector $\boldsymbol{x}$ into a high dimensional Euclidean space, using a nonlinear mapping function $\boldsymbol{\Phi}$. The given data can be separated by a hyperplane with the mapping function $\boldsymbol{\Phi}$ to a sufficiently high dimension. The decision boundary of the two-class problems takes the form of an optimal separating hyperplane, $\boldsymbol{w}\cdot\boldsymbol{\Phi}(\boldsymbol{x}) + b = 0$, obtained by solving the convex optimization problem:

$$\min_{w,b,\xi}\left(\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i\right) \tag{4}$$

subject to $y_i(w\cdot\Phi(x_i)+b)+\xi_i \geq 1$ and $\xi_i \geq 0$ ($i = 1,\cdots, N$),

where $\boldsymbol{x}_i$ is the ith component and $\boldsymbol{y}_i \in \{-1, +1\}$ is the class label. The parameter $\boldsymbol{w}$ is a weight vector, b is a bias, and ξ is the slack variable. A parameter C controls the size of $\boldsymbol{w}$ and the sum of ξ_i. For a large C value, a penalty is largely assigned to errors. The two points closest to the hyperplane strongly affect its orientation, leading to a hyperplane coming close to several other data points. For a smaller C value, those points move inside the margin; the orientation of the hyperplane is changed, resulting in a larger margin. Instead of such a method, there is a formulation of SVM using a parameter $0 < v \leq 1$. The parameter v regulates the fractions of support vectors and margin errors (v-SVM).

With the Lagrange multiplier α_i, the dual formation of (4) can be expressed as follows.

$$\max_{\alpha}\left(\sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{N}\alpha_i\alpha_j y_i y_j K(x_i,x_j)\right) \tag{5}$$

subject to $\sum_{i=1}^{N}\alpha_i y_i = 0$ and $0 \leq \alpha_i \leq C$, ($i = 1,\cdots, N$), where the kernel function is $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{\Phi}(\boldsymbol{x}_i)\cdot\boldsymbol{\Phi}(\boldsymbol{x}_j)$.

The hyperplane ($\boldsymbol{w}\cdot\boldsymbol{\Phi}(\boldsymbol{x}) + b = 0$) is expressed as

$$f(x) = \sum_{i=1}^{N}\alpha_i y_i K(x_i, x) + b\cdot \tag{6}$$

This equation shows the decision function for all unknown samples $\boldsymbol{x}$. If $f(\boldsymbol{x})$ is larger than 0, then the predicted class is assigned to +1 (-1 otherwise).

Acknowledgment

This study was funded by a Grant-in-Aid for Young Scientists (B) from the Ministry of Education, Culture, Sports, Science and Technology of Japan (KAKENHI, 22700466).

References

[1] M. A. Lebedev, and M. A. Nicolelis, "Brain-machine interfaces: past, present and future", Trends Neurosci., Vol. 29, 2006, pp. 536-546.

[2] S. Ioannou, A. Raouzaiou, V. Tzouvaras, T. Mailis, K. Karpouzis, and S. Kollias, "Emotion recognition through facial expression analysis based on a neurofuzzy network", Neural Networks, Vol. 18, 2005, pp. 423-435.

[3] M. M. Bradley, L. Miccoli, M. A. Escrig, and P. J. Lang, "The pupil as a measure of emotional arousal and autonomic activation", Psychophysiology, Vol. 45, 2008, pp. 602-607.

[4] M. Porta. "Vision-based user interfaces: methods and applications", International Journal of Human-Computer Studies, Vol. 57, 2002, pp. 27-73.

[5] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain-computer interfaces", Journal of Neural Engineering, Vol. 4, 2007, R1-R13.

[6] K. Kashihara, K. Okanoya, and N. Kawai, "Effects of emotional pictures on eye movements", Technical Report of IEICE, HCS 2010-8, HIP 2010-8, 2010, pp.41-46.

[7] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "International Affective Picture System (IAPS): Affective ratings of pictures and instruction manual (Technical Report A-6)", Gainesville: University of Florida, 2005.

[8] C. Tallon-Baudry, O. Bertrand, C. Delpuech, and J. Pernier, "Stimulus-specificity of phase-locked and non phase-locked 40-Hz visual responses in human", Journal of Neuroscience, Vol. 16, 1996, pp. 4240-4249.

[9] C. Cortes, and V. N. Vapnik, "Support vector networks", Machine Learning, Vol. 20, 1995, pp. 273–297.

[10] N. Cristianini, and J. Shawe-Taylor, "Introduction to support vector machines and other kernel-based learning methods", Cambridge University Press, Cambridge, 2000.

[11] G. Csurka, C. R. Dance, L. Fan, J. Williamowski, and C. Bray, "Visual categorization with bags of keypoints", Proceedings of the IEEE Workshop on Statistical Learning in Computer Vision (SLCV'04), 2004, pp. 1-16.

[12] R. O. Duda, P. E. Hart, and D. G. Stork, "Pattern classification", second ed., John Wiley & Sons, Singapore, 2001.

[13] M. Muja, and D. G. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration", in VISSAPP, 2009, pp. 331--340.

Koji Kashihara received the Ph.D degree in Engineering from Tokyo Institute of Technology, Japan, in 2001. He is an associate professor at the University of Tokushima.

Momoyo Ito is an assistant professor at the University of Tokushima, Japan (Ph.D in Engineering).

Minoru Fukumi is a professor at the University of Tokushima, Japan (Ph.D in Engineering).

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

A Discussion on Elliptic Curve Cryptography and Its Applications

Sharad Kumar Verma[1] and Dr. D.B. Ojha[2]

[1]Research Scholar, Mewar University, Rajasthan, India

[2]Professor, Department of Mathematics, Mewar Institute of Technology, Ghaziabad, UP, India

Abstract

Elliptic curve cryptography (ECC) is a kind of public key cryptosystem like RSA. But it differs from RSA in its quicker evolving capacity and by providing attractive and alternative way to researchers of cryptographic algorithm. The security level which is given by RSA can be provided even by smaller keys of ECC (for example, a 160 bit ECC has roughly the same security strength as 1024 bit RSA). In this paper, we will present some ECC algorithms and also gives mathematical explanations on the working of these algorithms.

Keywords: *Elliptic Curve, cryptography, cryptosystem, RSA.*

1. Introduction

Elliptic curve cryptography was introduced in the mid-1980s independently by Koblitz and Miller [3] as a promising alternative for cryptographic protocols based on the discrete logarithm problem in the multiplicative group of a finite field (e.g., Diffie-Hellman key exchange [5] or ElGamal encryption/signature [8]). ECC is a kind of public key cryptosystem like RSA. But it differs from RSA in its quicker evolving capacity and by providing attractive and alternative way to researchers of cryptographic algorithm. The security level which is given by RSA, can be provided even by smaller keys of ECC. For example, the 1024 bit security strength of a RSA could be offered by 163 bit security strength of ECC.

Elliptical curve cryptography (ECC) is a public key encryption technique based on *elliptic curve theory* that can be used to create faster, smaller, and more efficient cryptographic keys. ECC generates keys through the properties of the elliptic curve equation instead of the traditional method of generation as the product of very large prime numbers. The technology can be used in conjunction with most public key encryption methods, such as RSA, and Diffie-Hellman. According to some researchers, ECC can yield a level of security with a 164-bit key that other systems require a 1,024-bit key to achieve. Because ECC helps to establish equivalent security with lower computing power and battery resource usage, it is becoming widely used for mobile applications. ECC was developed by Certicom, a mobile e-business security provider, and was recently licensed by Hifn, a manufacturer of integrated circuitry (IC) and network security products. RSA has been developing its own version of ECC. Many manufacturers, including 3COM, Cylink, Motorola, Pitney Bowes, Siemens, TRW, and VeriFone have included support for ECC in their products.

An elliptic curve is not an *ellipse* (oval shape), but is represented as a looping line intersecting two axes (lines on a graph used to indicate the position of a point). ECC is based on properties of a particular type of equation created from the mathematical group (a set of values for which operations can be performed on any two members of the group to produce a third member) derived from points where the line intersects the axes. Multiplying a point on the curve by a number will produce another point on the curve, but it is very difficult to find what number was used, even if you know the original point and the result. Equations based on elliptic curves have a characteristic that is very valuable for cryptography purposes: they are relatively easy to perform, and extremely difficult to reverse.

For current cryptographic purposes, an *elliptic curve* is a plane curve which consists of the points satisfying the equation 1 along with a distinguished point at infinity (∞). This set together with the group operation of the elliptic group theory form an Abelian group, with the point at infinity as identity element. The structure of the group is inherited from the divisor group of the underlying algebraic variety.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

The entire security of ECC depends on the ability to compute a point multiplication and the inability to compute the multiplicand given the original and product points [4]. The ECC standards are specified in SEC, Standards for Efficient Cryptography [7].

1.1 Domain parameters

There are certain public constants that are shared between parties involved in secured and trusted ECC communication. This includes curve parameter **a**, **b**, a generator point **G** in the chosen curve, the modulus **p**, order of the curve **n** and the cofactor **h**. There are several standard domain parameters defined by SEC, Standards for Efficient Cryptography [6].

1.2 Point multiplication

Point multiplication is the central operation in ECC. In point multiplication a point P on the elliptic curve is multiplied with a scalar k using elliptic curve equation to obtain another point Q on the same elliptic curve. i.e. k*P=Q

Point multiplication is achieved by two basic elliptic curve operations

- Point addition, adding two points J and K using elliptic curve equation to obtain another point L i.e., L = J + K.
- Point doubling, adding a point J to itself using elliptic curve equation to obtain another point L i.e. L = 2J.

Here is a simple example of point multiplication.

Let P be a point on an elliptic curve. Let k be a scalar that is multiplied with the point P to obtain another point Q on the curve. i.e. to find Q = k*P.

If k = 23 then k*P = 23*P = 2(2(2(2P) + P) + P) + P.

In the ECC explanations given below upper case letter indicates a point in the elliptic curve and the lower case letter indicates a scalar.

1.3 One Way function in ECC

The security of ECC depends on the difficulty of Elliptic Curve Discrete Logarithm Problem. Let P and Q be two points on an elliptic curve such that k*P = Q, where k is a scalar. Q can be easily obtained from P and k but given P and Q, it is computationally infeasible to obtain k, if k is sufficiently large. k is the discrete logarithm of Q to the base P.

2. Elliptic Curve cryptography

An overview of EC cryptographic algorithms for key agreement and digital signature are explained below.

2.1 ECDH – Elliptic curve Diffie Hellman

ECDH, a variant of DH, is a key agreement algorithm. For generating a shared secret between A and B using ECDH, both have to agree up on Elliptic Curve domain parameters. An overview of ECDH is given below.

2.1.1 Key Agreement Algorithm

For establishing shared secret between two device A and B

Step1. Let d_A and d_B be the private key of device A and B respectively, Private keys are random number less than n, where n is a domain parameter.

Step2. Let $Q_A = d_A*G$ and $Q_B = d_B*G$ be the public key of device A and B respectively, G is a domain parameter
Step3. A and B exchanged their public keys
Step4. The end A computes $K = (x_K, y_K) = d_A*Q_B$
Step5. The end B computes $L = (x_L, y_L) = d_B*Q_A$
Step6. Since K=L, shared secret is chosen as x_K

2.1.2 ECDH - Mathematical Explanation

To prove the agreed shared secret K and L at both devices A and B are the same
From Step2, Step4 and Step5

$K = d_A*Q_B = d_A*(d_B*G) = (d_B*d_A)*G = d_B*(d_A*G) = d_B*Q_A = L$

Hence K = L, therefore $x_K = x_L$

Since it is practically impossible to find the private key dA or dB from the public key Q_A or Q_B, its not possible to obtain the shared secret for a third party.

2.2 ECDSA - Elliptic curve Digital Signature Algorithm

Digital Signature Standard (DSA) is a public key algorithm that is used for Digital Signature. The DSA

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

standard is specified FIPS 186-2, Digital Signature Standard [2]. ECDSA is a variant of the Digital Signature Algorithm (DSA). For sending a signed message from A to B, both have to agree up on Elliptic Curve domain parameters. Sender A have a key pair consisting of a private key d_A (a randomly selected integer less than n, where n is the order of the curve, an elliptic curve domain parameter) and a public key $Q_A = d_A*G$ (G is the generator point, an elliptic curve domain parameter). An overview of ECDSA process is defined below.

2.2.1 Signing

Consider the device A that signs the data M that it sends to B.

Step7. Let d_A be A's private key
Step8. Calculate m = HASH (M), where HASH is a hash function, such as SHA-1
Step9. Select a random integer k such that $0<k<n$
Step10. Calculate $r = x_1 \bmod n$, where $(x_1, y_1) = k*G$
Step11. Calculate $s = k^{-1}(m + d_A*r) \bmod n$
Step12. The signature is the pair (r, s)

2.2.2 Verification

Step13. Let M be the message and (r, s) be the signature received from A.
Step14. Let Q_A be A's public key. Since Q_A is public, B has access to it.
Step15. Calculate m = HASH (M)
Step16. Calculate $w = s^{-1} \bmod n$
Step17. Calculate $u_1 = m*w \bmod n$ and $u_2 = r*w \bmod n$
Step18. Calculate $(x_1, y_1) = u_1*G + u_2*Q_A$
Step19. The signature is valid if $x_1 = r \bmod n$, invalid otherwise.

2.2.3 ECDSA - Mathematical Explanation

From the verification step19, the signature is valid if $x_1=r \bmod n$ ---- [EX1]
But from Step18, x_1 is the x-coordinate of equation $u_1*G+u_2*Q_A$
From Step10 r is the x-coordinate of equation k*G
Thus to prove equation EX1, It has to prove that $u_1*G+u_2*Q_A = k*G$
Substituting the value of u1 and u2 from Step17 the first part of the above equation
$u_1*G+u_2*Q_A = (m*w \bmod n)*G + (r*w \bmod n)*Q_A$
But $Q_A=d_A*G$, therefore
$u_1*G+u_2*Q_A = (m*w \bmod n)*G + (r*w*d_A \bmod n)*G$, i.e.
$u_1*G+u_2*Q_A = (w*(m+r*d_A) \bmod n)*G$
But from Step16, $w=s^{-1} \bmod n$, i.e $s^{-1}\equiv w \bmod n$ therefore
$u_1*G+u_2*Q_A = (s^{-1}*(m+r*d_A) \bmod n)*G$ ----- [EX2]
but from equation Step11
$s=k^{-1}(m+d_A*r) \bmod n$, i.e. $k = s- 1(m+d_A*r) \bmod n$
Substituting k in EX2
$u_1*G + u_2*Q_A = k*G$
Therefore $x_1=r \bmod n$

3. Equations

$$y^2 = x^3 + ax + b \quad (1)$$

4. Conclusion

Elliptical curve cryptography (ECC) is a public key encryption technique based on *elliptic curve theory* that can be used to create faster, smaller, and more efficient cryptographic keys. This is paper gives a crystal clear picture of a comparative study between ECC and RSA, ECC's advantages and some application of ECC like ECDSA. The demonstration included some of the theoretical and practical aspects of ECC.

References

[1] Neal Koblitz. Elliptic curve cryptosystems. Mathematics of Computation, 48:203-209, 1987.

[2] FIPS PUB 186-2, Digital Signature Standard (DSS), January 2000.

[3] Victor S. Miller. Use of elliptic curves in cryptography. In H.C. Williams, editor, *Advances in Cryptology CRYPTO'85*, vol. 218 of *Lecture Notes in Computer Science*, pp. 417-426. Springer-Verlag, 1986.

[4] http://en.wikipedia.org/wiki/Elliptic_curve_cryptography.

[5] Whitfield Diffie and Martin E. Hellman. New directions in cryptography, *IEEE Transactions on Information Theory*, 22(6):644-654, 1976.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

[6] Certicom, Standards for Efficient Cryptography, SEC 2: Recommended Elliptic Curve Domain Parameters, Version 1.0, September 2000.

[7] Certicom, Standards for Efficient Cryptography, SEC 1: Elliptic Curve Cryptography, Version 1.0, September 2000.

[8] Taher ElGamal. A public key cryptosystem and a signature scheme based on discrete logarithms. *IEEE Transactions on Information Theory*, 31(4):469- 472, 1985.

Sharad Kumar Verma, received his Bachelor of computer application (BCA) degree from MCRPV, Bhopal(MP), INDIA in 2004, Master of computer application (MCA) degree from UPTU Lucknow(UP), INDIA in 2007, and currently pursuing Ph.D in computer science (Network Security) from MEWAR University, Rajasthan, INDIA. He has more than three years of teaching experience in Meerut Institute of Engineering & Technology, Meerut (UP) INDIA. He is the author/co-author of more than 5 publications in reputed journals. The research fields of interest are Coding Theory and Time Synchronization in Wireless network.

Dr. Deo Brat Ojha, Birth Place & date –Bokaro Steel City, (Jharkhand), INDIA on 05/07/1975. Ph.D from Department of Applied Mathematics, Institute of Technology, Banaras Hindu University, Varanasi (U.P.), INDIA in 2004. The degree field is Optimization Techniques In Mathematical Programming. The major field of study is Functional Analysis. He has more than seven year teaching experience as PROFESSOR & more than eight year research experience. . He is working at MEWAR Institute of Technology, Ghaziabad (U.P.), INDIA. He is the author/co-author of more than 50 publications in technical journals and conferences.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

A Hand Finding and Segmenting in a Partially Cluttered Scene

R. M. Farouk[1], Gamal F. Elhadi[2] and Abdalhakeem. T. Issa[3]

[1]Department of Mathematics, Faculty of Science, P.O. Box 44519,
Zagazig University, Zagazig Egypt.

[2]Department of Computer Science, Faculty of Computers and information,
University of menoufia, Egypt.

[3]Department of Computer Engineering, DCC, Shaqra University, KSA

Abstract

In this paper, we discuss a system for finding and segmenting a human hand, which is holding an object. This is a difficult computer vision problem, because of the many degrees of freedom, the flexibility, and the clutter problem of the hand. In order to find the correspondences, we have used a lateral similarity function. In the lateral excitation we have not only compare-ed different features at one point but also the features at that point and its neighbors (e.g. Gobar feature, and skin color feature). The system is an extension of the Elastic Graph Matching (EGM) algorithm [1]. EGM has been shown to be successful in numerous object vision tasks, its best performance being in face recognition.

Keywords: Hand detection, object detection, object recognition, lateral excitation, elastic graph matching, Gabor wavelets.

1. Introduction

Recognizing non-rigid objects like the human hand is a very hard computer vision problem. Different methods have been presented in the literature, whose main drawbacks are low robustness or high computational load in the analysis of cluttered scenes [1]. In order to achieve immersive human computer interaction, human body parts like the hand could be considered as a natural input which motivates the research of tracking, analyzing, and recognizing human body movements.

As an application example, the gestures could be used to represent some commanding inputs such as pointing, rotating, starting, stopping, etc. One of the important characteristics of human body parts is the skin color. Many good algorithms are based on the skin color to detect the hand in the scene [2], [3].

We have used the skin color feature in our algorithm, and we have found that the lighting and background (as in many other algorithms) play a big role in the response of the feature. The human hand has received attention from many researchers in the last decade. Many of the researchers have concentrated in hand tracking [4], finger-print identification [5], hand sign recognition, and hand segmentation [6]. In our system we have looked for the characteristics of the hand and tried to find the correspondences of these characteristics by Labeled Graph Matching [12].

The condition of perfect reconstruction of the signal from its wavelet coefficients, the fast implementation algorithm is given as well [16]. Efficient array architectures for multi-dimensional (m-D) Discrete Wavelet Transform (DWT), e.g. m=2, 3, are presented in [17]. An analysis of the effect of partial occlusion on facial expression recognition is investigated in [18]. Tracking people through partial occlusion is discussed in [19]. Rotation-discriminating template matching based on Fourier coefficients of radial projection with robustness to scaling and partial occlusion is presented in [20].

The rest of this paper is organized as follows. In section II we discuss the object representation with Gabor wavelets and describe the different features that we have used in our graph matching algorithm [7]. In section IV we explain the generation of the algorithm described in section III with the EGM algorithm. In section V we discuss our results and finally the advantages and disadvantages of our system.

2. Object representation

Let $f(x)$ be the distribution of the input image in the color space HSI (Hue, Saturation and Intensity), [15]. We have represented the components of the input image separately: for the intensity component $I(x)$, we have

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

convolved it with a family of Gabor kernels $\psi_k(x)$ as in the formula:

$$\mathrm{WI}(k,x_0)=\int_R \psi_k(x_0-x)I(x)d^2x \qquad (1)$$

the operator W symbolizing the convolution with a family of Gabor wavelet kernels which take the form of a plane wave restricted by a Gaussian envelope function as given by the formula

$$\psi_k(x)=\frac{k^2}{\sigma^2}\exp\left(\frac{-k^2x^2}{2\sigma^2}\right)\left[\exp(ik^Tx)-\exp\left(\frac{-\sigma^2}{2}\right)\right] \quad (2)$$

Which are parameterized by 2-dimensional vector $\vec{k}$. The first term in the square brackets determines the oscillatory part of the kernel, the second term compensates for the DC-value of the kernel, to avoid unwanted dependence of the filter response on the absolute intensity of the image. For sufficiently high values of σ the effect of the DC-term becomes negligible. The complex valued ψ_k combine an even (cosine-type) and odd (sine-type) part. The filter response of ψ_k in frequency space is given by:

$$(F\psi_k)(k_0)=\exp\left(\frac{\sigma^2(k_0-k)^2}{2k^2}\right)-\exp\left(\frac{\sigma^2(k_0+k)^2}{2k^2}\right) \quad (3)$$

Where F denotes the Fourier transform. The first Gaussian centered at the characteristic frequency k provides a band pass filter. The second exponential removes the DC-component of ψ_k. Equation (3) does not normalize the energy picked up by the kernel in the convolution. Consequently, this energy will be proportional to $\left|k^2\right|$. D. Field [8] noted that the power spectrum of natural images decreases like $\frac{1}{\left|k^2\right|}$. The energy resulting components of our image transform should therefore be roughly independent of $\left|k^2\right|$. The ψ_k form a family that is self-similar under the application of the group of translation, rotation, and scaling. The family is also known as "Gabor wave-lets". The wavelets are parameterized by the wave vector k, which controls the width of the Gaussian window and the wavelength and orientation of the oscillatory part. The parameter σ determines the ratio of window width to wavelength, i.e., the number of oscillations under the envelope function. The Gabor wavelets seem to be a good approximation to the sensitivity profiles of neurons found in visual cortex of higher vertebrates.

In order to construct a group of numerical values of Eq. (1), we employed a discrete lattice of k values

$$k_{\nu\mu}=\begin{pmatrix}k_\nu\cos(\phi_\mu)\\ k_\nu\sin(\phi_\mu)\end{pmatrix},\ k_\nu=\frac{k_{\max}}{f^\nu},\ \phi_\mu=\frac{\mu\pi}{D} \qquad (4)$$

where $\nu \in \{0, 1, ...,L-1\}$, L is the number of scales and $\mu \in \{0, 1, 2, ...,D-1\}$, D is the number of orientations. In our algorithm we have used L = 3, D = 6, $k_{\max}=\frac{\pi}{4.0}$, $f^\nu=0.5$ and $\sigma=\pi$, The wavelet transform as in Eq. (1) produces a complex numbers of oscillating phase. This leads to local similarity optima when matching stored wavelet features to the transform of an image. We call the vector of

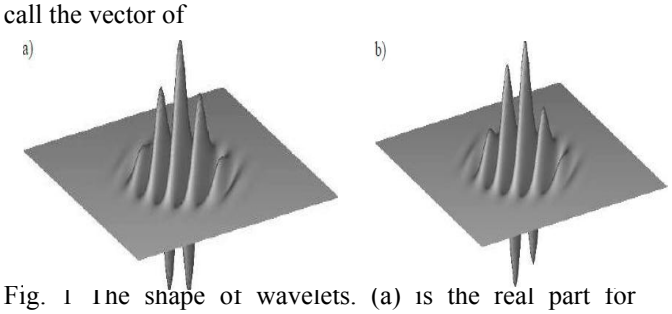


Fig. 1 The shape of wavelets. (a) is the real part for $\phi=45^\circ$ and $\left|\vec{k}\right|=0.72$ (b) is the imaginary part.

the stored wavelet features that taken from the intensity component $I(x)$ at a position x_0 Gabor jet and we can write it in the form:

$$(WI)(k,x_0)=J(k,x_0)\exp^{i\phi(k,x_0)} \qquad (5)$$

where J is the amplitude, which is slowly varying with position, and phase ϕ varying with the spatial frequency given by the characteristic wave vector $\vec{k}$. Gabor wavelets were chosen for their technical properties and biological relevance. Since they are DC-free, they provide robustness against varying illumination in the image. Robustness against varying contrast can be obtained by normalizing the jets. The limited localization in space and frequency yields a certain amount of robustness against translation, distortion, rotation, and scaling.

Only the phase changes drastically with translation, and can there to can be used for estimating displacement. [9, 10] provided more information about wavelets and its

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

properties. For two jets, J extracted at one point in an image, and J' at another point in the same image or a another image. There are two similarity functions [11]: the first uses only the absolute value in Eq. (5), and it is given by the formula:

$$A_{abs}(J,J') = \frac{J.J'}{\|J\|\|J'\|} \quad (6)$$

The second similarity function uses the phase and amplitude and defined by the formula:

$$A_{phase}(J,J') = \frac{J.J'\cos(\phi_1 - \phi_2)}{\|J\|\|J'\|} \quad (7)$$

The second feature that we have used is a local color average in HSI (for transform from RGB color space to the HSI color space [15]. Two image regions are compared by their color averages. We consider regions of the 8-neighbors to compare color regions with the formula:

$$A_{col}(J,J') = \frac{\langle \underline{\Gamma}(J), \underline{\Gamma}(J') \rangle}{\|\underline{\Gamma}(J)\| . \|\underline{\Gamma}(J')\|} \quad (8)$$

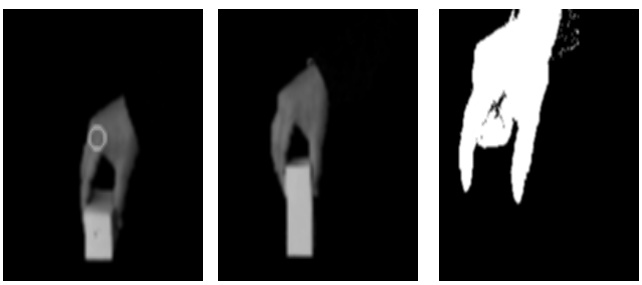


Fig. 2. In the left is the input image with a circle about the region that we try to find its correspondence in the second image, a segmented skin region according to Eq. (9), and Eq. (10) in the right respectively.

Where

$$\underline{\Gamma}(\Omega) = (\tau_H \Omega_H, \tau_S \Omega_S, \tau_I \Omega_I)^T$$

is the 8-neighbors averaged color. The result of this feature is displayed in Fig. 2 . The third feature that we have used is a skin color similarity image comparing each pixel in the original image with a prototype skin color (H_0, S_0) as in the equation:

$$B_{skin} = \chi\left(1 - \sqrt{\left(\frac{H - H_0}{\tau_H}\right)^2 + \left(\frac{S_H - S_0}{\tau_S}\right)^2}\right) \quad (9)$$

Where $\tau_H = 1$, and $\tau_S = 0.2$ are the normalization factors, and $\chi(x)$ is a threshold function defined at time x > 0 by the equation:

$$\chi(x) = \begin{cases} 1 & If \quad x \geq 0; \\ 0 & If \quad x < 0. \end{cases} \quad (10)$$

We have extracted the Gabor jet J at each pixel from the resulting image B skin. To compare two jets J extracted at one point in an image and J' at another point in the same image or a another image, we have used Eq.(7).

2.1 Jet of Multiple Features

We have extracted the three different features at each pixel in the database. In order to compare two jets with the multiple features. We have used the weighted sum of the different similarities features as in:

$$A_{total}(J,J') = \sum_n w_n A_n(J,J'), \quad \sum_n w_n = 1 \quad (11)$$

Where n is an index over the multiple features, and w_n are normalized weighting factors. Experiments as in Fig. 1, have shown that the skin color segmented feature plays the major role in finding the corresponding region.

3. Elastic Graph Matching

A visual pattern can be represented via a graph containing nodes labeled with local features and links encoding the topological relationship between the features [11]. Based on this representation, the problem of pattern recognition can be formulated as the one-to-one correspondence between the nodes of an input graph and one from the database. A good correspondence is one that respects the topological relationships between the nodes, and finds high similarity between the labels of the correspondence nodes. In the neural implementation of EGM, two separate neural layers represent the input and the stored graphs. Graph nodes and graph links are represented by neurons and excitatory connection between the neurons, respectively. Node labels are represented by receptive field profiles of the neurons. The correspondence between the two graphs is found through a dynamical process where at the final state each neuron in the input layer is mapped to its corresponding neuron in the stored model layer [10].

4. Lateral Excitation

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

The traditional measure of similarity A between two Gabor jets J, and J' is the cosine of the angle between the two jets is given by Eq. (6), and Eq. (7). This measure is useful as it provides robustness to variations in degree of contrast. The optimization of similarity between node labels $V = \{J_1, J_2, ..., J_n\}$ and $V' = \{J'_1, J'_2, ..., J'_n\}$ of two graphs, G and G', respectively, requires the definition of a cost function A_{graph}. A natural choice is the sum of all pair-wise jets similarities:

$$A_{graph}((G,G') = \frac{1}{n}\sum_{i=1}^{n} A_{abs}\left(J_i, J'(i)\right) \qquad (12)$$

The previous algorithmic implementations of graph matching for object recognition have used this cost function. An alternative cost function would be to use a count of nodes with a similarity beyond a threshold. While the relative performance of alternative cost function including the latter scheme have not been documented, the latter method seems to be inferior in that it requires defining a threshold for the similarity values to be accepted as high. In our new version of EGM, the similarity between two jets is computed in our algorithm according to Eq. (11). However, we augmented the graph similarity function with an element that emphasizes the topological coherence of the match. The new graph similarity function A_{lat} involves the enhancement of each pair-wise similarity value $\overline{A}_{total}$ by its neighboring similarity values as in the following formula.

$$A_{lat}(G,G') = \frac{1}{n}\sum_{i=1}^{n} \overline{A}_{total}\left(J_i, J'_c(i)\right) \qquad (13)$$

$$\overline{A}_{total}\left(J_i, J'_c(i)\right) = A_{total}\left(J_i, J'_c(i)\right) + A_{total}\left(J_i, J'_c(i)\right)\sum_k A_{total}\left(J_k, J'_{c(k)}\right)$$

Where k is the index of neighbors of J_i in the graph topology, and $c(k)$ is the index of the jet matched with J_k (for biological analog of this function [14]). The second term in Eq. (13) is the excitation received by $A_{total} = \left(J_i, J'_c(i)\right)$. This is consistent with the physiological findings in the visual cortex. It has been shown that the level of excitation induced by activating the horizontal inputs depends on the level of depolarization of the target cell: the more depolarized the cell is, the larger the excitatory postsynaptic potential, as the result of voltage-dependent sodium conductance. Thus one can think of the effect of horizontal connections as being state dependent and influenced by the level of activation of other inputs converging onto the cell. The motivation behind the similarity measure with lateral excitation was to fortify the algorithm against noise and clutter.

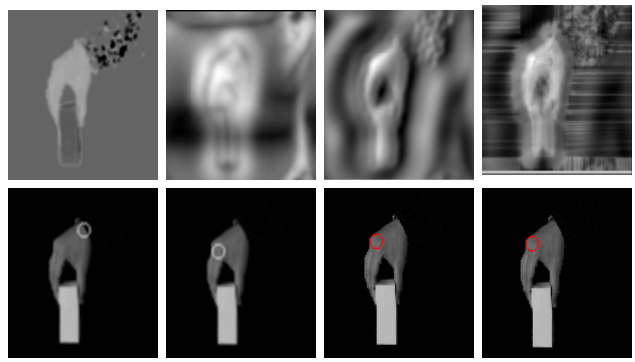


Fig. 3. The first row from left to right shows the color similarity distribution computed from Eq. (8), the Gabor jet similarity as defined in Eq. (6), the similarity of the segmented skin image Eq. (7), total similarity of different features computed with the form in Eq. (11), respectively. The second row marked the highest similarity to the circled region in the input image in Fig. 2.

5. Experiments and Results

We have tested our system on a database of pictures taken by a Sony fire wire camera with a constant back-ground, different lighting degrees, and complex back-ground (Fig. 3). In our test we have found that the system is sensitive against background and the illumination conditions as in many different algorithms.

Another condition that we have tried to control is the pose of the grip of the hand. Due to the flexibility of the hand, and different characteristics of each pose of the hand, we have found that the system is tolerant with regard to slight pose changes. The use of the lateral similarity in Eq. (13) is practically important, because the second term in Eq. (13) is the excitation received by $A_{total} = \left(J_i, J'_c(i)\right)$. As it can be seen, the amount of excitation directly depends on the value of $A_{total} = \left(J_i, J'_c(i)\right)$. The matching process declined when the pose difference became greater. In our experiment we have found that the matching depends upon the background, lighting, and the pose of the grip. Some of our results have been shown in Fig. 4. As it can be seen, the correspondence in the first example (the first row in Fig. 4) is perfectly matched because we have tried to

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

control the pose as good as possible, and then we have succeeded to segment the hand perfectly.

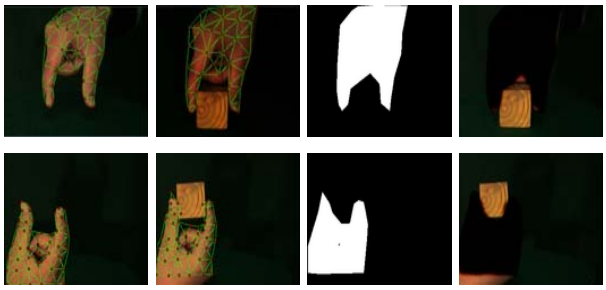


Fig. 4. In the first row from left to right: The grip of the hand with the model graph, the model graph after the matching process, segmented part of the hand and the separated part of the object after separating it from the hand. The second row shows another example with a different grip.

6. Discussion

The finding of non-rigid parts of the human body in our opinion is not a trivial computer vision tasks. The main aim from our system was how to find the hand which is partially occluded by another objects, and then separating the object from the hand by segmenting the hand. We have found that the task is not trivial because the hand has no fixed characteristics, but we have tried to extract features at the most important parts of the hand like finger tips, finger joints and other important positions. The system for the constant background, and the slightly pose gives a good results. The slighter the pose is, the better results are. The Elastic Graph Matching (EGM) [1] was introduced over a decade ago as a biologically inspired method for pattern recognition. Our results demonstrated that despite its age, EGM remains a competitive algorithm, even in comparison with the modern analytically developed statistical information processing methods. EGM with lateral excitations (similarity) in Eq. (13) reveals the following characteristic. Given equal similarity values across all the nodes in the graph for a given match, the nodes at the graph boundary receive less excitation from their neighbors than the nodes in the center because they have fewer neighbors. In other words, given random similarity values, the nodes at object boundary are weighted less than those in the center, and thus their contribution to the total graph similarity is reduced. Of course this argument only holds on average. In cases where a boundary node is surrounded by very high similarity values, or where a center node is surrounded by low similarity nodes, this will no longer be the case. However, this subtle weighting scheme can be important in situations where the objects are to be detected in clutter. The jets located at or near the boundary of the object are affected by the structure in the background, hence leading to low similarity with the boundary jets in a model graph. Lateral excitation in effect weights down the contribution from the boundary jets and imp-roves the robustness of graph matching. In the future work, we will discussions the complexity of our system and compare the result with the existing systems. The effect of lighting is only sensitive for the skin color features, but our system is an integrated system from different cues as it is clear in above sections.

ACKNOWLEDGMENTS

The authors would like to thank the developers of the FLAVOR software environment, which served as the platform for this research. The author also would like to thank Prof. Dr. C. von der Malsburg for his friendly discussions.

References

[1] L. Wiskott and Christoph von der Malsburg, A Neural System for the Recognition of Partially Occluded Objects in Cluttered Scenes, Int. J. of Pattern Recognition and Artificial Intelligence, Vol.7, No. 4, pp. 935-948,1(993).

[2] S. Thomas and Huang Ying Wu, Non-stationary color tracking for vision-based human computer interaction, IEEE Transactions on Neural Networks, Vol.13, No.4, (2002).

[3] J. Triesch, Vision-Based Robotic Gesture Recognition, Ph.D thesis, Berichte aus der Informatik, Shaker Verlag Aachen, (1999).

[4] S. Igi Akira Utsumi and Nobuji Tetsutani, Hand detection and tracking using pixel value distribution model for multiple-camera-based gesture interactions, Proceedings of the IEEE Workshop on Knowledge Media Networking, (2002).

[5] K. Anil, Jain Sharath Pankanti and Salil Prabhakar, On the individuality of fingerprints, In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.805-812, Hawaii, December 11-13 , (2001).

[6] Y. Cui and J. Weng, Hand segmentation using learning-based prediction and verification for hand-sign recognition, In in Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp.88-93, (1996).

[7] R. M. Farouk, Analytical Analysis of Image Representation by Their Discrete Wavelet Transform, International Journal of Computer Science 3; 4, pp. 216-221, (2008).

[8] D.Field, Relation between the statistics of natural images and the response properties of cortical cells, J.Opt.Soc. Amer.A, Vol.4, No.12, pp.2379-2394, (1987).

[9] Gerald Kaiser, Friendly Guide to Wavelets, Birkh¨auser, (1994).

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

[10] Rolf P. Wuertz, Multilayer dynamic link networks for establishing image point correspondences and visual object recognition, Ph. D thesis, volume 53 of Reihe Physik,Verlag Harri Deutsch,Thun,Frankfort a. main, (1995).
[11] L. Wiskott, Labeled Graphs and Dynamic Link Matching for Face Recognition and Scene Analysis, PhD thesis, volume 53 of Reihe Physik,Verlag Harri Deutsch,Thun,Frankfort a. main, (1995).
[12] Christoph von der Malsburg, Pattern recognition by labeled graph matching, Neural Networks, Vol.1, pp.141-148, (1988).
[13] J. Chamorro-Martinez, D. Sa'nchez, B. Prados-Su'arez, A Fuzzy Color Image Segmentation Applied to Robot Vision, (1998).
[14] L. B. Shams, Mark J. Brady and Stefan Schaal, Graph Matching vs mutual information maximization for object detection, Neural Networks, Vol.14, pp. 345-354, (2000)
[15] R. G. Gonzalez, R. E. Woods, Digital Image Processing, Addison-Wesley Publishing Company, (1992).
[16] Y. Qin, Baoping Tanga, J. Wang, Higher-density dyadic wavelet transform and its application, Mechanical Systems and Signal Processing, Vol. 24, 3, pp. 823-834, (2010).
[17] C. Xionga, Jian-hua Houa, Jin-wen Tianb, Jian Liub, Efficient array architectures for multidimensional lifting-based discrete wavelet transforms, Signal Processing, Volume 87, pp. 1089-1099, (2007).
[18] I. Kotsia , Ioan Buciua, Ioannis Pitasa, An analysis of facial expression recognition under partial facial image occlusion, Image and Vision Computing, Volume 26, pp. 1052-1067, (2008).
[19] J. LU, and An-ni CAIa, Tracking people through partial occlusions, The Journal of China Universities of Posts and Telecommunications, Volume 16, pp. 117-121, (2009).
[20] H. Yong Kim, Rotation-discriminating template matching based on Fourier coefficients of radial projecttions with robustness to scaling and partial occlusion, Pattern Recognition, Volume 43, pp. 859-872, (2010).

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

A Novel Seven Input Majority Gate in Quantum-dot Cellular Automata

Keivan Navi[1,2], Amir Mokhtar Chabi[1,2] , Samira Sayedsalehi[2]

[1] Faculty of Electrical and Computer Engineering, Shahid Beheshti University, G. C., Tehran, Iran

[2] Nanotechnology and Quantum Computing Lab., Shahid Beheshti University, G. C., Tehran, Iran

Abstract

A Quantum Cellular Automaton (QCA) is a nanotechnology which is an attractive alternative for transistor based technologies in the near future. A new seven input majority gate in quantum dot cellular automata is proposed in this paper. The basic elements in QCA are majority and inverter gates, therefore using a majority gate with more inputs in QCA circuit will cause reduction in cell count, latency and complexity. Furthermore, by using the proposed seven input majority gate we can design four inputs AND gate and OR gate in only two clock phases. By applying these kinds of gates QCA circuits could be simplified and optimized. In order to prove the functionality of the proposed device, QCADesigner tool and some physical proofs are utilized.

Keywords: *Quantum-Dot Cellular Automata, Majority gate, AND gate, OR gate, Nanoelectronic circuits.*

1. Introduction

CMOS technology has experienced serious problems such as short channel effects, doping fluctuations, increasingly difficult and expensive lithography at nano scale, high leakage current and speed limitation in GHz range. Many alternative technologies have been introduced in the Industry Technology Roadmap for Semiconductors (ITRS) [1]. A Quantum-dot cellular automaton (QCA) is one of these promising nanotechnologies that could be utilized instead of conventional transistor technology in the future.

Quantum-dot Cellular Automata provides new possibilities to achieve outstanding properties such as extremely high density and fast operation speed at Tera Hertz frequencies together with low power dissipation [2-4].

QCA circuits are implemented with two principle gates: majority and inverter gates. As the basic element in QCA is majority gate, the structure of this gate is a significant factor in designing circuits in QCA [5-7]. The very recent designs have been constructed with three input majority gate [8-11]. Although these designs work correctly, they can be reconstructed using new devices to reduce cell counts, clock phases and complexity. Recently five pins majority gates with a new compact full adder cell with lower area and cell counts have been proposed in [12-15].

This paper investigates novel 7-input majority design in QCA. This proposed design only works in two clock phases and in one layer. The input and output of this proposed design is located in two separate sides representing a powerful advantage for designing QCA circuit. By applying this form of majority gate we can simplify logical functions and achieve improved results.

The rest of this paper is organized as follows: in section 2, a brief review of QCA technology is provided and the implementation of the novel seven input majority gate is proposed in section 3. Then accuracy of the proposed design is verified by using physical relations. Finally, in section 4 simulation results are demonstrated and conclusions are drawn in section 5.

2. Review of QCA

In this section we propose a brief background of basic elements of QCA. The basic element of QCA is a cell. A QCA cell, shown in Fig 1, due to the Columbic interactions consists of four quantum dots at the corners of a square with two extra mobile electrons in two different configurations, thus we have two polarizations (p= +1 , p= -1) used for encoding binary information [5][16-17].

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

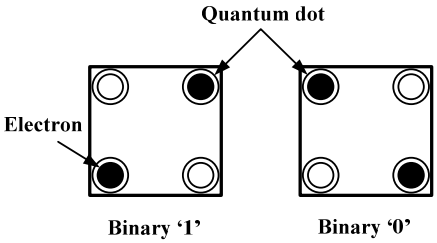


Fig. 1 Basic QCA cell and two possible polarizations

The Columbic interactions between two neighbor cells cause these cells to have the same polarizations. Therefore a series of QCA cells performs as wire in QCA (Fig 2) [18-19].

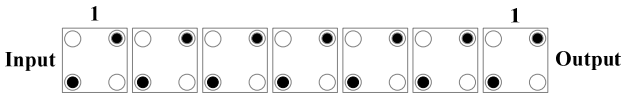


Fig. 2 A QCA Wire

The principle QCA logical circuits are majority and inverter gates. Fig 3 demonstrates two different inverter gates that were presented in [6].

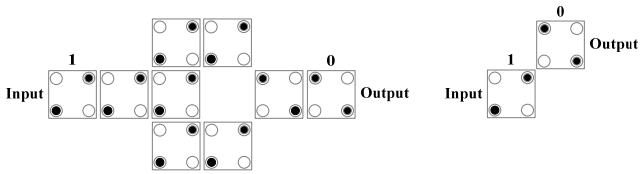


Fig. 3 QCA inverter gates

Up to now, most circuits have been implemented by three majority gate that act based on Eq. (1) (Fig 4) [7-9][20]. Recently a five input majority gates have been presented in [13-14]. Functionality of a five input Majority gate is shown in Eq. (2) and its QCA implementation is presented in Fig 5.

$$M(A,B,C)=AB+AC+BC \quad (1)$$

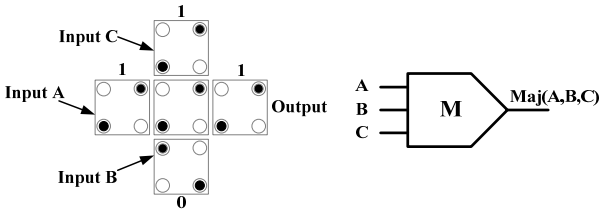


Fig. 4 A QCA three-input Majority voter

$$M(A,B,C,D,E)= ABC + ABD+ ABE+ ACD+ ACE+ ADE+ BCD+ BCE+ BDE+ CDE \quad (2)$$

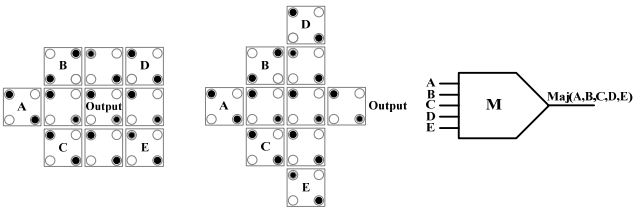


Fig. 5 Structures of Five-input Majority gates[13-14]

3. Proposed seven input Majority Gate

3.1 Design of novel seven input majority gate

A seven input majority gate comprises of seven inputs and one output. The design of this gate is based on five majority gate that was proposed in [14]. A truth table of the seven input majority gate based on sum of inputs is shown in Table 1 and the QCA implementation of majority voter can be presented as Fig 6.

Table 1: Truth table of seven-input majority gate based on sum of inputs

∑(A,B,C,D,E,F,G)	M(A,B,C,D,E,F,G)
0	0
1	0
2	0
3	0
4	1
5	1
6	1
7	1

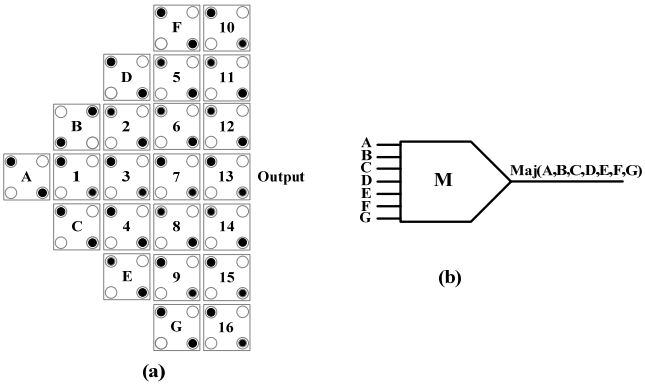


Fig. 6 (a) Proposed seven input majority gate (b) Schematic symbol for the majority gate

In this scheme, labels *A, B, C, D, E, F* and *G* are inputs and the output cell is shown by *output*. Middle cells labeled 1 through 16 are device cells and these middle cells are affected by neighbor cells (input cells or other middle cells). Through these effects, the majority decision

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

of inputs is transferred to the output and efficiently constructs the seven-input majority gate.
By using this device and fixing three inputs, we can implement four input AND gate and four input OR gate, which can reduce the complexity, area and latency of the complex QCA circuits instead of using conventional three input majority gates. Functionality of these gates is presented in Eq. (3a) and (3b).

M(A,B,C,D,0,0,0)=ABCD (3a)
M(A,B,C,D,1,1,1)=A+B+C+D (3b)

The schematic symbols of four input AND and four input OR gates are illustrated in Fig 7. In order to illustrate the usefulness of this type of gates, some logical functions such as XOR, MUX and compressor should be simplified.

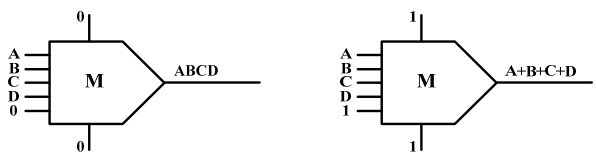


Fig. 7 schematic symbol of four input AND gate and four input OR gate.

3.2. Physical proof:

For physical proof, we assume that all cells are similar, the length of each one is (*a=18nm*), there is a space of x (*x=2nm*) between each two neighbor cells and only neighbor cells would affect each other.
It should be noted that in order to achieve more stability, electrons of QCA cell are arranged in such a manner that their potential energy reaches the minimum level. The potential energy between two cells is calculated using Eq. (4). In this equation, U is potential energy, k is fixed colon, q_1 and q_2 are electric charges and r is the distance between two electric charges. U_T is the summation of potential energies between two cells [21-23].

$$U = \frac{kq_1q_2}{r} \quad (4a)$$

$$kq_1q_2 = 9*10^9*(1.6)^2*10^{-38} = 23.04*10^{-29} = A = cte$$

$$U_T = \sum_{i=1}^{4} U_i \quad (4b)$$

As an example, the potential energy of neighbor cells (Fig 8) as well as other states can be calculated.
The Table 2 illustrates the computation results of potential energies for other states.

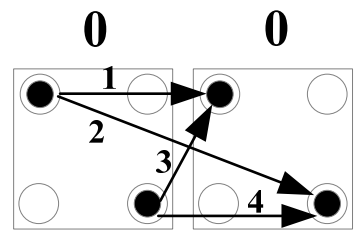


Fig. 8 A sample for computation of potential energy

$$U_1 = \frac{A}{r_1} = \frac{23.04*10^{-29}}{20*10^{-9}} \approx 1.152*10^{-20}(J)$$

$$U_2 = \frac{A}{r_2} = \frac{23.04*10^{-29}}{41.59*10^{-9}} \approx 0.553*10^{-20}(J)$$

$$U_3 = \frac{A}{r_3} = \frac{23.04*10^{-29}}{18.11*10^{-9}} \approx 1.272*10^{-20}(J)$$

$$U_4 = \frac{A}{r_4} = \frac{23.04*10^{-29}}{20*10^{-9}} \approx 1.152*10^{-20}(J)$$

$$U_T = 4.129*10^{-20}(J)$$

As in the above method, we can calculate the potential energy of all 128 different input states in seven majority gate to verify the accuracy of functionality. Here, two possible states (Fig 9) are proved.
In Fig 9(a), the input states are fixed (A=1, B= 1, C=0, D=1, E=0, F=0, G=0) and the value of the 16 device cells are computed in 2^16 possible states. First, we calculate the total potential energy with fixed zero value of device cells due to Table 2. In a similar manner, the total potential energy of other possible state of device cells can be simultaneously computed. The comparison of total potential energies in all states determines which state is more stable. The state which has the lower potential energy level is more suitable. According to computation results of potential energy, minimum energy achieved is $367.461*10^{-20}(J)$ when the value of output is zero. Similarly, the minimum total potential energy of input states (A=0, B=1, C=1, D=0, E=1, F=1, G=0) can be obtained as in Fig 9(b).
According to the above computing, the proposed structure for implementing seven input majority gate works accurately as the majority voter.
Further to physical proof, we can also check the proposed design using QCADesigner. Simulation of this majority gate is shown in the next section.

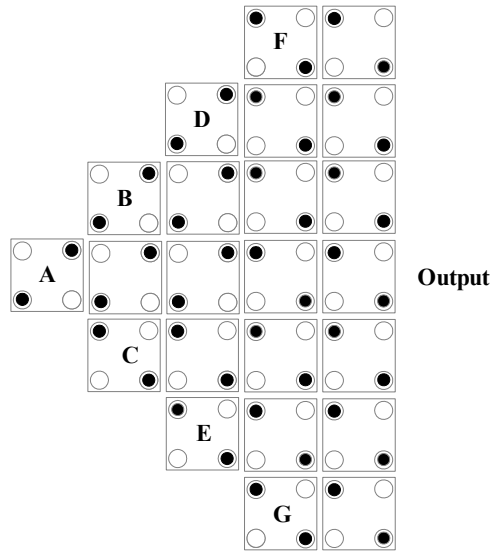
Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Table 2: The computation results of potential energies.

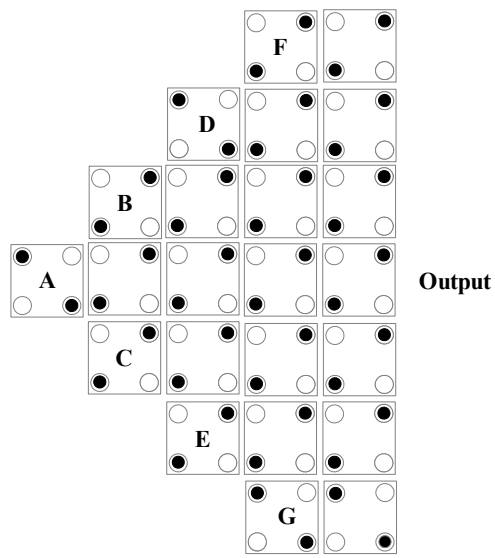
$*10^{-20}(J)$			
0 0	4.129	0 1	13.838
1 0	13.838	1 1	4.129
0 0	10.226	0 1	3.364
0 1	3.364	1 1	2.838
0 0	2.838	0 1	3.364
1 0	3.364	1 1	10.226
0 0	4.129	1 0	13.838
0 1	13.838	1 1	4.129

4. Simulation Results

The proposed seven input majority gate is implemented and simulated using QCADesigner version 2.0.3 [24]. The following parameters are used for a coherence vector options: cell size = 18nm, temperature= 1.000000 K, relaxation time= 1.000000e-015 s, time step= 1.000000e-016 s, total simulation= 7.000000e-011 s, clock high= 9.800000e-022 J, clock low = 3.800000e-023 J, clock shift = 0, clock amplitude factor = 2.000000, radius of effect = 80.000000 nm, relative permittivity = 12.900000, layer separation = 11.500000 nm. Most of the above mentioned parameters are default values in QCADesigner.



(a) Minimum Energy = 367.461* 10^{-20}(J)



(b) Minimum Energy = 359.547* 10^{-20}(J)

Fig. 9 Two possible states

The simulation result of new seven input majority gate is presented in Fig 10. For proving truth functionality of four input AND gate and OR gate, Figs 11 and 12 show the simulation results of these gates.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

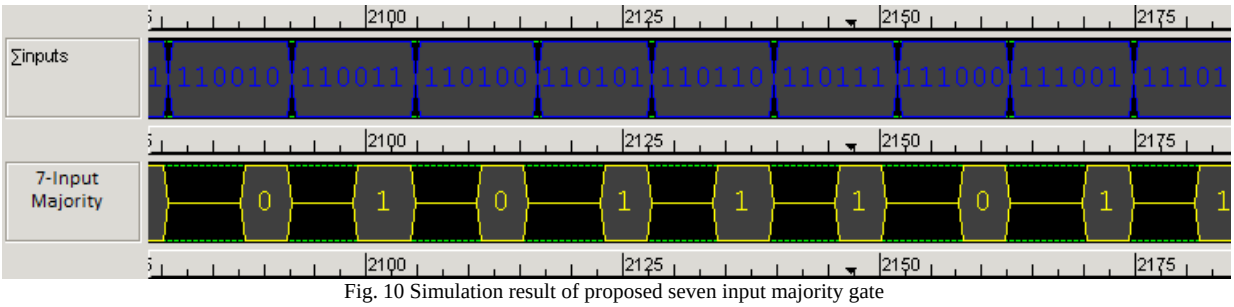


Fig. 10 Simulation result of proposed seven input majority gate

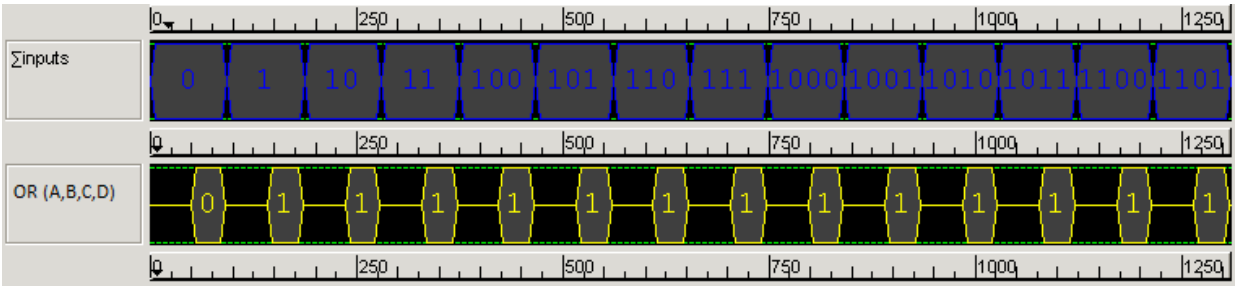


Fig. 11 Simulation result of proposed OR gate

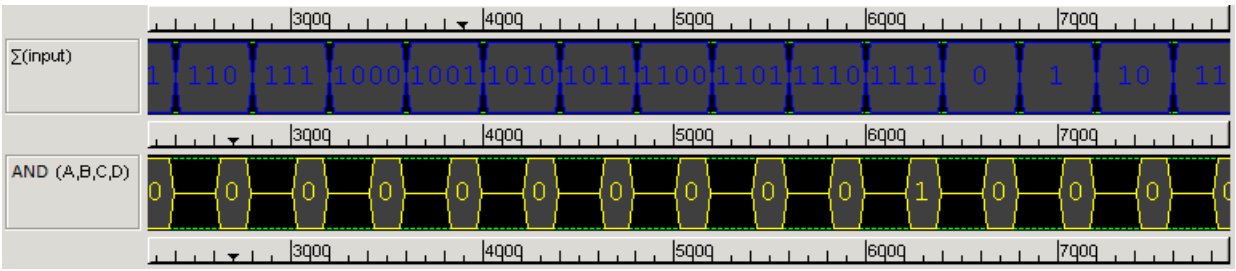


Fig. 12 Simulation result of proposed AND gate

5. Conclusion

A novel seven input majority gate for Quantum-Dot Cellular Automata is presented. For verifying the functionality, this proposed majority gate has been proved using physical relations and also verified in QCADesigner tools. By utilizing this seven input majority gate four input AND gate and OR gate can be achieved. Therefore most recent proposed designs in QCA can be simplified by using this majority device in order to reduce complexity, cell count and latency.

References

[1] International Technology Roadmap for Semiconductors. (ITRS) [Online]. Available: http://www.itrs.net.

[2] R.P. Cowburn and M.E. Welland, “Room Temperature Magnetic Quantum Cellular Automata,” Science, vol. 287, no. 5457, pp. 1466-1468, Feb. 2000.

[3] J.M. Seminario et al., “A Molecular Device Operating at Terahertz Frequencies: Theoretical Simulations,” IEEE Trans. Nanotechnology,vol. 3, no. 1, pp. 215-218, Mar. 2004.

[4] H. Qi et al., “Molecular Quantum Cellular Automata Cells. Electric Field Driven Switching of a Silicon Surface Bound Array of Vertically Oriented Two-Dot Molecular Quantum CellularAutomata,” J. Am. Chemical Soc., vol. 125, pp. 15250-15259, 2003.

[5] Lent C. S., Tougaw P. D., Porod W. and Bernstein G. H., "Quantum cellular automata,” Nanotechnology, vol. 4, no. 1, pp. 49-57, 1993.

[6] P.D. Tougaw and C.S. Lent, “Logical Devices Implemented Using Quantum Cellular Automata”, Journal of Applied Physics, Vol. 75(3), pp. 1818-1825, 1994.

[7] A.O. Orlov, I. Amlani, G.H. Bernstein, C.S. Lent and G.L. Snider, “Realization of a Functional Cell for

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Quantum-Dot Cellular Automata", Science, Vol. 277, pp. 928-930, 1997.
[8] H. Cho., and E. E. Swartzlander, "Adder Designs and Analyses for Quantum-Dot Cellular Automata," IEEE Transactions on Nanotechnology, Vol. 6, No. 3, pp. 374-384, 2007.
[9] H. Cho and Earl E. Swartzlander, "Adder and Multiplier Design in Quantum-Dot Cellular Automata", IEEE Transactions on Computers, Vol. 58, No. 6, 2009.
[10] A. Vetteth et al., "Quantum-Dot Cellular Automata Carry-Look-Ahead Adder and Barrel Shifter," Proc. IEEE Emerging Telecomm.Technologies Conf., Sept. 2002.
[11] W. Wang, K. Walus, and G.A. Jullien, "Quantum-Dot Cellular Automata Adders", Proc. Of IEEE Conference on Nanotechnology, 2003.
[12] M. R. Azghadi, O. Kavehei and K. Navi, "A Novel Design for Quantum-dot Cellular Automata Cells and Full-Adders", Journal of Applied Sciences 7, pp. 3460-3468, 2007.
[13] K. Navi,S. Sayedsalehi, R. Farazkish, M. Rahimi Azghadi, "Five-Input Majority Gate, a New Device for Quantum-Dot Cellular Automata", Journal of Computational and Theoretical Nanoscience, Volume 7, Number 8, August 2010 , pp. 1546-1553(8).
[14] K. Navi , R. Farazkish, S. Sayedsalehi, M. Rahimi Azghadi, "A new quantum-dot cellular automata full-adder", Elsevier, Microelectronics Journal, In Press, Corrected Proof, Available online 9 September 2010.
[15] S. Sayedsalehi, M. h. Moaiyeri, K. Navi, "Novel Efficient Adder Circuits for Quantum-dot Cellular Automata", accepted in Journal of Computational and Theoretical Nanoscience.
[16] C. S. lent, P. Tougaw, and W. Porod, "bistable saturation in coupled quantum dots for quantum cellular automata," Appl. Phys. Lett., vol. 62, pp. 714, 1993
[17] P.D. Tougaw and C.S. Lent, "Dynamic Behavior of Quantum Cellular Automata", Journal of Applied Physics, Vol. 80(8), pp. 4722-4736, 1996.
[18] C.S. Lent and P.D. Tougaw, "Lines of Interacting Quantum-Dot Cells: A Binary wire", Journal of Applied Physics, pp. 6227-6233, 1993
[19] A. Orlov et al., "Experimental Demonstration of a Binary Wire for Quantum-Dot Cellular Automata," Applied Physics Letters, vol. 74,no. 19, pp. 2875-2877, May 1999.
[20] J. Huang, M. Momenzadeh, M.B. Tahoori and F. Lombardi, "Design and Characterization of an And-Or-Inverter (AOI) Gate for QCA Implementation", GLSVLSI, pp 26-28, 2004.
[21] D. Halliday, and A. Resnick, "Fundamentals of Physics," 7th Edition, New York: John Wiley & Sons, Inc, Part 1 (Chapters 3-6), 2004.
[22] Mc Dermott and C. Lillian, "Research on Conceptual Understanding in Mechanics," Physics Today, Vol. 37 (7), 1984.
[23] I. Halloun and D. Hestenes, "Common sense concepts about motions". American Journal of Physics, 53(11). pp. 1056-1064, 1985.
[24] QCADesigner Home Page: www.atips.ca/projects/qcadesigner/

Keivan Navi received his M.Sc. degree in electronics engineering from Sharif University of Technology, Tehran, Iran in1990. He also received his Ph.D. degree in computer architecture from Paris XI University, Paris, France, in 1995. He is currently Associate Professor in Faculty of Electrical and Computer Engineering of Shahid Beheshti University. His research interests include Nano electronics with emphasis on CNFET, QCA and SET, Computer Arithmetic, Interconnection Network Design and Quantum Computing and cryptography. He is a senior member of IEEE.

Amir Mokhtar Chabi received his B. Sc. from Isfahan University of Technology, Isfahan, Iran in 2009 in hardware engineering and M.Sc. degree in computer engineering from Shahid Beheshti University, Tehran , Iran in 2011. His research interests include Nanoelectronics with emphasis on Emerging Technologies, Quantum-dot Cellular Automata, CNFET, Multi-Value Logic and Computer Architecture. He is currently is faculty of Nanotechnology and Quantum Computing Lab in Shahid Beheshti University, G. C., Tehran, Iran.

Samira Sayedsalehi received his B.Sc. from Islamic Azad University, Tehran, Iran in 2005 in hardware engineering, and her M.S. from Science and Research Branch of Islamic Azad University, Tehran, Iran in 2008 in computer architecture Engineering. She is currently working toward the Ph.D. degree in Science and Research Branch of Islamic Azad University computer architecture engineering at the Science and Research. Her research interests lie in quantum cellular automata and testing, design and fault tolerance issues in digital systems. She is a student member of IEEE.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

A Priority Based Bandwidth Reservation Protocol for Wireless Mesh Networks

[1]K. Valarmathi and [2]Dr.N. Malmurugan

[1]Research Scholar, Sathyabama University,
Chennai, India

[2]Principal Oxford Engineering College
Trichy, India

Abstract

In wireless mesh networks, when compared with traditional applications, multimedia application requires the need of inflexible Quality of services (QoS) for high speed digital audio and videos. Under such a condition, the need of proper resource allocation is extremely handy. In this paper, we propose to develop a Priority Based Bandwidth Reservation Protocol (PBRP) for wireless mesh networks. Our protocol consists of two phases namely Bandwidth Request phase and Bandwidth Reply phase. In the former Phase, a Bandwidth Request (BREQ) message is forwarded from the node that requests the admission of a new traffic flow to its destination. In the later Phase, a Bandwidth Reply (BREP) message proceeds backwards, hop-by-hop, from the destination node to the node that originated the request along the path laid down by the corresponding (BREQ) message. The destination node precedes the reply according to the priority of traffic classes and reserves the bandwidth on the reply path. By simulation results, we show that our proposed protocol achieves high bandwidth utilization and throughput with reduced delay, when compared with existing technique.

Keywords: *Wireless Mesh Networks (WMN), Quality of services (QoS), Priority Based Bandwidth Reservation Protocol (PBRP), Bandwidth Request (BREQ), access points (APs)*

1. Introduction

1.1 Wireless Mesh Networks

Wireless mesh networks (WMN's) contains several stationary wireless routers which are interlinked by the wireless links. Wireless routers acts as the access points (APs) for wireless mobile devices. Through the high speed wired links, some wireless routers act as a gateway for internet. Wireless mobile devices transfer data to the corresponding wireless router and further these data's are transferred in a multi-hop manner to the internet via intermediate wireless routers. The popularity of WMN's is due to their low cost and auto-organizing features [1].

Multi-channel wireless mesh network architecture requires topology discovery, traffic profiling, channel assignment and routing. This includes static aggregation nodes similar to the wireless LAN access points. For the construction of multi-channel wireless mesh networks MCWMN, 802.11b interface hardware is used because it can handle the bandwidth problem. Every node in a multi-channel wireless mesh networks MCWMN includes multiple 802.11 complaint NIC's and it is tuned to a particular radio channel for long duration such as hours or days [2].

1.2. Bandwidth Allocation in WMN

Bandwidth allocation refers to various methods used in the communication industry to design and assign frequency channels to various wireless applications. It can be defined as a process which allocates resources to various connections so that minimum rate requirements of every connection are satisfied.

In other words, it is a method of fairly allocating traffic bandwidth in shared telecommunication medium and between different users of that particular bandwidth on demand. This is referred to as bandwidth management where sharing of link adapts in some way to the instantaneous traffic demands of the nodes connected to the link.

In case of high speed digital audio and videos, multimedia application requires the source of inflexible Quality of Service (QoS), when compared with traditional application. The purpose of proper Bandwidth allocation is essential for these conditions. The crucial factors for QoS in real time multimedia applications are the Bandwidth allocation and delay factors [3].

The challenges faced in wireless media for Bandwidth allocation on multimedia application includes [4]

- Harmful effects such as Fading and co-channel interferences (CCI) causes distortion in channels.
- The resources such as Bandwidth and power are limited.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

- The data rates in multimedia application are irregular.
- System efficiency and fairness are the service factors which are given more importance among individuals.

The few Bandwidth Allocation techniques are:

- IntServ and DiffServ [5]
- QUOTA [6]
- SWAN [7]
- INSIGNIA [8]

In our previous paper [13], we have proposed a congestion control routing protocol along with multi-channel assignment. We have used a traffic aware metric in that protocol to provide quality of service. The proposed protocol can improve the throughput and channel utilization to very high extent because it provides solution for multi-channel assignment and congestion control. The proposed algorithm assigns the channels in a way that, congestion is avoided and co-channel interference levels among links with same channel are reduced.

As a continuation of our previous work, in this paper, we propose to develop a Priority Based Bandwidth Reservation Protocol (PBRP) for wireless mesh networks. Our protocol consists of two phases namely Bandwidth Request phase and Bandwidth Reply phase. In the former Phase, a Bandwidth Request (BREQ) message is forwarded from the node that requests the admission of a new traffic flow to its destination. In the later Phase, a Bandwidth Reply (BREP) message proceeds backwards, hop-by-hop, from the destination node to the node that originated the request along the path laid down by the corresponding (BREQ) message. The destination node precedes the reply according to the priority of traffic classes and reserves the bandwidth on the reply path.

2. Related Work

Bo Wang et al [6] have focused on the problem of providing QoS support for real-time flows, while allocating bandwidth to elastic flows fairly. They have proposed QUOTA (quality-of-service aware fair rate allocation), a framework that combines QoS support and fair rate allocation. Their proposed framework QUOTA provides higher priority to real-time flows than elastic flows by reserving the necessary bandwidth for the former and fairly allocating the left-over bandwidth to the latter.

Claudio Cicconetti et al [9] have introduced a Fair End-to-end Bandwidth Allocation (FEBA) algorithm. Their FEBA algorithm is implemented at the Medium Access Control (MAC) layer of single-radio, multiple channels IEEE 802.16 mesh nodes, operated in a distributed coordinated scheduling mode. FEBA negotiates bandwidth among neighbors to assign a fair share proportional to a specified weight to each end-to-end traffic flow. Thus the traffic flows are served in a differentiated manner, with higher priority traffic flows being allocated more bandwidth on the average than the lower priority traffic flows.

Ho Ting Cheng et al [10] have proposed a low-complexity intracluster resource allocation algorithm by considering the power allocation, sub-carrier allocation, and packet scheduling. The time complexity of their proposed scheme is on the order of $O(LMN)$, where L is the number of time slots in a frame, M is the number of active links, and N is the number of sub-carriers.

Ho Ting Cheng et al [11] have proposed an efficient intra-cluster packet-level resource allocation approach. Their approach considers power allocation, sub-carrier allocation, packet scheduling, and QoS support. Their proposed approach combines the merits of a Karush-Kuhn-Tucker (KKT)-driven approach and a genetic algorithm (GA)-based approach. Their proposed approach achieves a desired balance between time complexity and system performance.

Andre Herms et al [12] have addressed the problems of the reservation on a single hop. They have discussed the reasons for the inconsistencies in the existing approaches which lead to admission failures and present a protocol for preventing them. This allows for increasing the reliability of established communication links in WMNs. They have focused only on the local admission control and not the various searching strategies for finding a suitable path.

3. Bandwidth Reservation Protocol

3.1. Protocol Overview

Basically, our proposed protocol consists of two phases namely Bandwidth Request phase and Bandwidth Reply phase. In the Bandwidth Request Phase, a Bandwidth Request (BREQ) message is forwarded from the node that requests the admission of a new traffic flow to its destination. During this phase bandwidths are not reserved. The BREQ message consists of traffic flow specifications and the requested bandwidth.

Next in the Bandwidth Reply Phase, a Bandwidth Reply (BREP) message proceeds backwards, hop-by-hop, from the destination node to the node that originated the request along the path laid down by the corresponding (BREQ) message. The destination node precedes the reply

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

according to the priority of traffic classes and reserves the bandwidth on the reply path.

3.2. Determining Priority of Traffic Classes

The traffic flow specifications consists of traffic classes such as UGS, ertPS, rtPS, nrtPS, BE and the priority for the each traffic is given in the same order.

The equation for the priority for the traffic classes is given by;

$$PTC = RBW + SC \tag{1}$$

Where RBW is the requested bandwidth and is calculated by

$$\text{RBW} = 1 - \frac{\text{BWreq}}{\text{BWmax}} \tag{2}$$

BWreq is the node's requested bandwidth .

BWmax is the maximum bandwidth

where SC is the value assigned to each traffic service types as 1, 0.80, 0.60, 0.40 and 0.20 in order as per the priority of the traffic classes.

In case of a buffer overflow, the traffic classes are transmitted according to their priority by utilizing other channels having low priority traffic. In case of same priority value, there will be traffic congestion. So the proposed method makes use of earliest deadline technique to schedule the traffic using a Deadline Factor (DF). It is given by using the request's residual time R_t and request's execution time E_t . Thus the equation for DF is given by;

$$DF = \frac{1}{1 + e^{R_t - E_t}} \tag{3}$$

Thus our equation of earliest deadline technique takes into account both PTC and DF (eq 1 and 3) to allocate traffic for similar priority data.

$$PTC = PTC + DF \tag{4}$$

Algorithm for Earliest Deadline Technique

In incoming traffic flow

At $T = 1$

1. If $Ch < C_{th}$, then
   - 1.1 Channel is marked as good channel,
   - 1.2 Assign priority using (1)
   - 1.3 Schedule the traffic
2. Else if $Ch > C_{th}$, then
   - 2.1 Mark the channel as bad.
   - 2.2 Put each traffic in separate queues $q_1, q_2, \cdots, q_m$
   - 2.3. If $Q_1(i) > Q_{\max}$, where $i = 1,2,\cdots,k$, then
     - 2.3.1 Schedule the traffic using (4).
   - 2.4. End if.
3. End if
4. T = T + 1,
5. Repeat from 1.0

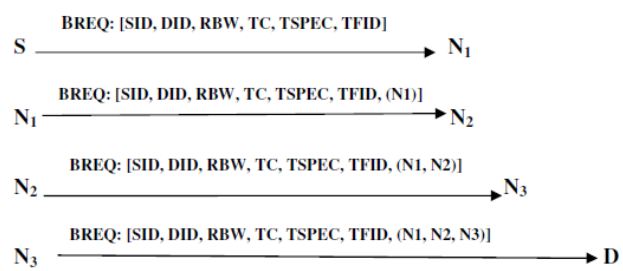
3.3. Bandwidth Request Phase

The node that originates the traffic flow conveys the admission request to the final destination node through a BREQ message, which is relayed by the intermediate nodes along the path. In addition to the requested bandwidth (RBW) and the traffic classes (TC) such as UGS, ertPS, rtPS, nrtPS, a BREQ message includes:

- Source and Destination node IDs (SID and DID)
- Intermediate node IDs
- TSPEC
- Traffic flow identifier (TFID)

In the Bandwidth Request Phase the bandwidths are not reserved and only the necessary messages are transmitted to the destination. The source is required to select the TFID of any new flow in such a way that the source, destination, TFID uniquely identifies the traffic flow in the network.

Let N1, N2, N3 are the intermediate nodes between the source S and the destination D. Then the bandwidth request process is illustrated as below:



3.4. Bandwidth Reply Phase

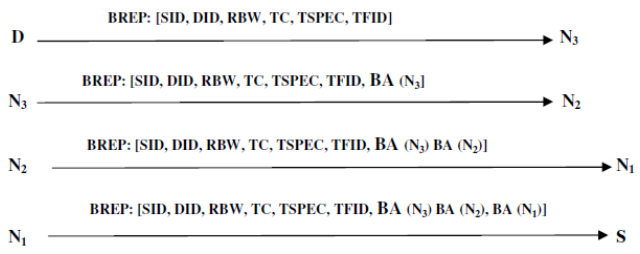
In this phase, the destination sends back to the source a BREP message and it is routed through the same path that has been enclosed by the BREQ message. This is obtained by using the list of intermediate node IDs included in the BREQ message. On receiving the BREP message, each node reserves the bandwidths according to the priority of the traffic. The traffic priority is determined as in 3.2.

If the nodes do not receive packets until the traffic flow is dropped for a particular amount of time TS, then the bandwidth remains allocated. The source generates probe packets to guarantee an established traffic flow state on

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

each node in the path to prevent premature termination of the traffic flow. Probe packets are the messages which include the information about their traffic and these packets are discarded by the receivers in the MAC layer. The generation interval of the probe packets must be smaller than the TS. Generally, by transmitting the probe packets it consumes the bandwidth which is already reserved for the traffic flow in the data sub-frame.

The bandwidth reply process is illustrated as below:



4. Simulation Results

4.1. Simulation Model and Parameters

We use NS2 [14] to simulate our proposed protocol. We use the IEEE802.16e simulator [15] patch for NS2 version 2.33 to simulate a WiMax Mesh Network. It has the facility to include multiple channels and radios. It supports different types of topologies such as chain, ring, multi ring, grid, binary tree, star, hexagon and triangular. The supported traffic types are CBR, VoIP, Video-on-Demand (VoD) and FTP. In our simulation, 20 mobile nodes are arranged in a topology of size 500meter x 500 meter region. All nodes have the same transmission range of 250 meters. In our simulation, the speed is set as 5m/s. A total of 3 traffic flows (two VoIP and one VoD) are used.

Our simulation settings and parameters are summarized in table 1.

Table 1: Simulation Settings

No. of Nodes	20
Area Size	500 X 500
Mac	802.16e
Radio Range	250m
Simulation Time	100 sec
Traffic Source	VoIP and VoD
VoD Packet Size	65536
VoD Rate	150Kb
VoIP Codec	GSM.AMR
No. of VoIP frames per packet	2
Topology Type	Ring
OFDM Bandwidth	10 MHz

4.2.Performance Metrics

The proposed Priority Based Bandwidth Reservation Protocol (PBRP) is compared with the FEBA [9] algorithm. The performance is evaluated mainly, according to the following metrics.

- **Average End-to-End delay**: The end-to-end-delay is averaged over all surviving data packets from the sources to the destinations.
- **Aggregated Throughput**: We measure aggregated throughput in terms of no. of packets received.
- **Bandwidth Utilization:** It is the ratio of bandwidth received into total available bandwidth for a traffic flow.

4.3. Results

A. Effect of Varying Rate

In the first experiment, the transmission rate is varied as 0.5, 1.0, 1.5 and 2Mb and the above metrics are measured.

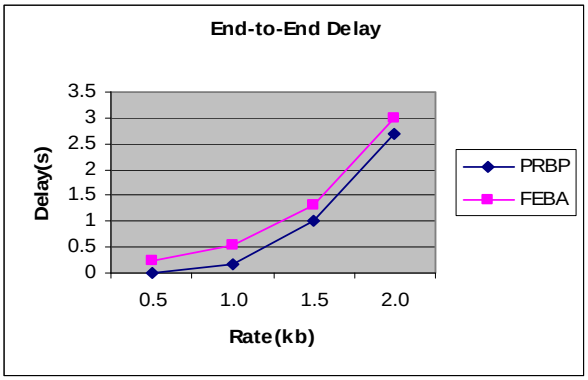


Fig .1 Rate Vs Delay

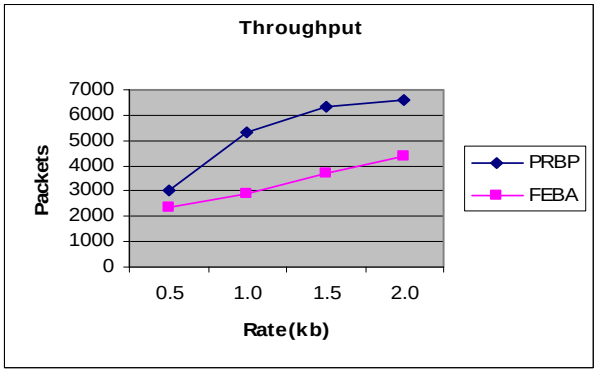


Fig .2 Rate Vs Throughput

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

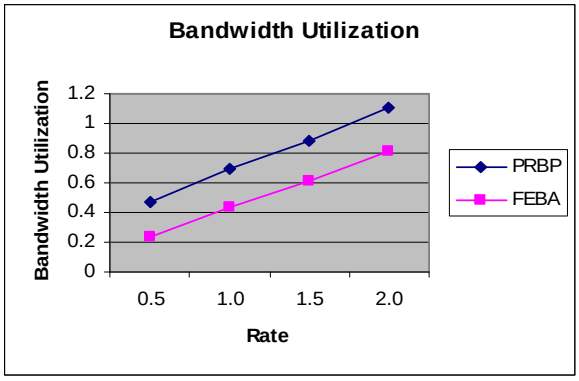


Fig .3 Rate Vs Bandwidth Utilization

Fig.1 shows the end-to-end delay values when the number of rate is increased. It is clear that PRBP has less delay when compared to FEBA algorithm.

Fig.2 shows the throughput values when the number rate is increased. From the figures, it can be seen that the throughput is more in the case of PRBP and outperforms the FEBA algorithm.

Fig.3 shows the bandwidth utilization obtained, when the number of rate is increased. It shows that PRBP utilizes more bandwidth than the FEBA algorithm.

B. Effect of Varying Flows

In the second experiment, we vary the number of data flows as 2,4,6,8 and10.

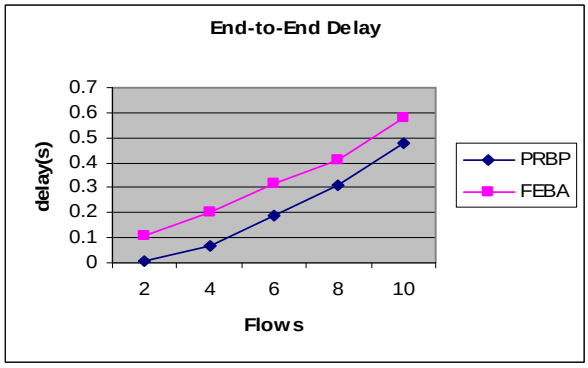


Fig .4 Flow Vs Delay

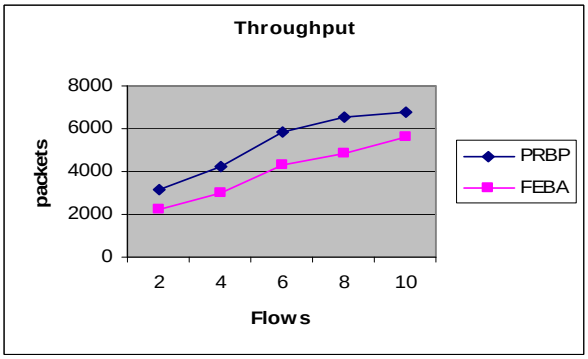


Fig .5 Flow Vs Throughput

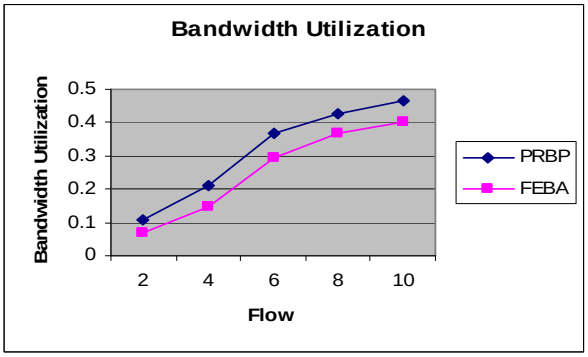


Fig .6 Flow Vs Bandwidth Utilization

Fig. 4 shows the end-to-end delay values when the number of flow is increased. It is clear that PRBP has less delay when compared to FEBA algorithm.

Fig. 5 shows the throughput values when the number flows are increased. From the figures, it can be seen that the throughput is more in the case of PRBP and outperforms the FEBA algorithm.

Fig.6 shows the bandwidth utilization obtained, when the number of flows are increased. It shows that PRBP utilizes more bandwidth than the FEBA algorithm.

5. Conclusion

In this paper, we have proposed a Priority Based Bandwidth Reservation Protocol (PBRP) for wireless mesh networks. Our protocol consists of two phases namely Bandwidth Request phase and Bandwidth Reply phase. In the former Phase, a Bandwidth Request (BREQ) message is forwarded from the node that requests the admission of a new traffic flow to its destination. The BREQ messages includes the requested bandwidth, traffic classes, Source and Destination node IDs, Intermediate node IDs, TSPEC, Traffic flow identifier (TFID). In the

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

later Phase, a Bandwidth Reply (BREP) message proceeds backwards, hop-by-hop, from the destination node to the node that originated the request along the path laid down by the corresponding (BREQ) message. The destination node precedes the reply according to the priority of traffic classes and reserves the bandwidth on the reply path. In case of same priority value, we have used the earliest deadline technique to schedule the traffic using a Deadline Factor (DF). In addition to this, to prevent premature termination of the traffic flow, the source generates probe packets to guarantee a constant refresh of the established traffic flow state on each node in the path. By simulation results, we have shown that our proposed protocol achieves high bandwidth utilization and throughput with reduced delay, when compared with existing technique.

Reference

[1] Letor, Blondia, Bouckaert, Moerman and Demeester, "A cluster driven channel assignment mechanism for wireless mesh networks", 5th IEEE International conference on Mobile Ad Hoc and Sensor Systems, Oct.2008

[2] Raniwala, Gopalan, and Chiueh, "Centralized Channel Assignment and Routing Algorithms for Multi-Channel Wireless Mesh Networks", in Mobile Computing and Communications Review, Vol.8, April 2004.

[3] Kaan Bur and Cem Ersoy, "Ad hoc quality of service multicast routing", Computer Communications, Volume 29 , Issue 1, 2005.

[4] Zhu Han and K. J. Ray Liu, "Resource allocation for wireless networks: basics, techniques, and applications", Cambridge university press, 2008, ISBN – 978 0 521 87385 7.

[5] Bosheng Zhou, Alan Marshall and Tsung-Han Lee, "A Cross-Layer Architecture for DiffServ in Mobile Ad-hoc Networks", International Conference on Wireless Networks, Communications and Mobile Computing, WIRELESSCOM 2005

[6] Bo Wang and Matt Mutka, "QoS-aware fair rate allocation in wireless mesh networks", Elsevier, Computer Communications 31, 2008

[7] Gahng-Seop Ahn, Andrew T. Campbell, Andras Veres and Li-Hsiang Sun, "SWAN: Service Differentiation in Stateless Wireless Ad Hoc Networks", In the Proceedings of the 21st Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM 2002.

[8] Seoung-Bum Lee, Gahng-Seop Ahn, Xiaowei Zhang, and Andrew T. Campbell, "NSIGNIA: An IP-Based Quality of Service Framework for Mobile ad Hoc Networks", Journal of Parallel and Distributed Computing, 2000.

[9] Claudio Cicconetti, Ian F. Akyildiz and Luciano Lenzini, "FEBA: A Bandwidth Allocation Algorithm for Service Differentiation in IEEE 802.16 Mesh Networks", IEEE/ACM Transactions on Networking (TON), Volume 17, Issue3, June 2009.

[10] Ho Ting Cheng and Weihua Zhuang, "Joint Power-Frequency-Time Resource Allocation in Clustered Wireless Mesh Networks", IEEE Network, vol. 22, no. 1, pp. 45-51, Jan/Feb 2008.

[11] Ho Ting Cheng and Weihua Zhuang, "Novel Packet-Level Resource Allocation with Effective QoS Provisioning for Wireless Mesh Networks", IEEE Transactions on Wireless Communications, Vol. 8, No. 2, February 2009

[12] Andre Herms and Georg Lukas, "Preventing Admission Failures of Bandwidth Reservation in Wireless Mesh Networks", IEEE/ACS International Conference on Computer Systems and Applications, 2008

[13] K.Valarmathi and N.Malmurugan, "Joint Design of Congestion Control Routing With Distributed Multi Channel Assignment in Wireless Mesh Networks", Networking and Internet Architecture, IJCSIS, Vol. 7 No. 3, March 2010.

[14] Network Simulator http://www.isi.edu/nsnam/ns

[15] "WiMsh: a simple and efficient tool for simulating IEEE 802.16 wireless mesh networks in ns-2", Proceedings of the 2nd International Conference on Simulation Tools and Techniques, 2009

K. Valarmathi received BSC (Computer Science), MCA degrees from University of Madras in the year 1991, 1998 respectively. She has also received ME (Computer Science) degree in 2004 from Sathyabama University. She has worked in several Public Limited companies for more than 7 years. She worked as a Lecturer in MNM Jain Engineering College for more than 2 years. Now she is working as Assistant Professor in **Panimalar Engineering College** from July 2003 onwards. During her teaching service she has received Best Teacher award in the year 2007 and received cash awards many times for her academic excellence. She is currently pursuing the PhD degree in the field of Wireless Mesh Networks under the guidance of Dr. N. Malmurugan, Principal of Oxford Engineering College, Trichy.

Dr. Malmurugan Nagarajan has around Twenty two years of both Teaching and Industrial experience. He worked extensively in Audio, Video Codec development for mobile platforms and well versed with Audio and Video standards. He is strong in developing tools & system applications and Modeling & Simulation. Currently he is developing new algorithms and IPs for the components in Wireless Broadband Physical Layer He is an Educationist and held various positions ranging from Lecturer to Principal in various educational organizations. His area of interest includes Wavelet variants based Signal and Image Processing, Multimedia Compression & Watermarking, Signal Processing algorithm development in Telecom domain. He is Member of IEEE, Fellow of IETE, Fellow of UWA and Editor of Journal of Simulation and Modeling

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Analysis of the Similarities and Differences between MPLS Label Distribution Protocols RSVP and CR-LDP

Muhammad Asif[1], Zahid Farid[2], Muhammad Lal[3], Junaid Qayyum[4]

[1]Department of Information Technology and Media (ITM), Mid Sweden University Sundsvall 85170, Sweden.

[2]School of Electrical and Electronics Engineering, Universiti Sains Malaysia (USM) Penang 14300, Malaysia.

[3, 4] Department of Information Technology, Gandhara University of Sciences Peshawar 25000, Pakistan.

Abstract

MPLS is a new technology that offers to open up the Internet by providing many additional services to applications using IP. MPLS forwards data using labels that are attached to each data packet. These labels must be distributed between the nodes that comprise the network. Many of the new services that ISPs want to offer rely on Traffic Engineering functions. There are currently two label distribution protocols that provide support for Traffic Engineering: Resource ReSerVation Protocol (RSVP) and Constraint-based Routed Label Distribution Protocol (CR-LDP). Although the two protocols provide a similar level of service, the way they operate is different, and the detailed function they offer is also not consistent. Hardware vendors and network providers need clear information to help them decide which protocol to implement in a Traffic Engineered MPLS network. Each protocol has its champions and detractors, and the specifications are still under development. Recognizing that the choice of label distribution protocol is crucial for the success of device manufacturers and network providers; this White Paper explains the similarities and important differences between the two protocols, to help identify which protocol is the right one to use in a particular environment. Data Connection's DC-MPLS family of portable MPLS products offers solutions for both the RSVP and CR-LDP label distribution protocols.

Keywords MPLS, ISPs, Traffic Engineering, Resource ReSerVation Protocol (RSVP), Constraint-based Routed Label Distribution Protocol (CR-LDP), Data Connection's DC-MPLS.

1. Introduction

CR-LDP is a set of extensions to LDP specifically designed to facilitate constraint-based routing of LSPs. Like LDP, it uses TCP sessions between LSR peers and sends label distribution messages along the sessions. This allows it to assume reliable distribution of control messages. [1] The basic flow for LSP setup using CR-LDP is as shown in Figure 1.

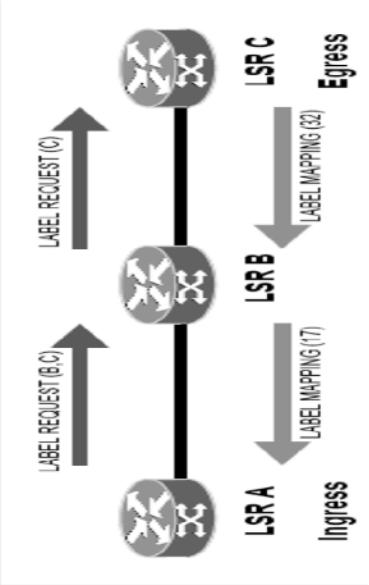


Figure 1 (CR-LDP LSP Setup Flow)

- The Ingress LSR, LSR A, determines that it needs to set up a new LSP to LSR C. The traffic parameters required for the session or administrative policies for the network enable LSR A to determine that the route for the new LSP should go through LSR B, which might not be the same as the hop-by-hop route to LSR C. LSR A builds

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

a LABEL_REQUEST message with an explicit route of (B,C) and details of the traffic parameters requested for the new route. LSR A reserves the resources it needs for the new LSP, and then forwards the LABEL_REQUEST to LSR B on the TCP session.

- LSR B receives the LABEL_REQUEST message, determines that it is not the egress for this LSP, and forwards the request along the route specified in the message. It reserves the resources requested for the new LSP, modifies the explicit route in the LABEL_REQUEST message, and passes the message to LSR C. If necessary, LSR B may reduce the reservation it makes for the new LSP if the appropriate parameters were marked as negotiable in the LABEL_REQUEST.

- LSR C determines that it is the egress for this new LSP. It performs any final negotiation on the resources, and makes the reservation for the LSP. It allocates a label to the new LSP and distributes the label to LSR B in a LABEL_MAPPING message, which contains details of the final traffic parameters reserved for the LSP.

- LSR B receives the LABEL_MAPPING and matches it to the original request using the LSP ID contained in both the LABEL_REQUEST and LABEL_MAPPING messages. It finalizes the processing at LSR A is similar, but it does not have to allocate a label and forward it to an upstream LSR because it is the ingress LSR for the new LSP.

Generic RSVP uses a message exchange to reserve resources across a network for IP flows. The Extensions to RSVP for LSP Tunnels enhances generic RSVP so that it can be used to distribute MPLS labels. RSVP is a separate protocol at the IP level. It uses IP datagrams (or UDP at the margins of the network) to communicate between LSR peers. It does not require the maintenance of TCP sessions, but as a consequence of this it must handle the loss of control messages. [2] The basic flow for setting up an LSP using RSVP for LSP Tunnels is shown in Fig.2 below.

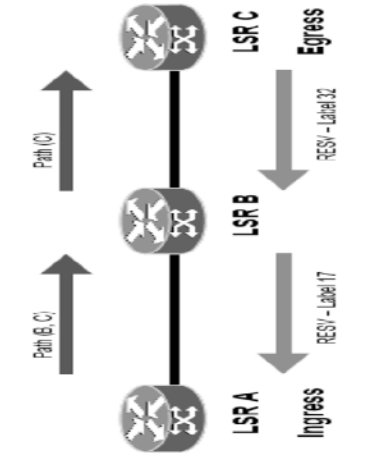


Figure 2 (RSVP LSP Setup Flow)

- The Ingress LSR, LSR A, determines that it needs to set up a new LSP to LSR C. The traffic parameters required for the session or administrative policies for the network enable LSR A to determine that the route for the new LSP should go through LSR B, which might not be the same as the hop-by-hop route to LSR C. LSR A builds a Path message with an explicit route of (B,C) and details of the traffic parameters requested for the new route. LSR A then forwards the Path to LSR B as an IP datagram.

- LSR B receives the Path request, determines that it is not the egress for this LSP, and forwards the request along the route specified in the request. It modifies the explicit route in the Path message and passes the message to LSR C.

- LSR C determines that it is the egress for this new LSP, determines from the requested traffic parameters what bandwidth it needs to reserve and allocates the resources required. It selects a label for the new LSP and distributes the label to LSR B in a Resv message, which also contains actual details of the reservation required for the LSP.

- LSR B receives the Resv message and matches it to the original request using the LSP ID contained in both the Path and Resv messages. It determines what resources to reserve from the details in the Resv message, allocates a label for the LSP, sets up the forwarding table, and passes the new label to LSR A in a Resv message.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

- The processing at LSR A is similar, but it does not have to allocate a new label and forward this to an upstream LSR because it is the ingress LSR for the new LSP.

2. Comparative Analysis

The key differences between CR-LDP and RSVP are the reliability of the underlying transport protocol and whether the resource reservations are done in the forward or reverse direction. From these points come many of the other functional differences. [3] The table below summarizes the main technical similarities and differences between CR-LDP and RSVP for LSP Tunnels. The sections that follow explain in greater detail the implications of these technical differences between the protocols.

	CR-LDP support	RSVP Support
Transport	TCP	Raw IP
Security	Yes[1]	Yes[1]
Multipoint-to-Point	Yes	Yes
Multicast Support	No[2]	No[2]
LSP Merging	Yes[3]	Yes[3]
LSP State	Hard	Soft
LSP Refresh	Not needed	Periodic, hop-by-hop
High Availability	No	Yes
Re-routing	Yes	Yes
Explicit Routing	Strict and loose	Strict and loose
Route Pinning	Yes	Yes, by recording path
LSP Pre-emption	Yes, priority based	Yes, priority based
LSP Protection	Yes	Yes
Shared Reservations	No[4]	Yes
Traffic Parm Exchange	Yes	Yes
Traffic Control	Forward Path	Reverse Path
Policy Control	Implicit	Explicit
Layer 3 Protocol Indicated	No	Yes
Resource Class Constraint	Yes	No

Notes:

1. CR-LDP inherits any security applied to TCP. RSVP cannot use IPSEC but has its own authentication. See “Security” below.

2. Multicast support is currently not defined for any of the existing label distribution protocols.

3. CR-LDP does not allow explicit sharing, but see “LSP Modification” below for details of changing the allocated resources.

2.1. Availability of Transport Protocol

The most obvious difference between CR-LDP and RSVP is the choice of transport protocol used to distribute the label requests. RSVP uses connectionless raw IP (or UDP packets at the margins of the network). CR-LDP uses UDP to discover MPLS peers, and uses connection-oriented TCP sessions to distribute label requests. Many operating systems are packaged with full IP stacks including UDP and TCP, but sometimes TCP is not available. On some platforms access to raw IP is restricted. [4] Some existing ATM switches might not already incorporate an IP stack at all and one must be added to support either CR-LDP or RSVP. The availability and accessibility of the transport protocols may dictate which label distribution protocol is used, but is unlikely to be a major factor in the choice made by most MPLS equipment suppliers. RSVP requires that all received IP packets carrying RSVP messages are delivered to the RSVP protocol code without reference to the actual destination IP address in the packet. This feature may require a minor modification to the IP implementation.

2.2. Security

TCP is vulnerable to denial of service attacks, where the performance of the TCP session can be seriously impacted by unauthorized access to the network. This could impact CR-LDP. Authentication and policy control are specified for RSVP. This allows the originator of the messages to be verified (for example using MD5) and makes it possible to police unauthorized or malicious reservation of resources. Similar features could be defined for CR-LDP but the connection-oriented nature of the TCP session makes this less of a requirement. TCP itself could make use of MD5. IPSEC is a series of drafts from the IETF to provide authentication and encryption security for packets transported over IP. If IPSEC support is available in the IP stack it can be used by CR-LDP simply as part of the normal TCP/IP processing. RSVP targets its Path messages at the egress LSR, not at the intermediate LSRs. This means that IPSEC cannot be used because the intermediate LSRs would find themselves unable to access the information in the Path messages.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

2.3. Multipoint Support

Multipoint-to-point LSPs allow label switched paths to merge at intermediate LSRs, reducing the number of labels required in the system and sharing downstream resources. This approach works particularly well in packet-switched networks, but requires non-standard hardware in cell-switched networks such as ATM to prevent interleaving of cells. CR-LDP and RSVP support multipoint-to-point LSPs. [5] Point-to-multipoint (multicast) IP traffic is not addressed by the current version of the MPLS Architecture, so it is not supported by CR-LDP or Labels RSVP. Generic RSVP was originally designed to include resource reservation for IP multicast trees, so it may be easier to extend to support multicast traffic in the future. However, this is an area for further study in both protocols.

2.4. Scalability

The scalability of a protocol should be considered in terms of the network flows it uses, the resources needed to maintain the protocol state at each node, and the CPU load on each node. All of this must be considered in the context of the way in which MPLS is to be used in the network. If trunk LSPs are to be used across the network to connect key edge points, there will be less demand on scalability than using one LSP per flow, or setting up LSPs based on the routing topology. [6] The ability to merge LSPs also has a clear impact on scalability requirements, because data flows may be able to share resource allocations, and the number of labels needed in the network is reduced.

2.5. High Availability

- Availability is a measure of the percentage of time that a node is in service. Equipment vendors typically claim high availability for their boxes when they attain availability levels in the region of 99.999% ("5-nines").

- High Availability is a matter of detecting failures and handling them in a timely manner without any – or with only minimal – disruption to service. Detection and survival of link failures is covered in the following sections. This section is concerned with detection of and recovery from local failures, specifically hardware and software fault-tolerance and the use of online software upgrades to minimize system downtime.

- Survival of LSPs across software failure, and provision of online software upgrades in an MPLS system, are software implementation issues and should be addressed by any vendor serious about the provision of networking solutions. Tolerance of hardware faults relies on hardware detection and reporting of failures, on the availability of backup hardware, and on a suitably designed software implementation.

- Because RSVP is designed to run over a connectionless transport, it lends itself well to a system that must survive hardware failures or online software upgrades. Any control steps that are lost during the failover to the replacement backup system can be recovered by the state refresh processing that is built into RSVP.

- CR-LDP, on the other hand, assumes reliable delivery of control messages and so is not well placed to survive failover. Additionally, it is particularly hard to make TCP fault tolerant (a problem familiar to BGP implementers), with the result that a failover to a backup TCP stack results in the loss of the TCP connections. [7] This is interpreted by CR-LDP as a failure in all of the associated LSPs, which must subsequently be re-established from the ingress LSR.

- Metaswitch is researching ways to extend CR-LDP to allow it to survive online software upgrades and hardware faults.

- Until such extensions are added to CR-LDP, RSVP implementations will be able to provide better solutions for highly available MPLS networks.

2.6. Link and Peer Failure Detection

Where two LSRs are directly connected using a point-to-point link technology, such as ATM, the failure of LSPs can usually be detected by monitoring the state of the interfaces for the LSP. If, for example, an ATM link suffers loss of signal, both CR-LDP and RSVP can use the interface failure notification to detect the failure of the LSP. If two LSRs are connected over a shared medium, such as Ethernet, or are indirectly connected over a WAN cloud, for example using an ATM PVC, they may not necessarily receive a link failure notification from the link hardware.[8] LSP failure detection then relies on techniques inherent in the signaling protocols. So long as normal signaling traffic is flowing nothing else is necessary, but in

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

stable state, additional processing is required to detect a failure.

- CR-LDP uses an exchange of LDP HELLO and KEEPALIVE messages to validate that the LSR peer and link are still active. Although TCP has a built-in keepalive system, this is typically too slow to respond to link and peer failures for the demands of MPLS LSPs.

- In RSVP, Path and Resv refresh messages serve to provide background traffic that indicates that the link is still active. However, to keep the per-LSP refresh traffic in a relatively stable network to a minimum, the refresh timer would be set quite high. To address this problem, an extension has been added to Labels RSVP so that RSVP HELLO messages can be exchanged to prove that the link and peer LSR are still active. The failure detection techniques and speed are therefore similar for both CR-LDP and RSVP, provided that RSVP uses the HELLO extensions. MPLS failure detection is much faster for directly attached LSRs.

2.7. Re-Routing

A strictly specified explicit route cannot be re-routed except by the ingress LSR (initiator). Consequently, failure at some point of an LSP must be reported to the ingress, effectively bringing down the whole LSP. However, a loosely specified portion of an explicit routed LSP, and any part of a hop-by-hop routed LSP, may be re-routed if

- a failure of a link or neighbor is detected (this is called Local Recovery)
- a better route becomes available
- the resources for the LSP are required for a new, higher priority LSP (this is called Pre-emption).

Re-routing is most easily managed from the ingress (including re-routing of strictly specified LSPs) and is supported by both CR-LDP and RSVP, though with slightly different characteristics. An LSR using RSVP can install a new route by simply refreshing the Path for an LSP to a different next-hop as soon as the alternate route is available/required. The old path can be left to time out because refreshes will no longer be sent. However, this wastes resources on the old path. “Make-before-break” is a mechanism whereby the old path is used (and refreshed) while the new path is set up, and then the LSR performing the re-routing swaps to using the new path and tears down the old path. This basic technique can be used to avoid double reservation of resources in both CR-LDP (using the modify value for the action flag on the LABEL_REQUEST) and RSVP (using shared explicit filters). Re-routing of loosely specified parts of LSPs at intermediate LSRs when a “better” route becomes available can lead to thrashing in unstable networks. To prevent this, a loosely specified part of a route may be “pinned”: In CR-LDP this is simply a matter of flagging the loose part of the explicit route as pinned. This means that once the route has been set up, it is treated as though it had been strictly specified and cannot be changed. In RSVP, pinning requires some additional processing. The initial route is specified with a loose hop. The Record Route object is used on the Path and Resv messages to feed back the selected route to the ingress. The ingress can use this information to re-issue the Path message with a strictly specified explicit route.

2.8. LSP Modification

LSP modification, for example, to change the traffic parameters for an LSP, is an equivalent operation to re-routing, though the change of route is optional for LSP modification. This means that the function is always present in RSVP and will be present in CR-LDP provided that the modify value of the action flag on a LABEL_REQUEST is supported by the implementations (this is a relatively recent addition to the CR-LDP drafts and so early implementations might not support modification of LSPs). Note that support for the modify value of the action flag in CR-LDP leads to increased data occupancy, bringing intermediate LSR occupancy up to a figure similar to that required at RSVP intermediate LSRs.

2.9. LAMBDA Networking

Lambda networking presents an interesting set of problems for an MPLS implementation. The full advantages of wavelength switching can only be encompassed if LSPs are switched in hardware without recourse to software. In this respect the lambda network is similar to an ATM network; the MPLS labels are identified with individual wavelengths. The number of wavelengths is, however, very small – too small for the likely number of LSPs transiting any one link. Additionally, the capabilities of an individual wavelength are far in excess of the normal

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

requirements of an LSP, so that such a one-to-one mapping would be highly wasteful of network resources.

2.10. Traffic Control

Significantly, CR-LDP and RSVP perform resource reservation at different times in the process of LSP setup. CR-LDP carries the full traffic parameters on the LABEL_REQUEST. This allows each hop to perform traffic control on the forward portion of LSP setup. The traffic parameters can be negotiated as the setup progresses, and the final values are passed back on the LABEL_MAPPING allowing the admission control and resource reservation to be updated at each LSR. This approach means that an LSP will not be set up on a route where there are insufficient resources. RSVP carries a set of traffic parameters, the Tspec on the Path message.[8] This describes the data that is likely to use the LSP. Intermediate LSPs can examine this information and could make routing decisions based on it. However, it is not until the egress LSR is reached that the Tspec is converted to a Flowspec returned on the Resv message, which gives details of the resource reservation required for the LSP. This means that the reservation does not take place until the Resv passes through the network, with the result that LSP set up may fail on the selected route because of resource shortage. RSVP includes an optional function (adspec) whereby the available resources on a link can be reported on the Path message. This allows the egress LSR to know what resources are available, and modify the Flowspec on the Resv accordingly. Unfortunately, not only does this function require that all the LSRs on the path support the option, but it has an obvious window where resources reported on a Path message may already have been used by another LSP by the time the Resv is received. A partial solution for RSVP LSRs lies within the implementation, which could make a provisional reservation of resources as it processes the Path message. This reservation can only be approximate since it is based on the Tspec not the Flowspec, but it can considerably ease the problem. CR-LDP offers a slightly tighter approach to traffic control especially in heavily used networks, but individual RSVP implementations can provide a solution that is almost as good.

2.11. Policy Control

RSVP is specified to allow the Path and Resv messages to carry a policy object with opaque content. This data is used when processing messages to perform policy-based admission control. This allows Labels RSVP to be tied closely to policy policing protocols such as COPS (Common Open Policy Service) using the Internet draft “COPS Usage for RSVP”. By contrast, CR-LDP currently only carries implicit policy data in the form of the destination addresses, and the administrative resource class in the traffic parameters.

2.12. Layer 3 Protocol

Although an LSP can carry any data, there are occasions when knowledge of the layer 3 protocol can be useful to an intermediate or egress LSR. If an intermediate LSR is unable to deliver a packet (e.g. because of a resource failure) it can return an error packet specific to the layer 3 protocol (such as ICMP for IP packets) to notify the sender of the problem. For this to work, the LSR that detects the error must know the layer 3 protocol in use. Also, at an egress, it may help the LSR to forward data packets if the layer 3 protocol is known. RSVP identifies a single payload protocol during LSP setup, but there is no scope within the protocol for CR-LDP to do this. Even RSVP is unable to help when more than one protocol is routed to a particular LSP. Recent discussions led by Metaswitch in the MPLS Working Group have considered options for identifying the payload protocol in CR-LDP, and for marking the payload packets so that their protocol can be easily determined.

2.13. QOS AND DIFF-SERV

CR-LDP and RSVP have different approaches to Quality of Service (QoS) parameters. The RSVP Tspec object carried on Path messages describes the data that will flow rather than the QoS that is required from the connection. Various RFCs and Internet drafts describe how to map from different QoS requirements to the Tspec (for example, RFC 2210 - The Use of RSVP with IETF Integrated Services). The CR-LDP specification is more explicit about how the information carried on a LABEL_REQUEST message is mapped for QoS. Support for Diff-Serv (IP Differentiated Services)

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

is addressed by an Internet draft (draft-ietf-mpls-diff-ext), which defines extensions to LDP, RSVP and CR-LDP. If implemented, this draft extends the full function of Diff-Serv to an MPLS network.

2.14. Provision of VPNs

Virtual Private Networks (VPNs) are an important feature of the service provided by ISPs to their customers. VPNs allow physically private networks to be extended to encompass remote sites by connecting them through the Internet. A customer in these circumstances expects to be able to preserve their IP addresses (which might not be globally unique) and to have the security of their data guaranteed. MPLS can provide an excellent solution as described in RFC 2547. Both CR-LDP and RSVP are suitable MPLS signaling protocols for VPNs over MPLS.

2.15. Voice Over IP and Voice Over MPLS

Voice over IP (VoIP) is an exciting development in Internet technology. The concept of a single infrastructure for voice and data, providing faster, cheaper and value-added services, is very attractive. MPLS is set to be a major component in VoIP networks, offering connection-oriented paths with resource reservation through the connectionless Internet. Voice over MPLS (VoMPLS) is the term given to the transfer of voice traffic over an MPLS network. This could involve establishing LSP Tunnels to act as trunks for multiple calls, or setting up LSPs for the duration of individual calls. Alternatively, VoMPLS could mean sending voice samples as labeled MPLS packets without including IP headers. Whichever approach is used, both CR-LDP and RSVP are suitable MPLS signaling protocols.

2.16. MIB Management

Traffic Engineered LSPs can be managed at their ingress and inspected at their egress through the MPLS Traffic Engineering MIB. This MIB is currently in an early stage which slightly favors CR-LDP, but new drafts will be produced that fully support RSVP and CR-LDP.

2.17. Acceptance/Availability

There is currently no clear “winner” between RSVP and CR-LDP in terms of market acceptance. Although generic RSVP has been available for a number of years from a variety of equipment vendors, and in that sense is an established network protocol, the changes required to a generic RSVP stack to add support for Labels RSVP are non-trivial, and hence Labels RSVP is in many respects a new protocol. CR-LDP is based on ideas that have been implemented in proprietary networks for as long as ten years, but as an IETF protocol it is very new and somewhat unproven.

2.18. Interoperability

There are two interoperability issues to be addressed. Do two implementations support a compatible set of options, and do they interpret the specifications in the same way? The option sets are functions of the flexibility of the protocol. RSVP has more implementation options than CR-LDP and so is perhaps at more risk. However, the protocol is specified to allow interworking between implementations that support different function sets. An IETF MPLS draft (draft-loa-mpls-cap-set) provides a list of capability sets to allow implementations to identify the functions that they provide. Interoperability testing is clearly the only way to prove that two implementations interwork correctly. Interoperability forums are being set up in many places including the University of New Hampshire InterOperability Labs (UNH IOL), George Mason University in Washington DC with the support of UUNET, EANTC in Berlin, NetWorld & Interop events. All of these forums will include work on RSVP and CR-LDP. Participation in interoperability events is a clear requirement for all MPLS software vendors. Testing for hardware vendors will be a combination of involvement in interoperability events, in-house testing with competitors’ equipment, and collaborative work with other vendors. Hardware vendors have a right to expect the support of their software suppliers during interoperability testing.

Summary

CR-LDP and Labels RSVP are both good technical solutions for setting up and managing Traffic Engineered LSPs. Early versions of both protocols had some functional omissions, but these are being fixed by subsequent Internet drafts so that the level of function provided by each protocol is similar. Some key differences in the structure of the protocols and the underlying transport mean that the support that the protocols can provide will

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

never converge completely. These differences and the differences in speed and scope of deployment will be the main factors that influence vendors when they are selecting a protocol. The choice between RSVP and CR-LDP should be guided by the function of the target system. What LSP setup model will be used? How stable are the LSPs – do they represent permanent trunks or short-duration calls? How large is the network and how complex is it? Is this a stand-alone network or must the components interwork with other hardware and other networks? A final consideration must be the robustness of the hardware solution. What level of fault tolerance is required? How important is high availability?

- Two informational Internet drafts may help guide the choice of protocol.
- Applicability Statement for Extensions to RSVP for LSP-Tunnels
- Applicability Statement for CR-LDP

References

[1] AT&T Knowledge Ventures. (2007, July 25). *Transitioning to an MPLS Network*. Retrieved November 19th, 2007, from http://www.business.att.com/nx_resource.jsp?repoid=Topic&rtype=Whitepaper&rvalue=eb_fpoc_navigating_to_mpls_enabled_networks&repoitem=vpns&segment=ent_biz

[2] AT&T Knowledge Ventures. (2007, August 31). *Understanding VPN Technology Choices:Comparing MPLS, IPSec and SSL*. Retrieved November 19th, 2007, from http://www.business.att.com/nx_resource.jsp?repoid=Topic&rtype=Whitepaper&rvalue=understanding_vpn_technology_choices&repoitem=vpns&segment=ent_biz&guid=4BFDAE84-C61B-416F-886A-F606E9678B1C;08905D72-1FE7-450C-8EA5-B5F1565DD558

[3] Cisco Systems, Inc. (2004). *Managed VPN – Van Wijnen and Versatel*. Retrieved November 19th, 2007, from http://www.cisco.com/en/US/netsol/ns465/networking_solutions_customer_profile0900aecd801aa3f5.html

[4] Cisco Systems, Inc. (2005). *From Frame Relay to IP VPN: Why to Migrate, Why to Out-Task.* Retrieved November 18th, 2007, from http://www.cisco.com/en/US/netsol/ns458/networking_solutions_white_paper0900aecd8017a894.shtml

[5] Layer 2 MPLS VPN. (2007, October 20). In *Wikipedia, The Free Encyclopedia*. Retrieved 18:03, November 18, 2007, from http://en.wikipedia.org/w/index.php?title=Layer_2_MPLS_VPN&oldid=165912463

[6] Multiprotocol Label Switching. (2007, November 7). In *Wikipedia, The Free Encyclopedia*. Retrieved 18:04, November 18, 2007, from http://en.wikipedia.org/w/index.php?title=Multiprotocol_Label_Switching&oldid=169803565

[7] Pseudo-wire. (2007, November 17). In *Wikipedia, The Free Encyclopedia*. Retrieved 18:01, November 18, 2007, from http://en.wikipedia.org/w/index.php?title=Pseudowire&oldid=172121304

[8] Juniper Networks. Traffic Engineering for the New Public Network. http://www.omimo.be/magazine/00q4/2000q4_p054.pdf

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Annotating Speech Corpus for Prosody Modeling in Indian Language Text to Speech Systems

Kiruthiga S[1] and Krishnamoorthy K[2]

[1] Faculty of CSE, Anna University Coimbatore
Coimbatore, Tamil Nadu, India

[2] Department of CSE, Sudharsan Engineering College
Pudukkottai, Tamil Nadu, India

Abstract

A spoken language system, it may either be a speech synthesis or a speech recognition system, starts with building a speech corpora. We give a detailed survey of issues and a methodology that selects the appropriate speech unit in building a speech corpus for Indian language Text to Speech systems. The paper ultimately aims to improve the intelligibility of the synthesized speech in Text to Speech synthesis systems. To begin with, an appropriate text file should be selected for building the speech corpus. Then a corresponding speech file is generated and stored. This speech file is the phonetic representation of the selected text file. The speech file is processed in different levels viz., paragraphs, sentences, phrases, words, syllables and phones. These are called the speech units of the file. Researches have been done taking these units as the basic unit for processing. This paper analyses the researches done using phones, diphones, triphones, syllables and polysyllables as their basic unit for speech synthesis. The paper also provides a recommended set of combinations for polysyllables. Concatenative speech synthesis involves the concatenation of these basic units to synthesize an intelligent, natural sounding speech. The speech units are annotated with relevant prosodic information about each unit, manually or automatically, based on an algorithm. The database consisting of the units along with their annotated information is called as the annotated speech corpus. A Clustering technique is used in the annotated speech corpus that provides way to select the appropriate unit for concatenation, based on the lowest total join cost of the speech unit.

Keywords: *Speech corpus, Indian languages, Concatenative Speech synthesis, Syllables, Prosody*

1. Introduction

The area of speech processing has gone to the extent of synthesizing natural speech that highly resembles a human voice. The idea behind is so simple. The system is fed with a human voice, processed and stored as smaller units. When the target text is given for converting into speech, these smaller units are concatenated to build a continuous speech. This technique is called as Concatenative Speech synthesis. The processing of human voice into smaller units is called the Speech Corpus development. The corpus can either be general purpose or application specific, based on the type of Speech Processing system. To build a corpus, a text file is first selected, which has phonetically rich and prosodically stable sentences. For some application specific Text to Speech synthesis systems, prosody rich text file may also be selected. The text file may be a News bulletin, forum interviews, every-day conversations in an organization, conversation in road traffic, etc. Then, with the help of a native speaker, who may be a news reader, this text file is made to be read and recorded. This recording is the corresponding speech file. A speech file is an audio file whose size generally ranges from several minutes to hours. This speech file is the phonetic representation of the selected text file. The speech file is processed in different levels viz., paragraphs, sentences, phrases, words, syllables and phones, which are called the speech units of the file. Many researches have been carried out using these units as the basic units of processing. Research in the area of speech synthesis have an importance in many applications, including information retrieval services over telephone such as banking services, public announcements at places like train stations and reading out manuscripts for collation, reading emails, faxes and web pages over telephone and voice output in automatic translation systems.

2. Speech Corpus for Indian Languages

The speech synthesis systems are developed for the Indian languages such as Hindi, Tamil, Telugu, Kannada, Bengali and Marathi. These Indian languages share phonological features. For Indian languages, building a speech corpus is a different task than that of the English speech corpus. Since prosodic processing such as duration and intonation prediction has to be done in the corpus-development stage itself, some more information has to be

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

annotated with the basic units after storing them in the corpus. Methodologies which include clustering are incorporated to improve the corpus, thus improving the entire system. Issues such as detection of mispronunciation, detection of pronunciation variants, un-transcribed speech are to be noted and addressed.

Phonetizers or the Grapheme to Phoneme converters are tools that convert the text corpus into its phonetic equivalent [7]. The phonetic nature of Indian scripts reduces the effort to building mere mapping tables and rules for the lexical representation. These rules and the mapping tables together comprise the Grapheme to Phoneme converters.

3. Concatenative Speech Synthesis

Concatenative speech synthesis uses phones, diphones, syllables, words or sentences as basic speech units. Speech is synthesized by selecting appropriate units from a speech database, called as a speech corpus, and concatenating them. Many researches have been made, selecting each separate unit as the basic unit. The size of the database differs, as the number and size of individual units differ. This also affects the quality of synthesized speech. If large units such as phrases or sentences are stored and used, the quality of synthesized speech is good, although the domain of synthesis is not unrestricted text. When small units such as phones are used, a wide range of words or sentences can be synthesized but with poor speech quality.

3.1 Phones as the basic unit

When phones are considered as basic units, the size of the database will be less than 50 units for Indian languages. Because, including allophones, the number of phone units of any Indian language is less than 50. The database is small, but phones provide very less co-articulation information across adjacent units, thus failing to model the dynamics of speech sounds with their large variability depending on context. Thus phones are found to be inefficient for speech synthesis.

3.2 Diphones and triphones as the basic unit

A diphone is made of two connected half phones and captures the transition between two phones by starting in the middle of the first phone and ending in the middle of the second one. It captures the co-articulation effects and minimizes the discontinuities at the concatenation points. Diphones are relatively bigger units than phones. There are about 1000 to 2000 diphones found in Indian languages. Unlike phones, they do not exhibit allophonic variations. i.e., each diphone has only one instance of pronunciation. Though diphone concatenation can produce a reasonable quality speech, a single example of each diphone is not enough to produce good quality speech. Moreover, diphone-based synthesizers need elaborate prosody rules to produce natural speech. In newer Speech Processing systems, one other unit was introduced, called triphones. A triphone provides more co-articulation information than diphones, since they involve in blending with both previous and next phones in the phrase. Compared to diphones, triphones are larger in number thus leading to a larger size of database. But some triphones rarely occur in the language model, thus making the database unnecessarily large. An attempt has been made for word and triphone based speech synthesis [4], where a full dictionary, fast dictionary and a Simple Breadth First Search Manager were experimented with. In scenarios where the vocabulary is limited, repeatability of sentences is more and speakers are limited, the word-based recognizer with trained voice on trained sentences is highly suitable. In the same scenario, when the vocabulary is high and speakers are limited, triphone based model is suitable.

3.3 Syllables as the basic unit

Indian languages are syllable centered, where pronunciations are mainly based on syllables. A Syllable can be the best unit for Indian language Speech synthesis systems. Intelligible speech synthesis is possible for Indian languages with syllable as the basic unit. Syllable units being larger in comparison to phones or diphones, can capture co-articulation better than phones. The number of concatenation points decreases when syllable is used as the basic unit. Syllable boundaries are characterized by regions of low energy, providing more prosodic information. A grapheme in Indian languages is close to a syllable. The general format of an Indian language syllable is C*VC*, where C is a consonant, V is a vowel and C* indicates the presence of 0 or more consonants. There are about 35 consonants and 18 vowels in Indian languages [12]. There are defined set of syllabification rules formed by researchers, to produce computationally reasonable syllables. A rule based grapheme to syllable converter was used for syllabification. Some of the rules used to perform grapheme to syllable conversion [9] are:

- Nucleus can be Vowel(V) or Consonant (C)
- If onset is C then nucleus is V to yield a syllable of type CV
- Coda can be empty or C
- If characters after CV pattern are of type CV then the syllables are split as CV and CV.
- If the CV pattern if followed by CCV then syllables are split as CVC and CV.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

- If the CV pattern is followed by CCCV then the syllables are split as CVCC and CV
- If the VC pattern is followed by V then the syllables are split as V and CV.
- If the VC pattern is followed by CVC then the syllables are split as VC and CVC

Researches were conducted to find which order of syllables is best acceptable for synthesis [8]. The following are the recommended combinations:

a) Monosyllables at the beginning of a word and bisyllables at the end.

b) Bisyllables at the beginning of a word and monosyllables at the end.

c) Monosyllables at the beginning and trisyllables at the end of a word.

d) Trisyllables at the beginning and monosyllables at the end of a word.

3.4 Polysyllables as the basic unit

An attempt exploring a Speech synthesizing system using polysyllabic units has been made [5]. Since polysyllable units are formed using the monosyllable units already present in the database, the synthesis quality can be improved without augmenting any new set of units. The system uses a large database, which consists of syllables, bisyllables and trisyllables. While synthesizing, the first matching trisyllable is selected followed by the bisyllable and monosyllable units, as needed. Picking up the largest possible unit in the database improves the quality of speech, since the number of co-articulation points greatly reduce.

Selection of an appropriate candidate unit set is carried out using search algorithms. Units which incur lowest total join cost for a word are preferred. Total join cost is the sum of selection cost of the candidate unit and the join cost of the selected candidate units [3].

$$C_{total} = C_{Sel} + C_{Join} \quad (1)$$

where, C_{total} is the total join cost, C_{Sel} is the selection cost and CJ_{oin} is the cost of join.

4. Clustering the syllables

The need for selecting the phonetically and prosodically best units for synthesis needs clustering the units in the database. An acoustic distance measure is defined to measure the distance between two units of the same phone type. Factors concerning prosodic and phonetic context are evaluated to form cluster units within a unit type. A decision tree is built based on questions concerning the phonetic and prosodic aspects of the grapheme. Ultimately the leaves of the decision tree are the list of database units that best suit the required aspects. At the time of synthesis, for each target, the appropriate decision tree is used to find the best cluster of candidate units. A search is then made to find the best path through the candidate units.

Pruning is performed to remove spurious atypical units which may have been caused by mislabeling or poor articulation in the original recording. It also removes those units which are so common that there is no significant distinction between candidates. Databases were tested with this clustering method in hand. The method [3] produced both extreme high quality examples and extremely low quality ones. Minimizing these bad examples was the important target.

A clustering technique is suggested in [2], which includes pre-clustering works that tags the syllables as begin, middle and end, depending on the occurrence of the syllable in the word. Further clustering is performed by tagging the syllables based on type of the syllable (v, c*v, vc*, c*vc*) and nature of the constituent vowels and consonants. Syllables of the same type were clustered using features like word length of the phrase, relative position of the syllable in the phrase, relative position of the parent phrase and the features of the preceding and the following syllables in the phrase. The acoustic distance measure is calculated using the mel frequency cepstral coefficients (MFCC). Using the feature set and the acoustic distance measure, the decision tree was built for each of the unique syllable in the database. Questions were used at the nodes to find the best set of candidate syllables. Morpheme tags are used for phrase prediction. This technique has improved the quality of the synthesized speech to a greater extent.

4.1 Prosody

The Speech synthesis system exhibits its naturalness when the corpus is annotated with prosodic information. Prosody modeling is subdivided into modeling the following constituents of prosody - phrasing, duration, intonation and intensity. Two major approaches for prosody modeling are the rule based approach and the corpus based approach. In the rule based approach, linguistic experts derive a complicated set of rules to model prosodic variations by observing natural speech. In the corpus based approach, a well-designed speech corpus, annotated with various levels of prosodic information is used. The corpus is analyzed automatically to create a prosodic model which is then made to synthesize a training data set, following which the test dataset is evaluated. Based on the performance on test data, the models are then improved. The syllables have sufficient duration information as it improves the quality of synthetic speech when used as a duration model. Thus syllables are identified as the best-suited processing units for Indian language Speech synthesis.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

5. Proposed Work

A prosodically stable set of sentences, such as a dictated full length text, is selected. The text is broken down to appropriate segmental or suprasegmental structure of utterances such as phones, syllables, or words. These structures are called speech units. A repository of speech units is created and each unit is annotated with prosodic features. i.e., each entry in the repository is a set which consists of a speech unit and its prosodic information. These speech units are henceforth being called as candidate units. During synthesis a candidate unit is selected, considering the lowest total join cost for a word. One of the important prosodic aspects is the prediction of phrase boundary. Some morpheme tags are identified as phrase boundary indicators. The work aims to automatically detect these morpheme tags in order to predict the phrase boundary. These phrase boundary details are annotated with the candidate units, as prosody information. The synthesized speech shall have to cater the issues of pitch normalization across speakers.

6. Conclusions

In this paper, the importance of annotating every individual candidate unit in the speech corpus with necessary prosody information is analyzed. The use of various units as basic units for developing a speech corpus for Indian language Text to Speech synthesis system is also sorted. The analysis significantly shows the need for Prosody modeling to obtain high quality intelligent speech. The appropriate speech unit for Indian languages is the syllable, which is the good constituent of prosodic features of the Indian languages. Selection of the appropriate candidate unit annotated with necessary prosodic information play a vital role in the development of Speech corpus, which obtains naturalness in the synthesized speech. A Clustering Technique is also proposed which selects the best unit when more than one candidate unit shows the lowest total join cost.

References

[1] M. Nageshwara Rao, Samuel Thomas, T. Nagarajan and Hema A. Murthy, “Text-to-speech synthesis using syllable-like units,” in National Conference on Communication, Kharagpur, India, Jan 2005, pp 277-280.

[2] G L Jayavardhana Rama, A G Ramakrishnan, R Muralishankar and Vijay Venkatesh. ”Thirukkural - A text to speech synthesis system”. Proc. Tamil Internet 2001, Kuala Lumpur 2001,92-97.

Hunt, A.J. and Black, A.W., “Unit selection in a concatenative speech synthesis system using a large speech database”, Proc. ICASSP, vol. 1,pp. 373–376, 1996.

[3] Alan W. Black and P. Taylor, “Automatically clustering similar units for unit selection in speech synthesis”, Proc. EUROSPEECH 97, Rhodes, Greece, 1997, Vol. 2, pp. 601-604. K. Elissa, “Title of paper if known,” unpublished.

[4] R. Thangarajan, A.M. Natarajan and M. Selvam, “Word and Triphone Based Approaches in Continuous Speech Recognition for Tamil Language”, in WSEAS Transactions on Signal Processing, Issue 3, Volume 4, March 2008.

[5] Vinodh M Vishwanath, Ashwin Bellur, Badri Narayan K, Deepali M Thakare, Anila Susan, Suthakar N M and Hema A Murthy,“Using Polysyllabic units for Text to Speech Synthesis in Indian languages,” Proceedings of National Conference on Communication (NCC),pp.1-5, 29-31 Jan. 2010

[6] Kishore Prahallad, Arthur R Toth, Alan W Black, “Automatic Building of Synthetic Voices from Large Multi-Paragraph Speech Databases”, in Proceedings of Interspeech, Antwerp, Belgium 2007.

[7] Anumanchipalli Gopalakrishna, Rahul Chitturi, Sachin Joshi, Rohit Kumar, Satinder Singh, R.N.V Sitaram andS.P. Kishore, "Development of Indian Language Speech Databases for Large Vocabulary Speech Recognition Systems", Proceedings of International Conference on Speech and Computer (SPECOM), Patras, Greece, Oct 2005.

[8] T.Jayasankar, R.Thangarajan, J.Arputha Vijaya Selvi, “Automatic Continuous Speech Segmentation to Improve Tamil Text-to-Speech Synthesis”, in International Journal of Computer Applications (0975 – 8887), Volume 25– No.1, July 2011.

[9] S. Saraswathi and T.V. Geetha, “Design of language models at various phases of Tamil speech recognition system”, International Journal of Engineering, Science and Technology Vol. 2, No. 5, 2010, pp. 244-257.

[10] Kiruthiga S, Krishnamoorthy K, “Design Issues in Developing Speech Corpus for Indian Languages”, 2nd International Conference on Computer Communication and Informatics, Jan 2012, Vol. 2, 978-1-4577-1581-5.

[11] Samuel Thomas, M. Nageshwara Rao, Hema A. Murthy and C.S. Ramalingam, “Natural sounding TTS based on syllable-like units,” in the Proceedings of the 14th European Signal Processing Conference, Florence, Italy, Sep 2006.

[12] G. L. Jayavardhana Rama, A. G. Ramakrishnan, R. Muralishankar and R Prathibha, “A Complete Text-To-Speech Synthesis System In Tamil”, in 0-7803-7395-2/02, IEEE proceedings 2002.

[13] Ashwin Bellur, K Badri Narayan, Raghava Krishnan K, Hema A Murthy, “Prosody Modeling for Syllable-Based Concatenative Speech Synthesis of Hindi and Tamil”, DOI: 10.1109/NCC.2011.5734737 , IEEE proceedings 2011.

[14] Grażyna Demenko, Stefan Grocholewski, Agnieszka Wagner, Marcin Szymanski, “Prosody Annotation for Corpus based Speech Synthesis”, Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9.

Kiruthiga S is a research scholar in Anna University Coimbatore. She obtained her Master of Engineering Degree in 2006 under Anna University Chennai. She completed her Bachelor of Engineering Degree under Bharathiyar University in the year 2003. She has worked as Lecturer in Sona College of Technology Salem for 4 years. Her research interest includes Text to Speech Synthesis system and other Natural Language Processing applications.

Krishnamoorthy K is a Professor in Sudharsan Engineering College, Pudukkottai. He is one of the recognized research

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

supervisor in Anna University Coimbatore. He obtained his Ph. D degree in 2007 and Master of Engineering Degree in 2003 both under Dayananda Sagar University, Bangalore. He has an experience as an academician for a span of a decade. His research interest includes Natural Language Processing and Digital Image Processing.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

A Novel Algorithm for Manets using Ant Colony

Javad Pashae Barbin `[1], Majid Taghipoor[2] and Vahid Hosseini [3]

[1]Islamic Azad University, Naghadeh Branch
Uromieh; Iran

[2]University of Applied Science Technology
Uromieh; Iran

[3] Islamic Azad University, Shabestar Branch
Uromieh; Iran

Abstract

Mobile Ad-hoc Networks have recently attracted a lot of attention in the research community as well as the industry. Quality of Service support for MANETs is an exigent task due to dynamic topology and limited resource. Routing, the act of moving information across network from a source to a destination. Conventional routing algorithms are difficult to be applied to a dynamic network topology, therefore modeling and design an efficient routing protocol in such dynamic networks is an important issue. It is important that MANETs should provide QoS support routing, such as acceptable delay, jitter and energy in the case of multimedia and real time applications. One of the meta-heuristic algorithms which are inspired by the behavior of real ants is called Ant Colony Optimization algorithm. In this paper we propose a new on demand QoS routing algorithm "Ant Routing for Mobile Ad Hoc Networks" based on ant colony. The proposed algorithm will be highly adaptive, efficient and scalable and mainly reduces end-to-end delay in high mobility cases.

Keywords: *MANET, Quality of Service (QoS), Ant Colony, Routing Protocol.*

1. Introduction

Mobile Ad hoc Networks (MANET) is a communication network of a set of mobile nodes, placed together in an ad hoc manner, without any fixed infrastructure that communicate with one another via wireless links. The devices used to form an Ad Hoc Network possess limited transmission range; therefore, the routes between a source and a destination are often multi hop. As there are no separate routers, nodes that are part of the network need to cooperate with each other for relaying packets of one another towards their ultimate destinations as they do not have central administration, it is easy to deploy and expand. This kind of network is very flexible and suitable for applications such as temporary information sharing in conferences, military actions and disaster rescues. However, multi-hop routing, random movement of mobile nodes and other features unique to MANETs lead to enormous overheads for route discovery and maintenance. Furthermore, compared with the traditional networks, MANETs suffer from the resource constraints in energy, computational capacities and bandwidth. [1]

With the increasing needs of QoS provisioning for evolving applications such as real-time audio/video, it is desirable to support these services in ad hoc networking environments. The network is expected to guarantee a set of measurable specified service attributes to the user in terms of end to-end delay, bandwidth, probability of packet loss, energy and delay variance (jitter). The role of a QoS routing strategy is to compute paths that are suitable for different type of traffic generated by various applications while maximizing the utilizations of network resources. [3] The major objectives of QoS routing are: [2]

1. To find a path from source to destination satisfying user's requirements
2. To optimize network resource usage and
3. To degrade the network performance when unwanted things like congestion, path breaks appear in the network.

The main problem to be solved by QoS routing algorithm is the Multi-Constraint Path problem. Algorithms to solve this family of problems are known to be heuristics which can reduce the complexity of the path computation. [2, 4] The path computation algorithm is at the core of QoS routing strategies. Instead of using a shortest path algorithm based on statically configured metrics, as in traditional routing protocols, the algorithm must select several alternative paths that are able to satisfy a set of constraints regarding, for instance, end-to-end delay bounds and bandwidth requirements. However, the algorithms to solve such a problem have been shown to have, in general, high computational complexity. [3]

Several approaches have been proposed to address the complexity of multi-constrained path computation problem. The paths that satisfy these multiple constraints

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

are called as feasible paths. The solution of this kind of multi constrained problem requires a path computation algorithm that finds paths satisfying all the constraints. Since the optimal solution of this type of problems with multiple additive and independent metrics is NP-complete, usually heuristics or approximation algorithms can be used to solve such kind of problems. [1] In this paper, I will present a new approach for an ad hoc routing algorithm, which is based on Ant Colony Optimization (ACO) algorithm. We show that for a wide range of different environments and performance metrics, ACO performs better than AODV, AOMDV.
The rest of the paper is organized as; section 2 discusses related ant based routing protocol. Section 3 describes the new proposed routing protocol. Section 4 discusses performance evaluation parameters and results of routing protocol followed by conclusions in section 5.

2. Ant colony optimization (ACO) algorithm

ACO algorithms have been inspired by the behavior of a real ant colony. The algorithm can find the optimum solution by generating artificial ants. Just as real ants search for food in their environment, the artificial ants search the solution space. The probabilistic movement of ants in the system allows the ants to explore new paths and to re-explore the paths visited earlier. The strength of the pheromone deposit directs the artificial ants toward the best paths and pheromone evaporation allows the system to forget old information and avoid quick convergence to suboptimal solutions. The probabilistic selection of paths allows the artificial ants to search for a large number of solutions. ACO has been applied successfully to discrete optimization problems such as the traveling salesman problem and routing [5, 6]. A number of proofs for the convergence to the optimum path of the ACO can be found in [8, 9].

2.1. Basic and Background

The basic idea of the ant colony optimization meta-heuristic is taken from the food searching behavior of real ants. When ants are on the way to search for food, they start from their nest and walk toward the food. When an ant reaches an intersection, it has to decide which branch to take next. While walking, ants deposit pheromone1, which marks the route taken. The concentration of pheromone on a certain path is an indication of its usage. With time the concentration of pheromone decreases due to diffusion effects. This property is important because it is integrating dynamic into the path searching process. [7]

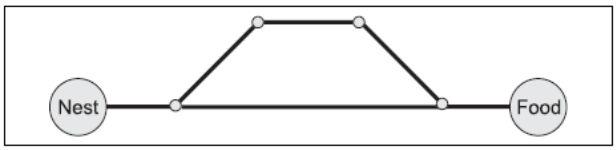


Fig. 1 All ants take the shortest path after an initial searching time. [7]

Figure 1 shows a scenario with two routes from the nest to the food place. At the intersection, the first ants randomly select the next branch. Since the below route is shorter than the upper one, the ants which take this path will reach the food place first. On their way back to the nest, the ants again have to select a path. After a short time the pheromone concentration on the shorter path will be higher than on the longer path, because the ants using the shorter path will increase the pheromone concentration faster. The shortest path will thus be identified and eventually all ants will only use this one.
This behavior of the ants can be used to find the shortest path in networks. Especially, the dynamic component of this method allows a high adaptation to changes in mobile ad-hoc network topology, since in these networks the existence of links are not guaranteed and link changes occur very often.[7]

2.2. Solving Network Routing Using ACO

Mobile ad hoc network routing is a difficult problem because network characteristics such as traffic load and network topology may vary stochastically and in a time varying nature. The distributed nature of network routing is well matched by the multi agent nature of ACO algorithms.[8] The given network can be represented as a construction graph where the vertices correspond to set of routers and the links correspond to the connectivity among routers in that network. Now network route finding problem is just finding a set of minimum cost path between nodes present in the corresponding graph representation which can be done easily by the ant algorithms.

3. Proposed Algorithm

This paper proposes an on demand QoS routing algorithm. Since the requirements for various applications may vary time to time, the approach for QoS routing may not be proactive. The proposed approach has two phases namely route discovery phase and route maintenance phase. When a source node has to pass data to a destination node with QoS requirements it starts with the route discovery phase. Once the route is found, the data transfer will take place. While data transmission is going on, it is also required to maintain the path to the destination. This is very much desirable and required in mobile ad hoc networks and

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

hence is done in the route maintenance phase. The proposed algorithm differs from other similar algorithms because routing from destination to source is done.

3.1. Route Discovery Phase

The outline of this phase is as follows:
Step 1: Let the source node S has data to send to a destination D with QoS requirements delay, energy, bandwidth and hop count. Each node has a routing table.
Step 2: D initiates a ForwardAnt to source S through all its neighbors which it has learned from periodic hello messages.
Step 3: While traveling to the source the ForwardAnt collects transmission delay of each link, processing delay at each node, the available capacity of each link, the number of hops visited and stores in routing table each node.
Step 4: When the ForwardAnt reaches the source, it will be converted as BackwardAnt and forwarded towards the original destination. The BackwardAnt will take the same path of the corresponding ForwardAnt but in reverse direction.
Step 5: For every BackwardAnt reaching an intermediate node or destination node, the node can find the delay, bandwidth, hop count from the received ant to the respective destination. Now the node can calculate the path preference probability to reach the source.
Step 6: If calculated path preference probability value is better than the requirements, the path is accepted and stored in memory.
Step 7: The path with higher path preference probability will be considered as the best path and data transmission can be started along that path.
In similar algorithms, Route discovery is achieved by flooding forward ants to the destination then the backward will take the same path of the corresponding forward but in reverse direction. Finally, the data packets sent from source to destination. While the proposed algorithm, the algorithm starts from the destination node and the data packet is sent with the backward.
The above stochastic strategy, described in Algorithm 1, establishes multiple paths between the source and destination. As a result, in contrast to regular position based routing algorithms which usually find a single route to the destination, proposed algorithm is a multipath routing algorithm (i.e. like other ACO routing algorithms).

Algorithm 1 routing algorithm

```
{S is the source node, D is the destination node and
C is the current node}
for each clock time do
if C = S then
Send one BackwardAnt from reverse direction
else
Send one ForwardAnt to neighbors
end if
for each message m in D's buffer do
if (m→type = ForwardAnt)  then
send m to NextHop
if NextHop = S then
m →type = Backward
end if
else if m→type = BackwardAnt then
find NextHop in C's BackRouting table
send m to NextHop
IncreasePheromone(NextHop,m)
if NextHop = D or NextHop = Intemediate node then
update averages of packet delays and remaining energy
drop m
end if
end if
end for
Evaporate()
/*calculate higher path preference probability will be
considered as the best path and data transmission can be
started along that path*/
end for
```

3.2. Route maintenance phase

As mentioned before, the values of pheromone trails are stored in a table at each node. Suppose that a data is currently residing in node N and this node has k neighbors H_1, H_2... Hk and φ_i is the amount of pheromone assigned to e_i and d_i. The data will select H_i as the next node with a probability p_i (ed) which is calculated using the equation 3.

$$P_i(e) = \frac{e_i}{\sum e_i} \qquad (1)$$

p_i (e) measure of energy consumption at node i is compared with the total energy consumption.

$$P_i(d) = \frac{1/d_i}{\sum 1/d_i} \qquad (2)$$

pi (d) measure of routing delay at node i is compared with the total routing delay.

$$P_i(ed)\frac{P_i(e) + \boldsymbol{\theta} P_i(d)}{(\sum P_i(e) + \boldsymbol{\theta} P_i(d))} \qquad \boldsymbol{\theta} < 0 \qquad (\mathbf{3})$$

Where:
e_i: energy level of node i.
d_i: latency of node i.
θ : Constant value $\theta < 0$

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Algorithm 2 Select Next Hop

Input: node N
for i = 1 to the number of N's neighbors do
$$P_i(ed)\frac{P_i(e)+\theta P_i(d)}{\sum(P_i(e)+\theta P_i(d))}$$
Return neighbor i with probability p_i
end for

Each node scans pi(ed) and selects the node with the highest preference and forwards the data packet (BackwardAnt) to this node. This process is repeated until the destination node is reached. The ants move from one node to another and carry table 1.

Table 1: route information

Node	Energy	Time	Estimated Waiting
X	Ex	t	Δt

4. Simulation and Performance analysis

VC++ 6.0 program is used to realize the simulation of the algorithm. For our results we assumed 50 mobile nodes communicating via IEEE 802.11. The nodes move inside a simulation area of 1500m*300m. The simulation time is 900 seconds. The nodes move with a maximal velocity of 10 m/s and according to the random waypoint mobility model.

4.1. Simulation and Performance analysis

A new routing algorithm should show its performance in comparison with existing and known algorithms. The simulated traffic is Constant Bit Rate (CBR). We compare our protocol with AODV and AOMDV. We evaluate mainly the performance according to the metrics packet delay, overload and packet delivery.

Our proposed algorithms are named MPAC (Multi Path Ant Colony). We will first discuss the robustness of the routing protocols. Figure 2 shows the delivery rate, i.e., the part of packets a certain routing protocol was able to deliver properly. This value is important, since it describes the performance which transport protocols will see, i.e., the throughput is restricted by this value.

In situations with very high dynamics MPAC shows the best performance. With less dynamic, down to 300 seconds of pause time, AODV has best performance. Because AODV uses particular route, over time this will destroy the path and find a new path will lead to delay. But in the proposed protocol does not caused a problem because use of different routes.

Figure 3 shows the overhead of MPAC. All results are very close through all simulation scenarios. This shows that MPAC creates much less routing overhead for all considered mobility scenarios.

Figure 4 depicts the needed number of packets to perform the routing job for all three routing algorithms. In the cases with high mobility it is obvious that MPAC and AOMDV create the least overhead. Especially MPAC shows here a better performance than AOMDV in high dynamic. With less, the overhead of MPAC and AODV are very close. AOMDV shows here again their poor performance by creating large numbers of routing packets.

Figure 5 show average remaining energy for all three routing algorithms. Remaining energy of MPAC is high than AODV and AODV. That is because optimum tour is established by the way of flooding, costing more time in MPAC, and lead to increasing consumption energy of network.

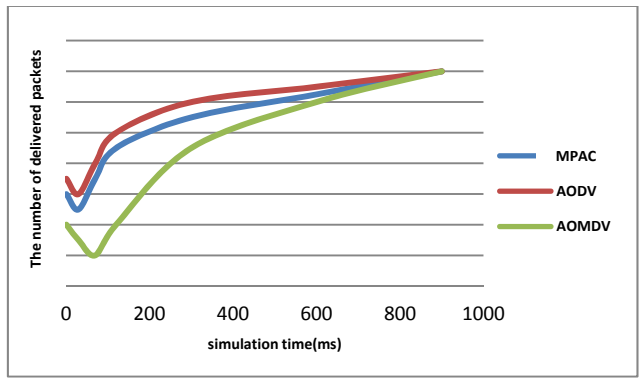


Fig. 2 Delivery rate

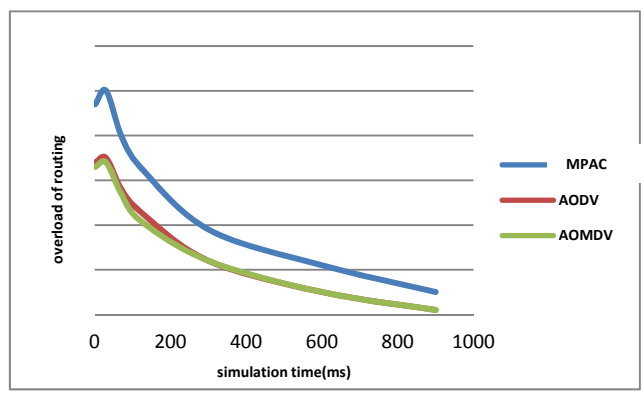


Fig. 3 Overhead

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

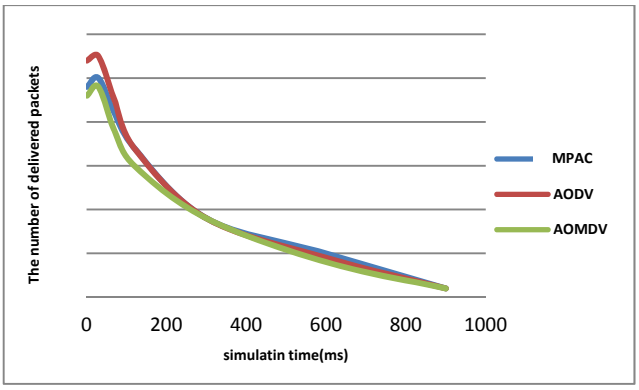


Fig. 4 Comparison of three protocols by the number of needed routing packets

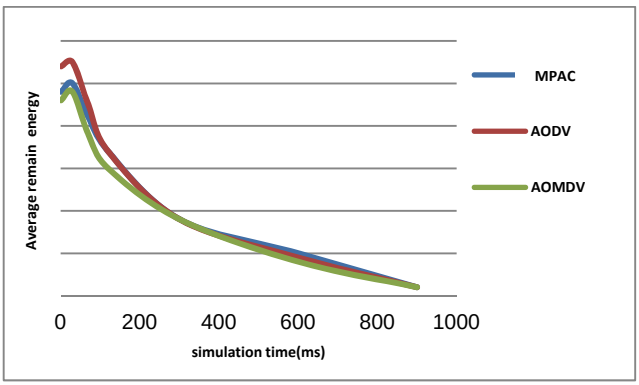


Fig. 5 Average remaining energy

5. Conclusion

In this paper, we have proposed an ant based routing protocol for mobile ad hoc networks. ACO based algorithms have specialized to provide adaptive and efficient solutions to network routing. AMPC provides multiple paths with comparatively less overhead in the network. The simulation results indicate that proposed scheme can perform better than AODV and AOMDV under high mobility because of alternate route maintenance scheme. In future, we have planned to investigate the performance of the algorithm for real-time multimedia data using various mobility models.

6. Reference

[1] Chlamtac, M. Conti, and J. Liu, "Mobile ad hoc networking: imperatives and challenges", Ad Hoc Networks, No. 1, 2003.

[2] Z.Liu , M. Z. Kwiatkowska and C. Constantinou , " A biologically inspired QoS routing algorithm for ad hoc network " , International Conference on Advanced Information Networking and Applications, AINA 2005 1: 426-431, 2005

[3] Dr. Shuchita Upadhayaya and Charu Gandhi," QUALITY OF SERVICE ROUTING IN MOBILE AD HOC NETWORKS USING LOCATION AND ENERGY PARAMETERS", International Journal of Wireless & Mobile Networks (IJWMN), Vol 1, No 2, November 2009.

[4] Frederick Ducatelle, ``Adaptive Routing in Ad Hoc Wireless Multi-hop Networks", Ph.D Thesis, Universitμa della Svizzera italiana Faculty of Informatics.

[5] A. Sabari, K Duraiswamy, ``Ant Based Adaptive Multicast Routing Protocol (AAMRP) for Mobile Ad Hoc Networks" International Journal of Computer Science and Information Security Vol 6, No 2, 2009.

[6] FJ Arbona Bernat, ``Simulation of Ant Routing Protocol for Ad-hoc networks in NS-2", Faculty of Electrical Engineering, Mathematics and Computer Science Network Architectures and Services Group.

[7] Claudia Lorenz,Patrizia Mottl ``Ant Algorithm - Phd thesis", Swiss Federal Institute of Technology, Zurick

[8] P.Deepalakshmi1 , Dr.S.Radhakrishnan2, "Ant Colony Based QoS Routing Algorithm For Mobile Ad Hoc Networks", in proc International Journal of Recent Trends in Engineering, Vol. 1, No. 1, May 2009.

[9] Chunxue Wu, Fengna Zhang, Hongming Yang," A Novel QoS Multipath Path Routing in MANET",International Journal of Digital Content Technology and its Applications Volume 4, Number 3, June 2010.

[10] Jafari, Saeid. M., Taghipour, M. and Meybodi, M. R. "Bandwidth Allocation in Wimax Networks Using Reinforcement Learning", World Applied Sciences Journal Vol. 15, No. 4, pp. 525-531, 2011.

Javad Pashaei received B.E. degree from Shabestar Azad University in 2006. He received M.S. degree from Shabestar Azad University in 2010. His research interests include network management as well as WiMAX and ad hoc networks.

Majid Taghipoor teaches in university of applied science technology. His research interests include network management as well as WiMAX, ad hoc, Manet networks and NOC.

Vahid Hosseini received M.S. degree from Shabestar Azad University in 2010.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

An efficient algorithm for generating AoA networks

Nasser Eddine Mouhoub [1], Abdelhamid Benhocine[2]

[1] Computer science department Bordj Bou Arreridj University, Algeria
Tel.: +213 771 64 98 17 Fax: +213 35 66 63 10

[2] Information System department, Qassim University, Saudi Arabia

Abstract

The activities, in project scheduling, can be represented graphically in two different ways, by either assigning the activities to the nodes '*AoN*' directed acyclic graph (dag) or to the arcs '*AoA* dag'.

In this paper, a new algorithm is proposed for generating, for a given project scheduling problem, an Activity-on-Arc dag starting from the Activity-on-Node dag using the concepts of line graphs of graphs.

Keywords: *Line graph of graph, project scheduling, AoN directed acyclic graph, AoA directed acyclic graph, PERT network.*

1. Introduction

Scheduling of project has received a great attention by researchers these last decades. Several fundamental techniques are largely applied and supported by many commercial systems in scheduling. Among these techniques we find Gantt diagram, PERT/ CPM etc …

The Project Evaluation and Review Technique (PERT) has been used as a tool for project management for over four decades. It consists of planning, designing, and implementing a set of activities to accomplish a particular goal or task. For many years, two of the most popular approaches to project management have been the Critical Path Method (CPM) and the Project Evaluation and Review Technique (PERT). J.E. Kelly of Remington-Rand and M.B. Walker of Dupont developed the CPM in the 1950's to assist in scheduling maintenance shutdowns of chemical processing plants. PERT was developed shortly after by the U.S. Navy to manage the development of the Polaris missile [1].

While CPM is a technique to manage projects with deterministic times, PERT is a technique used in projects that has activities with stochastic times [2]. However, practitioners now commonly use the two names interchangeably, or combine them into the single acronym PERT/CPM.

The original versions of PERT and CPM had some important differences. However, they also had a great deal in common, and the two techniques have gradually merged further over the years. In fact, today's software packages often include all the important options from both original versions.

Another similar method to the two quoted above, was developed in France by Roy [3], and called the 'Méthode des Potentiels Metra' (MPM), with the main difference that each activity is represented in the network by a node.

The activities of a project are often constrained by conditions such as "activity v cannot start until activity u has finished". Assuming that no activity is repeated, we can define a precedence relation $\prec$ on the activities, so that $u \prec v$ means that u must finish before v starts. According to Micha et al. [4], the relation $\prec$ can be represented graphically in two different ways, by either assigning the activities to the nodes or to (a subset of) the arcs. In either case, a *directed acyclic graph (dag)* is defined. In an *activity on node* (*AoN*) dag or (MPM network), each activity corresponds one-to-one with a node, and we say that $u \prec v$ is *represented* if there is a directed path of arcs leading from v's node to u's node. Thus, an *AoN* dag is unique except for possible transitive arcs. In an *activity on arc* (*AoA*) dag or (PERT network), each activity v corresponds to an arc, where parallel arcs that share the same start and terminal nodes are permitted. We say $u \prec v$ is represented in an *AoA* dag if there is a path from t_u the terminal node of the arc for u, to s_v, the start node of the arc for v (the path is empty if $t_u = s_v$,).

Experts prefer to work with the PERT network rather than the MPM. This is why according to Fink et al. [5], it is more concise. Furthermore, Hendrickson et al. [6] explains that it is close to the famous Gantt diagram. According to Cohen et al. [7], the structure of the PERT network is much more suitable for certain analytical techniques and optimization formulations. However, the major disadvantage of this method is in the existence of dummy arcs (see Fig. 1 and 2). Their number is likely to be significantly high especially if the size of the network is too large. The *AoA* dag is not unique.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

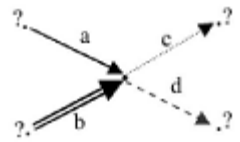


Fig. 1. The representation problem in *AoA* dag.

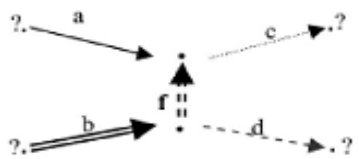


Fig. 2. Introduction of the dummy arc f and the representation in the *AoA* dag

The following example, with only precedence constraints, in Table 1 (the designations are not registered in the schedule table) presents a new technique of generating the *AoA* dag. It is rather a pedagogical method. Generally, the practitioner found difficulties to draw the graph and to mount this difficulty we note that even the experts of project scheduling can not draw correctly. This example which was taken from [5] presented a false *AoA* dag where the schedule table is not respected. The new proposed method uses successive stages in the construction of the *AoA* dag. This is done by scanning the table according to the codes column and the corresponding task to the current scanned line is added taking into account the principle of anteriority (Fig.3.). All the vertices are organized in levels.

We will see later another construction technique of PERT/CPM graph with a bipartite by using the results on the line graphs.

Table 1 : Schedule table

Codes	Durations	Predecessors
α	0	-
A	2	α
B	2	α
C	2	H
D	3	α
E	4	B,G
F	2	C,I
G	3	A,D
H	4	B,D
I	5	H
J	3	C
ω	0	E, J, F

Table 2: The vertices organized in levels

Level	Vertices
level I	α
level II	A , B, D
level III	G, H
level IV	C, E, I
level V	F, J
level VI	ω

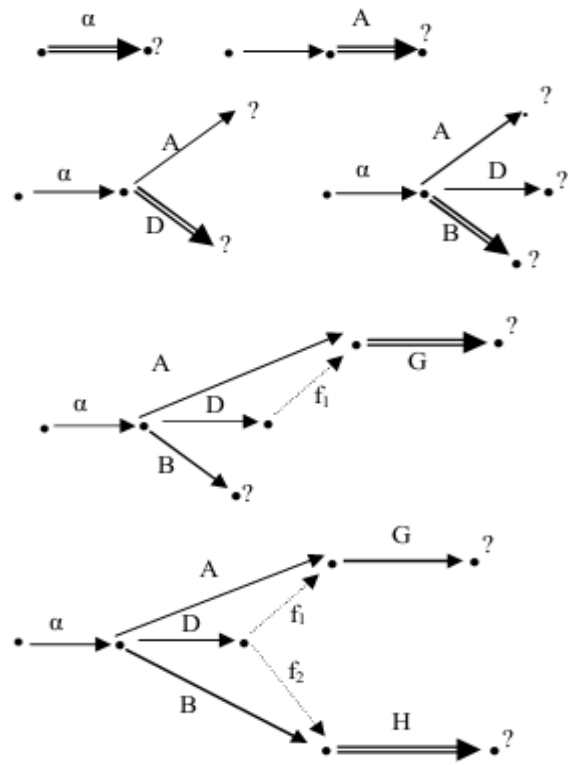


Fig. 3 (a) Constructing *AoA* dag.

and so on until the last graph :

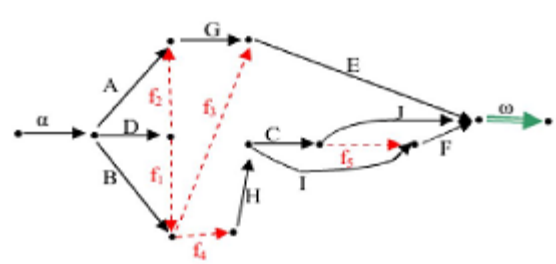


Fig. 3. (b) The *AoA* dag of the Table 1.

We can remind that the durations are not mentioned on the

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

different graphs. These durations can be uncertain. For this precise case, there are more details in [8], [9], [10].
By using the concepts of line graph of graph, we build an *AoA* dag starting from an *AoN* dag. This idea is transformed into an algorithm using graph theory.

2. Line graph

Let G = (X, U) a simple or multiple digraph. We build starting from G a graph or line graph noted L (G), called line graph or line digraph of G as follows:

- The nodes of L (G) are in bijective mapping with the nodes of G for simplicity reasons; we give the same name to the arcs of G and the corresponding nodes of L (G).
- Two nodes *u* and *v* of L (G) are connected by an arc of *u* towards *v* if and only if the arcs *u* and *v* of G are such as the final end of *u* matches with the initial end of *v*, i.e. $T(u) = I(v)$ [11].

2.1 Example

Let G the following directed acyclic graph be (Fig 4.):
By definition, any directed graph G admits a unique line graph L (G). On the other hand, two non isomorphs directed acyclic graphs can have the same line graph.

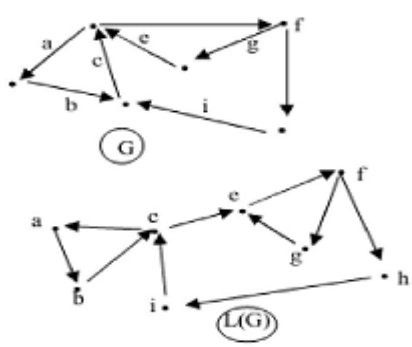


Fig. 4. A graph G and his line graph L(G).

2.2 The opposite problem

We suppose the following opposite problem:
Being given a directed acyclic graph H, is it the line graph of any directed acyclic graph? In other words, does there exist a graph G such as L (G) is isomorphs with H, where H = L (G)?

2.3 Configuration "Z "and "Δ"

G admits a configuration "Z" if G contains four nodes *a, b, c* and *d* such as if *(a, c), (b, c)* and *(b, d)* are arcs of G, then *(a, d)* is not an arc of G. With an only aim of simplicity, one will give the name of bar of "Z" the arc *(b, c)*.
Configuration "Z" appears when two nodes have common successors and no common successors or by symmetry when two nodes have common predecessors and no common predecessors.

G admits a configuration "Δ" if G contains arcs *(a, b), (b, c)* and *(a, c)* (Fig. 5.) [11].

Fig. 5. The configuration "Z" and "Δ".

2.4 Some characterizations of the line graphs

The line graphs have been studied but we will present, in this section, the features in which we are interested and obtained from [12]

1. H is the line graph of a directed acyclic graph if and only if H does not contain any "Z" configuration.
2. H is the line graph of a directed acyclic graph G if and only if arcs of H can be partitioned in a complete bipartite $B_i = (X_i, Y_i)$, i=1..., m, such as $X_i \cap X_j = \phi$ and $Y_i \cap Y_j = \phi \quad \forall \; i \neq j$.
The bipartite B_i of H are then in a bijection with the nodes also noted B_i which are neither sources nor well, two nodes B_i and B_j of G being connected by an arc from B_i towards B_j if and only if the complete bipartite B_i and B_j of H are such as $Y_i \cap X_j = \phi$ (Fig. 6).
3. H is the line graph of a directed acyclic graph without loops if and only if H does not contain any configuration Z or Δ.
4. H is the line graph of a directed acyclic graph if and only if any pair of nodes having common successors has all their common successors.
5. H is the line graph of a directed acyclic graph if and only if any pair of nodes having common predecessors has all their common predecessors.

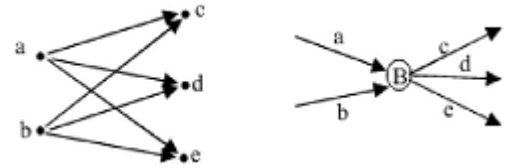


Fig. 6. A complete bipartite B of G and the star of G.

Thus, H is not the line graph of any directed acyclic graph if is only if there is a pair of nodes having common successors and no common successors or common predecessors and no common predecessors [12].

So, we want to know how to transform H in order to get a new graph which is a line graph of a graph.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

3. Passage from an *AoN* dag to an *AoA* dag

Because of the facility of the use of *AoA* dag, we must concentrate our efforts on the study of the possibility of transforming the *AoN* dag (a significant number of arcs) to *AoA* dag (a reduced number of arcs).
So, we want to know how to transform the graph H (which is an *AoN* dag) in order to get a new graph which is the line graph (*AoA* dag). The difficulty which arises is to know if H does contain Z configurations or not? If it does not contain Z, it is a line graph and the transformation is immediate. But if it contains Z, we have to eliminate the bar from each Z preserving the constraints of succession. Let us study each case apart:

3.1 The *AoN* dag is a line graph

Let us build the *AoA* dag starting from the *AoN* dag, assuming that it is a line graph. Under the terms of the results of paragraph 2.4, we proceed as follows:
We partition arcs of the *AoN* dag in a complete bipartite $B_i = (X_i, Y_i)$.

In the *AoA* dag which we wants to build, each B_i is represented by a node still noted B_i and will be the centre of the star.

3.2 The *AoN* dag is not a line graph

The construction of *AoA* dag is however more complex in general where the *AoN* dag is not a line graph: it does not admit a partition of the complete arcs in bipartite. It is in this case that one must modify it in order to transform it into a directed acyclic graph associated by preserving the constraints of anteriorities.

Let us suppose that the activities $a_1, \ldots, a_m$ precede the activities $b_1, \ldots, b_n$. In the *AoN* dag, these constraints of anteriority are represented by a complete bipartite. In *AoA* dag, they are represented by a star (see Fig. 7).

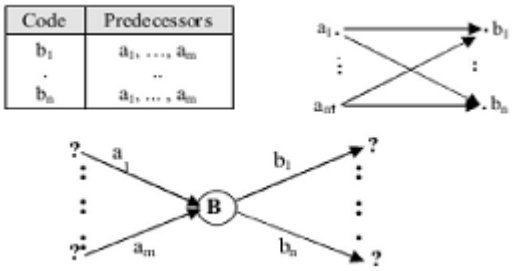


Fig. 7. The sub-table of anteriorities, the complete bipartite B in the line graph and the node B of the corresponding *AoA* dag.

Let us return to the problem of dummy arc in *AoA* dag. If there are for example 4 activities have *a, b, c* and *d* with the following constraints of anteriority: c is preceded by *a* and *b*, but *d* is preceded by *b* only.

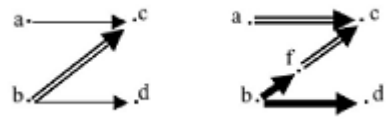


Fig. 8. "Z" configuration, his transformation in *AoN* dag and the partition of completes bipartite.

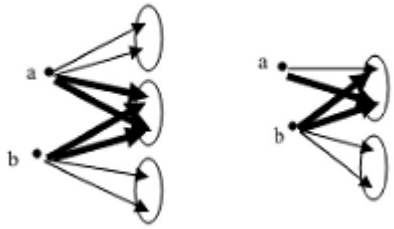
In *AoN* dag, there is no problem for representing these activities. We do it as in figure 8. But, if there is sub-graphs Z in *AoN* dag (which is considered like the line graph *H*), we are obliged to eliminate all "Z" configuration. We then introduce, in the *AoN* dag, a dummy arc *f* in every "Z" (Fig. 8).

The introduction of the dummy arcs aims at eliminating all "Z" configurations from the *AoN* dag, the constraints remain unchanged. We should recall that the dummy arcs are not necessary in the *AoN* dag but are introduced only to build *AoA* dag.

Kelley [13] notes that it is advantageous to reduce the length of calculations to build an *AoA* dag having the minimum number of nodes and dummy activities.
We then pose the problem of looking for "Z" in the *AoN* dag, i.e. nodes having common successors and no common successors or nodes having common predecessors and no common predecessors.

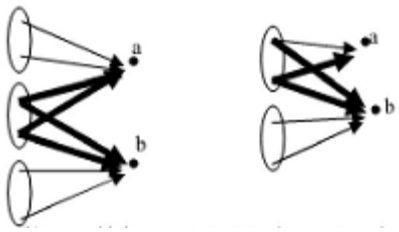
A first technique of elimination of Z in G_v consists of replacing the bare *(b, c)* of every "Z" by two arcs *(b, f)* and *(f, c)*, according to fig. 8, it is the simplest but the worst, the number of "Z" can be arbitrarily large.

Another technique more effective than the preceding one eliminates several "Z" at the same time by regrouping bars having the initial end or the final end in the same bipartite complete (Fig. 9 and 10).



(a) *a* and *b* have common successors and no common successors.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.



(a) *a* and *b* have common predecessors and no common predecessors.

Fig. 9. Grouping of the "Z" bars having the same initial extremity and belonging to the same complete bipartie

"Z" are destroyed as follows (see Fig. 10):

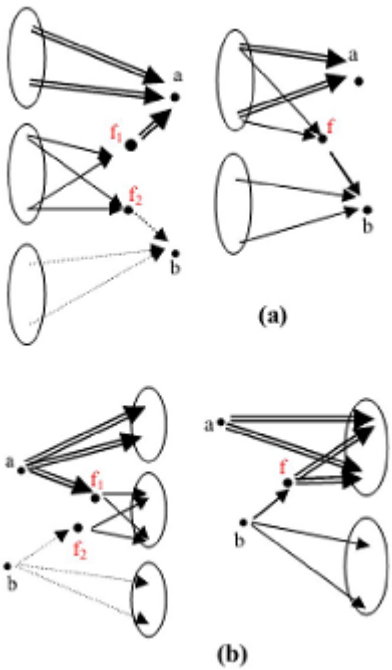


Fig. 10. Grouping of "Z" bars, dummy activities and the arcs partitioning in a complete bipartite.

The problem raised by Kelley [13] was approached by some other authors; more details in [14], [15], [7].

4. Algorithm

Let G_v be an *AoN* dag which must be directed, valued, connected and without circuit. With G_v being a conjunctive graph, organized in levels. We want to build the corresponding *AoA* dag which is called G_e.

Begin
if G_v contains Z configurations ***then***
Identify the Z_i *(i = 1, 2, …, m)*
For every bipartite containing 2 and more bars of Z ***do***

- Create dummy nodes corresponding to the number of bipartite.
- Regroup the bars having same initial extremity or same terminal extremity to the same complete bipartite.

Endfor
Endif

- Identify again the bipartite in G_v
- Represent every bipartite B_i in G_v by a node B_i in G_e.
- Represent every arc such as :
  An arc is drawn between 2 nodes B_i and B_j in G_e if and only if the 2 bipartite B_i and B_j in G_v are such as $Y_i \cap Y_j = \phi$.

End.

4.1. Example

Let us consider the following constraints table (Table 3) [16] according to the associated *AoN* dag (Fig. 11. (a)). The designations and the durations are not registered in the schedule table.

Table 3: Schedule table

Code	Predecessors
α	-
A	α
B	α
C	A, B
D	A, B
E	B
F	D
G	D
H	D,E
I	C, F
J	C, F,G
K	G, H
L	J, K
ω	I, L

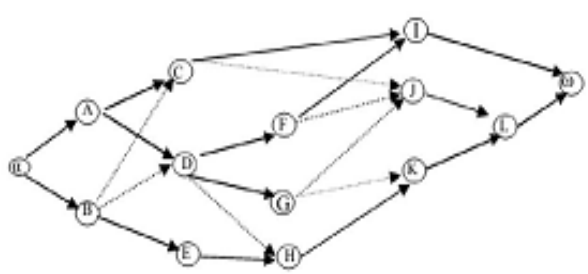


Fig. 11. (a) The *AoN* dag with "Z" bars in dots.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

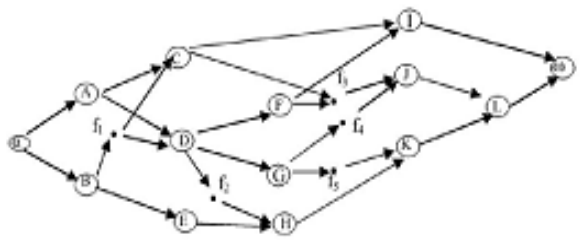


Fig. 11. (b) The modified *AoN* dag by introduction of dummy arcs f_i.

Fig. 11. (c) The *AoN* dag modified with reorganization of the activities in levels and the arcs partitioning in a complete bipartite.

$B_1 = (\{\alpha\}, \{A, B\})$, $B_2 = (\{B\}, \{F, E\})$,
$B_3 = (\{A, f_1\}, \{C, D\})$, $B_4 = (\{(C, F\}, \{F, G, f_2\})$,
$B_5 = (\{E, f_2\}, \{H\})$, $B_6 = (\{C, F\}, \{I, f_3\})$,
$B_7 = (\{G\}, \{f_4, f_5\})$, $B_8 = (\{f_3, f_4\}, \{J\})$,
$B_9 = (\{f_5, H\}, \{K\})$, $B_{10} = (\{J, K\}, \{L\})$,
$B_{11} = (\{I, L\}, \{\omega\})$.

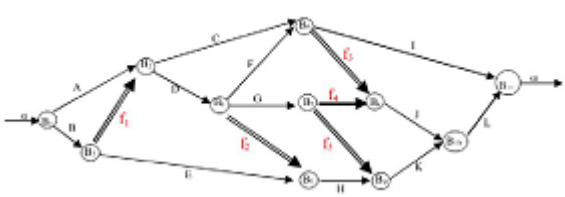


Fig. 11. (d). Construction of an *AoA* dag starting from an *AoN* dag of fig. 11. (a).

4.2. Discussion

The algorithm finishes since the loop is carried out only when there is Z, and the number of Z in G_v is known and finite. The rest of the algorithm is a succession of simple instructions. The complexity of the algorithm is polynomial ($O(n^4)$).

5. Conclusion

This paper presents a new approach for generating PERT networks. Indeed, by applying some results of line graphs of graphs, the algorithm generates a PERT network with the total respect of the constraints in schedule table.

Our new approach is very simple to be applied. It was programmed in C++ Builder and gives an *AoA* dag in a very short time since its complexity is polynomial $O(n^4)$. The techniques used in the algorithm can be exploited in other fields by specialists in the graph theory. The experimental results are very positive even when networks are of a very large size. Another major benefit that is worth noting is the fact that the algorithm works without any problem in the presence of transitive arcs in *AoN* dag. The new algorithm, due to its new characteristics, is more practical and more user friendly to use by project managers and other practitioners.

This work gives the possibility of new perspectives, such as the treatment of the temporal constraints as well as the research of minimal *AoA* dag in terms of nodes and dummy arcs number.

Bibliography

[1] W. A. Haga, Tim O'keefe, Crashing PERT networks: a simulation approach, 4th International conference of the Academy of Business and Administrative Sciences, Quebec City, Canada, July 12-14, 2001.

[2] S. M. Campos, Tomas E. S. Fierro and A. P. Solis, Modeling An Assembly Line Combining Pert /Cpm To Study The Flexibility Of A Mexican Automotive Plant Using Simulink, Proceedings of the 8th Annual International Conference on Industrial Engineering – Theory, Applications and Practice, Las Vegas, Nevada, USA, November 10-12, 2003.

[3] B. Roy, Algèbre moderne et théorie des graphes, tome 2, fascicule 3, Problèmes d'ordonnancement et ensembles de potentiels sur un graphe, DUNOD, Paris, 1970.

[4] D. Michael, J.Kambourowski and M. Stalim, On the minimum dummy-arc problem, Recherche opérationnelle/Operations Research, vol. 27, no 2, 1993, p. 153 2 168.

[5] G. Fink and J. Fourier, Operational research and networks, Wiley-ISTE, 2008.

[6] C. Henderckson, T. AU, Project Management for Construction, Department of Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh, PA l52l3, Version 2.2, 2008.

[7] Y.Cohen, A. Sadeh, A New Approach for Constructing and Generating AOA Networks, Journal of computer science, Volume 1, Issue 1, 2007

[8] N. E. Mouhoub, H. Belouadah & A. Boubetra, Towards a Better Assessment of the Durations in PERT

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Method, (IJCSIS) International Journal of Computer Science and Information Security, ISSN :1947-5500, Vol. 8, No. 9, December 2010.

[9] P. Morris, N. Muscettola, and T. Vidal, Dynamic Control of Plans with Temporal Uncertainty, In *Proceedings of the 17th International Joint Conference on Artificial Intelligence, IJCAI-01*, pages 494–502. Morgan Kaufmann, 2001.

[10] I. Tsamardinos, T. Vidal, and M. E. Pollack, CTP: A New Constrained-based Formalism for Conditional Temporal Planning, *Constraints journal, special issue on Planning*, 8(4):365–388, October 2003.

[11] C. Heuchenne, Sur une certaine correspondance entre graphes, Bull. Soc. Roy. Sci. Liege 33(1964) 743-753.

[12] N. E. Mouhoub, S. Akrouf, Approche optimisée dans les problèmes d'ordonnancement de projet, ANL'07, Actes du colloque international sur l'analyse non linéaire, Sétif university, Algeria, 2007.

[13] J. E. Kelley, Critical path planning and scheduling-Mathematical basis, Operations research, vol. 9, n° 3, 1961, p. 296-320.

[14] S. E. Elmaghraby, W. S. Herroelen, On measurement of complexity in activity networks, *European Journal of Operational Research*, 5, 223-234, 1980.

[15] M. Syslo, Optimal Construction of Event-Node Networks, *RAIRO*, 15, 241-260, 1981.

[16] N. E. Mouhoub, A. Benhocine & H. Belouadah, A new method for constructing a minimal PERT network, (APM) Applied Mathematical modelling, Elsevier ISSN: 0307904X, Vol. 35, Issue: 9, 4575-4588, 2011.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Modify the µCS-51 Architecture to SIMD, VLIW and Superscalar µC

Ph.D: Abdelmoneim M. Fouda [1], Assem Badr [2] and Prof: Abdelsamie kodb [3]

[1] Computer Eng. Department, Modern Academy for Engineering & Technology
Cairo, Egypt

[2] Computer Eng. Department, Modern Academy for Engineering & Technology
Cairo, Egypt

[3] Electrical Eng. Department, Faculty of Engineering, Al-Azhar University
Cairo, Egypt

Abstract

Computer architects have always strived to increase the performance of general purpose computers and special functional CPUs. The high performance may depend on the hardware chips manufacturing, however this trend has constrains, because there are limits of the physical technology, the other trend to increase the performance depend on the parallelism (multiple-processing). Recently the superscalar processing is the latest up-to-date technique of innovations aimed for producing ever-faster microprocessors, they are capable of executing more than one instruction in one clock cycle, in addition to perform multiple independent operations simultaneously.

In this paper we modify the architecture of the conventional microcontroller (µC-51) using VHDL, also we develop some innovated special programmed instructions which utilize parallelism (via multiple processing units). Thus we present a modified superscalar processor using single instruction multiple data (SIMD) with very long instruction word (VLIW) architectures. The proposed modified microcontroller (MµC) has impact in reducing number of clock cycles per instruction, improvement of overall µC performance, and increasing transmission/reception data rate of the µC serial/parallel ports.

Keywords: *µCS-51, ISA, VHDL, SIMD, ILP, Superscalar, VLIW, Data rate, Iron law.*

1. Introduction

Embedded applications are becoming ever more diverse and complex; processors targeting such applications have an increased tendency to attain a desirable performance via highly specialized instructions tailored for the needs of their targets[1].

Consequently, conventional microcontroller architectures vary widely, offering the possibility for implementing various calculating platforms, capable to execute applications in different fields such as signal processing, or communication, command and control. The limitations introduced by the standard architectures become a problem when a high performance and low cost implementation is demanded (for a specified application to a standing context). So it is required to develop the conventional µC (CµC) in such a way to satisfy a trade-off between reaching the specified application and costs[2]. Parallelism is one of the best solutions to achieve high speed of processing, and lowest power consumption for overall speeding application. Parallel processing reduces the execution time taken by any program. The execution time taken by any program is determined by three factors: First, the number of instructions executed, second, number of clock cycles needed to execute each instruction and the third is the length of each clock cycle [3].

There are two forms of parallelism that can be exploited by modern computing machinery to achieve higher performance, the first form of parallelism is called Thread Level Parallelism or TLP, it is software capability to execute independent programs simultaneously using different flows of execution, called threads. The illusion of multiple thread execution is often achieved on a single conventional processor by running each thread successively for short intervals. The second form is Instructional Level Parallelism (ILP), it is hardware architecture to reduce program runtime by overlapping the execution time, it mean that execute many instructions in same machine cycle[4].

The conventional 8051 is scalar processor at which each instruction manipulates single data items at a time, while vector processor manipulates many data simultaneously. The superscalar processor is mixture of both types, its CPU architecture implements a form of parallelism (ILP) within a single processor. Moreover the superscalar processor contains multiple redundant functional units within each CPU, which allows multiple instructions processing on separate data items concurrently.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

The ILP techniques can be achieved either by superscalar processor or multi-core processor. The superscalar processor executes more than one instruction during a single clock cycle by simultaneously dispatching multiple instructions to redundant functional units on the processor, each functional unit is not a separate CPU core but an execution resource within a single CPU (such as an arithmetic logic unit, a bit shifter, or a multiplier). The P5 Pentium was the first superscalar x86 processor; the Nx586, P6 Pentium Pro and AMD K5 were among the first designs which decode x86-instructions [5]. While as multi-core processor has two or more independent actual processors called "cores". Therefore the superscalar processor allows faster CPU throughput than would otherwise be possible at a given clock rate.

Practically it is found that most of the µCs 8051 family transmit their data in the low data rate (LDR) and medium data rate (MDR) which less than 3Mbit/s, therefore it isn't suitable to interface with higher data rate equipments (such as Ethernet, fast Ethernet, and gigabit Ethernet which have data rates of 10Mbit/s, 100Mbit/s and 1Gbit/s respectively). So we develop and modify the internal architecture for the CISC (complex instruction set computer) microcontroller family µCS-51 and their instruction set architecture (ISA) by using SIMD type of superscalar techniques and with Very Large Instruction Word (VLIW) architectures to improve their performance and increase the transmission rate.

The paper is organized as follows; the next section-2 describes the overall design steps of the developed µC, including the ISA instruction set modification in section-2-1, and the processor architecture modification in section 2-2. While as section-3 discuss the different numerical results and processor performance. Finally, section 4 presents an overall conclusion.

2. Design Processor

To develop a novel µC based on a conventional one, different modifications are required including instruction set modification, and architecture modification. This section introduces the necessary basic principles to design modified VHDL code for the internal architecture of the conventional CISC µC 8051. The developed VHDL code is obtained by adding the so called "modified instruction decoder" (MID) in addition to the conventional instruction decoder (CID). We can alternate between either MID or CID using the so called "selected instruction decoder flag (SIDF)". Another modification is obtained by inserting two advanced VHDL codes represent two individual programmable input/output data sequencer, the first one called programmable output digital parallel/serial sequencer (PODPSS), the other one called programmable input digital parallel/serial sequencer (PIDPSS), both sequencers depend on single instruction and multiple data (SIMD).

2.1 Instruction set modification

The ISA can be adapted or extended to meet the modern application requirements [7]; this section is dedicated to discuss the processor's ISA modification. Many methodologies can used to increase the number of instructions, either increasing the number of bits assigned for op-codes, or using two/more successive locations to store instructions in the µC program memory. The selected methodology must keep the characteristics of the µC's memory, internal data and address bus avoid from any change.

From literature survey, and guided with Intel µCS-51 family's data sheets and its manual of instructions set [6], it's found that the machine code "A5H" is a reserved code which is not used for any operations or tasks, as shown in Fig.1. Utilizing this reserved code, we can expand the number of µCS-51family's ISA and adding sophisticated instructions customized for specific multiple data applications. The suggested assembly syntax code "InovConv" uses the reserved machine code "A5H", this code will have toggle alternative action (TAA), namely, when the "InovConv" code is written by the user it is compiled into "A5h", and alternate as toggling action between either CID (when SIDF = "0") or MID (when SIDF = "1") as shown in Fig.2a. Now the developed µC can be used in two states either in its conventional mode of operation (256 instructions) or in the modified mode of operation (255 instructions) therefore the number of instructions for the CµC-51 will be expanded from 255 to 511 instructions.

Hex Code	Number of Bytes	Mnemonic
A0	2	ORL
A1	2	AJMP
A2	2	MOV
A3	1	INC
A4	1	MUL
A5	1	Reserved
A6	2	MOV

Fig.1 Partial of conventional µCS-51 instruction set.

Mnemonic	Hex code	SIDF	Enabled instruction decoder
InovConv	A5	0	Conventional Instruction Decoder (CID)
		1	Modified Instruction Decoder (MID)

Fig.2a Toggling stages of A5h.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

HEX code	conventional (ISA)		modified (ISA)		
	No. of Bytes	Assembly code	No. of Bytes	Assembly code	SIDF
A5	1	Reserved code	1	InovConv ***	0*
A5	1	Reserved code	1	InovConv ***	1**
A6	2	MOV @R0,Direct	6	OUTSeqB N,B,P,S	N.C
A7	2	MOV @R1,Direct	6	INSeq N,B,P,S	N.C
A8	2	MOV R0,Direct	1	HLTSeqB	N.C
A9	2	MOV R1,Direct	1	LdBKifDon	N.C

*** Boolean alternative action (BAA) ** Causes innovation codes
N.C ..... Not Care * Causes conventional codes

Fig.2b Partial of modified µCS-51 instruction set (ISA).

A sample of modified instruction set of the developed µC-51 is shown in Fig.2b; both modified instructions "OUTSeqB" and "INSeqB" are SIMD instructions with programmable process they can be executed simultaneously. For illustrating the execution of the new instruction "OUTSeqB N,B,P,S", it is assigned for output sequence of data bytes (maximum 8 bytes per sequence) retrieved from address 10h to 17h in the µC's main memory (RAM). The generated sequence may be transmitted throughout serial/parallel port according to the flag B (1-bit), if B="1" then the serial port will transmit data, while if B="0" the parallel port will transmit data. The operand "P" (3-bits) is assigned to allocate the number of transmitted data bytes per sequence, if "P" equal "000" it means that sequence has only one data byte (the contents of RAM location 10h), else if "P" equal "001" it means that the sequence has 2 data bytes (the contents of RAM location 10h, 11h), and so on, if "P" equal "111" it means that sequence has 8 data bytes (the contents of RAM location from 10h to 17h). The generated sequence may be repeated if the instruction "NLOPOUTS" is de-activated, and the repeated loop is terminated immediately if the instruction "HLTSEQB" is activated.

Moreover to control the transmission rate of the manipulated data and with the help of using simulator "Modelsim SE 6.5" we derived the following empirical relation (1):

$$AdjustedData\ Rating_{parallel} = \frac{f_{cp}}{1+N_{operand}} \qquad [MB/S] \quad (1)$$

Where f_{cp} is the frequency of the global clock pulses for the VHDL µC, to calculate the data rate performance, assume that $f_{cp} = 50\ MHz$ and the operand "N" has its minimum value N= "000000000h" then the maximum data rating equal to 50-MBps (i.e. output byte each 20 ns), else if operand "N" increased to (000000001), then the data rate is 25-MB/S (i.e., output byte each 40 ns) and so on, if N has its maximum value "7FFFFFFFFh" then very low data rate is obtained (Byte each 11.45 minutes).

Similarly, the instruction "INSeqB N,B,P,S" is developed for receiving sequence of consecutive 8-data bytes (stored in RAM data memory from address 20h to 27h) through the PIDPSS unit, and is controlled by the status of the 4 operands "B", "P", "S", "N" as previously mentioned. As a conclusion we described a sample of two innovated instructions, they have full duplex mode of transmitting/receiving data while the µC process other instructions simultaneously. In the next section we will explain the VHDL realization for the mentioned instruction modifications.

2.2 Processor's Architecture modification

Adapting the old architecture to a more flexible implementation empowering further developing[2], High-level design tools and field-programmable gate arrays (FPGAs) significantly reduce the effort, cost and risk of hardware implementation. These technologies can be incorporated into a manageable and affordable prototyping framework a VLSI-scale “breadboard” for exploring and evaluating new microprocessor designs[8]. For this reasons a new architecture will be delivered to modify the CµCs-51 by the VHDL over the FPGA technique. The basic block diagram for the internal architecture of the MµC-51 is shown in Fig.3. The shaded blocks represent the traditional units of the CµC-51, and the other blocks represent the created units for the MµC-51.

The 1st modification is obtained by adding MID-unit and modified instruction register (MIR-unit) beside conventional instruction register (CIR-unit) and the CID-unit, all of them is controlled by SIDF.

Initially the SIDF is set to '0', so the CID is active, so the traditional 8051 instruction set will be fetched, decode, and then executed. When using the code "InovConv" (denominated by user's program) the SIDF is set to '1', so the MID is activated and the modified 8051 instruction set will be fetched, decode, and then executed. Furthermore, when the "InovConv" code is written again by user, it toggles the SIDF to logic '0' causes enable for CID again, and so on.

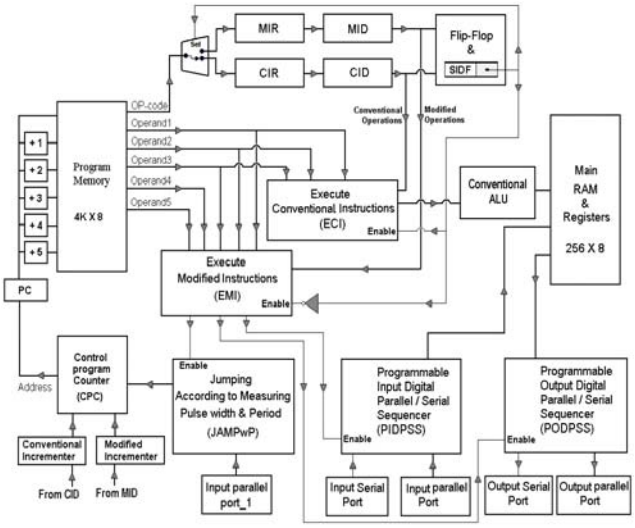


Fig.3 The basic of internal architecture for the MµC-51

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

The 2nd modification is particular for the program memory's interface; it is performed by increasing the number of data bus associated with program memory (from single bus to 6 data bus). The instructions of the CµC-51 represented by maximum 4 bytes in the program memory as show in Fig.4a, each byte fetched one by one through single bus, while the instructions of the modification version needs maximum 6 sequencing bytes stored in the program memory as show in Fig.4b. Such developed construction is capable for fetching all adequate bytes simultaneously. The recommended memory location fetched starting from the value stored in the program counter (PC), and the desired number of fetched bytes specified by the modified unit designated as control program counter (CPC), the increment in the program counter related to the absolute values stored in the CID or MID.

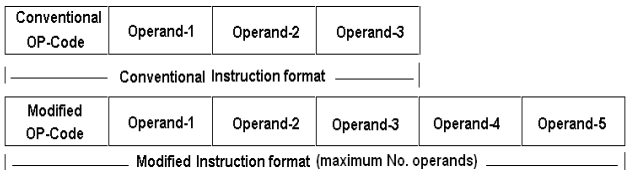


Fig.4a The conventional and modified VLIW instruction formats

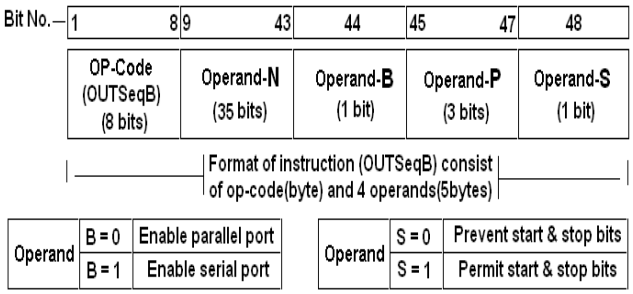


Fig.4b The format of modified instruction (OUTSeqB)

In addition to the original RAM data bus, the 3rd modifications adds additional 8-input data bus particular for the RAM locations (with address 10h to 17h), and adding additional 8-output data bus for the RAM locations (with address 20h to 27h). Such modifications allow RAM-data bus to read/write its specific memory locations concurrently according to the specified address.

The 4th modification includes the creation of so the called "execute modified instruction EMI" which is responsible for controlling both two modified data sequencer units "PODPSS" and "PIDPSS". They are providing response for executing the other logical, arithmetic, and transferring modified operations. Each sequencer attached with µC internal/external ports for handling a generated sequence of data with the µC's peripherals. The PODPSS is particularized for transmitting data through specific µC external output port, while as the PIDPSS is particularized for transmitting data through specific µC external input port. Each sequencer handles data with different data rates and adequate number of data bytes.

Another modification includes different types of adjustable counters to establish the operation of PODPSS or PIDPSS. The adjustable counter_1 is responsible for adjusting the data rate (related to the detected operand "N"). The adjustable counter_2 is used to adjust the number of bytes per each generated sequence of pulses (according to the operand "P"). Moreover; the adjustable counter_3, is allocated for serial port transmission, which is responsible for inserting start/stop bits for the data sequence (according to the operands "B" and "S").

Firstly; as show in Fig.5 the adjustable counter_1 consists of logical counter, register and logical comparator. The basic idea of the adjustable counter is loading operand "N" value into the register, and counting the received clock pulses, so if the counting value of the counter equal to the stored operand "N" value, the comparator becomes active output to clear the counter, and the two comprised values becomes not equal, so the counter will restarting again and so on.

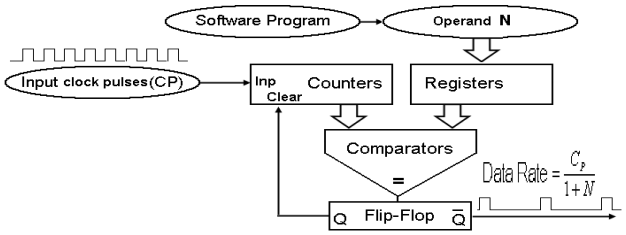


Fig.5 The block diagram for the adjustable counter_1

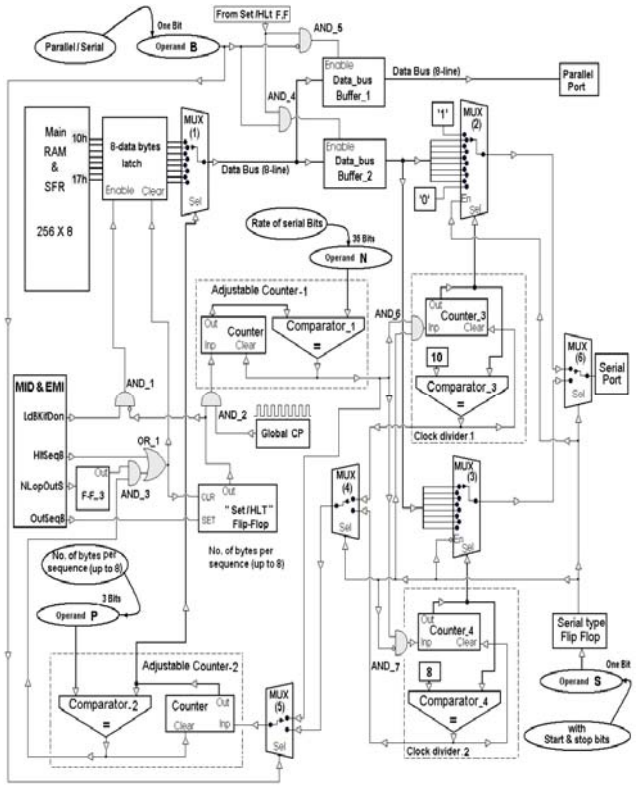


Fig.6 Architecture of Programmable Output Digital Parallel / Serial Sequencer (PODPSS)

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

The internal structure of PODSS is shown in Fig.6 for more illustrations assume the following scenario; it is required to transmit (through parallel port) a sequence of 3-data bytes with values AAh, CCh, and F0h, stored in RAM locations with addresses 10h, 11h, and 12h respectively. The data transmission rate will equal to the global frequency of the µC, and. Initially, the "Set/HLt " Flip-Flop has output '0' for halting the PODPSS by blocking the gate AND-2 (block global clock pulse Cp), and permit the 1st modified instruction "LdBkifDon" for passing through AND_1 to latch the contents of 8 RAM locations (starting from address 10h) to the multiplexer MUX_1 via 8-output RAM bus. The modified instruction "NlopOutS" set FF-3 to "1" thus it's ANDED with the output of comparator_2 (via AND-3) to prevent looping of generated data sequence (allow only one sequence of data bytes). For more clarifications, the SIMD instruction "OutSeqB 0,0,2,0" is used to set the PODPSS, the operand "N=0" causes maximum data rate, the operand "B=0" drive "data_bus_buffer_1" through AND_5, so the 1st data byte (AAh) will be transmitted via the parallel port. The operand "P" adjusts the number of transmitted data bytes per one sequence, in case "P=2" sequence has 3 data bytes (AAh, CCh, F0h) respectively. The 1st data byte (AAh) still on the parallel port until the rising edge from comparator_1 increment adjustable counter_2 through MUX_5, thus causing change in the selection pins of MUX_1 to select the 2nd data byte (CCh). Moreover; the next rising edge from comparator_1 increments counter_2 to select the 3rd data byte (F0h) , the next rising edge from comparator_1 will increment the counter_2, so the comparator_2 will drive '1' which will AND with the logic '1' from "FF_3" to halt the PODPSS. Note that the operand "S" not effected in the parallel port.

Now for serial data transmission, another SIMD instruction "OutSeqB 0,1,2,1" is used. The operand "B=1" enables "data_bus_buffer_2" through AND_4. The 1st data byte (AAh) will be transferred to the serial port through the MUX_2 which enabled by operand "S=1". Each rising edge of the adjusted data rate signal from comparator_1 will increment the counter_3 causes transmission of start bit and the consequence 8-data bits (starting with least significant bit LSB) and terminated with the stop bit consecutively. After 10 rising edge of the adjusted data rate signals, the comparator_3 increment the counter_2 through the MUX_4 and MUX_5 causes select the next data byte to be transmitted (CCh) of the input of MUX_1 and so on.

Note that the both modified instruction "HLtSeqB", "NlopOutS" can used at any time by the programmer, the "HLtSeqB" for immediate halting for the transmitting data, but the "NlopOutS" for halting the transmitting data after finishing the current sequence (don't loop again)

3. Numerical results

Form the simulation point of view; assume the previously described task for transmitting data bytes (retrieved from 3-RAM locations) through serial and parallel ports. To compare the behavior of both modified and CµCs assuming the measuring parameters including data rate, time processing. Firstly; using the CµC-AT89C52, and Proteus 7.5 sp3 simulator package (Labcenter.com) to draw its schematic circuit, then guided with simulator package Prog-Studio-5 (Batronix.com) to edit and compile program code for particular task as shown in Fig.7. With aid of the Proteus digital timing analyzer we can follow up the behavior of AT89C52 for the desired task as shown in Fig.8 and Fig.10. Secondly; using package "Xilinx ISE 9.2i" to edit and synthesize in VHDL code the MµC. With aid of the Modelsim-SE 6.5 (Mentor-graphic) we can follow up the behavior of MµC, as shown in Fig.9 and Fig.11.

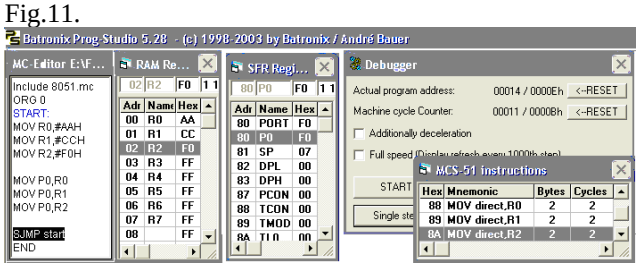


Fig.7 Assembly and machine codes for 3 sequence patterns

In the case of parallel transmission, for the CµC, as shown in Fig.7, the left form indicate the editor of the assembly code to transfer 3 data bytes (AAh, CCh, F0h) from 3-RAM locations (R0,R1,R2) respectively via the parallel port (P0). The two forms in the middle indicate the contents of data memory and the contents of ports during execution. The top right form indicate the control form illustrate the address of program counter and the number of machine cycles. In the bottom right form partial of µCs-51 instruction set. It is found that; the output data instructions occupy two bytes in the program memory and it is executed in two machine cycles.

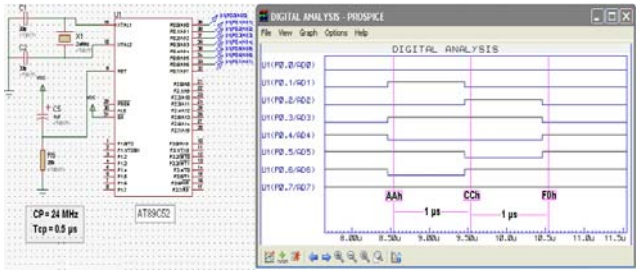


Fig.8 µC AT89C52 circuit diagram and output timing diagram (1us between bytes)

In Fig.8 the parallel output timing diagram of CµC's port P0, it is found that the smallest difference time between any two successive output data bytes is 1µs, this is very

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

large time and causes low data rate transmission. It is concluded that, this low data rate is due to the number clock pulses and number of cycles taken for µC-51 instructions which response for transmitting the parallel bytes (as MOV P0,R1), it has 2-machine cycles each cycle has 12-Cps for execution. Since that the maximum Cp for µC-51 equal to 24MHz, then the delay between any two successive output data bytes or the minimum duration of the conventional parallel output (Tc_{PP}) is determined by "Iron Law" (2) [9][10]

$$\frac{Time}{program} = \frac{No.of\ instructions}{program} \times \frac{No.of\ cycles}{instruction} \times \frac{time}{cycle} \qquad (2)$$

From 8051 data sheet, the duration time for one machine cycle T_{MC} is equal to 12 of clock pulse (Cp) duration and defines by the equation (3)

$$T_{MC} = 12\ T_{cp} = \frac{12}{f_{cp}} = \frac{12}{24 \times 10^{6}} = 0.5\ \mu s \qquad (3)$$

(3)

From equation (2) and (3) then

$$Tc_{PP} = 1 \times 2 \times 0.5 = 1\ \mu s \qquad (4)$$

Now applying Iron law to calculate the execution time (Tc_{exe}) for the 3 conventional instructions (MOV Port, Memory) which response to out the 3 data byte (AAh, CCh, F0h) from 3-RAM locations (R0,R1,R2) as observed in the digital analyzer in Fig.8 we get eq.(5)

$$Tc_{exe} = 3 \times 2 \times 0.5 = 3\ \mu s \qquad (5)$$

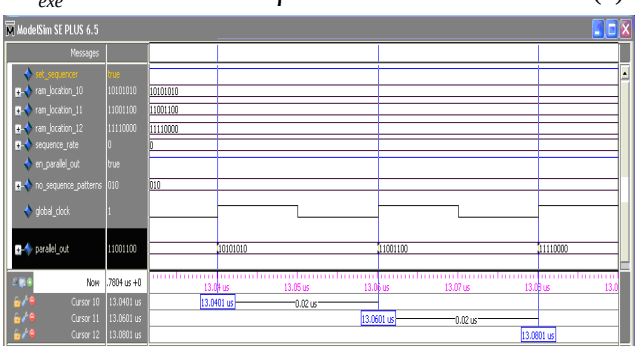


Fig.9 Timing diagram "Modelsim SE 6.5" showing the sequences of 3 parallel patterns

Consequently, for the MµC as shown in Fig.9, the SIMD instruction "OutSeqB 0,0,2,0" (as we explained before) will transmit 3-bytes via parallel port and with maximum data rate. Our MµC execute the instruction in one Cp for each output. The time difference between any two successive data bytes (AAh, CCh, F0h) is (Tm_{PP}), which equal to the duration time of modified machine cycle (T_{MMC}), and equal to one Cp duration, assume that the operating frequency of MµC is f_{Mcp} equal to 50 MHz, so the Tm_{PP} will define as eq. (6):

$$Tm_{PP} = T_{MMC} = T_{Mcp} = \frac{1}{f_{Mcp}} = \frac{1}{50 \times 10^{6}} = 0.02\ \mu s \qquad (6)$$

Therefore the modified executed time (Tm_{exe}) for the 3-data bytes using the SIMD "OutSeqB 0,0,2,0" takes 3 clock pulses as observed in the digital analyzer in Fig.9 is determined by "Iron Law" defined in eq. (7)

$$Tm_{exe} = 1 \times 3 \times 0.02 = 0.06\ \mu s \qquad (7)$$

It is noted that the execution time of MµC outperform the traditional one, the speed up of the MµC will obtained as equation (8) [9]:

$$Speed\ \ up(M_{Parallel}) = \frac{\left(time / program \right)_C}{\left(time / program \right)_M} = \frac{Tc_{exe}}{Tm_{exe}} = \frac{3\ \mu s}{0.06\ \mu s} = 50\ times \qquad (8)$$

Moreover, for CµC-51, the data rate of parallel transmission (parallel channel signaling rate PCSR) can be determined by the following eq. (9) [10]:

$$PCSR = \sum_{i=1}^{m} \frac{\log_2 n_i}{T_i} \qquad [Mb/s] \qquad (9)$$

Where m is the number of parallel channels, n_i is the number of significant conditions of the modulation in i-th channel, and T_i is the unit interval, expressed in seconds, for i-th channel. For parallel transmission with equal unit intervals and equal numbers of significant conditions (two- condition modulation) on each channel, the conventional parallel channel signaling rate ($PCSR_C$) will reduces to eq. (10):

$$PCSR_C = \frac{8}{Tc_{PP}} = \frac{8}{1 \times 10^{-6}} = 8\,Mb/s = 1\,MB/s \qquad (10)$$

the parallel transmitting data rate for MµC-51 determine by parallel channel signaling rate (PCSR) law, and the modified parallel channel signaling rate ($PCSR_M$) will reduces to eq. (11):

$$PCSR_M = \frac{8}{Tm_{PP}} = \frac{8}{0.02 \times 10^{-6}} = 400\,Mbps = 50\,MB/s \qquad (11)$$

$PCSR_M$ Is meeting with our empirical parallel data

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

rating in equation (1), in case of operand (N=0), the maximum data rate is defined as eq. (12):

$$Adjusted\ Data\ Rating_{parallel} = \frac{f_{cp}}{1+N}$$

$$= \frac{50\times10^6}{1+0} = 50\,MB/s \qquad (12)$$

Comparing between the results of eq.(10) and eq.(11) or eq.(12), the parallel data rate of the modified type µC is superior than that of conventional type. The data rate enhancement factor can be obtained as relation (13):

$$PDRE = \frac{\mathrm{PCSR_M}}{\mathrm{PCSR_C}} = \frac{50MBps}{1MBps} = 50\ times \qquad (13)$$

Furthermore the second task is output group of 3 data bytes through the serial port for both conventional and MµC, the Fig.10 indicates the editor of the assembly code designed for output 3 data bytes (AAh, CCh, F0h) from 3 RAM locations (R0, R1, R2) with start and stop bits, and simulated in simple AT89C52's circuit diagram using "Proteus 7.5" as show in Fig.11. The maximum serial data rate (mode1) of CµCs-51 is direct proportional with Cp ($f_{cp} = 24MHz$) in relation (14) [6].

$$Maximum\,Data\,Rate_{mode1} = \frac{f_{cp}}{192}$$

$$= \frac{24\times10^6}{192} = 0.125\ \ [Mb/s] \qquad (14)$$

Moreover, for CµC-51, the data rate of serial transmission (serial channel signaling rate SCSR) can be determined as follows in eq. (15) [10]:

$$SCSR = \frac{\log_2 n}{T} \qquad [Mb/s] \qquad (15)$$

As shown in Fig.11 the minimum duration for one bit ($Tc_{SP} = 8\,\mu s$), and with a two-condition modulation (n=2) the $SCSR_C$ will reduce to eq. (16):

$$SCSR_C = \frac{1}{Tc_{SP}} = \frac{1}{8\times10^{-6}} = 0.125\,Mb/s \qquad (16)$$

We notice that the simulation result of the CµC-51 data rates (mode1) is meet with the $SCSR_C$. The total time to transmit start bit, 8 data bits, and stop bit is ($10Tc_{SP} = 80\,\mu s$), so the conventional executed time (Tc_{exe-S}) for out the 3 data bytes (AAh, CCh, and F0h) serially is define by eq. (17):

$$Tc_{exe-S} = 30 * Tc_{SP} = 240\,\mu s \qquad (17)$$

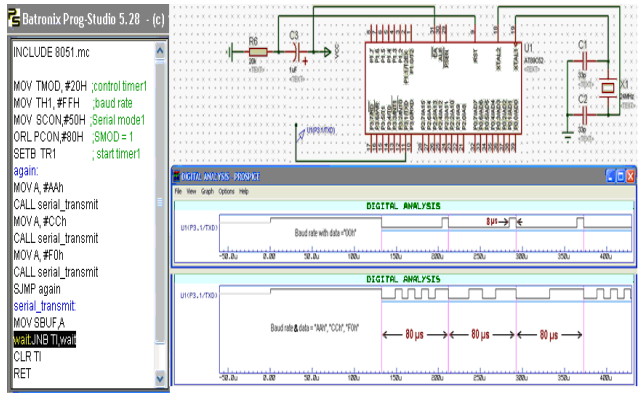


Fig.10 Circuit diagram, timing diagram and assembly code of AT89C52 for 3 sequence serial bytes

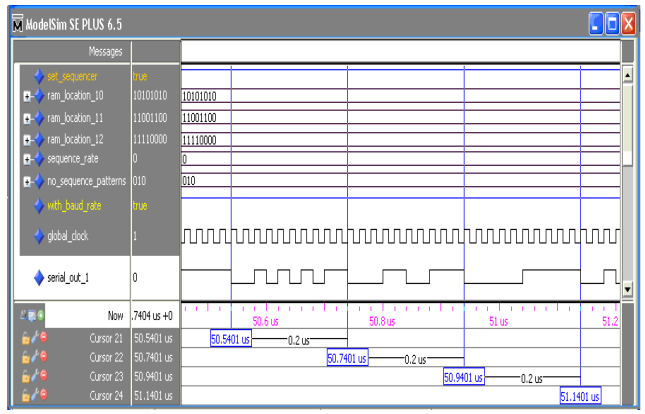


Fig.11 Timing diagram showing the generated sequences of 3 serial bytes with start and stop bits

Consequently the simulation using "Modelsim SE 6.5" as show in Fig.11, for the SIMD instruction "OutSeqB 0,1,2,1" (as we explained before) will transmit 3 bytes in serial transmitting with maximum data rate, we can observe that the minimum duration of the modified the total time for transmit any byte (AAh, or CCh, or F0h) is 0.2 µs, the duration time for any serial output bit (Tm_{SP}) or the time difference between any two successive bits is only 0.02 µs, which equal to one of clock pulse duration (assume f_{Mcp} = 50 MHz) is define by eq. (18)

$$Tm_{SP} = T_{Mcp} = \frac{1}{f_{Mcp}} = \frac{1}{50\times10^6} = 0.02\,\mu s \qquad (18)$$

Therefore the modified executed time (Tm_{exe-S}) for SIMD "OutSeqB 0,1,2,1" which response to out the 3 data bytes (AAh, CCh, F0h) serially (with start and stop bits) from 3 RAM locations (R0, R1, R2), each bit need one clock pulse (totally 30 Cps) as observed in the digital analyzer in Fig.11 is determined by "Iron Law" in equation (19).

$$Tm_{exe-S} = 1\times30\times0.02 = 0.6\,\mu s \qquad (19)$$

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Comparing results of eq. (17) and eq. (19) it is found that the MµC is more superior to that the traditional type in the case of serial transmissions. The speed up of the modified superscalar µC for the modified SIMD "OutSeqB" in the case of serial data transmission is obtained from relation (20) [9]:

$$Speed\ up(M_{Serial}) = \frac{Tc_{exe-S}}{Tm_{exe-S}} = \frac{240\ \mu s}{0.6\ \mu s} = 400\ times \quad (20)$$

Moreover the serial transmitting data rate for MµC-51 determine by serial channel signaling rate (SCSR) law, so the modified serial channel signaling rate ($SCSR_M$) will reduces to eq. (21):

$$SCSR_M = \frac{1}{Tm_{SP}} = \frac{1}{0.02\times10^{-6}} = 50\,Mbps = 5.0\,MB/s \quad (21)$$

The $SCSR_M$ will equal to $6.25\ MB/s$ in case of transmitting serially without both start and stop bits.

The $PCSR_M$ is meeting with our empirical parallel data rating in eq.(1), in case of operand (N=0), the maximum serial data rate (without both start and stop bits) will define in equation (22) , and for maximum serial data rate (with start and stop bits) will define in equation (23).

$$Adjusted\ Data\ Rating_{Serial_8bit} = \frac{f_{cp}}{8\times(1+N)} = \frac{50\times10^6}{8\times(1+0)} = 6.25\ MB/s \quad (22)$$

$$Adjusted\ Data\ Rating_{Serial_10bit} = \frac{f_{cp}}{10*(1+N)} = \frac{50\times10^6}{10\times(1+0)} = 5.0\ MB/s \quad (23)$$

Moreover, in all cases the modified type is superior to the traditional µC, the serial data rate enhancement (SDRE) of the modified superscalar µC as follows:

$$SDRE = \frac{SCSR_M}{SCSR_C} = \frac{5MBps}{0.0125MBps} = 400\,times \quad (24)$$

Finally, for the purpose of comparisons between the traditional µC-51 and the developed type shows the maximum data rate, enhancement data rate, speed up, at different applied clock frequencies of FPGA chips as shown in Fig.12.

4. Conclusions

Recently the family of µC-51 has numerous applications in different fields. Utilizing the reserved bits of the traditional µC-51 we extended its internal architecture and expanded its instruction set along the main dimensions of parallelism. By using the VLIW and SIMD types on superscalar architecture processor we can enhance the performance of the conventional µCs. we can execute multiple instructions in same clock cycle, and we can perform multi-operations like fast Fourier transform, matrix multiplication, finite/infinite impulse response filters. The modified µC is more superior to the traditional µC from many points of view of performance parameters including data rate, execution time, and speed up.

	Parallel Outputs (Bytes)					Serial Outputs (start+8data+stop) bits				
f_{cp} of FPGA [MHz]	50	100	150	200	250	50	100	150	200	250
Clock Ratios	2.08	4.17	6.25	8.33	10.42	2.08	4.17	6.25	8.33	10.42
Max Bit Rate [Mb/s]	400	800	1200	1600	2000	50	100	150	200	250
Data Rate Enhancement	50	100	150	200	250	400	800	1200	1600	2000
Speeding up	50	100	150	200	250	400	800	1200	1600	2000

Fig.12 most comparison levels for several clock frequencies

References

[1] Jonghee M. Youn, Sechul Shin, "A New Addressing Mode for the Encoding Space Problem on Embedded Processors", 7th Symposium on Application Specific Processors (SASP), IEEE, 2009.
[2] Elena Roxana Buhus, "A System-On-Chip Approach in Designing a Dedicated RISC Microcontroller Unit Using the Field-Programmable Gate Array", 5th International Conference on Systems, computer society IEEE, 2010.
[3] M.Suaib, A.Palaty, K.Sambhav, "Architecture of SIMD Type Vector Processor ", International Journal of Computer Applications, Volume 20, April 2011.
[4] Website,www.realworldtech.com/page.cfm, last accessed on the 18th of Aug 2011, hour 16:30
[5] Web site, "wikipedia.org/wiki/Superscalar" last accessed on the 10th of Sep 2011, hour 13:40
[6] Web site, www.Philips.com, "80C51 family programmer's guide and instruction set" last accessed on the 17th of Aug 2011, hour 02:03
[7] Sajjan G.Shiva, "Computer Organization, Design and Architecture", Marcel Deker,NY, third edition, revised and expanded, pp.181-450,2000
[8] Joydeep Ray, "High-Level Modeling and FPGA Prototyping of Microprocessors ", IEEE, 2003.
[9]A.R.Alameldeen, D.A.Wood, "IPC Considered Harmful for Multiprocessor Workloads", computer society IEEE, Aug 2006.
[10] H. Weik Martin, "Fiber Optics Standard Dictionary", Hall book, 3rd edition, 1997.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

A Strategic Framework for Designing E-Learning System with Focus on University Entrepreneurship

Masoud Asgarimehr[1], Babak Shirazi[2] Meisam Jafari Eskandari[3] and Salman Rostami[4]

[1] **Department of Information Technology Engineering, Mazandaran University of Science and Technology, Babol, Iran**

[2] **Department of Industrial Engineering, Mazandaran University of Science and Technology, Babol, Iran**

[3] **Department of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran**

[4] **Department of Applied Mathematics, Tarbiat Modares University, Tehran, Iran**

Abstract

One of the most important issues in the current e-learning systems, is lack of attention to university entrepreneurship topics. This article offers a strategic framework for designing e-learning system with focus on university entrepreneurship. The main goal is strategic conformity of e-learning and university entrepreneurship. In this regard, in addition to considering various aspects of e-learning system design, we will emphasize on cooperation, overlap, and matching three factors includes: government, industry and educational system to create an entrepreneur system. This framework is a strategic framework that without addressing the details, offers the overall road map for e-learning system design and this is one the advantages of this framework. Moreover, according to the second academic revolution and attention to issues of university entrepreneurship, the provided framework has special look at the production and transfer of knowledge and move toward university entrepreneurship.

Keywords: *e-learning, university entrepreneurship, designing e-learning system, strategic framework.*

1. Introduction

In the current era that is the era of knowledge and information, the need for education and learning new technologies is inevitable. Meanwhile, e-learning plays an important role in education of the public and remove existing restrictions on the traditional educational system. Also, the growing and expanding application of information and communication technologies, has contributed to the elimination of international borders and globalization. The field of education did not exclude so that the various universities in the world are looking for teachers and attract students from around the world and gaining international market in this area.

E-learning is subset and common denominator of IT and educational technologies. In e-learning, providing courses and learning is performed through electronic technologies. This type of learning is based on technologies such as Internet, intelligent learning systems, computer-based training systems and multimedia [1].

Entrepreneur university is a university that produce knowledge for society. Original pattern for entrepreneur university is creation and implementation of new and transferable knowledge. Moreover, entrepreneur university plays a key role in economic and social development [2].

In recent years the role of universities and educational systems, have been dramatic changes. In fact, universities in addition to education and research that is their main mission, will play another role that surely is the economic and social development. In other words in addition to education and research, entrepreneurship has been as the main component of the universities roles in recent years [3]. Meanwhile, the frameworks and many models have been presented in order to e-learning system design. Also, models and strategies have been presented in line with creating entrepreneur university. But in the field of e-learning, haven't performed so much activities for designing entrepreneur system. The main mission of the paper is to provide a strategic framework for designing e-learning system with a focus on entrepreneurship. In fact, the main goal is strategic integration and conformity of e-learning and entrepreneurship.

Thus, the first section of this paper, will review the works done in the field of e-learning and discusses about the frameworks and models presented in this context. It will also review existing models and strategies for creating entrepreneur university. The next section, explains the desired framework for designing entrepreneur e-learning system. And finally, the last section of this paper will pay attention to analysis the components of system.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

2. Background Work to Pedagogical Framework Design

Obviously, for designing any system, moving toward a particular model or framework, lead to systematic work and will achieve better results. This also applies for designing e-learning systems. Using a help form specifically a defined framework, can determine e-learning system requirements, and make it easier to design [4]. In this regard many models and frameworks are presented, which we will review some of them.

2.1 A Framework for Web-Based Learning

Badrul khan (2001) presented a framework for designing web-based pedagogical system. He believes that with the growth of the Internet, the web has become the powerful, global, interactive, dynamic, and economic interface for distance education and learning. He found that numerous factors help to create a meaningful learning environment, and many of these factors are systemically interrelated and interdependent. He clustered these factors into eight dimensions: *institutional*, *management*, *technological*, *pedagogical*, *ethical*, *interface design*, *resource support*, and *evaluation*. Each dimension has several sub-dimensions. Each sub-dimension consists of items focused on a specific aspect of a Web-based learning environment [5].

2.2 The CSALT Networked Learning Model

The CSALT networked learning model (2001) is developed by Peter Goodyear and his colleagues at Lancaster University. The model is aimed particularly at tutors in higher education and includes a pedagogical framework as well as providing an overview of the broader issues surrounding networked learning. The pedagogical framework defined here introduces four levels of pedagogy: *philosophy*, *high-level pedagogy*, *strategy* and *tactics*. The upper two levels are considered as declarative or conceptual and the lower two levels are regarded as procedural or operational. The model suggests a distinction between the *tasks* designed by the tutor and the *activities* carried out by the learner. The model is also sensitive to organizational context. In addition, the model has strong focus on collaborative learning. This model demonstrates how learning outcomes can be associated with specific supported learner groups and their activities need to be designed with these outcomes in mind [6].

2.3 The European CANDLE Project

The CANDLE consortium are a European IST[1] funded group exploring collaborative and network distributed learning environments. Part of their deliverables has included a pedagogical framework that focuses upon the interactions that take place between tutors and learners. This framework is based on activity theory and rhetorical structure theory. The activity theory emphasizes on the essential unity of systems, users and their goals, and contexts, including the community in which an activity occurs. The interrelationships between all the elements of an activity (described as the structure of an activity) are key to describing complex systems of behavior such as pedagogy. Rhetorical structure theory is concerned with the choice of particular forms of expression to realize discursive goals. Six dimensions are refined from these two theories: *the purpose of an activity*, *the structure of an activity*, *the context of an activity*, *tools used in the activity*, *objects used in the activity*, and *roles for the participants in the activity* [6].

2.4 The UNITE Pedagogical Framework

UNITE[2] was a 30-month project, partially supported by the European Community under the information society technologies priority of the sixth Framework Programme for R&D. The main objective of the project was to provide innovative approaches to the education of young Europeans by integrating into the curricula of a number of schools in Europe-wide network different state-of-the-art technologies in e-Learning, also taking into consideration innovation in technology and pedagogy. The project was successfully completed in July 2008. The derived framework consists of five components [4]:

- Pedagogical framework context: This includes: an emphasis on the empowerment of learners through the increased interaction/communication that is supported by the system; an interest in the affordances of mobile devices inside and outside the classroom and an awareness of the need to understand traditional practices in order to challenge them.
- Pedagogical approaches: UNITE project intended to promote active learning in collaborative and individual contexts with an emphasis on student autonomy.
- Assessment techniques: Four types of assessment were considered: computer-based assessment, self-assessment, peer-assessment and tutor-assessment.
- Current pedagogical practices implemented in national curricula and national specifics.
- Teacher training: Teacher training was also included in the pedagogical framework design process in order to help teachers get started and support their ongoing work.

Each of the presented models and frameworks, focuses on a particular aspect of e-learning. In addition they have also many commonalities. Table 1 presents the summary of these frameworks and models.

Table 1: Summary of E-Learning Models and Frameworks

[1] Information Society Technologies

[2] unified e-Learning environment for the school

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

	Framework or Model	*Year*	*Specifications*	*Components*
1	Framework for Web-based Learning	2001	Considering all aspects of educational system design	Institutional, management, technological, pedagogical, ethical, interface design, resource support, and evaluation.
2	The European CANDLE project	2003	Based on activity theory and rhetorical structure theory	The purpose of an activity, the structure of an activity, the context of an activity, tools used in the activity, objects used in the activity, and roles for the participants in the activity.
3	The CSALT Networked Learning Model (V1.3a)	2004	Strong focus on collaborative learning - distinction between the *tasks* designed by the tutor and the *activities* carried out by the learner	Philosophy, high-level pedagogy, strategy, tactics and organizational context
4	The UNITE pedagogical framework	2009	Design, implement and validate a pedagogical framework for e- and m-learning in secondary schools	pedagogical framework context, pedagogical approaches, assessment techniques, Current pedagogical practices, teacher training

3. Entrepreneur University

As also mentioned earlier, entrepreneur university should be producer of knowledge for society. The main aspects in entrepreneurial pattern of university are creation and implementation of transferable new knowledge. All these contain topics related to R&D funding, inventorship and ownership of patents. Patenting alone is not the evidence of entrepreneurial behavior of the academia, but this is one of the first steps targeted to implement created new knowledge in real business [2].

Tonis Mets (2010) emphasizes on commercialization universities' research. He believes that the previous missions of universities (education and research), have been complemented by a third, economic and social development mission mentioned also as serving society, innovation, or in narrower meaning – technology transfer (TT) activity. The adoption of the third mission is referred to as the second academic revolution and active universities in that process are called entrepreneurial universities [2].

He believes that for creating entrepreneurial university, having more tight links and overlaps between government, industry, and academia as well as between three domains of academia (education, research, and entrepreneurship) is critical. In this context, he has offer the Business model of R&D commercialization for entrepreneurial university in the framework of University-Industry-Government.

The model shows that how university can create value from its own research [3].

Salehi Amiri et al. (2009) have offered five strategies for creating entrepreneur university: *entering the investigation field*, *expanding the activities of industry offices and universities*, *insurance services*, *offering services related to credit*, and *business activities*. They also believe that universities can create opportunities for technology and productivity for people, and guide them to use these opportunities effectively [7].

Clearly, the trends of higher education have experienced great changes in the world. Today, the prestigious universities in addition to providing proper educations and emphasis on effective researches, also have special attention to development and economic growth and university entrepreneurship is an inseparable part of the new educational systems. Figure 1 shows the main actors and domains of entrepreneurial university [3]. For design of entrepreneur e-learning system, should pay special attention to these actors and domains and this is one of the cases that is considered in the proposed framework.

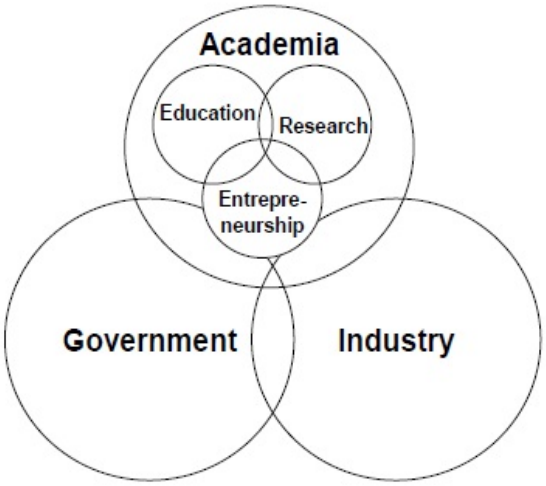


Figure 1: Main Actors and Domains of Entrepreneurial University [2]

Totally, according to the presented models and strategies in order to achievement of the entrepreneur university, and to create the entrepreneur university, the following components are as effective components:

- Investment in research and development
- Special attention to inventions and their commercialization
- Conducting Business
- Providing entrepreneurship trainings

4. Research Method

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

This research is development and operational in purpose and is descriptive and survey in performance. In fact, by reading documents (discussed models and frameworks in the field of e-learning and the entrepreneur university) as well as polls of experts and professionals in universities, the existing and effective components were extracted and analyzed and finally, the intended framework was designed. The statistical populations of this research are professors and Ph.D students in related areas of research.

5. Framework of Entrepreneur E-Learning System Design

As also mentioned previously, to design effective e-learning system, moving toward a specific framework, is determined the educational system requirements for designer and the system design work will be easier. Also the current frameworks haven't special attention to the issues of university entrepreneurship, and are emphasized on the main components of e-learning systems. According to the mentioned cases, we've tried to provide a strategic framework in order to entrepreneur e-learning system design. In fact the main goal is integrating e-learning and university entrepreneurship and strategic conformity of both of them. The purpose of the strategic framework is to designing overall road map for e-learning system without addressing the details.

Two main aspects are considered to design presented framework .The first one has been reviewed of the models, frameworks, strategies, and works done in the field of e-learning and entrepreneur university. As you previously observed in the second part of this paper, four important models and frameworks on e-learning, were reviewed. Each of these frameworks and models has looked at the e-learning from the special aspect, and haven't considered all the components. Also we discussed about the models and strategies of creating entrepreneur university. In the presented framework in this article, we have intended to combine these models and attention to all the important components in order to e-learning system design. The second aspect has been used of experts opinions in the field of e-learning and entrepreneurship. In this regard, we appointed a separate questionnaire to the professors, students and hard-working experts in the areas of e-learning and the university entrepreneurship. The questionnaires were related to effective components within e-learning system design and strategies for creation the entrepreneur educational system. In consider to these two aspects and achieved results, we offered the desired framework. Figure 2 shows the presented framework.

As can be seen in Figure 2, the presented framework consists of various components that we will explain each of them.

5.1 Rules and Practices of Higher Education

On the tip of the pyramid of presented framework, the relevant laws and rules and practices of higher education located in that country is placed. It means that before do anything you must have special attention to existing rules and methods. In fact, this component is as a mover for e-learning system design. In this regard should be considered various cases. Issues such as: national curriculum and education, higher education policies, educational methods and techniques available in universities, technical infrastructure in the country, and current assessment techniques for students. It must be noted that, in addition to the mentioned cases and consider them in the design process, we should try to change some existing attitudes and also use modern educational methods and approaches [4].

5.2 Context

The other cases in the beginning stages of design should be considered, is attention to the existing and considered context. The context involved a wide range of concepts, some of which include: society cultural backgrounds, technology, motivational factors, Student's financial constraints, and the contexts used by the design team (e.g., interest in the use of mobile learning or special attention to the increasing interaction between students and professors). In fact, the context has determined effective cases on desired framework and is discussed as a basis for system development [4,8,9].

5.3 Mission and Goals

Before designing any educational system, we should determine the mission and goals of the Institute. For example, is the institution looking for money or it is looking to increase the level of access to education at any time and any place. This will affect directly on the educational system design.

5.4 Education, Research, Entrepreneurship

The most important part of the presented framework is the design of its main components namely education, research, and entrepreneurship. In traditional education systems, only the two components are considered: education and research. But now serving the community and generate and transfer knowledge and technology and entrepreneurship in a word as an inseparable part of the educational system is considered. Each these three components to a suitable form should be developed for the design of educational system. In addition to this issue, we also note that these three components linked together and have an interaction to each other and design and development of each of them will require attention to other components. Below we examine each of these three components:

5.4.1 Education

Education traditionally is considered as the most important part of any educational system. In presented framework, there are five major strategies for design of the education component: *learner education, teacher training, educational content, learner assessment, and entrepreneurship trainings.*

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

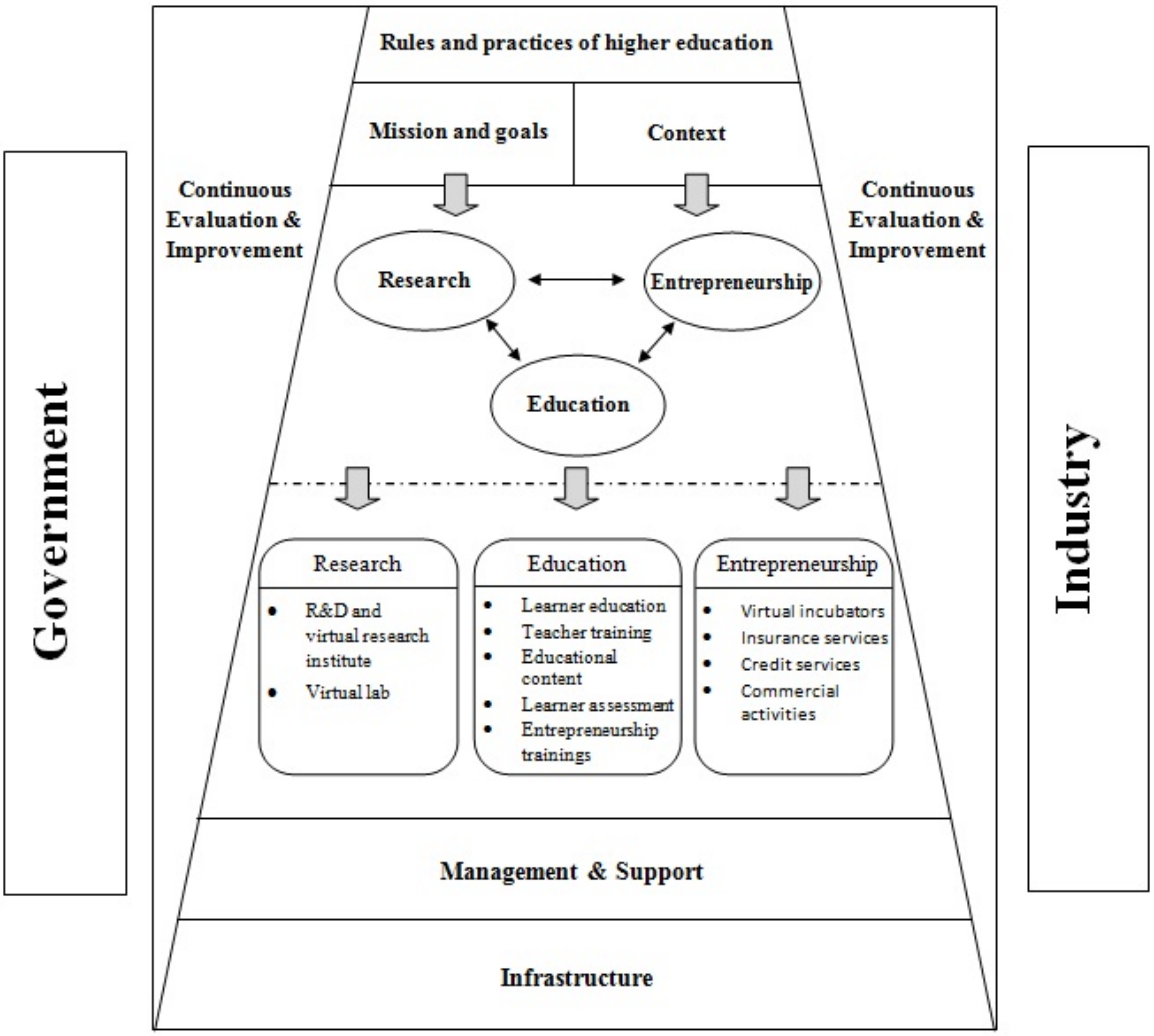


Figure 2: Strategic Framework for Designing Entrepreneur E-Learning System

5.4.1.1 Learner Education

The purpose of learner education is select and design of educational approaches and methods. Various approaches are considered for learner education in electronic learning systems. Among them are: *blended learning, collaborative learning and active learning* [4]. The purpose of blended learning is combining traditional learning methods (e.g., classroom) and online learning. The term of collaborative learning refers to efforts and activities among a group of learners and professors. There are various approaches to collaborative learning that all emphasize to form the groups of students as a mutual cooperation to search and review issues and problems. The purpose of active learning is use of teaching methods that involve learners in the learning process [4]. E-Learning system designer according to the backgrounds and objectives of the Institute and also considering the rules and practices of higher education, should be selected the appropriate approach or even a combination of approaches.

5.4.1.2 Teacher Training

Teachers training refers to training of educational new skills for teachers, learning how to work with designed system, training of new skills in information technology, and face to face meetings and briefings with the teachers [4].

5.4.1.3 Educational Content

Educational content is a very important component, in any educational system especially in e-learning systems. The role of Teachers in e-learning isn't direct training the students but the teacher has the role of facilitator in the process of knowledge transfer and training. Therefore, in these systems, learners will engage in learning process, by taking advantages of educational content. In e-learning, educational content will be presented in electronic form and through the Web, CD, simulations, etc. Also with the help of social networks, wikis, video conferences, blogs, etc. the desired information can be shared and an interactive learning environment can be created. In this context, creating a separate part for educational content is very important and

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

necessary [6,10]. Moreover, the importance of using appropriate Learning Content Management System (LCMS) would be felt too. Learning content management system is used for converting the information into collaborative learning content and content presentation in different ways for different educational purposes and various audiences.

5.4.1.4 Learner Assessment

Learner assessment is one of the cases that considered in the presented framework. Traditionally, assessment of learner progress is considered as an inseparable part of each educational system. In the presented framework, four type of assessment is considered: Computer-based assessment, self assessment, peer assessment, and Tutor assessment. In this framework the purpose of computer-based assessment is using of online exams in form of multiple choice questions or anatomical questions. We can held these exams during the semester and on several occasions and also use them at the end of the semester for final assessment of the learners. Self-assessment includes involving the learners in their assessment process. Thus the questions can be raised for them and the learners after answer to the questions, can assess the answers based on learning content and they will know the status of their course. Peer assessment is very similar to the self-assessment. Using peer assessment, the students can participate in the assessment process and actually help each other. By the use of forums, they will express their views about other learners works. Tutor assessment is expressing the traditional assessment form with new technologies. The Tutors can give research projects or practical exercises to learners and eventually send feedback to them by e-mail [4,11].

5.4.1.5 Entrepreneurship Trainings

In the presented framework, entrepreneurship trainings will be presented during the educational course. In addition, entrepreneurship short-term and professional courses and entrepreneurship workshops, should be held. In these courses, learners are familiar with the basic concepts of entrepreneurship, the importance of its role in the creation of small and medium companies and economic development, and learning how to create jobs outside government. These trainings will be presented by electronic platforms and especially with the help of the Internet [7].

5.4.2 Research

Particular attention to research is one of the most important trends in higher education in the world. Today, various Universities support professors and students in their research activities and provide appropriate infrastructure [7].The main foundation of entrepreneurship is production and transmission of knowledge, and this will be possible with special attention to research and investment in this sector. In the presented framework two major strategies are considered for achieving the goals of the research: *investment in research and development (R&D) and creating virtual research institutes, creating virtual labs.*

5.4.2.1 Research and Development and Creating Virtual Research Institutes

As mentioned earlier, research is one of the most important trends in higher education in the world. Traditionally, Research is one of the main pillars of education and university systems. In addition, it can be named as executable processes in educational environment for the growth of entrepreneurship. In an entrepreneur educational system, research plays a vital role [2,7]. By creating virtual research institutes or virtual research environment, we can focused on research activities and give them direction. Virtual research environment, is a Web-based environment in which various researchers from around the world are working together on issues related to their research topics [12]. In the presented framework, the purpose is creating a virtual research environment as a virtual research institute. Among these, we should pay special attention to applied sciences. In addition, we can support the teachers and learners research activities. It must be noted that all the mentioned cases should be done through the Internet and the Web. Internet provides an interactive environment to facilitate interaction and collaboration between teachers and learners. In other words, different teachers and students and learners around the world can exchange knowledge with each other. Thus, in the context of the Internet we can provide a favorable environment for research and development and exchange of knowledge and information and this is the main goal of the presented framework. Obviously, for achieving the research goals, moving toward research-oriented education, will be another effective strategy.

5.4.2.2 Virtual Lab

As we know, many courses in engineering and basic sciences, need to laboratory environment to complete the students learning. Also researchers can test and research on various research topics in lab environments. In fact, using lab is common point between education and research. It is not possible to use physical lab in e-learning environments. One way to overcome this problem is using virtual labs. In fact, the virtual lab is a computer software that simulates experimental experiences in various sciences and make it possible for learners to engage in simulated tests. This lab can be a web-based program or an independent program in the learners personal computer [13]. For designing e-learning system, design and development of virtual laboratories is very important and the presented framework has special attention to this subject.

5.4.3 Entrepreneurship

Entrepreneur educational system must produce knowledge and technology for society. In this regard, in addition to investment in research, it's also essential to plan appropriate mechanisms for entrepreneurship. In the provided framework there is a special attention to entrepreneurship issues. In this regard there are four major strategies: *creation of virtual incubators and technology transfer, providing insurance services to entrepreneurs, providing credit services, and commercial activities.*

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

5.4.3.1 Virtual Incubators and Technology Transfer

The university incubator is an environment that will be created in order to support of entrepreneurship. In fact, the main purpose is development of new institutions and companies based on research and technology and relying on the university facilities. Connecting research, technology, and capital will be the main role of university. Moreover, it acts as the driving force for flowering entrepreneurship talents, new companies development based on technology, and accelerate technology commercialization. In fact, appropriate capacities for research and industry connection will be created [14,15]. In this regard, can be brought together the entrepreneurs, professors, university graduates, students etc. , and provide them appropriate software, hardware, workshop, and educational facilities and equipments to allow them to test and implement their ideas [7].It is not possible to create a physical location for the growth centers in virtual and electronic learning environments. Therefore, a virtual environment should be designed for it. This environment is designed in the context of the Internet and World Wide Web, and the entrepreneurs, teachers and learners can exchange the information with each other and various facilities will provide for them in form of virtual. Virtual lab is one of these facilities. The presented Framework has special attention to creation of virtual incubators. As a result, one of the main components of the presented framework is designing appropriate virtual environment for incubators. Also various Internet technologies such as e-mail, video conference, and virtual groups can help to establish the possibility of the information exchange and effective collaboration among entrepreneurs, teachers and students. In addition, the process of transferring new technology to the market will be facilitated by creation the virtual enterprises based on research and technology and cooperation with industrial companies. We should also pay special attention to patents and the results of researches and intellectual property. Alongside the patents, should be turned to creation of Spin-off companies. The term Spin-off, refers to businesses those are under a certain brand but in terms of production strategy and business management are completely independent and just work for the company under specific brand [2]. It must be noted that these Spin-off companies, will be virtual with distributed resources.

5.4.3.2 Insurance Services

Obviously, in order to design and implement an entrepreneur system, the necessity of entrepreneurs adequate support is inevitable. In addition to spiritual supports, material supports must be considered for entrepreneurs. In this context, the special entrepreneurship insurances must be considered for entrepreneurs [7].

5.4.3.3 Credit Services

In order to meet financial needs and to support entrepreneurship, can be engaged in credit and financial activities. Some of the effective strategies in this area are creating funding agency, private banks, and providing financial facilities [7]. It must be noted that the goals and mission of the institution are effective in this case. For example, if the institution is as a state institute and free from financial concerns, and its main purpose is to promote access to education at any time and any place, using of credit strategies will be less important.

5.4.3.4 Commercial Activities

Educational systems, can have an important contribution in the commercial activities. In fact, considering to market needs and participation in national and international projects, can help significantly to promote university entrepreneurship and national development and progress. Presently, there are many undiscovered and disabled markets in national and international levels that we can use them appropriately. Today, information and communication technology have become to one of the most important and the most efficient sciences in developing human societies. Also learners have unique interest to this side. Thus, with providing an appropriate context for growth of potential talents in these areas, and enough support of learners, can be involved in further growth of this industry and this will ultimately led to the development of entrepreneurship in this sector [7]. This is one of the goals of presented framework.

5.5 Management & Support

Appropriate management and support of e-learning process, will have a positive impact on results. The purpose of management and support is to design the management processes and to create the support team. Also it is necessary to supply or design the appropriate learning management system. Learning Management System is a software package that is responsible for management of teaching and learning. In fact, it is a strategic solution for planning, presentation and management of all educational events. The learning management system, focus on managing the learners and tracking the progress of them in educational activities. Learning Management System is a platform to achieve the set of distributed resources and educational activities. This tool will automate the educational planning management. According to the mentioned definition, the main tasks of the learning management system are: tasks related to registration and learner's information, tasks related to course management, tasks related to tracking the progress of learners, and tasks related to reports. Another task in this layer is design and implementation of a digital reference library, that it can help us to access to the online articles, e-books, etc., through the internet.

5.6 Infrastructure

One of the basic conditions for designing any system, is providing the suitable infrastructure. In the presented framework, the infrastructures have been also considered as a substructure of the system. In the presented framework, the purpose of infrastructures is to create a suitable infrastructure for communicational networks in order to easy communication between learners and the system. In this regard, providing access to broadband internet is essential. Clearly, government and

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

industry are responsible for this. These two sectors together with effective collaboration, will provide a suitable infrastructure of communicational networks.

5.7 Continuous Evaluation & Improvement

In provided framework, has been special attention to continuous evaluation and improvement, so that is involved in all stages of e-learning system design. The main goal is to improve the system quality. In all stages of system design, and even after delivery and run, it is necessary to have adequate attention to the system evaluation and its improvement. This is very important and we can design a separate system for this purpose. In this regard, the various indicators of different aspects of the system must be considered. Performance and quality indicators, educational, research, entrepreneurship, networking, software, and learners satisfaction, are such important cases. As we mentioned, continuous evaluation & improvement is one of the most important parts of the presented framework, and we should design a separate system for it.

5.8 Government and Industry

To create the entrepreneur educational system, cooperation, overlap, matching and connecting three factors includes government, educational system, and industry, is necessary [2]. So the government and industry are two main wings in the presented framework that play a key role in all stages of system design and specially after performing the system. In fact, the government by providing and performing the appropriate laws, support, monitoring and evaluation of all matters, adoption of the appropriate policies to facilitate the transfer of knowledge and technology, as well as financial support, will support the university entrepreneurship. [2,7]. Also the industry with appropriate investment, the use of experiences, and using his ability to create different and diverse markets, can provide the conditions to establish the small and medium enterprises with high productivity [7]. Obviously, achieving the objectives of entrepreneurship will be possible with the effective collaboration of these three factors and this is one of the main goals of the presented framework.

6. Components Analysis of the Proposed Framework

As mentioned earlier, in order to design the presented framework, first we extracted related components by study and review the models, frameworks, strategies and work done in the areas of e-learning and university entrepreneurship and also by experts and technical specialists assistance. After analysis of the significant dimensions by researchers and experts, the most important and the most applied components were identified. Then some questionnaires were provided to identify the rate of effect and operational level of these components. Factors and effective components, the number of respondents, the average scores, and reliability of research questions, are shown respectively in table 2.

As can be seen in Table 2, to determine the reliability of the questionnaires, the Cronbach's alpha coefficient is used.

After the analysis of the components and designing the desired framework, in order to validate the proposed framework, a questionnaire was prepared and the opinion of 30 experts, were collected. The statistical results obtained from the data collected, represents that totally 86.3 percent of the experts, have diagnosed that the proposed framework is very good and well, and hence, the above framework has been approved by experts.

Table 2: Effective Component of the Presented Framework

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Effective components	*Number of respondents*	*Average scores*	*Standard deviation*	*Cronbach's Alpha*	*Effective components*	*Number of respondents*	*Average scores*	*Standard deviation*	*Cronbach's Alpha*
Rules and practices of higher education				0.928	**Contexts**				0.876
Higher education policies and national curriculum and education	85	3.45	1.19		Cultural background	85	3.50	1.38	
Technical infrastructure in the country	85	3.30	1.08		Motivational factors	85	3.57	0.96	
Current assessment techniques for students	85	3.48	1.15		Student's financial constraints	85	3.48	1.26	
Mission and goals				0.830	**Research**				0.878
Institute mission	85	3.37	1.02		R&D and Virtual research environment	85	3.80	1.009	
Institute goals	85	3.42	1.03		Virtual lab	85	3.67	1.01	
Education				0.899	**Entrepreneurship**				0.877
Learner education	85	3.44	1.36		Virtual incubators	85	3.95	1.09	
Educational content	85	3.94	1.12		Insurance services	85	3.87	1.07	
Learner assessment	85	3.57	1.20		Credit services	85	3.78	1.10	
Entrepreneurship trainings	85	3.87	1.16		Commercial activities	85	3.98	1.06	
Management & support				0.804	**Infrastructure**				0.813
Development of a learning management system	85	3.63	0.79		communication networks	85	3.47	0.97	
Design and implementation of digital reference library	85	3.43	0.80		information infrastructure	85	3.22	1.14	
Continuous evaluation & improvement				0.820	**Government and Industry**				0.891
Development of an operational evaluation system	85	3.60	0.86		Support for SMEs	85	3.45	1.05	
Development of evaluation indicators	85	3.35	0.85		Appropriate investment	85	3.42	1.004	

7. Conclusions

Advances in information and communication technology, had great impacts on various sectors including education, so that in the recent two decades, e-learning and virtual education as a modern educational system and away from the limitations of traditional systems, is well-known among the various social groups. One of the significant issues in current e-learning systems is lack of attention to university entrepreneurship issues. In this article we offered a strategic framework in order to design

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

the entrepreneur e-learning system. The main goal has been strategic conformity of e-learning and university entrepreneurship. In this regard, we emphasized on various aspects of e-learning system design. The main axis of this framework was design of three core components of the educational system, namely education, research, and entrepreneurship. In addition, we learn that the government and industry are as two major wings in design of entrepreneur e-learning system. This framework has a strategic vision and will provide the overall road map for designers, managers and policymakers of e-learning systems and this is one the advantages of this framework. In addition, the presented framework has special attention to production and transfer of knowledge and technology and commercialization of research activities. In summary, achievements of this research are:

- Integrity of entrepreneurship and e-learning
- Design a strategic framework for e-learning
- Provide a new framework for entrepreneur e-learning system design

In the future, development of the presented framework and attention to the details of the components will be considered.

References

[1] Moore, J.L., Dickson-Deane, C. and Galyen, K., "E-Learning, Online Learning, and Distance Learning Environments: Are They the Same?", The Internet and Higher Education, 2010, doi: 10.1016/j.iheduc.2010.10.001.

[2] Mets, T., "Entrepreneurial Business Model for Classical Research university", Inzinerine Ekonomika-Engineering Economics, Vol. 21, No.1, 2010, pp. 80-89.

[3] Mets, T., "Creating business model for commercialization of university Research", Management of Organizations: Systematic Research, No.51, 2009, pp. 83-94.

[4] Granic, A., Mifsud, C., and Maja, C., "Design, Implementation and Validation of a Europe-wide Pedagogical Framework for e-Learning", Computers & Education, No. 53, 2009, pp. 1052-1081.

[5] Khan, B. H., A Framework for Web-based Learning. In B. H. Khan (Ed.), Web-based Training, Englewood Cliffs, New Jersey: Educational Technology Publications, 2001.

[6] de Freitas, S. and Mayes, T.," JISC e-Learning Models Desk Study Stage 2: Review of e learning theories, frameworks and models"[Internet] London, 2005, JISC,Available from: http://www.jisc.ac.uk/ uploaded_documents/Stage%202%20Learning%20Models %20(Version%201).pdf [Accessed 15th January 2010].

[7] Salehi Amiri, S. R., Kavousy, E., and Shah Husseini, A., "Strategies for Creating Entrepreneur University", European Journal of Social Sciences, Vol. 11, No. 3, 2009, pp. 496-506.

[8] McCombs, B. L., and Vakili, D., "A Learner-Centered Framework for E-Learning", Teachers College Record, Vol. 107, No. 8, 2005, pp. 1582-1600.

[9] Goodyear, P. and Jones, C., Pedagogical frameworks for DNER (the Distributed National Electronic Resource), Deliverable DC1, EDNER Project. Lancaster: Centre for Studies in Advanced Learning Technology, Lancaster University, 2004.

[10] Drexler, W., "The networked student model for construction of personal learning environments: Balancing teacher control and student autonomy", Australasian Journal of Educational Technology, Vol. 26, No. 3, 2010, pp. 369-385.

[11] Jamornmann, U., "Techniques for Assessing Students' elearning Achievement", International Journal of The Computer, the Internet and Management, Vol. 12, No. 2, 2004, pp. 26-31.

[12] Keraminiyage, K., Amaratunga, D., and Haigh, R., "Achieving Success in Collaborative Research: The Role of Virtual Research Environments", Journal of Information Technology in Construction, Vol. 14, 2009, pp. 59-69.

[13] Omar, N., Zulkifli, R., and Hassan, R., "Development of a Virtual Laboratory for Radiation Heat Transfer", European Journal of Scientific Research, Vol. 32, No. 4, 2009, pp. 562-571.

[14] McAdam, M., and McAdam, R., "High tech start-ups in University Science Park incubators: The relationship between the start-up's lifecycle progression and use of the incubator's resources", Technovation, No. 28, 2008, pp. 277-290.

[15] Yadollahi Farsi, J. and Nikraftar, T., "Contextual Dimension of University Incubator Organizations in Iran", European Journal of Scientific Research, Vol. 48, No. 3, 2011, pp. 341-351.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

EQNM²L: Towards a Metamodel for Extended Queuing Networks

Abdelhabib Bourouis[1] and Brahim Belattar[2]

[1] Department of Mathematics and Computer Science, University of Oum El Bouaghi, Oum El Bouaghi, 4000, Algeria

[2] Department of Computer Science, University of Batna Batna, 5000, Algeria

Abstract

Queuing theory is the mathematical study of waiting phenomena that allows several performance measures derivation and calculation. Furthermore, queueing systems could be simulated, especially in case of complex models including non-classical concepts. The theory has been applied successfully in diverse fields and the development of a Domain Specific Modeling Language (DSML) for extended queueing systems seems to be a necessity and of great interest. This paper focuses on the cornerstone of this process by providing a minimal but clear and extensible metamodel called EQNM²L. Our aim is to induce discussion on and contributions for elaborating a whole mature DSML. The proposed metamodel and the XML-based exchange format are the first step in enhancing interoperability between analytical solvers and simulation tools. Based on the Model Driven Engineering (MDE) approach principles and tools, modeling environments and simulation/analytical code are generated automatically and could be maintained easily.

Keywords: *Extended Queueing Systems, Metamodeling, Domain Specific Modeling Language, Discrete Event Simulation, Interoperability.*

1. Introduction

Several Domain Specific Modeling Languages (DSMLs) have been developed in the last decade. Their use gained a lot of success and a growing popularity. However, developing a DSML is still a challenging and time-consuming task. Generally, Domain specific modeling (DSM) includes also automating the code generation directly from models. Automatic construction and maintenance of source code increases significantly developer's productivity. The reliability of automatic generation compared to manual coding reduces notably the number of defects in the resulting code thus improving quality.

Queueing network models have been used extensively as a modeling paradigm for deriving analytical as well as simulation based performance measures. They are commonly used to model a wide range of discrete-event systems. The Kendall notation is a mean for describing queueing networks especially in case of simple systems. For complex ones, a graphical notation with textual annotations is used instead. To analyze a model either by simulation or by mathematical analytic tools, the model is commonly coded and saved directly in a proprietary tool file format. Although it is the same formalism, there are always some ambiguities and disagreements on a number of concepts as well as on the exchange format. Therefore, tools are not interoperable and models can't be reused nor interchanged easily. In addition to rebuilding models from scratch every time the tool is changed, the development of modeling environments and model transformations, including code generation, are time-consuming, very expensive and hard to validate and maintain. The best solution to these problems seems to be based on MDE concepts including metamodeling and models manipulation in a generative manner.

The next section addresses key motivations and objectives of our work. A clear development methodology is presented in section 3. Section 4 discusses the basic domain concepts used for building the metamodel presented itself in section 5. The concrete syntax and the exchange format are discussed in section 6 while implementation is addressed in section 7. Finally, conclusions and improvements are given in section 8.

2. Motivations and objectives

The development and use of a common metamodel and a file exchange format is motivated by:

1. Lack of interoperability: performance evaluation tools of discrete event systems based on queueing

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

networks use different file formats for describing their models. There is no standard to ensure interoperability and reuse of conceptual models. Importing and exporting models is still hard to achieve due to the multitude of file formats. The problem is beyond simple technical solutions. XML-based standard exchange format seems to be the best solution.

2. Lack of expressiveness: unfortunately the expressive power of modeling languages depends on the tool purposes and capabilities. Tools don't model what they can't handle. In our opinion, a modeling language must be tool independent and focuses particularly on offering all necessary constructs for describing the largest possible range of systems.

3. Difficulty of the design and development of modeling environments: building modeling environments from scratch is hard, complex and time consuming. It is also difficult to maintain as the modeling language evolves. MDE seems offering the solution by automating the development based on metamodels.

4. Difficulty of code generation: the translation of the conceptual model into a simulation model is complicated and needs careful verification in addition to programming skills. Besides the manual translation, the classical automatic translation alternative is barely coded in simulation environments using programming languages. So translation rules used in this case are not clear and not easy to modify as the used formalism evolve.

The first initiative for developing such a metamodel and an XML-based exchange format named QNML was undertaken by the Center for Computer integrated Manufacturing enterprises [1]. It focuses only on analytical resolution of queueing networks and simulation was not considered. In addition, the expressiveness power of the formalism is limited and notably identical to that of the analysis tool RAQS developed at the same laboratory [2], so only concepts supported by this tool are considered.

Our aims behind this work are to explore basic concepts of Queueing theory, extract common terminology and clear semantics for designing a metamodel and an XML-based exchange format. An XML-schema will serve as a mean to define the markup language for describing queueing networks. Designing of an UML metamodel based on considered concepts that will serve as a basis for applying MDE techniques, especially, creating an integrated modeling environment and generating low level code from conceptual models. Taking into account the language extensibility, since this work constitutes only the starting version, other new concepts may be included. Modeling environment and automatic simulation code generation would evolve easily also. The A large acceptance of this formalism facilitates greatly the task of experts and researchers in the domain of systems performance evaluation.

3. Development methodology

Several methodologies for developing Domain Specific Languages (DSLs) have been proposed in [3], [4], [5] and [6]. A study conducted in [7], identified clearly and in detail the development process of DSLs independently of the used tools and languages. We can summarize the different stages of development:

1. Decision: the adoption of an existing DSL is the best solution and the decision to create a new one must be justified. The developing of a new DSL will have an economic impact to be considered (tools used, time spent, effort to develop, learning … etc.).
2. Analysis: the domain is clearly identified and basic concepts are collected. This requires the cooperation of domain experts and computer scientists. The domain analysis is often done informally, but clear methodologies have been proposed and can effectively serve as DARE [8] DSSA [9] and FODA [10].
3. Design: two important elements in this phase are considered, the relationship of the new DSL with an existing language and the degree of formality of the design description. The second element is the design specification. Once the image of the DSL is clear, it can be done informally using natural language for example, easy to produce but more difficult to handle, or formally, using known techniques such as regular expressions and formal grammars to specify syntax and rewrite systems, finite state machines and attribute grammars to specify the semantics.
4. Implementation: the implementation of the design is subject to various choices made about the nature of execution of the DSL. A non-exhaustive list of alternatives includes interpretation, compilation, pre-processor, embedding and extension (as a library of modifications made to the compiler/interpreter for the host language, etc...).
5. Deployment: the effort devoted to the deployment of the DSL and its acceptance is reduced by the success of the previous phases. In addition, the ease of use, adaptation to the domain, efficiency, expressiveness, access to the DSL, involvement of a large community for testing and the impact on productivity are key points to consider.

The first step towards the development of the Extended Queueing Networks Modeling and Markup Language (EQNM^{2}L) has been presented in the previous section where motivations and objectives are discussed. The analysis of the domain knowledge is the critical phase that guides the design and implementation which are discussed in the next sections. For the deployment stage, an online and open source project has been created to

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

involve a large community and easing access to this DSML.

4. Domain knowledge

Most of basic concepts discussed in this paper are based on [11], [12] and [13] where a queueing network is considered as a set of service stations that are visited by jobs. Terms as transaction, customer, client and job refer to the same concept which characterizes any entity that moves through the network, from one station to another to acquire services.

4.1 Job classes and priorities

In queueing networks, jobs could belong to different classes. Job classes may differ in their service and arrival distributions, the number of resource units needed to accomplish services and in their routing schemes leading to different life cycles.

A class independent priority mechanism may be considered to organize the whole population of jobs. Hence, jobs may be scheduled according to their priorities. Within the same priority category, another scheduling discipline could be indicated to distinguish between multiple jobs with equal priorities. It is possible also to consider High Value First (HVF) or Low Value First (LVF). A priority may also be static or dynamic. While a dynamic priority changes its value over time, a static one assigned to a job remains unchanged. A dynamic priority may be time-dependent in the form of *Pr(t)* or be a function of some system parameters. It is also possible that a job changes its class or priority when it moves inside the system.

4.2 Service stations

Service centers, nodes, stations refer to the same concept. Each station is distinguished by a unique identifier and characterized by a number of parallel servers, usually identical, a waiting queue with limited or unlimited capacity, class independent scheduling disciplines, and a service probability distribution for each job class. For a particular class of jobs the service is identified by a probability distribution. Rates are generally static, but it is also possible to apply dynamic rates varying over time, especially in terms of some system indicators.

Several service disciplines are used to serve jobs including FCFS (First Come First Served), LCFS (Last Come First Served), Service in Random Order (SIRO), Processor Sharing (PS), Shortest Job First (SJF), Round Robin, and Shortest Remaining Time First (SRTF). If Round Robin is used, a time slice is indicated. A variant of this discipline called Weighted Round Robin; consider job classes with weighted slices. Other specific disciplines may be considered.

The study of a system is influenced by its initial state which is described by the number of jobs in each station. In contrast to the case where the system is initially empty, realistic cases take into account the exact number of jobs in the system. This is important especially for closed systems where jobs are permanent components.

4.3 Asymmetric stations

If servers are identical, the station is qualified to be simple, ordinary or symmetric. Otherwise, it is said to be asymmetric where parallel servers present different characteristics. We consider here those having different service distributions or various vacation/failure schemes. In real world, the heterogeneity models, for example, machines with different ages or manufacturers. This category of stations requires more information about job affectation in case of multiple idle servers. For example:

1. Random: the task is assigned to a free server at random.
2. Fastest Service: the task is assigned to the fastest free server (the greatest service rate or the smallest average time of service calculated from the log).
3. Longest idle time: the task is assigned to the server which remained idle for the longest period of time from the free servers set.

It is possible to define other allocation strategies. In addition, sometimes the allocation process can't distinguish a single server and the use of another criterion is required.

4.4 Passive stations

If we consider a computer system, then a job (a process or a program) may use two or more resources at the same time, memory space and CPU for example. We consider here passive resources which are simple stations with a number of tokens (identical units) and allocate queues. A passive station has no service to offer except allocating its tokens, so the service duration is generally assumed to be nil. These passive nodes are of two types. The first is devoted to allocation of tokens and the other for release.

When arriving at an allocate station, a job requests a number of tokens. If it gets them, it can continue visiting other nodes of the network; otherwise it must wait at this node. When arriving at the corresponding release node, it releases all its tokens. These tokens are then available for other jobs [11].

4.5 Decision stations

A task, during its life cycle, may have to make decisions based on the state of the system. Since a discrete event

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

system is described by a set of variables known as state variables, a choice structure may be assimilated to a service station in which the task takes a decision. The decision is based on a Boolean expression containing one or more state variables of the system among those predefined or user-defined.

The evaluation of the Boolean expression can have only two possible values. Hence, a decision structure has only two directions or issues. The task is then directed upon the result of the evaluation. This process is not time consuming.

4.6 Service types

Service could be performed in several ways including:

1. Batch processing: the simplest way is to serve only one job at time, but it is possible to do so with a batch. Class dependent batch processing is possible by indicating the minimum and the maximum batch sizes. In this case, the whole batch is treated as one job. It is clear that jobs of a batch are homogeneous and belong to the same class.
2. Pre-emption: Some scheduling disciplines may cause preemption of the processed job when a new job arrives. A preempted job returns to the queue and may resume later starting at the point where it was stopped (preemptive resume) or restarting its service from the beginning (preemption without resume or preemptive repeat). For example, with Priority or LIFO scheduling, preemption could be enabled or disabled. It is also possible to use preemption in case of server failure. Generally, preempted job returns to the front of the queue, but other strategies may be considered.
3. Blocking: it may arise in a station when its queue has a limited capacity. Several blocking models could be distinguished such as:
   a) Rejection: the blocked job is forced to leave the system.
   b) Blocking after service (BAS): the blocked job will be forced to wait in its origin station until the next station is able to accept it. Hence, the origin station will be blocked (still busy) and stops servicing other jobs.
   c) Repetitive service (RS): the blocked job is forced to repeat service by joining an Orbit until the next station is able to accept it. The blocked job may also choose a different destination according to the routing probability.
   d) Waiting queue (WQ): The job that has just been served from a station is rooted to a temporary waiting queue until next station can accept it. This method doesn't block the origin station.
4. Load dependent service: In some situations, service duration may depend on the station load, expressed by the number of jobs inside the station. In this case, for each range of the station load, a service distribution is indicated.
5. Server failure and vacation: In practical queueing systems, service stations are not always reliable. They may become unavailable for a period of time for a variety of reasons (vacation, maintenance, failure ...). Servers failures are assumed to be independent identically distributed random variables specified by a distribution function, but some models requires different types of failure and follow different distributions. Failure may be synchronous (all servers fail at the same time) or asynchronous (servers fail independently). Two probability distributions are useful for modeling the mean time between failures (MTBF) and mean time to repair or recover (MTTR).
   Vacation is slightly different from failure. While the latter is accidental, the former is scheduled and conscientiously undertaken [14]. It is a process governed by a policy that explicitly specifies:
   a) Staring rules: exhaustive when beginning vacation only if queue (system) is empty, otherwise it is non-exhaustive and can begin at any time.
   b) Termination rule: determines when resuming. Multiple vacation policy if keeping vacation until having a job to serve. Single vacation policy if resuming immediately and begin servicing if customers are waiting or stay idle otherwise.
   c) Vacation duration: symmetric (independent and identically distributed) or asymmetric (different distributions).

   For multi-server systems, starting and termination policies are more complicated (synchronous or asynchronous). Other policies are also possible like guarantee of minimum service availability.

4.7 Sources

A service station may or not receive jobs from outside of the system (exogenous arrivals), but must have at least an output job flow towards another station or outside the system.

An input from outside denotes a source of jobs and has a unique identifier. It is associated to an exogenous job flow characterized by a probability distribution, a class and a priority to describe the corresponding arrival process. It is also possible to specify batch arrivals by indicating the distribution of the batch size. Arrival rate for a specific job class is generally static, but we may consider dynamic rates. Some sources have a limited population which is important to consider.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

4.8 Job behavior

In a queuing model, the behavior of jobs is an important feature and three major situations may arise:

1. Balking: Some customers decide not to join the queue because of its length or insufficient waiting space. This behavior results in the discouragement of customers for not joining an improper or inconvenient queue. It could be modeled using decision stations.
2. Reneging: Once in a queue, a job may choose to leave it if it has waited too long ("timeout reneging"). It requests a resource and wants to leave before acquiring it, for example because it could get quicker service by another resource with a shorter queue ("conditional reneging"). Reneging pertains to impatient customers. After being in queue for some time, few customers become impatient and may leave the queue.
3. Jockeying: Jockeying is a phenomenon, in which customers move from one queue to another queue with a hope that they will receive quicker service in the new position.

Job's behavior is station dependent and is more suited to be indicated for each station.

4.9 Orbits

In many situations, jobs leaving a service station join a region called "Orbit". After a delay they retry their queries. So, an orbit is a source of jobs fed by the system itself and allows its jobs to retry queries after a random amount of time. In queueing theory, this is a key feature for retrial queues [15].

4.10 Outputs

Each station may have one or more outputs. A routing strategy must be defined for each job class. An output is identified by a reference to the destination station and routing information according to the routing strategy.

It is possible that a job changes its class after a service. This situation is common in manufacturing systems where a given machine transforms received products to new ones. An output normally needs to indicate the associated resulting job class for every received one.

4.11 Routing strategy

After service completion, a job is routed according to a routing strategy which may be for example:

1. Probabilistic: job moves to the next station with a given probability.
2. Round robin: all possible destinations for a job class are chosen in a cyclic manner.
3. Shortest queue length: job moves to the next station having the shortest queue.
4. Smallest response time: job moves to the next station having the smallest response time calculated from the log.
5. Smallest utilization: job moves to the next station having the smallest average utilization calculated from the log.
6. Fastest service: job moves to the next station having the smallest average service time calculated from the log.

4.12 Regions

The concept of regions is very powerful and denotes a set of stations in the system with controlled capacity. It allows expressing global constraints on a group of stations as finite capacity for specific job classes. It is also useful in global statistics computation. Hence, a *Finite Capacity Region* (FCR) is a region of the model or a set of stations, where the number of jobs is under control [16]. It is possible to define an FCR capacity by setting upper bounds for the number of customers in the region for specific classes or globally regardless of the job classes.

5. Metamodel

Models are created using a modeling language called a metamodel. The modeling language is itself is described in another language called meta-metamodel. The philosophy of DSM favors the creation of a new language for a specific task, and hence there are naturally new languages designed as meta-metamodels. DSM environments can considerably reduce the cost of obtaining tool support for a DSML, since a well-designed DSM environment will automate the creation of application modules costly to build from scratch, such as domain-specific editors, compilers and models transformers. The domain experts only need to specify the domain specific constructs and rules, and the DSM environment provides a modeling tool tailored for the target domain.

It has been established that a language is characterized by its syntax and semantics. The syntax describes the different language constructs and their arrangement rules while semantics refers to the relationship between a signifier (program or model) and a signified (mathematical object) to give meaning to each of the constructs of the language. So the relationship between semantics and syntax is the same as between the content and the form.

The metamodel represents the abstract syntax and is the heart of the language, since it captures the whole domain concepts and their relationships. It acts as a pivot between the concrete syntax description and the semantics

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

description. Concrete syntax contains information on how the concepts in the metamodel are to be represented to the language user. This is sometimes amended by a mapping in which each abstract concept is related to a textual notation or a graphical symbol.

The EQNM^{2}L metamodel illustrated in Figure 1 is expressed in UML class diagram. It represents the abstract syntax and a part of the static semantics. The rest of the static semantics could be expressed using OCL.

An extended queueing network (EQNET) is an aggregate of nodes, decision stations, sources, orbits, outputs, outsides, sink and finite capacity regions. It may contain a description which consists of a set of meta-information about the model as name, used tool, version and author. A node is an abstract concept that models a variety of specific stations as simple, passive and asymmetric stations. In contrast of an open network, a closed network does not need a sink which models the system's environment.

It is important to note that some domain concepts discussed in previous sections are not considered in this minimal metamodel. The metamodel in its actual state lacks of many concepts but is extensible to include new concepts in order to gain more maturity.

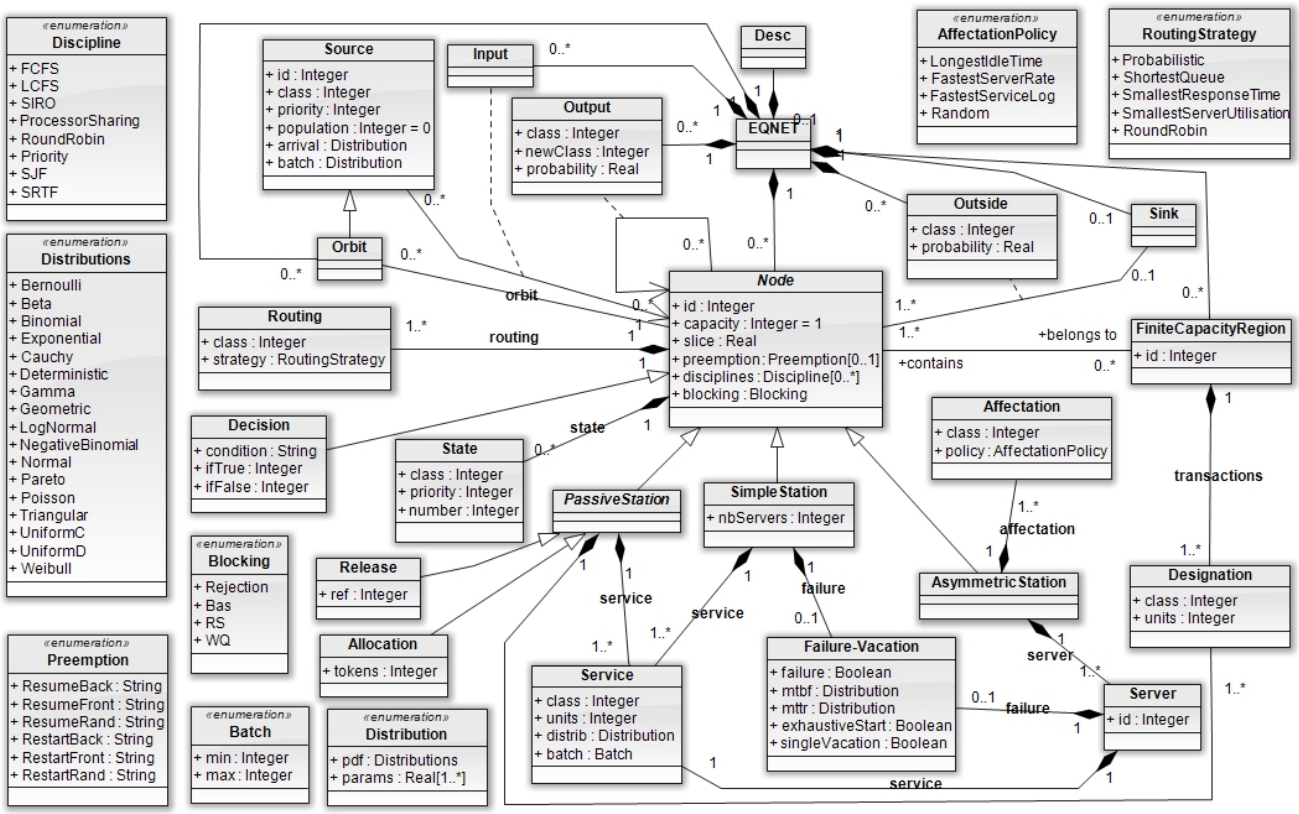


Fig.1 EQNM^{2}L Metamodel.

The static semantics is part of the metamodel. It consists of a set of rules to ensure that it is well formed (well-formedness rules). Rules allow expressing constraints and thus reducing the overall valid models. We present these constraints in natural language while trying to be concise and clear. Subsequently, these rules are expressed in OCL during the implementation phase in respect to the chosen implementation tool. In addition to the cardinality constraints already present in the class diagram, the other rules are:

1. All identifiers are strictly positive integers which include nodes, sources, orbits and finite capacity regions (*id > 0*).
2. The identifier of a node is unique in the model.
3. A source of jobs directs its job flow to only one node.
4. A job flow of the same class and priority can't be directed more than once from one node to another, to the outside or to an Orbit. It means that identical job flows from the same node could not be directed to the same target (Node, Orbit or Outside).

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

5. For a job class, the network of queues can be either open or closed.
6. Sources and orbits of the same station must have distinct identifiers.
7. A probability is a real number comprised between 0 and 1 ($0 \leq pr \leq 1$).
8. If the routing strategy in a given station for a specific job class is probabilistic, then the sum of probabilities is equal to one ($\sum pr = 1$).
9. Job classes are positive integers (for Routing, Designation, Output and Outside), but for a Source or an Orbit, it is strictly positive (the class number 0 is reserved to represent any class not explicitly mentioned).
10. Only one Sink is authorized in a model (for open networks) for which at least one source must exist (no sink without sources).
11. An Orbit manipulates only one job class, so it must receive only this class from the corresponding node to which, as a Source, it directs jobs.
12. Finite Capacity Regions have distinct identifiers (identifiers are unique).
13. A node may belong to different finite capacity regions and a finite capacity region may contain different nodes.
14. The population attribute of a source of jobs is a positive integer (population $\geq$ 0). Zero means unlimited.

6. Concrete syntax and exchange format

Some DSL developers consider that in modern visual (or graphical) language environments there is no need to be very specific about concrete syntax. At the low level, information can be exchanged between different tools in a textual concrete syntax using XML, and at the high level, humans can input linguistic utterances into a tool using graphical components. In consequence, the only thing that remains is the need for renderings of models which are meaningful to humans. However, because a language is a means of communication that must preserve interoperability, all tools developers need to agree on the XML schema for interchange as well as on the symbols to be used in rendering. In addition, both the XML schema and the set of symbols for rendering are considered concrete syntaxes of the language. If no agreement on concrete syntaxes is done, lack of interoperability and confusion would arise. It is clear that a mapping from abstract syntax to a concrete syntax is as important as the inverse mapping.

It is possible to define several concrete syntaxes for the same abstract syntax. We adopted the derivation illustrated in Figure 2, matching each element of the abstract syntax with a visual (graphic) appearance. It is important to provide visual elements as distinct (or similar) as their abstractions. The visual syntax must be as close as possible to the domain graphical notation. This will facilitate the task of domain experts, ensure large deployment and encourage acceptance of the DSML.

Abstract syntax	Visual (concrete) syntax
Simple Station	
Asymmetric Station	
Allocation Station	
Release Station	
Source	
Sink	
Finite Capacity Region	
Decision	
Orbit	
Routing	
State	
Server	
Input/Output/Outside	
Distribution	
Service	
Affectation	
Designation	
Failure-Vacation	

Fig. 2 Mapping between abstract and visual concrete syntax.

For simple languages and in most of the current DSM tools, concrete syntax representation is directly mapped to the abstract syntax. Logical entities are always visualized as nodes, and logical relationships as edges. Nevertheless the experience has shown that using more complex metamodels, especially those conceived for automated model transformations, not only results in visual models being too complicated to overview, but it can also drain system resources heavily. Consequently, modern approaches make use of a separate visualization metamodels, which describe the structural appearance of diagrams. The last technique allows hiding superfluous details; however it is still limited in the sense that classes can only be mapped to nodes and references to edges.

The problem with EQNM^{2}L formalism is the exchange and reuse of developed models between various simulation and analysis tools. It is necessary to refer to an exchange format which must be open and promotes maximum interoperability.

Various textual exchange formats exist and most of them are based on XML. For example, QNAT software [17] uses the Mathematica format and RAQS software [2] uses a specific ASCII format. For JMT [16], it uses XML as a mean to describe models. The syntax is specified by the *JMTmodel* schema which serves as the unique format

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

used by the simulation and the analytical engines. The main problem here is not only the multiple used formats, but even if XML is used, the solution suffers from the lack of expressiveness and agreement on a standard format. The exchange format for EQNM^{2}L [18] is a proposed draft which may serve as a basis for developing more adequate, stable and mature templates. It represents a metamodel described as an XML-schema. Each element of the metamodel is projected as an XML element. For the inheritance concept, XML provides restrictions and extensions that are similar and useful. In addition, some constraints could be expressed easily, especially those of uniqueness.

7. Implementation

One of the advantages of using the MDE approach is the gain in productivity where the task of developing tooling for domain-specific languages in a cost-effective manner is possible.

The modeling environment must offer at least visual syntax-aware editor to assist the user in model construction. Actually, several tools support domain metamodeling and domain application models, such as MetaEdit+ from the MetaCase Company [19], GME from ISIS laboratory of Vanderbilt University [20] , Microsoft DSL Tools [5], AToM3 [21] as well as Eclipse EMF/Ecore [22] and Eclipse GMF [23].

Eclipse GMF and GME are appropriate tools to implement the EQNM^{2}L modeling environment. Main advantages are that both are open source, freeware, mature, well documented and offer a sufficient set of constructs to define most aspects of the DSML including model validation and transformation. Actually, our project is based mainly on Eclipse platform, together with its EMF and the GEF plugins. The latter is a basic diagram drawing engine. In addition, the static metamodel mapping-driven GMF platform is the solution for linking the two previous plugins and complete tool building environment.

The domain model illustrated in Figure 3 covers only a subset of EQNM^{2}L. OCL constraints are inserted in the EMF/Ecore model as shown in Figure 5 to allow model validation.

The Eclipse DSL Toolkit is based mainly on EMFs capabilities, including diagram definitions, transformation definitions and code-generation templates, in addition to model serialization and persistence. Many of these capabilities are developed using EMF models. For domain-specific modeling surfaces generation, Graphical Modeling Framework (GMF) is based on a collection of EMF models that are considered as DSLs themselves. GMF allows providing a graphical concrete syntax where the proposed graphical notation is mapped to the abstract syntax.

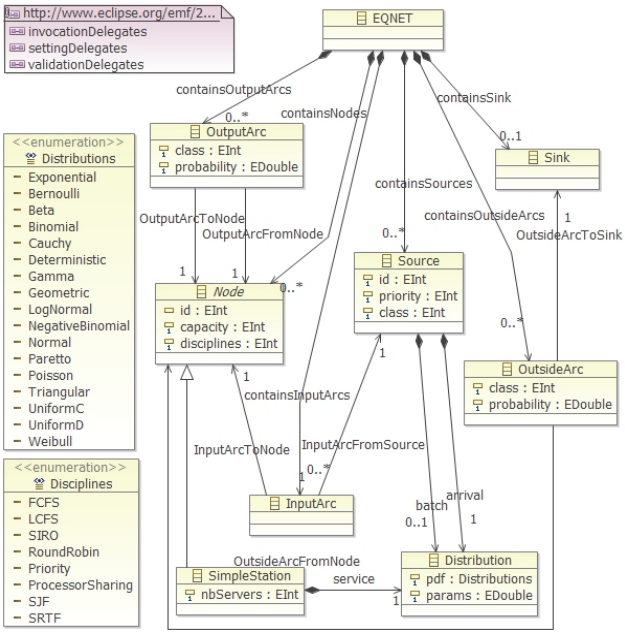


Fig. 3 Domain model as an Ecore diagram.

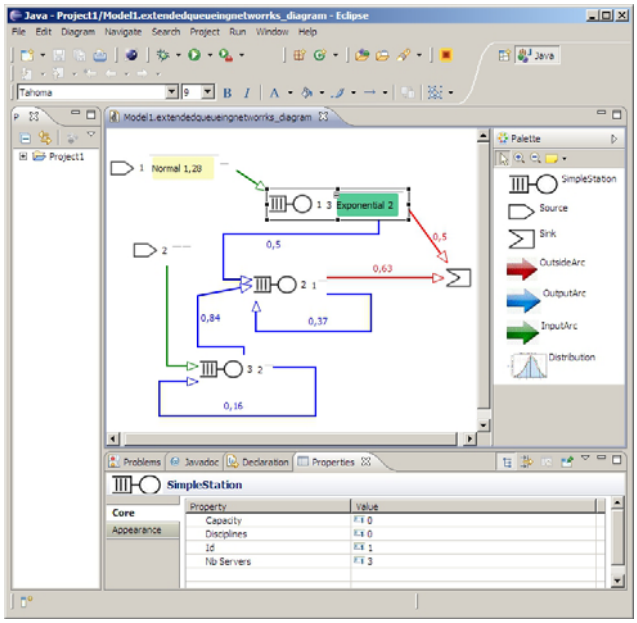


Fig. 4 EQNM^{2}L diagram editor in Eclipse.

The EQNM^{2}L diagram editor illustrated in Figure 4, offers a complete environment for graphical editing and model validation. The project is an Eclipse plugin and could be integrated easily.

Instead of coding model transformation rules, for code generation, directly in the programming language of

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

the application, as most software does, it is more practical to separate them and to lean on hard background of a formal theory.

```
36  class SimpleStation extends Node
37  {
38      invariant positiveNbServers: self.nbServers >= 0;
39      attribute nbServers : ecore_0::EInt[1] = '3';
40      property service : Distribution[1] { composes };
41  }
42  class Source
43  {
44      invariant positivePriorities: self.priority > 0;
45      invariant UniqueSourceIDs:
46      Source.allInstances()->forAll(s1, s2 | s1 <> s2 implies s1._'id' <> s2._'id');
47      invariant positiveClasses: self._'class' > 0;
48      invariant positiveIDs: self._'id' > 0;
49      attribute _'id' : ecore_0::EInt[1] = '1';
50      attribute priority : ecore_0::EInt[1] = '1';
51      attribute _'class' : ecore_0::EInt[1] = '1';
52      property batch : Distribution[?] { composes };
53      property arrival : Distribution[1] { composes };
54  }
55  class OutsideArc
56  {
57      invariant positiveClasses: self._'class' > 0;
58      invariant positiveProbabilities: self.probability >= 0 and probability <= 1;
59      attribute _'class' : ecore_0::EInt[1] = '1';
60      attribute probability : ecore_0::EDouble[1] = '0.25';
61      property OutsideArcToSink : Sink[1];
62      property OutsideArcFromNode : Node[1];
63  }
```

Fig. 5 EQNM^{2}L OCL constraints.

Dynamic semantics of a DSML may be specified by the various model transformations. Operational semantics is well suited for directly executable or simulated models.

For EQNM^{2}L, a model is either translated into a simulation/programming language to be executed or into an analysis tool specific format in order to be resolved. In the first case, the semantics domain is changed and transformation rules specify a denotational semantics.

The code generation process is a model transformation operation. More accurately, it is considered as model-to-text (M2T) transformation. The Xpand template language is an increasingly popular template engine, used extensively by the GMF project. It is originated with the OpenArchitectureWare component within GMT but has graduated to the M2T project.

Transformation rules are established according to the source and the target metamodels. EQNM^{2}L models could be transformed into discrete event simulation code, expressed in a general purpose programming language (GPL) or a simulation language.

Several discrete event simulation libraries have been developed to offer a mean to assist programmers in writing customized and efficient simulation code. In the actual state of the work, a Java simulation code is generated automatically from EQNM^{2}L conceptual models using Xpand templates illustrated in Figure 6. The Japrosim simulation library [24] has been chosen for code generation.

```
1  «IMPORT ExtendedQueueingNetworrks»
2
3  «EXTENSION template::GeneratorExtensions»
4
5  «DEFINE main FOR EQNET»
6  «FILE "QueueNetwork.java"»
7  import uoeb.japrosim.kernel.SimProcess;
8
9          public class QueueNetwork {
10             public static void main(String[] args){
11                 SimProcess.sched.start();
12                 «FOREACH containsSources AS s»
13 «FOREACH this.containsInputArcs.typeSelect(InputArc) AS i»
14 «IF i.InputArcFromSource.id == s.id»
15 «LET i.InputArcToNode.id AS iid»
16                 new Transaction«s.id»_«iid»().beginAfter(0.0);
17 «ENDLET»«ENDIF»«ENDFOREACH»
18 «ENDFOREACH»
19             }
20         }
21 «ENDFILE»
```

Fig. 6 Xpand-template For QueueNetwork class.

In order to facilitate understanding these rules, it is necessary to have an idea of how Japrosim works. For each model, a class named "*QueueNetwork*" is needed for system initialization by creating first events and then launching the simulation. The corresponding Xpand template is illustrated in Figure 6. Another class named "*Transaction*" which must extend the predefined "Entity" class is required. It is the place for jobs (or transactions) to declare needed shared resources as arrivals, services, routing schemes, stations, queues and shared behavior. Template in Figure 8 allows generating the previous Java class. In addition, each source of jobs requires its own Java class, named "*Transactioni_j*" that implements their specific behavior in the system. These Java classes are generated using the template shown in Figure 7.

```
68 «DEFINE javaClass FOR Source»
69 «FOREACH ((EQNET)this.eContainer).containsInputArcs.typeSelect(InputArc) AS i»
70 «IF i.InputArcFromSource.id == this.id»
71 «LET ((SimpleStation)i.InputArcToNode).id AS iid»
72
73 «FILE "Transaction"+id+"_"+iid+".java"»
74         public class Transaction«id»_«iid» extends Transaction {
75             public Transaction«id»_«iid»() {
76                 super();
77             }
78
79             public void body() {
80                 new Transaction«id»_«iid»().beginAfter(arrival«id».sample());
81                 intoStation(
82 «FOREACH ((EQNET)this.eContainer).containsInputArcs.typeSelect(InputArc) AS i»
83 «IF i.InputArcFromSource.id == this.id»
84 «((SimpleStation)i.InputArcToNode).id»);«ENDIF»«ENDFOREACH»
85             }
86         }
87 «ENDFILE»«ENDLET»«ENDIF»«ENDFOREACH»
88 «ENDDEFINE»
```

Fig. 7 Xpand-template For Transactioni_j classes.

The same work has been done in [25] based on the exchange format. The source model is expressed in XML which allows using XSLT for expressing transformation rules and helps maintaining and reusing the translation software for future versions or other simulation languages.

XSLT is a powerful XML-dialect for manipulating data in XML documents. It provides a set of operations and manipulators, while XPath provides precision in

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

locating elements and attributes. So in order to translate the conceptual model into an executable simulation model, an XSLT stylesheet and an XSLT processor are used.

The XML model description is transformed into the corresponding set of Java classes according to the Japrosim library as discussed previously. The code generation module is extensible for other simulation languages or analysis tool format easily by associating the corresponding XSLT stylesheets. The user will be free to choose his target tool.

```
23 «EXPAND javaClass FOREACH sources()»
24 «FILE "Transaction.java"»
25 import uoeb.japrosim.kernel.*;
26 import uoeb.japrosim.random.distributions.*;
27         public class Transaction extends Entity {
28 «FOREACH containsSources AS s»
29           static «s.arrival.pdf» arrival«s.id» = new «s.arrival.pdf»(«s.arrival.params.first()»
30           «FOREACH s.arrival.params.withoutFirst().toList() AS p»,«p»«ENDFOREACH»);«ENDFOREACH»
31 «FOREACH containsNodes.typeSelect(SimpleStation) AS ss»
32           static «ss.service.pdf» serv«ss.id» = new «ss.service.pdf»(«ss.service.params.first()»
33           «FOREACH ss.service.params.withoutFirst().toList() AS r»,«r»«ENDFOREACH»);
34           static Queue queue«ss.id» = new Queue("Queue 0«ss.id»");
35           static Resource server«ss.id» = new Resource("Station«ss.id»", «ss.nbServers»);
36           static int destination«ss.id»;
37           static Discrete outputs«ss.id» = new Discrete(new double[]{
38 «FOREACH ((EQNET)(ss.eContainer)).containsOutputArcs.typeSelect(OutputArc) AS o»
39 «IF ss.id == o.OutputArcFromNode.id»«o.OutputArcToNode.id», «ENDIF»«ENDFOREACH»0}, new double[]{
40                 «FOREACH ((EQNET)(ss.eContainer)).containsOutputArcs.typeSelect(OutputArc) AS o»
41 «IF ss.id == o.OutputArcFromNode.id»«o.probability», «ENDIF»«ENDFOREACH»
42 «FOREACH ((EQNET)(ss.eContainer)).containsOutsideArcs.typeSelect(OutsideArc) AS o»
43 «IF ss.id == o.OutsideArcFromNode.id»«o.probability»«ENDIF»«ENDFOREACH»});«ENDFOREACH»
44             public Transaction() {
45                 super();
46             }
47             public void intoStation(int i){
48 «FOREACH containsNodes.typeSelect(SimpleStation) AS ss»
49                 if(i == «ss.id»){
50                     queue«ss.id».insert(this);
51                     while (server«ss.id».getAvailability() < 1) {
52                         passivate();
53                     }
54                     seize(server«ss.id», 1);
55                     queue«ss.id».remove(this);
56                     hold(serv«ss.id».sample());
57                     release(server«ss.id», 1);
58                     int c = (int)outputs«ss.id».sample();
59                     if(c!=0){
60                         intoStation(c);
61                     }
62                 }«ENDFOREACH»
63             }
64         }
65 «ENDFILE»
```

Fig. 8 Xpand-template For Transaction class.

8. Conclusions

We presented EQNM^{2}L, an extended queueing networks modeling and markup language where models are simply queueing networks with an extension for new concepts that help modeling more accurately a wider range of discrete event systems. Obtained models may be suited to analytical solutions, approximations, or simulation techniques.

As a DSML, it is tool-independent, extensible and well-established if we consider that it constitutes only a starting version. Its main advantages include interoperability enhancement at the conceptual modeling level throw the proposed metamodel and the XML-based exchange format. A modeling environment as an Eclipse plugin is created guided by the proposed metamodel and the adopted visual syntax. Based on the MDE concepts, automation allows reducing the development time and effort and makes validation easier.

It is possible to define a common visual (concrete) syntax, but even if several rendering schemes are adopted by different tools, the metamodel and the exchange format remain unchanged.

The metamodel presented in this paper is a minimal version and several concepts are not yet considered. It could be extended easily, whenever a new concept has a clear semantics. In our opinion, it is much easier to conclude an agreement on a DSML than a specific tool. Therefore, EQNM^{2}L modeling environment could be developed using any other tool that support domain metamodeling.

Domain experts must be involved in order to improve different aspects of this DSML. This includes more complete metamodel, more concise semantics and more adequate visual or concrete syntax. Furthermore the proposed exchange format must also reflect those efforts.

To ensure acceptability and large deployment, in addition to the adhering to the open source initiative to imply a large community, several code and model generators for the most used tools are to be considered.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Future work will focus also on the expressiveness improvement by including new concepts.

References

[1] U. M. Chalavadi, "Automatic configuration of queueing network models from business process descriptions". M. S. thesis. Oklahoma State University, USA, 2004.

[2] M. Kamath, "Recent developments in modeling and performance analysis tools for manufacturing systems", in Computer Control of Flexible Manufacturing Systems, 1994, pp. 231-263.

[3] P. Hudak, "Modular domain specific languages and tools", in the Fifth International Conference on Software Reuse (JCSR98), 1998, pp. 134-142.

[4] D. Spinellis, "Notable design patterns for domain-specific languages", The Journal of Systems and Software, Vol. 56, 2001, pp. 91-99.

[5] S. Kelly, and J. P. Tolvanen, Domain-Specific Modeling: Enabling Full Code Generation. John Wiley and Sons, Hoboken, New Jersey, 2008.

[6] S. Cook, G. Jones, S. Kent, and A. C. Wills, Domain-Specific Development with Visual Studio DSLTools. Addison-Wesley, 2007.

[7] M. Mernik, J. Heering, and A. M. Sloane, "When and how to develop domain-specific languages". Report Software Engineering SEN-E0309, Stichting Centrumvoor Wiskundeen Informatica (CWI), Amsterdam, the Netherlands 2003.

[8] W. Frakes, R. Prieto-Diaz, and C. Fox, "DARE: Domain analysis and reuse environment". Annals of Software Engineering, 1998, Vol. 5, pp. 125-141.

[9] R. N. Taylor, W. Tracz, and L. Coglianese, "Software development using domain-specific software architectures", in ACM SIGSOFT Software Engineering Notes, 1995, Vol. 20, N° 5, pp. 27-37.

[10] K. C. Kang, S. G. Cohen, J. A. Hess, W .E. Novak, and A. S. Peterson, "Feature-oriented domain analysis (FODA) feasibility study". Technical Report CMU/SEI-90-TR-21, Software Engineering Institute, Carnegie Mellon University 1990.

[11] G. Bolsh, S. Greiner, H. de Meer, and K. S. Trivedi, Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications, Second Edition. John Wiley & Sons, 2006.

[12] E. D. Lazowska, J. Zahorjan, G. S. Graham, and K. C. Sevcik, Quantitative System Performance: Computer System Analysis Using Queueing Network Model. Prentice-Hall Inc, Englewood Cliffs: New Jersey, 1984.

[13] R. B. Cooper, Introduction to Queueing Theory. Elsevier North Holland, Inc. 1981.

[14] N. Tian, and Z. G. Zhang, Vacation Queueing Models: Theory and Applications. International Series in Operations Research & Management Science, Frederick S. Hillier, Series Editor, Stanford University Springer Science + Business Media, LLC. 2006.

[15] R. Artalejo Jess, and A. Gmez-Corral, Retrial Queueing Systems: A Computational Approach. Springer-Verlag: Berlin Heidelberg. 2008.

[16] M. Bertoli, G. Casale, and G. Serazzi, "JMT: performance engineering tools for system modeling", ACM SIGMETRICS Performance Evaluation Review, 2009, Vol. 36, Issue 4, New York, USA, March. ACM press, pp. 10-15.

[17] D. Manjunath, D. M. Bhaskar, H. Tahilramani, S. K. Bose, and M. N. Umesh, "QNAT: A Graphical Tool for the Analysis of Queueing Networks", in TENCON '98. 1998 IEEE Region 10 International Conference on Global Connectivity in Energy, Computer, Communication and Control, 1998, Vol. 2, pp. 320-323.

[18] EQNM^{2}L: http://sourceforge.net/projects/eqnm2l/. Accessed on November 15th 2011.

[19] M. Rossi, and S. Kelly, "Construction of a CASE tool: the case for MetaEdit+", in Proceedings of The First International Symposium on Constructing Software Engineering Tools (CoSET'99), 17-18 May 1999, Los Angeles.

[20] A. Ledeczi, M. Maroti, A. Bakay, G. Karsai, J. Garrett, C. T. IV, G. Nordstrom, J. Sprinkle, and P. Volgyesi. "The Generic Modeling Environment", In Workshop on Intelligent Signal Processing, Budapest, Hungary, May 17, 2001.

[21] J. de Lara, H. Vangheluwe, "AToM3 as a Meta-CASE environment (DFD to SC)", in 4th International Conference On Enterprise Information Systems. Universidad de Castilla-La Mancha-Ciudad Real -Spain- 3-6 April, 2002.

[22] F. Budinsky, D. Steinberg, Ed.Merks, R. Ellersick, and T. J. Grose, Eclipse Modeling Framework: A Developer's Guide. AddisonWesley, 2003.

[23] Gronback. Richard C. Eclipse Modeling Project: A Domain-Specific Language Toolkit. Addison-Wesley Professional, 1st ed, 2009.

[24] A. Bourouis, and B. Belattar. "JAPROSIM: A Java Framework for Discrete Event Simulation". Journal of Object Technology, Vol. 7, No. 1, January-February 2008, pp. 103-119.

[25] A. Bourouis, and B. Belattar, "Using XML in Simulation Modelling: automatic code generation for XML-based models", in Proceedings of the CARI 08, Tanger, Morocco, October 23-25, 2008, pp 101-108.

Abdelhabib BOUROUIS received his BS degree in Computer science from the University of Constantine (Algeria) in 1999, his MS degree and Ph.D. degree respectively in 2003 and 2011 both from the University of Batna (Algeria). He is a Senior Lecturer at the University of Oum El Bouaghi (Algeria) since 2003. His research interests include Artificial intelligence, Model Driven Engineering, performance evaluation, parallel and distributed simulation.

Brahim BELATTAR received his BS degree in Computer science from the University of Constantine (Algeria) in 1981 and his MS and PhD degrees from the University Claude Bernard of Lyon (French) respectively in 1986 and 1991. He is an Associate professor at the University of Batna since 1992. He has also taught at the University of Constantine from 1982 to 1985. His research interests include simulation, databases, semantic web, Model Driven Engineering and Artificial intelligence.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Evaluation of coordination contracts for a two stage Supply Chain under price dependent demand

Noureddine Labiad[1], Zitouni Beidouri[1], Otmane Bouksour[1]

[1] Laboratoire de Mécanique Productique & Génie Industriel (LMPGI) ; Université Hassan II Ain Chock, Ecole Supérieure de Technologie, Km 7 Route El Jadida, Casablanca, Maroc.

Abstract

Supply chains are characterized by many activities and actors that generally pursue conflicting objectives. Coordination between them may be then necessary to align the individual objectives with the global supply chain objective and achieve optimal performance.

In this paper, we propose a model that aims to assess the relative performance of three well known Coordination Contracts for a two level Supply Chain under price dependent demand. It is shown that a suitable design of the contracts could secure global system efficiency and improve the profit of all the Supply Chain actors.

Keywords: *Supply Chain Management, Coordination, Contracts, Revenue Sharing, Buyback Contracts, Quantity discount*

1. Introduction

A Supply Chain system is usually comprised of many organizations that are often separate and independent economic entities. Although centralized decisions may lead to optimal global Supply Chain performance, these decisions are not always in the best interest of every individual member of the Supply Chain. Independent Supply Chain members are usually more keen in optimizing their individual objectives rather than that of the entire system. This could result in a poor global performance for the Supply Chain as a whole [1] (Whang, 1995).

Thus a key issue in SCM is to develop mechanisms that motivate the independent actors to achieve coordination and optimize Supply Chain Global Performance.

Research has initially focused on centralized Supply Chain. In this setting, the Supply Chain control is assured by a single decision-maker who is given all the contractual powers. In practice, the Supply Chain involves many independent decision makers who pursue generally different objectives.

Coordination contracts are among the main tools that actors of a decentralized Supply Chain can use to reach coherent decisions among them. [2] (Ilaria Giannoccaro*, Pierpaolo Pontrandolfo (2004)).

Coordination contracts have been widely used in practice to increase overall Supply Chain performance and to share risk and rewards between Supply Chain actors.

This paper focuses on evaluating relative performance of different SC contracts. These contracts are used to coordinate a two-stage SC facing a stochastic price dependent customer demand. The issue of the desirability of the contract to the SC actors is also considered.

The paper is organized as follows. In Section 2, a review of the SC contracts is given. We then propose a reference model for evaluating the performance of the different SC contracts in Section 3. In section 4, SC Contracts performance and desirability are discussed through a numerical application.

2. Literature Review

Many forms of Supply Chain contracts have been implemented in industries and analyzed or studied by researchers. Most models are based on the Single Period Newsvendor Problem where coordination is reached through setting the optimal order quantity that maximizes the overall profit of the Supply Chain.

Many extensions to this model have been explored by researcher and usually consider a subset of the following classes of parameters:

- *Decision-Making Process :* Two aspects are mainly considered in literature :
  - Distribution of decision among Supply Chain actors (ex: Centralized vs Decentralized SC)
  - Compliance regime: This issue is related to the right of the SC actors to comply partially or totally with coordination contract terms (Ex: Under a forced compliance regime, the supplier must fulfill totally his customer order.) [3] Cachon, G.P. and Lariviere, M. A., 2001
- *Supply Chain structure:* (Number of tiers and SC network nodes). Most contracts models addressed in literature are two-tier SC (bilateral monopoly) composed of a single supplier and a single retailer serving a final demand. Some research papers have addressed Coordination issues of three level Supply Chain with single actors at each level [4] (Ding Ding, Jian Chen 2008). Some others studied the coordination problem in a context of a Two

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

level Supply Chain with single vendor and multiple retailer [5] (Darwish & Odah 2010)

- *Certainty/Uncertainty of demand:* The uncertainty of Supply Chain environment refers generally to market demand. The two broad categories are deterministic and probabilistic market demands.
- *Environment dependence of Supply Chain decision:* Market demand is usually sensitive to product selling price and marketing efforts. In this case, Coordination between Supply Chain actors is considered with regard to these internal Supply Chain parameters that impact the market demand. He and all studied Coordination of a two tiers two node Supply Chain in a context of effort and price dependent demand. [6] He, Y., Zhao, X., Zhao, L., and He, J., 2009. [7] Petruzzi and Dada (1999) examine an extension of the newsvendor problem in which stocking quantity and selling price are set simultaneously in context of price dependent demand.
- *Contract period:* The contract period is the duration of time that the contracting agents are assumed to uphold the contract. Distinction can be made between single and multiple replenishment periods.
- *Risk Aversion:* The Supply Chain agent can be risk neutral or show an aversion to risk. [8] Xianghua & all (2004) explore the coordination of SCM with risk averse agents.
- *Information structure:* It pertains to the agents 'knowledge in comparison to the collective knowledge of agents in the Supply Chain. When all the information about Supply Chain is simultaneously known by every agent, the information structure is said to be complete or symmetric. On the other hand, if some agents have some information that the other agents do not, the information structure is incomplete or asymmetric.

In a recent publication, Behzad and all (2010) [15] provide a detailed overview of coordinating contract in literature and present the state of art research in this field.

Two broad classes of coordination contracts have been identified in literature: Quantity dependent Contracts (Quantity Discounts, Quantity flexibility) and Price dependent Contracts (Wholesale price, Buyback or Return policies, Revenue Sharing, Sales rebate, Quantity Discount). The scope of this paper is restricted to three of the well known contracts:

- **Buy Back Contract:** Under this type of contracts, a manufacturer, or an upstream distributor sets the wholesale price and commit to refund a downstream channel member for excess inventory return at the end of the season. The most generous policy promises to refund the full wholesale price for all returned products. [9] (Pasternack, B. A. (1985)) shows that a policy allowing for unlimited returns at partial credit will be system optimal for appropriately chosen wholesale price and refund value. [10] Cachon 2003 shows also that partial return and partial compensation can achieve coordination. [11] Padmanabhan, V., & Png, conclude that, in a context of a single upstream and many downstream actors and a price dependent demand, Return policies is beneficial to the upstream actor when production costs are low and demand is not highly uncertain and competition intensifies at the downstream level.
- **Revenue Sharing:** Under this agreement, the downstream agent (Distributor, retailer) commits to share its revenue with the upstream agent (manufacturer, Distributor) in exchange of a smaller wholesale price. The successful application of this type of contracts has been reported in the American Video rental industry [12] (Cachon and Lariviere, 2005). Cachon and lariviere show that in the case of a single distributor and a single retailer, Revenue Sharing Contracts coordinate the Supply Chain and can allocate profits arbitrarily between actors. [2] Ilaria Giannoccaro*, Pierpaolo Pontrandolfo developed a Revenue Sharing mechanism that coordinate efficiently a three stage Supply Chain and improves profits of the stakeholders.
- **Quantity Discount:** Distinction can be made between incremental and All unit discount. The Quantity Discount is a coordination mechanism that consists of offering a price discount to the buyer so that he orders the global optimal quantity, and improve the profit of both supplier and buyer. This contract has mainly been studied in contexts of deterministic demands. [13] (Jianli Li and Liwen 2006) explore how to use an all unit discount coordination mechanism to achieve coordination within a multi-period supplier–buyer system selling one type of product and facing a probabilistic customer demand. [10] Cachon 2003 shed the light on the coordinating ability of Quantity Discount contract under the assumptions of a single period newsvendor problem.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

- **Quantity flexibility:** This coordination mechanism allows downstream actor to change his order quantity within a predefined range as he gains more insight of market demand. On the other side the upstream commits to guaranty product availability within the same range. [14] Tsay 1999 shows that under certain conditions Quantity flexibility contracts can be considered to share risk of market uncertainty between Supply Chain actors and achieve system wide optimal outcome.

3. Reference Model

Assumptions and notations

In this model, we consider a two stage Supply Chain composed of an upstream and a downstream firm. We will refer to the upstream firm as the manufacturer and the downstream firm as the retailer

There is one selling season with stochastic demand and a single opportunity for the retailer to order inventory from the manufacturer before the selling season begins.

The manufacturer produces a single short life cycle product at marginal cost c_m and sells it to a Retailer at a wholesale price w.

The Retailer decides the retail price p and order quantity q. He also incurs marginal variable cost c_r. In the event of stock out, unmet demand is lost, resulting in the margin being lost to the retailer, but without additional penalty (such as a loss of good-will). The retailer can return full-unsold inventory back to the supplier. All The Unsold items are salvaged at price **s** by the retailer at the end of the season. The manufacturer and retailer are both in a monopoly position to serve the end customer. The firms are assumed risk neutral and thus each firm pursue profit maximization.

Information is symmetric for all parties i.e. all the information regarding demand and cost parameters is common knowledge to both parties.

When a contract is accepted, the manufacturer must fulfill the order placed by the retailer.

At the end of the selling season, the market demand is realized, which is stochastic, downward sloping in the retail price. We assume market demand for the product D is stochastic price-dependent as $D(p)= d(p) + \varepsilon$, where $d(p)$ is a deterministic decreasing function of p that captures the dependency between demand and price. $d(p)$ is a linear curve in p and is represented by the following function : $d(p) = a - bp \ (a > 0, b > 0)$. ε is a random variable with known distribution that shows randomness in demand and is price-independent, we assume that ε follows Normal distribution with mean μ and standard deviation σ, i.e., $\varepsilon \sim N(\mu, \sigma^2)$ and that the variance of σ^2 is small relative to a . Let f be the probability density function of ε and F its cumulative distribution function where f is strictly positive.The expected value of demand for a given price p is $d(p)$.

It is also assumed that the retailer sets his price at the same time as his stocking decision and the price is fixed throughout the season.

All these assumptions are moderately reasonable and standard in the literature.

Notations:

c_m Marginal Cost of the manufacturer
c_r Marginal Cost of the retailer
c Total Marginal Cost ($c_u + c_d$)
p Unit selling price
w Unit wholesale price charged by supplier
s Unit salvage value at the downstream side
Q_r^* Retailer optimal quantity
Q_{sc}^* Supply Chain optimal quantity
π_m Expected profit of the manufacturer
π_r Expected profit of the retailer
q Quantity ordered by retailer
$S(q,p)$ Expected sale for an order quantity of q and a retail price of p
π_i^j The profit realized where i = {r:retailer; m: manufacturer} and j ={NC: Market Setting, SC: optimal case, B: buyback contract, RS: Revenue Sharing contract, QD: Quantity Discount contract}

Buy Back Contract:
r: Unit Buyback price offered by the manufacturer to the retailer for unsold items

Revenue Sharing Contract:
Φ Fraction of the retailer revenue transferred to upstream actor

Additional Assumptions

It is assumed that $p > w > s$; $c_m < w$, $w + c_r < p$; $w > r + s$ and $c_r + c_m > s$

4. Detailed model

In a first step, the Supply Chain performance is assessed for two basic cases:

The first case corresponds to Market-like setting, where the retailer sets his price at the same time as his stocking decision independently of the manufacturer. In the second case, a unique decision maker determine the optimal stocking and pricing policy that maximizes the global Supply Chain profit given the market demand.

In a second step, the Supply Chain performance is evaluated under 3 different types of coordination contracts (Buy Back Contracts, Revenue Sharing, and Quantity Discounts), relatively to the basic cases.

4.1. Centralized setting :

A unique decision regarding the stocking policy and the retail price is made for the whole Supply Chain. The

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

optimal Buyer Order quantity and the retail price are determined by differentiating Equation (5).

The profit of the downstream actor is:

$$\pi_r\ (q,p) = \begin{cases} p * D(p) - (c_r\ + w) * q + s * \big(q - D(p)\big) & if\ D(p) \le q \\ p * q - (c_r\ + w) * q & if\ D(p) > q \end{cases}$$
(1)

The profit of the upstream actor is :

$\{\pi_m(q,p) = \big(w - c_m\big)q$ **(2)**

Let's consider a new variable z such **z = q –d (p)**, the new retailer profit formula becomes :

$$\pi_r\ (z,p) = \begin{cases} p * [d(p) + \varepsilon] - \big(c_r\ + w\big)[d(p) + z] + s * [z - \epsilon] \\ \qquad if\ \varepsilon \le z \\ \big(p - c_r\ - w\big) * [d(p) + z] \qquad if\ \varepsilon > z \end{cases}$$
(3)

The Supply Chain profit is in this case:

$$\pi(z,p) = \begin{cases} p * [d(p) + \varepsilon] - c * [d(p) + z] + s * [z - \epsilon] & if\ \varepsilon \le z \\ (p - c) * [d(p) + z] & if\ \varepsilon > z \end{cases}$$
(4)

The optimal stocking and pricing policy consists of selling at price p^*_{SC} and ordering q^*_{SC}=z^*_{SC}+d(p^*_{SC}) where z^*_{SC} and p^*_{SC} maximize expected profit

The expected profit of the Supply Chain is :

$$E\big(\pi(z,p)\big) = \int_{-\infty}^{z} (p * [d(p) + u] - c * [d(p) + z] + s * [z - u])f(u)du + \int_{z}^{+\infty} \big((p - c) * [d(p) + z]\big)f(u)du$$

$E\big(\pi(z,p)\big) = (p - c)[d(p) + \mu] - (c - s)\int_{-\infty}^{z}(z - u)f(u)du - (p - c)\int_{z}^{+\infty}(u - z)f(u)du$ **(5)**

We refer in the remaining of this section, to the solution approach proposed by Petruzzi and Dada 1999.
Defining $\Lambda(z) = \int_{-\infty}^{z}(z - u)f(u)du$, $\theta(z) = \int_{z}^{+\infty}(u - z)f(u)du$, $\psi(p) = (p - c)[d(p) + \mu]$, we can write the expected Supply Chain profit as:

$E\big(\pi(z,p)\big) = \psi(p) - (c - s)\Lambda(z) - (p - c)\theta(z)$

Let's consider the first and second derivative with respect to z and p:

$\frac{\partial E(\pi(z,p))}{\partial z} = -(c - s) + [p° - s - \frac{\theta(z)}{2b}][1 - F(z)]$ (15)

$\frac{\partial^2 E(\pi(z,p))}{\partial z^2} = -(p - s)f(z)$ (16)

$\frac{\partial E(\pi(z,p))}{\partial p} = 2b(p° - p) - \theta(z)$ (17)

$\frac{\partial^2 E(\pi(z,p))}{\partial p^2} = -2b$ (18)

By extension of Petruzzi and Dada solution to the case where **ε** has an unbounded probability distribution function, there is one optimal solution to the maximization problem of $E\big(\pi(z,p(z))\big)$ if the following conditions hold:

a) If F(.) is a distribution function satisfying the conditions 2r(z)²+dr/dz>0 where r(.) =f(.)/[1-F(.)] is the hazard rate
b) $\lim_{+\infty} \frac{\partial E(\pi(z,p(z)))}{\partial z} < 0$
c) If the conditions for (a) and (b) are met and $\lim_{-\infty} \frac{\partial E(\pi(z,p(z)))}{\partial z} > 0$, there is a unique z^*_{SC} that satisfies $\frac{dE(\pi(z,p(z))}{dz} = 0$

The optimal pricing an stocking policy consists then of stocking q^*_{SC}=d(p^*_{SC})+ z^*_{SC}units and to sell them at price $p^*_{SC} = p° - \frac{\theta(z^*_{SC})}{2b}\ where\ p° = \frac{a+bc+\mu}{2b}$

Given the assumptions of our model, the conditions a) and b) are fulfilled and the equation (5) has a unique optimal solution (cf. Proof).

Proof:

i. As ε has a normal distribution, It has a strictly increasing failure rate dr/dz>0. It follows that 2r(z)²+dr/dz>0 and then a) is satisfied
ii. $\lim_{+\infty} \frac{\partial E(\pi(z,p(z)))}{\partial z} = -(c - s)$

As $c > s$, it follows that $\lim_{+\infty} \frac{\partial E(\pi(z,p(z)))}{\partial z} < 0$

iii. By rewriting eq (5) as :

$$\frac{\partial E\big(\pi(z,p(z))\big)}{\partial z} = \{-(c - s) + [p° - s][1 - F(z)]\} + \left\{-\frac{\theta(z)}{2b}[1 - F(z)]\right\}$$

When $z \to -\infty$ the first term converges to $-(c - s)$

Let's consider the second term:

$$-\frac{\theta(z)}{2b}[1 - F(z)] = -\frac{\theta(z)}{2b}\int_{z}^{+\infty} f(u)du = -\frac{1}{2b}\left[\int_{z}^{+\infty} uf(u)du - z\int_{z}^{+\infty} f(u)du\right]\int_{z}^{+\infty} f(u)du$$

When $z \to -\infty$, $\int_{z}^{+\infty} uf(u)du \to \mu\ and\ \int_{z}^{+\infty} f(u)du \to 1$

Then the second term converges to +∞ when $z \to -\infty$

From ii and iii it follows that There exist some A and B such that:

$for\ z < A\ , \lim_{-\infty} \frac{\partial E(\pi(z,p))}{\partial z} > 0$ And $for\ z > B\ , \lim_{+\infty} \frac{\partial E(\pi(z,p))}{\partial z} < 0$

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

As a corollary of Petruzzi and Dada proof for the case of bounded Probability Distribution Function it can be deduced that there exists a unique optimal solution to the maximization problem of $E(\pi(z,p(z)))$.

4.2. Market like setting (Non Coordinated case):

The retailer and manufacturer act independently and each aims to maximize his profit. The optimal retailer profit is determined by differentiating eq.(6)

$E(\pi_r\ (z,p)) = (p - c_r\ - w)[d(p) + \mu] - (c_r\ + w - s)\int_{-\infty}^{z}(z-u)f(u)du - (p - c_r\ - w)\int_{z}^{+\infty}(u - z)f(u)du$ **(6)**

Eq(6) is similar to eq(5) if c is substituted by $c_r\ + w$ in eq(5). The conditions a), b) and c) are satisfied as in the coordinated case and the optimal pricing and stocking decision (z_{NC}^*, p_{NC}^*) are determined by differentiating (6).

As $c_m < w$, it can be shown that $(z_M^*, p_M^*) \neq (z_{SC}^*, p_{SC}^*)$. Thus, the expected profit of the downstream actor is reduced if the Supply Chain is centrally coordinated.

Eq(2) shows that the upstream actor's profit is proportional to the Buyer order quantity and then he expects more profit if more quantity is ordered by the Downstream actor.

Even if Central Coordination scheme yields the highest global profit, it is not in the best interest of the downstream actor to be part of it as he will experience a reduction of his profit.

4.3. Revenue Sharing Contract :

Under this contract, the retailer benefits from a reduced wholesale price w_{RS}while sharing a fraction $\boldsymbol{\Phi}$ of the revenue he generates with the manufacturer.

From (1) and (2), the expected profit of the global Supply Chain and the retailer actor can be written as follows:

$E(\pi(q,p)) = (p - s)S(q,p) - (c - s)q$ **(7)**

$E(\pi_r^{RS}(q,p)] = \boldsymbol{\Phi}(p - s)S(q,p) - (c_r\ + w_{RS} - \boldsymbol{\Phi}s)q$ **(8)**

If the wholesale price and the sharing factor are set such that:

$w_{RS} = \boldsymbol{\Phi}c - c_r\ \ where\ \boldsymbol{\Phi} > \frac{c_r}{c}$ **(8')**

Eq.8 changes into :

$E(\pi_r^{RS}(q,p)) = \boldsymbol{\Phi}(p - s)S(q,p) - \boldsymbol{\Phi}(c - s)q$

$E(\pi_r^{RS}(q,p)) = \boldsymbol{\Phi}\mathrm{E}(\pi(q,p))$**(9)**

Then optimal stocking and pricing policy is then optimal for the downstream actor.

The upstream actor profit is:

$E(\pi_r^{RS}(q,p)) = E(\pi(q,p)) - E(\pi_r^{RS}(q,p)) = (1 - \boldsymbol{\Phi})\mathrm{E}(\pi(q,p))$**(10)**

The Downstream actor profit is an increasing function of $\boldsymbol{\Phi}$ while the upstream actor profit is decreasing in $\boldsymbol{\Phi}$. This revenue contract coordinates the Supply Chain and arbitrarily allocates its profit.

This Revenue Sharing contract should be desirable if the Chain partners obtain higher profits than in the Market like Settings:

$\frac{E(\pi_i^{RS}(q_{RS}^*, p_{RS}^*))}{E(\pi_i^{M}(q_{NC}^*, p_{NC}^*))} \geq 1\ where\ \mathrm{i} =$ {r: Retailer; m: manufacturer} **(10')**

4.4. Buy Back Contract :

Under this contract, the retailer salvages the leftover inventory and gets a credit of ***r*** for each unsold unit at the end of the season. It is assumed that the manufacturer is able to monitor the level of the leftover inventory and that the retailer should not profit from it $(w \geq r + s)$.

With a buyback contract the downstream actor profit is :

$E(\pi_r^B(q,p)) = (p - s - r)S(q,p) - (w_B\ - r + c_r\ - s)q$ (11)

A necessary condition with respect to price for the contract to coordinate the Supply Chain at the downstream level is that:

$\frac{\partial E(\pi_d^B(q,p))}{\partial p} = S(q,p) + (p - s - r)\frac{\partial S(q,p)}{\partial p} = 0$ (12)

A comparison of the first order condition -with respect to price- for the global Supply Chain

$\frac{\partial E(\pi(q,p))}{\partial p} = S(q,p) + (p - s)\frac{\partial S(q,p)}{\partial p} = 0$ (13)

Implies that eq.12 holds if r=0

By comparing the necessary first order condition -with respect to quantity- for the retailer and the global Supply Chain, $w_B = c_m$ should hold for the Buyback contract to coordinate the Supply Chain.

Eq.11 changes into:

$E(\pi_r^B(q,p)) = (p - s)S(q,p) - (c - s)q = E(\pi(q,p))$

Under these conditions, the retailer monopolizes the whole Supply Chain profit. This situation cannot be acceptable to the supplier.

As a consequence, the Buy Back contract doesn't coordinate the Supply Chain when the demand is price dependent.

4.4.1. Special case (Price is set first):

When the retailer sets the price before the order quantity is decided, Buy-Back Contracts coordinates the Supply Chain if the wholesale price w_B and the repurchase price r are defined in the following way:

$$\begin{cases} r = (1-\alpha)(p-s) \\ w_b = r + \alpha(c-s) - (c_r - s) \end{cases} where\ \ 0 < \alpha < 1$$

By substituting b and w_b in (11), the retailer profit can be expressed as a fraction of the Supply Chain profit:

$E(\pi_r^B(q,p)) = \boldsymbol{\alpha}\mathrm{E}(\pi(q,p))$

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

The retailer's profit is increasing in **α** and the supplier's profit is decreasing in **α.** So the parameter **α** acts to allocate the Supply Chain's profit between the two firms.

The feasible values of **α** are determined with respect to the constraints

$$\begin{cases} \dfrac{E\left(\pi_i^B(q_B^*, p_B^*)\right)}{E\left(\pi_i^M(q_{NC}^*, p_{NC}^*)\right)} \geq 1 \; where \; \mathrm{i} = \begin{Bmatrix} \text{r: Retailer;} \\ \text{m: manufacturer} \end{Bmatrix} \\ w_b \geq r + s \; equivalent \; to \; \alpha > \dfrac{c_r}{c-s} \end{cases}$$

4.5. Quantity Discount Contract :

Many Quantity Discount scheme has been addressed in literature. Distinction can be made between incremental and all unit Quantity Discount. We consider all unit Quantity Discount in what follows. With Quantity Discount contract, the wholesale price is a decreasing function of the ordered quantity.

The retailer profit function has the form:

$E(\pi_r^{QD}(q,p)) = (p-s)S(q,p) - \left(w_{QD}(q) + c_r - s\right)q$ (14)

Let $p_{SC}^*(q)$ be the optimal stocking policy for a given q for the global Supply Chain.

As

$$\frac{\partial E(\pi_r^{QD}(q,p))}{\partial p} = S(q) + (p - v - b)\frac{\partial S(q,p)}{\partial p} = \frac{\partial E(\pi \quad (q,p))}{\partial p}$$

It follows that

$$\frac{\partial E(\pi_r^{QD}(q,p_{SC}^*(q))}{\partial p} = \frac{\partial E(\pi \quad (q,p_{SC}^*(q))}{\partial p} = 0$$

So the first optimality condition with regard to p is satisfied .

Let's replace p by $p_{SC}^*(q)$ in (14), the retailer profit is reduced to a function of the single variable q:

$$E(\pi_r^{QD}(q,p_{SC}^*(q))) = (p-s)S(q,p_{SC}^*(q)) - \left(w_{QD}(q) + c_r - s\right)q$$

By choosing $w_{QD}(q)$ such that :

$$w_{QD}(q) = ((1-\alpha)((p_{SC}^*(q_{SC}^*) - s)\left(\frac{S\left(q,p_{SC}^*(q_{SC}^*)\right)}{q}\right) + \alpha(c-s) - (c_r - s) \text{ Where } \alpha \geq 0 \quad \textbf{(15)}$$

The retailer profit can be written as:

$$E\left(\pi_r^{QD}\left(q,p_{SC}^*(q)\right)\right) = (p-s)S\left(q,p_{SC}^*(q)\right) - \alpha(c-s)q - (1-\alpha)(p_{SC}^*(q_{SC}^*) - s)S(q,p_{SC}^*(q_{SC}^*))$$

Let's evaluate the expected retailer profit at the global Supply Chain optimal price$p_{SC}^*(q_{SC}^*)$:

$$E\left(\pi_r^{QD}\left(q,p_{SC}^*(q_{SC}^*)\right)\right) = (p_{SC}^*(q_{SC}^*) - s)S\left(q,p_{SC}^*(q_{SC}^*)\right) - \alpha(c-s)q - (1-\alpha)(p_{SC}^*(q_{SC}^*) - s)S\left(q,p_{SC}^*(q_{SC}^*)\right)$$

$$E\left(\pi_r^{QD}\left(q,p_{SC}^*(q_{SC}^*)\right)\right) = \alpha(p_{SC}^*(q_{SC}^*) - s)S\left(q,p_{SC}^*(q_{SC}^*)\right) - \alpha(c-s)q$$

$$E\left(\pi_r^{QD}\left(q,p_{SC}^*(q_{SC}^*)\right)\right) = \alpha E\left(\pi\left(q,p_{SC}^*(q_{SC}^*)\right)\right)$$

Thus given that the optimal global Supply Chain price is chosen, the Quantity Discount contract coordinates the Supply Chain and allocate the Supply Chain profit between the retailer and manufacturer.

The positivity condition for the optimal wholesale price and the sharing factor, results in the following inequality constraints for the sharing factor α:

$0 < \alpha < 1 \; and \; (B-A)\alpha > (C-A)$ **(16)** where $A = (p_{SC}^*(q_{SC}^*) - s)\left(\frac{S\left(q_{SC}^*,p_{SC}^*(q_{SC}^*)\right)}{q_{SC}^*}\right)$, $B = (c-s)$ and $C = (c_r - s)$

The desirability issue of the Quantity Discount contract can be addressed through determining the feasible values of α that let the profit of both manufacturer and retailer higher than in the Non Coordinated case:

$\frac{E(\pi_i^{QD}(q_{QD}^*,p_{QD}^*))}{E(\pi_i^M(q_{NC}^*,p_{NC}^*))} \geq 1 \; where \; i =$ {r: Retailer; m: manufacturer} (17)

5. An application

The proposed model has been applied to a numerical example and various scenarios were simulated using Matlab Software.

	Parameter	Value
Salvage value	s	20
Retailer marginal cost	c_r	15
Manufacturer marginal cost	c_m	50
Wholesale price (Market like Setting)	w	65
Demand = *D = a-bp + ε*	a	2000
	b	20
$\varepsilon \sim$ Normal (μ,σ)	μ	0
	σ	30

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Eight scenarios have been considered to illustrate the results of coordination contracts simulation:

Contract	Scenario	Description	Decision Variables
Centralized Setting	**S1**	Centralized Control	
Market like Setting	**S2**	Non coordinated Supply Chain	
Revenue Sharing	**S3**	Minimum profit for retailer	Φ_{min} = 0,28, w = 3,21
	S4	Maximum profit for retailer	Φ_{max} = 0,498, w= 17,38
Quantity Discount	**S5**	Minimum profit for retailer	α_{min} = 0,28
	S6	Maximum profit for retailer	α_{max}= 0,498
Buy Back contract (Special Case $p = p^*_{SC}$)	**S7**	Minimum profit for retailer	(r = 41,28,w=60,28)
	S8	Maximum profit for retailer	(r = 31,07,w=58,48)

The different scenarios have been designed such that the manufacturer and retailer profits are higher than in the Market like setting scenario, and constraints are fulfilled for each coordination contract type:

- Scenarios S3 resp. (S5,S7) corresponds to the situation where the retailer gets exactly the same profit as in the Market like setting, under Revenue Sharing contract resp. (Quantity Discount, Buy back contract)
- Scenarios S4 resp. (S6,S8) corresponds to the situation where the retailer gets the maximum profit under each of the coordination contracts : Revenue Sharing resp. (Quantity Discount, Buy back contract)

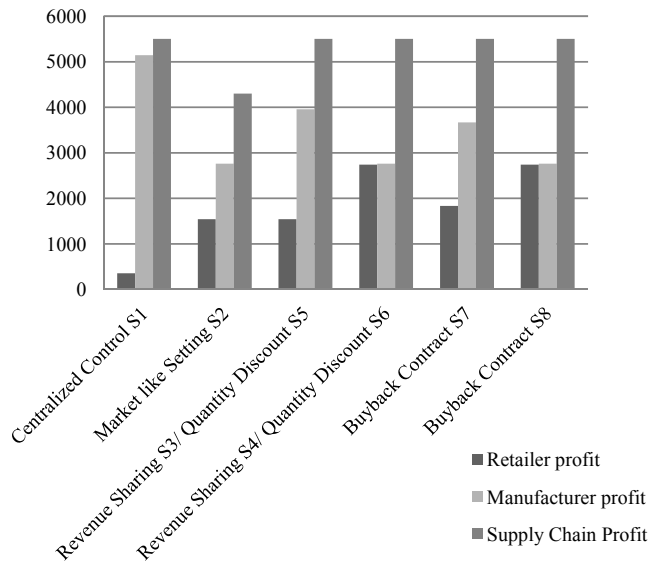


Fig. 1 Actor's profit for different coordination scenarios

Market like Setting

The order quantity and price are determined by differentiating eq.(6) ($p^*_{NC} = 89.11$; $Q^*_{NC=}184$).

When channel coordination is not pursued, Supply Chain efficiency $\frac{E(\pi^{SC})-E(\pi^{RS})}{E(\pi^{SC})}$ amounts to 12%. However Retailer profit is higher than in the centralized setting (1544 vs 355).

Unless, some mechanism is implemented to achieve coordination and allocate the extra Supply Chain profit between the retailer and manufacturer, the retailer will find no interest in ordering the optimal order quantity $Q^*_{SC=}343$.

Centralized Control

By differentiating eq.(5), the optimal order quantity and price amounts to ($p^*_{SC} = 81.92$; $Q^*_{SC=}343$).

The Supply Chain profit is optimal in this case. This setting is more attractive to the manufacturer who captures the biggest part of the Supply Chain profit (94%) compared to the retailer (6%).

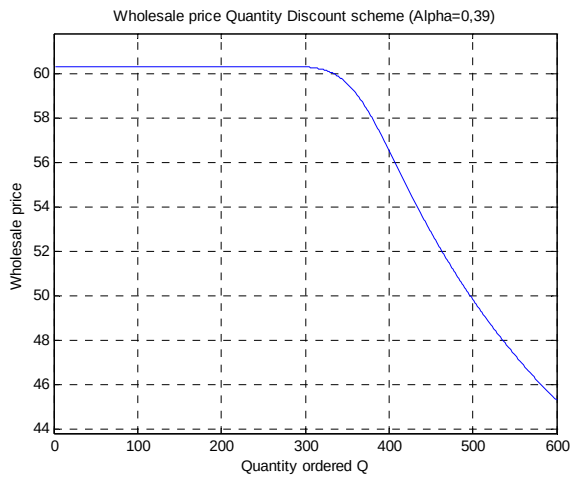
Revenue Sharing

The proposed coordination contract achieves coordination. The minimum and maximum values of the sharing ratio Φ are determined using the inequalities (10') and (8'). The Wholesale price w_{RS} can be deduced from (8'). The ranges of the sharing ratio and wholesale price are respectively [28%; 49.8%] and [3.21; 17.38]. The two ranges suggest a contract design that not only ensures channel coordination, but also an improvement of the profit of all actors.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Quantity Discount

For the chosen Quantity Discount contract scheme, the retailer's share of Supply Chain profit ranges from 28% to 49,8% using Inequalities (16) and (17). The stakeholders' negotiation capabilities can decide the allocation of the profit between the retailer and Manufacturer. The wholesale price is a function of the order quantity and is expressed through eq.(15) (cf. Graph for a fraction of 39%)



Buy Back Contract (Special Case)

The Buyback contract fails to coordinate Supply Chain when quantity ordered and retail price are set simultaneously. However if price is set first to the optimal Supply Chain price $p = p_{SC}^{*}$ before order quantity is defined, this contract coordinates Supply Chain and allocate extra profit between retailer and manufacturer. The expected profits gained by SC actors for different choices of the contract parameters are compared to those obtained under the market-like setting. For this application, the retailer (resp Manufacturer) Supply Chain profit share falls in the range [33%;49.8%] resp [50,2%;67%].

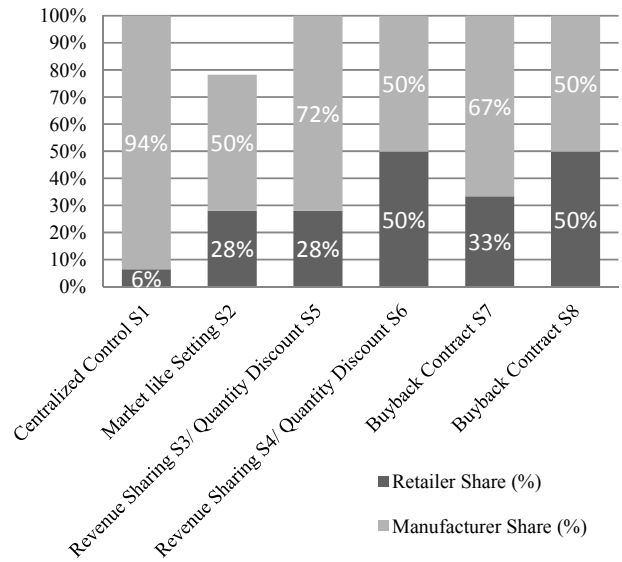


Fig .2 Actors ' share of optimal Supply Chain profit

6. Conclusions and Discussion

This paper has addressed the problem of coordination under the hypothesis of decentralized control and price dependent demand for a two level Supply Chain. Our comparative analysis of 3 coordination contracts shows that a suitable design of the Revenue Sharing and Quantity Discount contracts can lead the Supply Chain actors to select the price and order quantity that are optimal for the global Supply Chain. It also demonstrates that a correct choice of the contract parameters improves the individual profit of all the stakeholders.

Buy Back contract shows its limits in coordinating the Supply Chain when demand is price dependent. However it can still achieve coordination and divide extra profit between Supply Chain actors if the later agree to choose the optimal price as a retail price.

Through an application, it has been shown that a proper design of the coordination contracts could be a convincing incentive for the Supply Chain actors to adopt them.

This research work has shed the light on the potential benefits of coordination contracts. The model studied can be developed by making further assumptions regarding some parameters such as Supply Chain structure, contract periodicity, information structure and external factors affecting the supply chain (ex: Effort dependent demand,…).

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

References

[1] Whang, S., 1995, Coordination in operations: taxonomy. Journal of Operations Management 12, 413–422.

[2] Ilaria Giannoccaro*, Pierpaolo Pontrandolfo, Supply Chain coordination by Revenue Sharing contracts, Int. J. Production Economics 89 (2004) 131–139)

[3]Cachon, G.P. and Lariviere, M. A., 2001. Contracting to assure Supply: How to share demand forecasts in a Supply Chain. Management Science, 47(5):629–646

[4] Ding Ding, Jian Chen, Coordinating a three level Supply Chain with flexible return policies Omega 36 (2008) 865 – 876

[5] Darwish & Odah, Vendor managed inventory model for single-vendor multi-retailer Supply Chains; European Journal of Operational Research 204 (2010) 473–484

[6] He, Y., Zhao, X., Zhao, L., and He, J., 2009. Coordinating a Supply Chain with effort and price dependent stochastic demand. Applied Mathematical Modeling, 33(6):2777–2790

[7] Petruzzi and Dada, Pricing and the newsvendor problem: A review with extensions, Operations Research Vol. 47, No. 2, March–April 1999

[8] Xianghua Gan Suresh P. Sethi, Houmin Yan, Coordination of Supply Chains with Risk-Averse Agents Production and Operations Management POMS Vol. 13 No. 2, Summer 2004, pp. 135–149

[9] Pasternack, B. A. (1985), Optimal pricing and return policies for perishable commodities. Marketing Science, 4, 166–176

[10] Cachon, G. P., 2003, Handbooks in Operations Research and Management Science: Supply Chain management, chapter Supply Chain Coordination with Contracts

[11] Padmanabhan, V., & Png, I. P. L. (1997). Manufacturer's returns policies and retail competition. Marketing Science, 16, 81–94

[12] G.P. Cachon, M.A. Lariviere Supply Chain Coordination with Revenue-Sharing Contracts: Strengths and Limitations, Management Science, Vol. 51, No. 1, January 2005, pp. 30–44

[13] Jianli Li and Liwen Liu, Supply Chain coordination with Quantity Discount policy, Int. J. Production Economics 101 (2006) 89–98)

[14] Tsay, The Quantity Flexibility Contract and Supplier-Customer Incentives, Management Science/Vol. 45, No. 10, October 1999 pp. 1339–1358

[15] Behzad.H and Wiesław.K, Coordinating Contracts in SCM:A Review of Methods and Literature, Decision Making in Manufacturing and Services, Vol. 4 , 2010, No. 1‑2, pp. 5‑28

Noureddine Labiad Master's of Science in Industrial Engineering & Operations Research (2001-2002), North Carolina State University (NC,USA) ; Bachelor in Control Engineering (1992-1997), Ecole Mohammadia des Ingénieurs(Rabat, Morocco)
Supply Chain Manager (2010-2011) (Ciments de l'Atlas), Consultant Manager (2004-2009) (Capital Consulting/ Omnidata)

Zitouni Beidouri Engineer-Phd Ecole Mohammadia des Ingénieurs (Rabat, Morocco), Professor at Ecole Superieure de Technologie, Casablanca, Morocco

Othmane Bouksour Engineer-Phd Ecole Mohammadia des Ingénieurs (Rabat, Morocco), Professor at Ecole Supérieure de Technologie, Casablanca, Morocco

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Steady State and Transient Study of The Electron Transport In N-Type GaSb Using Three Valley Monte Carlo Model

Youcef Belhadji [1], Benyounes Bouazza[2]

[1]Unit of research of materials and renewable energies, faculty of science, Abou bekr Belkaid University, PB 119 Tlemcen 13000, Algeria.

[2] Unit of research of materials and renewable energies, faculty of science, Abou bekr Belkaid University, PB 119 Tlemcen 13000, Algeria.

Abstract

In this work, an investigation of the steady state and transient electron transport, at high electric field and high temperature, in n-type GaSb material was presented. We applied a Monte Carlo model considering the three valleys of the conduction band (Γ, L, X), isotropic and non-parabolic centered on the important symmetry point of the first Brillouin zone. This model provides a detailed description of the electronic dynamic and the electrons behavior at high electrical fields and high temperatures in these materials in each considered valleys. The calculation is made for different values of applied electrical field ranging from 50 to 600KV at different values of temperature. The scatterings rates taken into consideration are coulomb scattering acoustical and optical scattering, intra and inter-valleys scattering.

Keywords: *Gallium Antimonide, Electron Transport, Band Structure, Monte Carlo Method, Polar Scattering.*

1. Introduction

Over the last several years, the III-V binary compounds started showing considerable interest towards the study and use in microelectronic manufacturing [8, 19]. The Gallium Antimonide (GaSb) is one of these materials, with a direct bandgap that can be grown with high purity. The III–V ternaries and quaternaries (AlGaIn) (AsSb) lattice matched to GaSb seem to be the obvious choice and have turned out to be promising candidates for high speed electronic and long wavelength photonic devices [18].GaSb based structures are promising candidates for a variety of military and civil applications, notably for laser diodes with low threshold voltage, Photo-detectors, high frequency devices, and a promising material for optoelectronic semiconductor devices in the near infrared [4].

Because it's narrow bandgap (0.7ev), the GaSb-based structures are of great interest for application in mechanically stacked tandem solar cell, in thermo-photovoltaic [1] and in rechargeable lithium batteries [15]. As the unintentionally doped GaSb is invariably p-type in nature [14]. In our work we choose to study the n-type GaSb substrate. Recent investigations of the electron transport in the n-type GaSb are proposed in literature [4, 5, 16, 14]. In this work we investigate the steady state and transient electron transport properties, at high field and high temperature. It's known that the electron transport at high fields and high temperatures has reached an important stage in the last years and is dominated by the complex conduction-band structure well above the lowest band edge [8]. For the n-type GaSb the transport properties are more complicated due to the contributions from Γ, L and X conduction bands [18]. However, the electronic transport and phonons scattering is always described by models based on the resolution of the Boltzmann transport equation (BTE) who characterizes the modification of the function distribution of particles, caused by various actions, in space and time. The most used models are: the drift diffusion models, Hydro-dynamical models and Monte Carlo Models. This last is computationally intensive in fact that we can simulate the trajectory of a large number of electrons in physical and momentum space. Nowadays the microscopic transport model based on the Monte Carlo method appears to be adequate for studying of electronic transports characteristic in bulk and semiconductor devices [20]. In what follows, we give, in section 2, a detailed description of our MC code which uses analytic descriptions for both the electron bands and the phonon dispersion, and the results of transient and steady-state transport simulations are discussed in the last section.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

2. Monte Carlo Model Description

Including the charges transport in semiconductors the Monte Carlo method is a statistical method, as a popular and effective mathematical tool for the study and analysis of the physical phenomena. This method was first applied to the study of transport by Kurosawa in 1966 [10] and is widely used since 1981 [11, 12]. The Monte Carlo method provides an additional advantage in which can be used as a theoretical laboratory in that parameters can be varied at will and their effects on observables assessed [17].
The principle of the method is to follow over time one or more electrons in real space and reciprocal space wave vector k. After the identification of symmetry L valleys which have minimum related lower to X valleys [11], this model has been used by many authors [6,8] and more recently by H. Arabshahi [16] in the studies of the electron transport in GaSb and other materials[17].Our model is developed considering the deformation potential of each considered valley, the effective masses, the dielectrics constant, the energy gap structures, and the main scattering mechanisms in details [3]. In general terms, the model is based on a process of drawing lots from the laws of probability associated with the electrons undergoing scatterings during their movements within the crystal lattice. The number of electrons simulated is 20000 electrons subjected to a varying field, at different temperature. Mainly, the simulation is done using the following steps [9, 7]: for each simulated electron "p" we associate an initial position "r", and an initial wave vector k, and using the method "individuals-dispersing" or "self-scattering", we introduce a fictitious scattering which aims to establish a time distribution.

$$\overrightarrow{K_p}(t), \overrightarrow{r_p}(t), \overrightarrow{E_p}(t) = \overrightarrow{E}(\overrightarrow{K_p}) \quad (1)$$

We performing free flight (no scattering) time period Δt constant, and low before the average time between two consecutive scatterings, such as:

$$\begin{cases} \overrightarrow{K_p}(t+\Delta t) = \overrightarrow{K_p}(t) + \dfrac{e\overrightarrow{E}}{\eta} \\ \overrightarrow{E_p}(t+\Delta t) = \overrightarrow{E_p}(\overrightarrow{K_p}(t+\Delta t)) \\ \overrightarrow{r_p}(t+\Delta t) = \int v_p(t)\,dt \end{cases} \quad (2)$$

During a free flight the electron is subject only to the electric field E, which changes the wave vector k according to the relationship [21]:

$$\frac{\partial \overrightarrow{K}}{\partial t} = \frac{q\overrightarrow{E}}{\hbar} \quad (3)$$

The integration of this equation, give the time evolution of the wave vector, and the knowledge of the band structure form $\varepsilon\left(\overrightarrow{K}\right)$, permit to deduce the carrier's energy such as:

$$\varepsilon\left(1+\alpha_i\varepsilon_i\right) = \gamma(k) = \frac{\hbar^2 k_i^2}{2m_i^*} \quad (4)$$

α_i denote the non-parabolicity coefficient of the valley "i". m_i^* is the relative effective masse in bottom of the valley. The density of state, by volume unity, associated to this model has the following form:

$$N(\varepsilon) = \frac{\left(2m_i^*\right)^{\frac{3}{2}}.\sqrt{\varepsilon\left(1+\alpha_i\varepsilon\right)}}{4\pi^2\hbar^3}\left(1+2.\alpha_i\varepsilon\right) \quad (5)$$

The expressions of the energy and the density of state permits the formulation of the scattering probability that directly dependant to the energy. During the simulation, when scattering occurs, we place at time t + Δt, and we determine subsequently, its nature by lot, and we modify accordingly the energy and electron wave vector. Each process of scattering characterized by a density of probability S(k, k') characterizing the probability, for which an electron will pass from a state of energy K to another state K'[6].

$$\lambda_j\left(\vec{K}\right) = \int_{over\ .all\ .possible\ .states\ .K'.} S_j\left(\vec{K},\vec{K}'\right) \quad (6)$$

$$\lambda(K) = \frac{V}{8\pi^3}.2\pi\int_0^{\pi}\left[\int_0^{\infty} S\left(\vec{K},\vec{K}'\right)K'^2.dK'\right]\sin\beta.d\beta \quad (7)$$

$\frac{V}{8\pi^3}$. denote the density of states in momentum space.

The factor 2π reflects the symmetry of revolution around all possible states. β represent the deviation angle. In Monte Carlo studies, the obtained results are strongly related to many physical parameters characterizing studied materials, which are sometimes badly known.

Therefore, the treatment of different information obtained, their analysis as function of the material parameters entered in the simulator, and compared to available experimental results, permit a real knowledge of physical effects and their consequences that occurs in the material. The basic and band structure parameters of GaSb material used in our simulation are shown in table1.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Table 1: Basic parameters of GaSb material [16, 18, 22]

Parameters	*Values*
Relative mass for gamma-valley	0.039
Relative mass for the L-valley	0.22
Relative mass for the X-valley	0.52
High-frequency permittivity	14.2
Low-frequency permittivity	15.5
Energies separation between gamma and L valley $\Delta_{\Gamma-L}$ [eV]	0.04
Energies separation between gamma and L valley $\Delta_{\Gamma-X}$ [eV]	0.33
Crystal density[g/cm^3]	5.613
Sound velocity [cm/s]	6070
Phonon energy (mev)	29

3. Results and Discussions

The Monte Carlo model adopted in this work is used for calculating and retrieval features of electronic transport in the n-type GaSb in the transitional regime. Interestingly, before describing the results and interpret them, establish a function of electric field and the valley occupied a hierarchy between the different scatterings experienced by the user. The results achieved by this model are given in the figures below. In '*Γ*' Valley, if the electron energy is very weak, then the scatterings predominate with a fairly low probability. According to Figure 1, the predominant scatterings for electrons in the valley *X* and L are those with polar optical phonons, with low wave vector. These scatterings are accompanied by an absorption or emission of a phonon. In this case the electron energy varies by an amount equal to the phonon energy $\hbar w$.

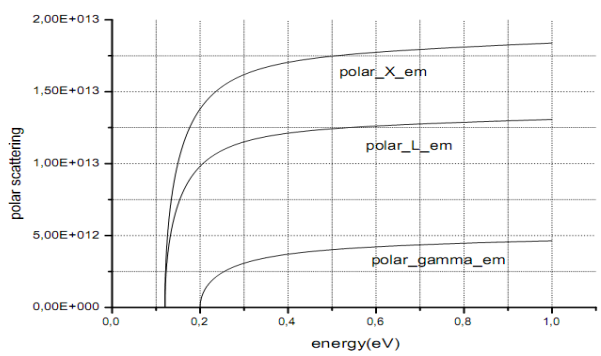


Fig. 1 Polar emission scattering rate

When the electron energy is sufficient, the emission of a phonon is a much more likely that absorption. The beginning of the phonon emission is easily identified in the curve of the valley from 40 mev. Also, the scatterings probabilities are greater in emission than that for absorption, and their value increases significantly when the electron energy increases.

Because of the dispersion relation of phonons is not constant; in this case the scatterings rate remains more complicated to calculate, and they occur when the ions through the crystal vibrate in the same direction.

According to Figure 2, due to the high value of effective mass, these scatterings are important in the X valley and increase more and more as energy increases. For high enough fields, the acoustic scatterings become dominant and subsequently the electron energy will be important. But as soon as the electron acquired energy greater than the energy separation $\Delta_{\Gamma L}$ and $\Delta_{\Gamma X}$ the transfer to the L and X valley become possible.

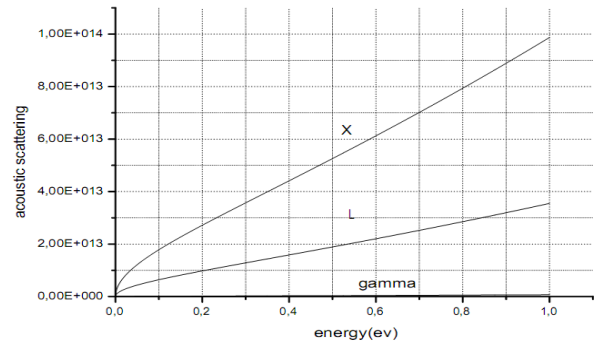


Fig. 2 Acoustic scattering rate .

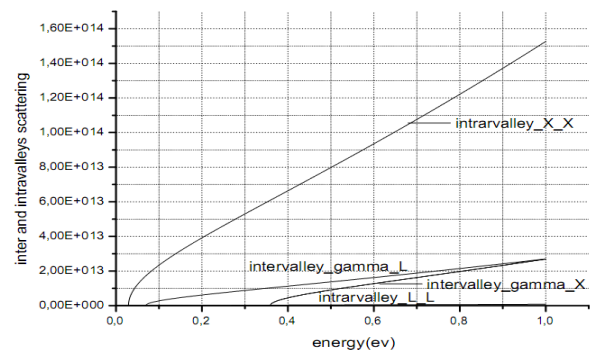


Fig. 3Piezoelectric inter and intra-valley scattering rate

Thus, when energy increases, we can see the beginning of the other type of scattering inter and intra- valley (figure 3). We can say that from energy for which transfers to the L and X valleys are possible, these processes are predominant. Piezoelectric interaction types are resilient and they are important in low field especially after the first transfers to valley level. Its main effect is decrease the value of mobility at low field. For acoustic interactions are the most favored among the elastic scatterings, they are mainly caused by the electron doping effects.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

The figure below give the electronic population as a function of time in each considered valley gamma, X and L for an applied electric field of 400KV/cm. This result shows that the inclusion of the satellite valleys in the simulation is very important [16]. When the applied field is significant and exceeds the threshold value, the inter-valley scattering becomes important into the satellite valleys L and X. This can be explained by the fact that electrons closed to the valley minimum have small kinetic energies, then more scattered. Thus, due to the combined effect of a lower gamma effective mass (0.039), lower separation energy ($\Delta_{\Gamma-L}$ and $\Delta_{\Gamma-X}$), and the lightly lowers phonon scattering rate within the central valley gamma, the intervalley scattering is substantially larger in GaSb over the range of applied electric fields.

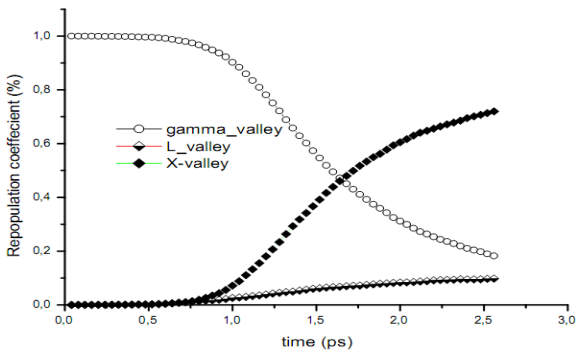


Fig. 4 The electronic population of the valleys depending on the time

The evolution of the electron drifts velocity as a function of field at room temperature was studied (see figure 5). There is a pronounced peak drift velocity located to $25x10^6$cm / s that occurs at a significantly lower electric field [16]. This result is the same obtained by H. Arabshahi [16]. In Central Valley Γ, the electrons have small effective masses (0.039 m_0), as soon as one applies a lower energy than the phonon scatterings there are only at the Γ valley.

By increasing the applied field, speed increases and reaches its maximum value near the threshold field (~1KV/cm), at which time the electron energy will be higher than the phonon, we have in this case a transition to the holder's side valleys L and X. When increasing the electric field, the speed decreases because the electrons are heavier side valleys. The interactions involved in this case the interactions are polar optical, intra and intervalley interactions.

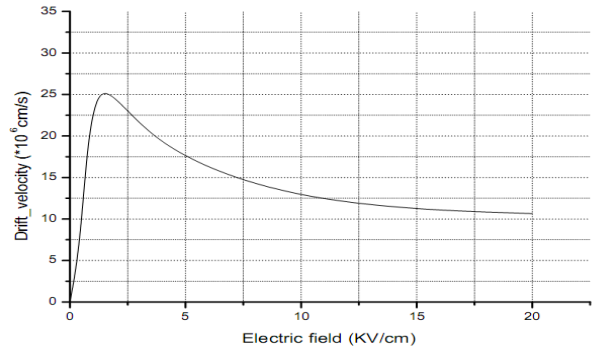


Fig. 5 Electron drifts velocity in GaSb bulk

When the electrons are subjected to fields largely varying in time (non-stationary mode), a phenomenon of overshoot velocity appears. This phenomenon was often studied by many authors, which did not fail to underline the practical interest of this phenomenon, in particular on the level of the submicronic components. Figure 6 give the look of the average electrons velocity. The calculation is made for electric fields ranging from 50 to 600KV/cm. We note a notable maximum velocity for the strong fields. Due to the low effective mass in valley, this maximum is important at high field ($9.5x10^7$ cm/s for a value of 600KV/cm of the electric field and a donor concentration of 8.10^{17} cm^{-3}). The maximum velocity is situated between 0.5 and 4 ps. More energy relaxation becomes effective as the distribution function and the drift velocity decreases.

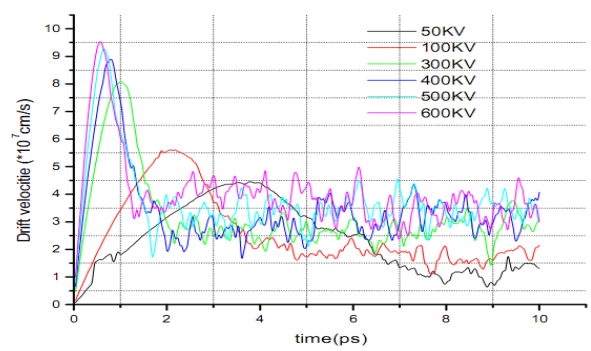


Fig. 6 Electron drifts velocity in GaSb bulk at 300K for different value of the applied electrical field

The temperature effect on the electrons velocity is also studied. The relating curves to drifts velocity at different values of temperature, for an electric field of 400KV/cm are presented.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

When the temperature increases the peak of velocity decreased. This last is due to the effect of acoustic scattering that becomes important as temperature increases and to the increase in the lattice vibrations (result of the increase in the temperature) supporting thereafter the transitions inter-valley and intra-valleys in another side. Nevertheless, the electron saturation velocity is significant in the GaSb what explains the increasing interest of this material within the industrial framework. Thereafter, polar optical interactions become dominant among all elastic scatterings.

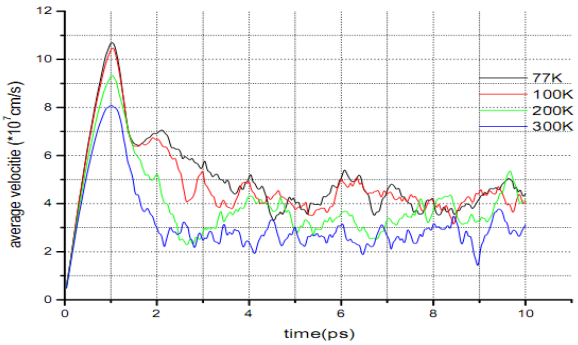


Fig. 7 Electron drifts velocity in GaSb bulk at different values of the temperature for an applied electrical field of 300KV

4. Conclusion

The electron transport in the n-type GaSb was studied. We adapted the Monte Carlo model written in FORTRAN, which is accurate, reliable and sometimes even simpler than the conventional numerical techniques. The transport characteristics and properties of GaSb obtained by this model, are satisfactory compared to other calculations in literature, and inform us about electronic dynamics as well as the behavior electric high fields and high temperatures in this material. The obtained results show that GaSb have a peak velocity little weak compared to other material. Therefore, we note an important overshoot velocity at low temperature (~$11*10^7$cm/s at 77K). The overshoot velocity phenomenon is major, particularly, in the determination of executions in the frequency and power components made from the material studied. However, analysis of results as the drift velocity of electrons demonstrates the remarkable effect of temperature and applied field on the dynamics of holders within the lattice as a result of interaction mechanisms.

References

[1] V.M. Andreev, V.P. Khvostikov, V.D. Rumyantsev, S.V. Sorokina, M.Z. Shvarts, V.I. Vasil'ev, GaSb/InGaAsSb-based PV cells for solar and therphotovoltaic

[2] L.N. Nikitina, S.V. Obukhov, V.G. Tyuterev Russian Physics Journal, 52, 747,2009

[3] Y. Belhadji, B. Bouazza, A. Guen-Bouazza, N.E. Chabane Sari, Proceedings ICEEDT'09,2009

[4] B.PŐDÖR, Acta physica polonica A,108, No.5, 2005.

[5] Y. Zheng, P.D. Moran, Z.F. Guan, S.S. Lau, and al., Journal of electronic materials ,32,2A,402, 2002.

[6] A. hamdoune, B. Bouazza, B. Guen-Bouazza, A. Lallam, N.E. Chabane Sari, Afrique sciences,03(2),189, 2007

[7]] B. Bouazza, B. Guen-Bouazza, L. Amer, C. Sayeh, N.E. Chabane Sari, C. Gontrand, Afrique sciences,01 (1),57, 2005.

[8] S. Singhal, A.K. Saxena, S. Dasgupta, Pramana journal of physics,69(4),687,2007

[9] L. Tirino, Transport properties of wide band gap semiconductors, Georgia institute of technology,2004

[10] K. Kurosawa, Journal of the Physical Society of Japan, 21, 424-426, 1966.

[11] R. W. Hockney, J. W. Eastwood, Computer simulation using particles, McGraw-Hill International Book Company, 1981.

[12] M. V. Feschetti, Monte Carlo simulation of electron transport in Si: the first 20 years, IBM Report, RC 20474, Physical Sciences, 1996.

[13] A. Arabshahi, African physical review,2, 131,2008

[14] L.G.O. Messias, E. Marega Jr. Brazilian journal of Physics, 32, 402, 2002

[15] R. Ahmed, F. Aleem, S. Javad Hashemifar, H Rashid, H. Akbarzadeh, Physical Properties of III-Antiminodes a First Principles Study,52,527, 2009

[16] H. Arabshahi, International journal of science and advanced technology, 1, 19-23, 2011

[17] S. Golafroz, H. Arabshahi, International journal of the physical sciences, 5(6),690-695,2010

[18]] P.S. Dutta, H.L. Bhart, Journal of applied physics, 81(9), 5821- 5870, 1997

[19] L.G.O. Messias, E. Marega. Jr., Brazilian journal of physics, 32(2A), 402-404, 2002

[20] A.M. Gerami, H.R. Soleimani, H. Arabshahi, M.R. Khalvati, Advances in Natural and Applied sciences, 3(3), 357-362, 2009

[21]] P. Hesto, "simulation Monte Carlo du transport non stationnaire dans les dispositifs submicroniques : importance du phénomène balistique dans GaAs à 77 K", Theses, Université de Paris Sud, Centre d'Orsay, 1984.

[22] H. Arabshahi, African physical review, 1, 131-135, 2008

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Comparison of Phase Only Correlation and Neural Network for Iris Recognition

Yogendra Kumar Jain[1] and Manoj Kumar Verma[2]
[1]Department of Computer Science and Engineering, SATI
Vidisha, M.P. 464001, India

[2]Department of Computer Science and Engineering, SATI
Vidisha, M.P. 464001, India

Abstract

This paper compares two different techniques of iris recognition and explains the steps of extracting iris from eye image palette formation and conditioning of palette for matching. The focus of the paper is in finding the suitable method for iris recognition on the basis of recognition time, false detection rate, conditioning time, algorithm complexity, bulk detection, database handling. In this paper comparison has been performed two methods POC (Phase Only Correlation) and Neural Network. Although both methods are practically proven by many researchers, still a comprehensive comparison is missing, we hope that the results drawn in this paper will be helpful for the peoples working in same field; the complete algorithm is developed in Matlab.

Keywords: *Iris recognition, POC (Phase Only Correlation), Neural Network.*

1. Introduction

Iris recognition [1] is a method of biometric authentication that uses pattern-recognition techniques based on high-resolution images of the irises of an individual's eyes. Iris recognition uses camera technology, with subtle infrared illumination reducing specular reflection from the convex cornea, to create images of the detail-rich, intricate structures of the iris. Converted into digital templates, these images provide mathematical representations of the iris that yield unambiguous positive identification of an individual [2].

Iris recognition [7] efficacy is rarely impeded by glasses or contact lenses. Iris technology has the smallest outlier (those who cannot use/enroll) group of all biometric technologies. Because of its speed of comparison, iris recognition is the only biometric technology well-suited for one-to-many identification. A key advantage of iris recognition is its stability.

The human iris is the annular part between the dark pupil and the white sclera; it is unique and stable throughout life. In an iris recognition system, iris localization is one of the most important steps, and has great influence on the subsequent feature extraction and classification. It aims to find the parameters, the centers and radii, for both the iris inner and outer boundaries. There are many iris localization methods; more algorithms were developed to improve the performance of iris localization. More recently, other algorithms have been presented by many researchers. However, these methods have promising performance; they need to search the iris boundaries over large parameter space exhaustively, which takes more computational time. Moreover, they may result in circle detection failure, because some chosen threshold values used for edge detection cause critical edge points being removed. In this paper, we compare two algorithms for iris recognition from iris template images. The paper also explains the iris extraction from eye image. The paper is arranged in following way, section 2 explains the method of iris extraction by using Hough circle method, and then the conversion of circular iris image into rectangular template, section 2.3 explains the methods of iris recognition then after section 4 shows the simulated results and conclusion on the basis of results.

Most of the researchers in the field of iris recognition use iris images from the following databases, which are available freely online:

- The Chinese Academy of Sciences database (CASIA)[16]
- The Bath database, produced by the university of Bath[17]

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

2. PRE PROCESSING FOR IRIS RECOGNITION

2.1 Iris Boundaries Detection

Iris is the area between pupil and white section of eyes, as shown in figures 2.1 the area between two white circles.

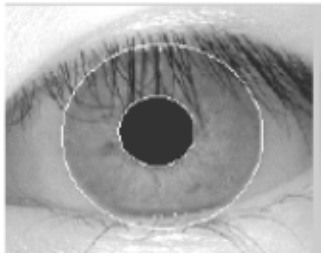


Figure 2.1 Eye and iris

It is clear the iris is a circular strip hence both iris boundaries are circular. Therefore most of the techniques [3][4][5][6] utilizes the shades of pupil, iris and other sections, but in proposed technique we preferred Hough circle[3][4] and threshold technique method for detection of correct iris boundaries. Following are the steps used in this paper for detection of correct iris boundaries.

1. Perform the histogram equalization in the iris image.
2. Convert it into edge image using sobel filter.
3. From each point of edge image draw a circle of diameter greater than 10 percentage of image height up to 90 percent of height.
4. Counts the edge travelled by each periphery of each circle and find normalized edge density of each periphery using the formula.

   D_n = edges/total points
5. Now find out the best circle with highest edge density with diameter from 10 to 30 percent, this will denote the inner boundary of the iris, repeat sane for diameter range 50 to 90 percent this will denote the outer boundary.
6. Now crop the image between these circles is the required iris.

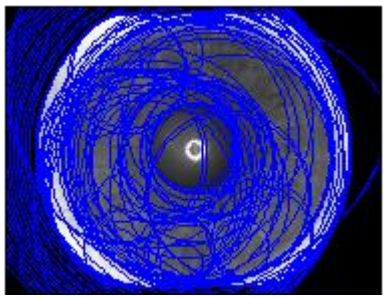


Figure 2.2 Circles drawn all over the image.

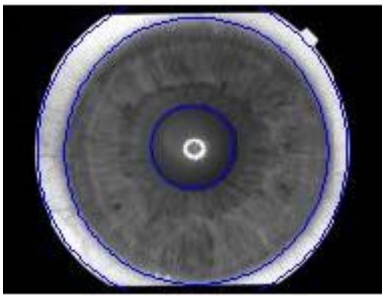


Figure 2.3 Selected circles.

2.2. Circular to rectangular conversion

From figure 2.2 we can easily crop iris but for recognition of iris it's much better to deal with Cartesian coordinates than circular, hence a conversion from circular to rectangular is performed, since the image is digital hence inner edge having lower number of points than upper edge. Now the conversion is performed by using formula,

$$x = r\cos\theta$$
$$y = r\sin\theta \qquad (1)$$

Where θ varies from 0 to 360 degree and r varies from inner circle radius to outer circle radius.

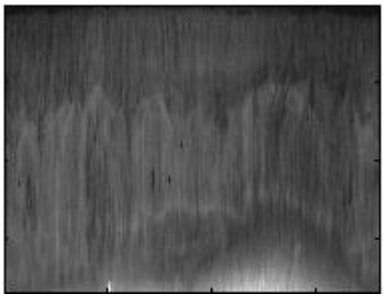


Figure 2.4 Rectangular iris image.

Figure 2.4 shows the rectangular converted image from of iris, since not all image content have important information,

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

therefore we can filter out some unwanted components figure 2.5 shows the filtered image [8].

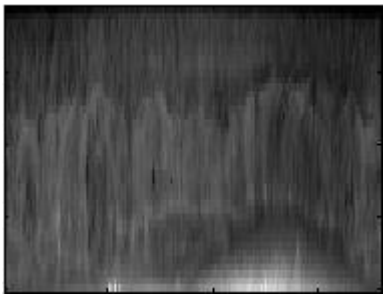


Figure 2.5 Rectangular filtered iris image.

Now this is called the standard template can be used for matching by any method.

2.3. Iris Matching Methods

2.3.1. Hamming distance

The Hamming distance gives a measure of how many bits are the same between two bit patterns. Using the Hamming distance of two bit patterns, a decision can be made as to whether the two patterns were generated from different irises or from the same one. In comparing the bit patterns X and Y, the Hamming distance, HD, is defined as the sum of disagreeing bits (sum of the exclusive-OR between X and Y) over N, the total number of bits in the bit pattern [19][15] .

$$HD = 1/N \sum_{i=1}^{N} X_j \,(XOR)\, Y_j \qquad (2)$$

If two bits patterns are completely independent, such as iris templates generated from different irises, the Hamming distance between the two patterns should equal 0.5. If two patterns are derived from the same iris, the Hamming distance between them will be close to 0.0.

2.3.2. Weighted Euclidean Distance

The weighted Euclidean distance (WED) can be used to compare two templates, especially if the template is composed of integer values. The weighting Euclidean distance gives a measure of how similar a collection of values are between two templates. This metric is specified as

$$WED(k) = \sum_{i=1}^{N} [(f_i - f_i^{(k)})^2 / (\partial_i^{(k)})^2] \qquad (3)$$

Where f_i is the ith feature of the unknown iris, and is the ith feature of iris template, k, and $\partial_i^{(k)}$ is the standard deviation of the ith feature in iris template k. The unknown iris template is found to match iris template k, when WED is a minimum at k.

2.3.3. Normalized Correlation

Normalized correlation between the acquired and database representation for goodness of match. This is represented as

$$\sum_{i=1}^{n} \sum_{j=1}^{m} (p_1[i,j] - \mu_1)(p2[i,j] - \mu_2) / nm\, \sigma_1 \sigma_2 \qquad (4)$$

Where p1 and p2 are two images of size nxm, μ1 and σ1 are the mean and standard deviation of p1, and μ2 and σ2 are the mean and standard deviation of p2. Normalized correlation is advantageous over standard correlation, since it is able to account for local variations in image intensity that corrupt the standard correlation calculation.

2.3.4. POC (Phase Only Correlation)

POC is a method for finding the correlation between two different images in frequency domain; it gives much better results than normal time domain correlation [8] [14].

This section introduces the principle of phase based image matching using the Phase Only Correlation (POC) [1][13] function. Consider two $N_1 X\, N_2$ images $f(n_1, n_2)$ and $g(n_1, n_2)$.

$$F(k_1,k_2) = \sum_{n_1=-M_1}^{M_1} \sum_{n_2=-M_2}^{M_2} f(n_1,n_2) W_{N_1}^{k_1 n_1} W_{N_2}^{k_2 n_2}$$

$$= A_F(k_1,k_2) e^{j\Theta_F(k_1,k_2)} \qquad (5)$$

$$G(k_1,k_2) = \sum_{n_1=-M_1}^{M_1} \sum_{n_2=-M_2}^{M_2} g(n_1,n_2) W_{N_1}^{k_1 n_1} W_{N_2}^{k_2 n_2}$$

$$= A_G(k_1,k_2) e^{j\Theta_G(k_1,k_2)} \qquad (6)$$

Here $F(k_1,k_2)$ and $G(k_1,k_2)$ are the 2D DCT transform of function F and g, now the cross-phase spectrum R_{FG}(k1; k2) between F(k1; k2) and G(k1; k2) is given by

$$R_{FG}(k_1,k_2) = \frac{F(k_1,k_2)\overline{G(k_1,k_2)}}{|F(k_1,k_2)\overline{G(k_1,k_2)}|}$$

$$= e^{j\{\Theta_F(k_1,k_2) - \Theta_G(k_1,k_2)\}} \qquad (7)$$

Where G(k1; k2) denotes the complex conjugate of G(k1; k2). The POC function r_{fg}(n1; n2) is the 2D Inverse DFT (IDFT) of R_{FG}(k1; k2) and is given by

$$r_{fg}(n_1,n_2) = \frac{1}{N_1 N_2} \sum_{k_1=-M_1}^{M_1} \sum_{k_2=-M_2}^{M_2} R_{FG}(k_1,k_2)\, W_{N_1}^{-k_1 n_1} W_{N_2}^{-k_2 n_2} \qquad (8)$$

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

When two images are similar, their POC [13] function gives a distinct sharp peak. When $f(n_1, n_2) = g(n_1, n_2)$, the POC function $r_{fg}(n_1, n_2)$ becomes the Kronecker delta function $\delta(n_1, n_2)$ If two images are not similar, the peak value drops significantly. The height of the peak can be used as a good similarity measure for image matching and the location of the peak shows the translational displacement between the two images.

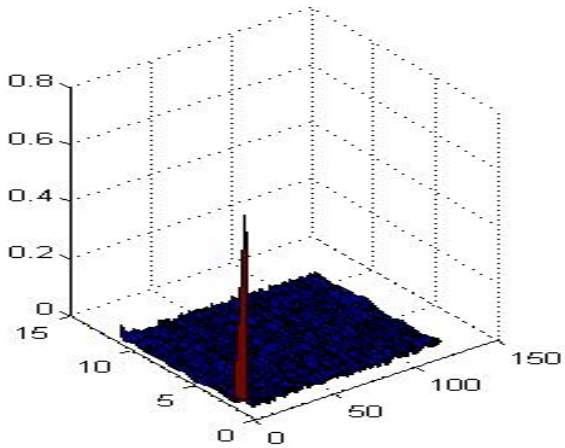


Figure 2.6 POC peak when iris image matched.

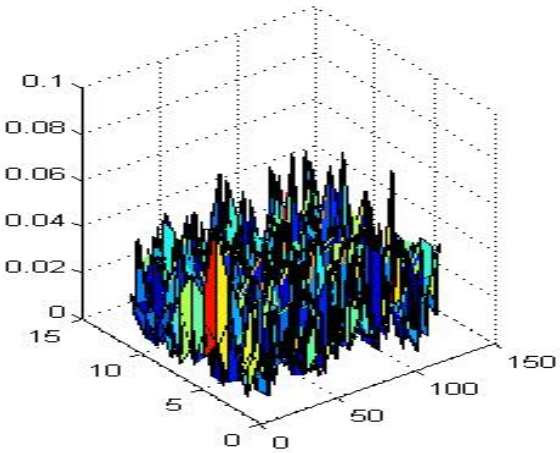


Figure 2.7 POC when iris image not matched.

2.3.5 Neural Network

The term **neural network** [15] was traditionally used to refer to a network or circuit of biological neurons. The modern usage of the term often refers to artificial neural networks, which are composed of artificial neurons or nodes. Thus the term has two distinct usages:

1. Biological neural networks are made up of real biological neurons that are connected or functionally related in the peripheral nervous system or the central nervous system. In the field of neuroscience, they are often identified as groups of neurons that perform a specific physiological function in laboratory analysis.

2. Artificial neural networks are composed of interconnecting artificial neurons (programming constructs that mimic the properties of biological neurons). Artificial neural networks may either be used to gain an understanding of biological neural networks, or for solving artificial intelligence problems without necessarily creating a model of a real biological system. The real, biological nervous system is highly complex: artificial neural network algorithms attempt to abstract this complexity and focus on what may hypothetically matter most from an information processing point of view. Good performance (e.g. as measured by good predictive ability, low generalization error), or performance mimicking animal or human error patterns, can then be used as one source of evidence towards supporting the hypothesis that the abstraction really captured something important from the point of view of information processing in the brain. Another incentive for these abstractions is to reduce the amount of computation required to simulate artificial neural networks, so as to allow one to experiment with larger networks and train them on larger data sets.

An **artificial neural network (ANN)**, usually called **neural network (NN)** [15], is a mathematical model or computational model that is inspired by the structure and/or functional aspects of biological neural networks. A neural network consists of an interconnected group of artificial neurons, and it processes information using a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. Modern neural networks are non-linear statistical data modeling tools. They are usually used to model complex relationships between inputs and outputs or to find patterns in data.

An iris-recognition algorithm first a discrete cosine transform is used in order to extract the spatial frequency range that contains the main abstract of iris image. The result is a set of complex numbers that carry local amplitude and phase information for the iris image which is used as training input for neural network.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

In this paper the matrix of discrete cosine transform of Edge of iris images has been used for Iris Recognition using Probabilistic neural networks following figure shows the structure of Iris Recognition system that has been design in this paper. The output of neural network is the class of iris which is name of the person. This system can classify the noisy iris image very well. The sobel edge detection has been used in this paper. Neural network that has been used in this paper is LVQ. LVQ, or Learning Vector Quantization, is a prototype-based supervised classification algorithm.LVQ can be understood as a special case of an artificial neural network.LVQ is one of the competitive neural networks. Following figure shows the structure of neural network:

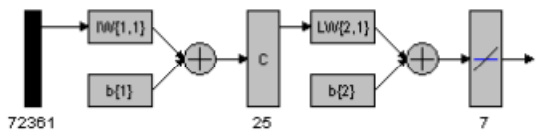


Fig. 2.8 Structure of Neural Network

2.3.6 Matching Score Calculation

In this step, we calculate the POC function $r^{f_{eff} g_{eff}}$ (n1, n2) between the aligned images f(n1, n2) and g(n1,n2) and we evaluate the matching score [19]. In the case of genuine matching, if the displacement between the two images is aligned, the correlation peak of the POC function should appear at the origin n1, n2 =(0,0) Thus, we calculate the matching score between the two images as the maximum peak value of the POC function within the small window O centered at the origin. The matching score is given by

$$\text{Matching score} = \underset{(n_1, n_2) \in O}{Max} \left\{ r_{fg}^{K_1 K_2}(n_1, n_2) \right\} \quad (9)$$

In our experiments, the size of the window O is 11x11 pixels. When multiple sub regions are extracted at the "Effective region extraction" process, as shown in figure 4.4b, the matching score is calculated by taking an average of matching scores for six sub regions.

3. Results

For result analysis in this paper, we are considering some parameters shown in table. Average segmentation time refer to average time taken by algorithm to partitioning a iris image into multiple segment of pixel. Segments accuracy rate refer as percentage of images correctly segmented by algorithm. Recognition accuracy means the probability that an authorized person is correctly recognized. It is the percentage of how much time a true iris image is recognized correctly.

Recognition time in our algorithm is referring as time taken by the algorithm to match the captured image with true imaged stored in database. A parameter we are considering for analysis is bulk detection. These parameters reveal the strength of LVQ Based algorithm. This tells us how our algorithm wills response in heavy load in term of high capturing rate and rich database. The reason behind good performance in heavy load is basic characteristic of neural network that is after a proper training it will response quickly. Last parameter we are considering here is algorithm complexity. Because of neural network LVQ Based algorithm is little bit complex.

Table 1:1 Result analysis

Parameters	POC Based	LVQ Based	Improvement with LVQ
Avg. Segmentation time	0.18 Sec.	0.14 Sec.	22 %
Segmentation Accuracy Rates	98.01	98.62	0.62 %
Recognition Accuracy	98.88	99.25	0.37 %
Recognition time	0.17 Sec.	0.07 Sec.	58 %
Algorithm Complexity	Low	Moderate	
Bulk Detection	Low	Best Suitable	

Experiment result analysis is shown in table 1.1.It is very clear with result that LVQ based algorithm has better performance than POC based method. Average segmentation time is reduced by 22 percent with LVQ based method. Similarly LVQ based iris recognition give better Segmentation Accuracy Rates and Recognition time. LVQ is best suitable where bulk detection is required. Because Learning Vector Quantization is a prototype-based supervised classification algorithm based on artificial neural network so once it is trained, it provides accurate and fast result. This is clearly shown by experiment. In figure 3.2, It is shown that the performance of ANN (LVQ) based algorithm is independent of input image load (number of iris need to process). Means LVQ based algorithm gives much better result in bulk detection. On the other hand performance is exponentially decrease as input image load is increase in POC based method. Similarly result can be found in figure 3.3 where matching time in LVQ based

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

method is considerable lower than POC based method (approximate 58% better than POC based method).
The simulation results of the implemented algorithm, all the simulation is performed with Intel P4 1.4 GHz processors with 1 GB of RAM, the OS is windows and simulation software is MATLAB 7.5(R2007b).

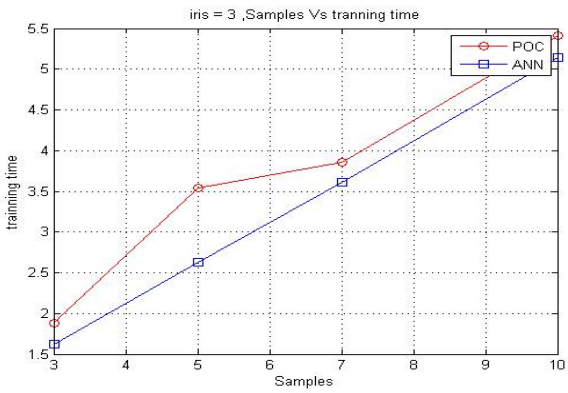


Figure 3.1 Samples vs. Training Time

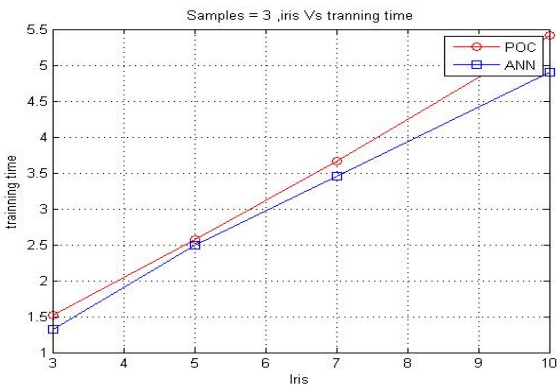


Figure 3.2 Number of Iris vs. Training Time

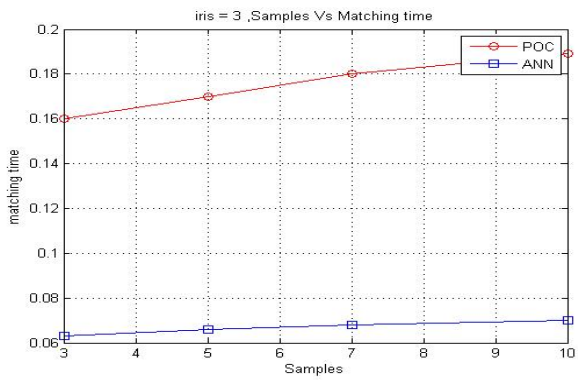


Figure 3.3 Samples vs. matching time

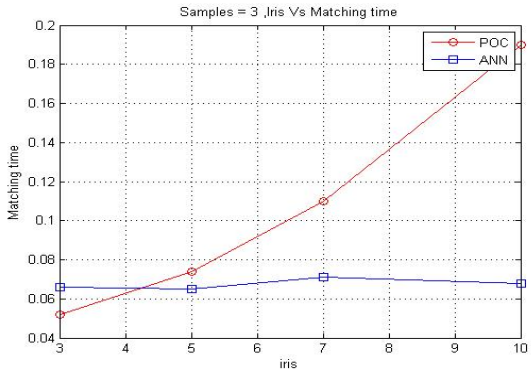


Figure 3.4 number of iris vs. matching time

4. Conclusions

Using neural networks a personal iris recognition system is presented in this paper. A fast iris matching method is proposed. Using this method, iris segmentation is performed in short time. Average time for iris segmentation is obtained to be 0.14 sec. Accuracy rate of iris segmentation 98.62% is achieved. The located iris after pre-processing is represented by a feature vector. Using this vector as input signal the neural network is used to recognize the iris patterns. The recognition accuracy for trained patterns is 99.25%. Matching time in LVQ based method is considerable lower than POC based method (approximate 58% better than POC based method). Proposed method also works considerable well for blurred or noisy images because of improved training method of Learning Vector Quantization (LVQ) neural network. This training makes learning process in such a way that it recognized correct images with lower error ratio. Once a system is proper trained then it improve recognition speed by 58%. So LVQ based method is responsible for both making system more robust (work well with noisy data) and improve accuracy. These characteristic make this system more appropriate for those application where bulk detection is required.

References

[1] C. L. Tisse, L. Martin, L. Torres, M.Robert, "Person identification technique using human iris recognition" ST Journal of System Research Current Issue 2003.

[2] R.Plemmonsa, M.Horvatha, E.Leonhardta, P. Paucaa, S. Prasadb, “Computational Imaging Systems for Iris Recognition”, Processing of Spie 2004

[3] L. Ma, U. Wang, T. Tan, "Iris Recognition Based on Multichannel Gabor Filtering”,The 5th Asian Conference on Computer Vision, 23-- 25 January 2002, Melbourne, Australia

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

[4] R. Narayan Swamy, P. E. X. Silveir, "Extended Depth-of-Field Iris Recognition System for a Workstation Environment" Proceedings of the Spied Biometric Technology for Human Identification II, Vol. 5779, pp. 41-50, Orlando, FL, 2005.

[5] Z. Wei, T. Tan and Z. Sun, "Synthesis of Large Realistic Iris Databases Using Patch-based Sampling", IEEE 2008.

[6] Czajka, A. Pacut, "Replay Attack Prevention for Iris Biometrics" ICCST 2008, IEEE.

[7] R. Wildes, "Iris Recognition: An Emerging Biometric Technology," Proc. IEEE, vol. 85, no. 9, pp.348- 1363, Sept. 1997.

[8] B.V.K. Vijaya Kumar, C. Xie, and J. Thornton, "Iris Verification Using Correlation Filters," Proc. Fourth Int'l Conf. Audio- and Video- Based Biometric Person Authentication, pp. 697-705, 2003.

[9] Z. Sun, T. Tan, and X. Qiu, "Graph Matching Iris Image Blocks with Local Binary Pattern," Advances in Biometrics, vol. 3832, pp. 366-372.

[10] C.D. Kuglin and D.C. Hines, "The Phase Correlation Alignment Method," Proc. Int'l Conf. cybernetics and Soc. '75, pp. 163 - 165, 1975.

[11] K. Takita, T. Aoki, Y. Sasaki, T. Higuchi, and K. Kobayashi, "High- Accuracy Subpixel Image Registration Based on Phase- Only Correlation," IEICE Trans. Fundamentals, vol. E86-A, no. 8, pp. 1925-1934, Aug. 2003.

[12] K. Takita, M.A. Muquit, T. Aoki, and T. Higuchi, "A Sub-Pixel Correspondence Search Technique for computer Vision Applications," IEICE Trans. Fundamentals, vol. 87-A, no. 8, pp. 1913-1923, Aug. 2004.

[13] Kazuyuki Miyazawa and Takafumi Aoki, An Effective Approach for Iris Recognition Using Phase-Based Image Matching. IEEE transactions on pattern analysis and machine intelligence, vol. 30, no. 10, October 2008.

[14] K. Ito, A. Morita, T. Aoki, T. Higuchi, H. Nakajima, and K. Kobayashi, "A Fingerprint Recognition Algorithm Using Phase- Based Image Matching for Low-Quality Fingerprints," Proc. 12th IEEE Int'l Conf. Image Processing, vol. 2, pp. 33-36, Sept. 2005.

[15] Zhou Zhiping, Hui Maomao, Sun Ziwen, "An Iris Recognition Based on 2DWPCA and Neural Network", 978 1-4244-2723- 9/09/$25.00© 2009 IEEE.

[16]CBSR,2005.Center for biometrics and security research. http://www.cbsr.ia.ac.cn/english/Databases.asp

[17] CASIA Iris Image Database, Inst. Automation, Chinese Academy of Sciences, http://www.sinobiometrics.com/, 2008.

[18]Univ. of Bath Iris Image Database, 2007, http://www.bath.ac.uk/elec- eng/research/sipg/irisweb.

[19] J. Daugman, "Biometric Personal Identification System Based on Iris Analysis", U.S. Patent No. 5,291,560, 1994.

Dr. Yogendra Kumar Jain presently working as Head of the Department, Computer Science & Engineering at Samrat Ashok Technological Institute Vidisha M.P India. The degree of B.E. (Hons) secured in E&I from SATI Vidisha in 1991, M.E. (Hons) in Digital Tech. & Instrumentation from SGSITS, DAVV Indore (M.P), India in 1999. The Ph. D. degree has been awarded from Rajiv Gandhi Technical University, Bhopal (M.P.) India in 2010.
Research Interest includes Image Processing, Image compression, Network Security, Watermarking, Data Mining. Published more than 50 Research papers in various Journals/Conferences, which include 30 research papers in International Journals.

Manoj Kumar Verma received the B.E. Degree in Computer Science and Engineering from MITS, Gwalior in 2001.He obtained M.Tech. Degree in Computer Science and Engineering from SATI, Vidisha, Rajeev Gandhi Technical University, Bhopal, India in 2011. Research interest includes Pattern recognition, Image Processing, Biometrics, Multimedia applications and Web Designing. He is working as a PGT (Computer Science) in KV, Trivandrum, India.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

A modified Kernelized Fuzzy C-Means algorithm for noisy images segmentation: Application to MRI images

Abdenour Mekhmoukh[1], Karim Mokrani[2] and Mohamed Cheriet[3]

[1] LTII Laboratory, Electrical Engineering department, University of Bejaia, Algeria

[2] LTII Laboratory, Electrical Engineering department, University of Bejaia, Algeria

[3] Department of Automated Manufacturing Engineering , ETS, Montréal, Québec, Canada

Abstract

Image segmentation is a low-level processing operation; it is the basis for many applications such as industrial vision and medical imaging. Segmentation provides a partition of the image by gathering pixels with similar grey levels in the same class. The main problem of this algorithm is that it does not take into account the image topology; it is based only on the pixels values. Thus, it is very sensitive to noise and inhomogeneities in the image moreover, it remains dependent on the initialization of the cluster centers.

In general the clustering algorithm chooses the initial centers in a random manner but the cluster centers initialization, using "Expectation Maximization" algorithm allows an optimal choice of these centers. To account for the topology of the image, the statistical parameters of a window around the pixel are considered. These a priori are used in the optimization of the cost function. The application to MRI images from the Brain Web public database, with different noise levels, shows the performance of the proposed approach.

Keywords: *image segmentation, FCM, KFCM, clustering, MRI images.*

1. Introduction

The magnetic resonance imaging (MRI) is one of the most significant advances in medicine; it allows, for the physician, the diagnosis and the monitoring of various diseases [1]. Given the size of the data to be analyzed, the use of image processing techniques can provide an aid to the diagnosis. Moreover, MRI images suffer from inhomogeneities and noise. Due to these problems, classical, or net, approaches do not succeed to correctly segment these images. Classification techniques, known from unsupervised artificial intelligence, fuzzy logic, probability and statistics, know a great success. In this paper, we propose to apply these techniques to segment MRI brain images.

2. Segmentation by unsupervised classifiers

2.1 Net segmentation

Classification methods are used to group objects into clusters or classes of homogeneous objects. The objects sets share common features; they are similar but are clearly distinguished from objects of other classes. The items listed are the pixels of the image that allows us to have groups. Generally, the classification methods lead to more or less different results. K-Means algorithm [2] is used because of its implementation simplicity.

2.1.1 K-Means algorithm

The K-means algorithm aims to collect N points in C or primitive groups whose number is predetermined. Given a set of points $X = x_1, x_2, \ldots, x_N$, each xj is characterized by n variables, and thus can be represented in an n-dimensional space $x_j \in R^n$.

As part of the clustering, we generally seek to partition the space into compact classes that are separated from each other. The K-Means algorithm aims to minimize the intra-class variance, which results in the minimization of the following objective function:

$$J(X,V) = \frac{1}{2}\sum_{i=1}^{C}\sum_{j=1}^{N}\left\|x_j - v_i\right\|^2 \qquad (1)$$

Where:

- v_i is the center of class i.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

- $\|x_j - v_i\|^2$ represents the Euclidean norm (clusters) which measures the dissimilarity between a pixel and a center.

2.1.2 Choice of the number of classes

The classical K-Means algorithm leaves a free parameter, the number of classes which, in the case of image segmentation, is the number of pixel intensities used to represent the image. Usually the choice of K is made empirically by selecting the value that minimizes Eq 1. This segmentation method considers that the regions are subsets of the net set of pixels of the image. Reconstruction of the regions is made by assigning each pixel to a single region. This approach has limitations when we do not have an a priori knowledge on the pixels allocation.

Fuzzy segmentation methods assume that the pixels belong more or less to these subsets and consider that the pixels can be grouped into fuzzy sets of pixels. These methods are based on the concepts of fuzzy logic.

2.2 Fuzzy segmentation

Fuzzy segmentation is build on the concepts of fuzzy logic [3] by considering the pixels as fuzzy sets. Methods such as C-Means and Fuzzy C-Means (FCM) [4], Possibilistic C-Means (PCM) [5] and Possibilistic Fuzzy C-Means (PFCM) [6] are often used.

2.2.1C-Means algorithm

Let $E = \{x_1, x_2,\ldots,x_k\}$ be the set of samples of the training set, x_k is the kth attribute vector x. The sets of prototypes or cluster centers are $\{V_1, V_2,., V_c\}$ where c denotes the number of classes. The C-Means algorithms minimize any objective function J (sum of the distances intra-classes). The Euclidean distance d is calculated between the attribute data vector and the group prototypes. The objective function J has the form:

$$J = \sum_{i=1}^{c} \sum_{k=1}^{n} \mu_{jk} d^2(x_k, V_i) \tag{2}$$

Where:

- v_i is the center of class i.
- $\mu_{jk} = P(c_j/x_k)$: usually represents the membership probability of x_k to class c_j

2.2.2 Fuzzy C-Means algorithm

Let E a set of vectors of attributes, $\{V_1, V_2,., V_c\}$ the centers of the classes, where c denotes the number of classes. The degree of membership of an element x_k in a group V_i is noted $\mu_i(x_k)$. In this approach, a given attribute vector can belong to several groups. U is the matrix of membership degrees (also called C-fuzzy partition matrix) whose size is $c*n$, where c is the number of classes and n the number of items to classify. The unsupervised clustering Fuzzy C Means (FCM) algorithm is a method of partitioning based on the total Picard iteration through necessary conditions for optimizing a sum of squared errors objective function (Jm) [7]. It is based on the sum of the distances between the vector and weighted classes center by the membership functions:

$$J_m(U, V) = \sum_{i=1}^{c} \sum_{k=1}^{n} \mu_{ik}^m d^2(x_k, V_i) \tag{3}$$

Where:

- $d = \|x_k - V_i\|$ is the Euclidean norm and V_i represents the center of class i.
- m is any real number greater than 1.

$$\forall i, j \quad \mu(x_k) \in \{0,1\} \quad U = [\mu_{ik}(x_k)] \tag{4}$$

$$U = \begin{pmatrix} \mu_{11} & \mu_{12} & . & . & \mu_{1n} \\ \mu_{21} & . & . & . & \mu_{2n} \\ . & . & . & . & . \\ . & . & . & . & . \\ \mu_{c1} & . & . & . & \mu_{cn} \end{pmatrix} \tag{5}$$

Three constraints are defined for the membership matrix of degrees U:

$$0 \le \mu_{ik} \le 1 \quad 1 \le i \le c \quad 1 \le k \le n \tag{6}$$

$$\sum_{i=1}^{c} \mu_{ik} = 1, \forall k \in [1, n] \tag{7}$$

$$0 \le \sum_{i=1}^{c} \mu_{ik} \le 1, \forall i \in [1, c] \tag{8}$$

The FCM is based on the update of the membership function. During the algorithm iterations, the FCM changes the partition (U matrix) by minimizing the objective function J_m (Algorithm 1).

$$\mu_{ik} = \frac{1}{\sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}} \right)^{\frac{2}{m-1}}} \tag{9}$$

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

where :

- μ_{ik} is the degree of membership of x_i in the cluster k

The cluster centers V_i are calculated by

$$V_i = \frac{\sum_{i=1}^{n} \mu_{ik}^m x_k}{\sum_{k=1}^{n} \mu_{ik}^m} \quad (10)$$

The process stops when $\left| J_m^{t+1} - J_m^t \right| < \varepsilon$ or a predefined number of iterations is reached.

Algorithm 1 FCM implementation

Step 1 :
– Set the number of class: c
– Set m ,
– Choose ε to stop the test.
– Initialize (t = 0) matrix $U^t = 0$ by random values.
– Choose the norm d_{ik} .

Step 2: Compute the new cluster centers using Eq (10)

Step 3 : Update U^t using :

$$\mu_{ik}^{t+1} = \frac{1}{\sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}} \right)^{\frac{2}{m-1}}} \quad (11)$$

Compare J^{t+1} to J^t if $\left| J_m^{t+1} - J_m^t \right| < \varepsilon$ then stop the process, else t = t + 1 and return to step 2.

2.2.3 Kernelized Fuzzy C-Means algorithm

By using the popular 'kernel method', we construct a kernel version of FCM (KFCM), where the original Euclidian distance in FCM is replaced with the following kernel-induced distance measures [8] :

$$d(x, y) = \left\| \Phi(x) - \Phi(y) \right\| \quad (12)$$

$$d(x, y) = \sqrt{K(x, x) - 2K(x, y) + K(y, y)} \quad (13)$$

Here $\boldsymbol{\Phi}$ is a nonlinear function which maps $\boldsymbol{x_k}$ from the input space $\boldsymbol{X}$ to a new space $\boldsymbol{F}$ with higher or even infinite dimensions. $\boldsymbol{K(x, y)}$ is the kernel function which is defined as the inner product in the new space $\boldsymbol{F}$ with:

$$\boldsymbol{K(x, y)} = \boldsymbol{\Phi(x)^T \Phi(y)}, \text{ for } \boldsymbol{x}, \boldsymbol{y} \text{ in input space } \boldsymbol{X}.$$

An important fact about kernel function is that it can be directly constructed in the original input space without knowing the form of $\boldsymbol{\Phi}$. That is, a kernel function implicitly defines a nonlinear mapping function. There are several typical kernel functions, e.g. the Gaussian kernel:

$$\boldsymbol{K(x, y) = exp\left(-\frac{\|x-y\|^2}{\sigma^2}\right)} \quad (14)$$

By replacing the Euclidean distance in Eq (3) with the Gaussian kernel-induced distance, the new objective function is:

$$J_m(U, V) = \sum_{i=1}^{c} \sum_{k=1}^{n} \mu_{ik}^m \left\| \Phi(x_k) - \Phi(V_i) \right\|^2 \quad (15)$$

$$\boldsymbol{J_m(U, V) = 2 \sum_{i=1}^{c} \sum_{k=1}^{n} \mu_{ik}^m (1 - K(x_k, V_i))} \quad (16)$$

The membership degrees terms become

$$\mu_{ik} = \frac{(1-K(x_k,V_i))^{\frac{-1}{m-1}}}{\sum_{j=1}^{c}(1-K(x_k,V_j))^{\frac{-1}{m-1}}} \quad (17)$$

And cluster centers

$$\boldsymbol{V_i = \frac{\sum_{k=1}^{n} \mu_{ik}^m K(x_k, V_i) x_k}{\sum_{k=1}^{n} \mu_{ik}^m K(x_k, V_i)}} \quad (18)$$

2.2.4 Modified Kernelized Fuzzy C-Means algorithm

Although KFCM can be directly applied to image segmentation like FCM, we propose to modify the algorithm by taking into account the image topology. For the Modified KFCM (MKFCM) we propose:

1. To initialize the cluster centers by using the "Expectation Maximization" (EM) algorithm [9] for an optimal choice of the centers.

2. To take into account the image topology; the statistical parameters of a window around the pixel are considered.

For an image y, of size $(N \times M)$, and a sliding window of size $(p \times p)$, the four features extracted from the window

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

centered at pixel *(n, r)* are given by the following equations:

$$Me = \frac{1}{MN}\sum_{i=-\frac{p-1}{2}}^{\frac{p-1}{2}}\sum_{j=-\frac{p-1}{2}}^{\frac{p-1}{2}} y(n+i, r+j), \quad (19)$$

$$V = \frac{1}{MN}\sum_{i=-\frac{p-1}{2}}^{\frac{p-1}{2}}\sum_{j=-\frac{p-1}{2}}^{\frac{p-1}{2}} (y(n+i, r+j) - Me)^2 , \quad (20)$$

$$Sk = \frac{1}{MN}\sum_{i=-\frac{p-1}{2}}^{\frac{p-1}{2}}\sum_{j=-\frac{p-1}{2}}^{\frac{p-1}{2}} (y(n+i, r+j) - Me)^3, \quad (21)$$

$$Ku = \frac{1}{MN}\sum_{i=-\frac{p-1}{2}}^{\frac{p-1}{2}}\sum_{j=-\frac{p-1}{2}}^{\frac{p-1}{2}} (y(n+i, r+j) - Me)^4, \quad (22)$$

Where, *Me, V, SK* and *Ku*, are the mean, the variance, the skewness, and the kurtosis respectively. Note that *p* must be odd for a window centered on each pixel. Thus, we obtain a matrix *H* that contains the extracted features. H is obtained as shown in the following figure, pixel by pixel, from left to right and top to bottom.

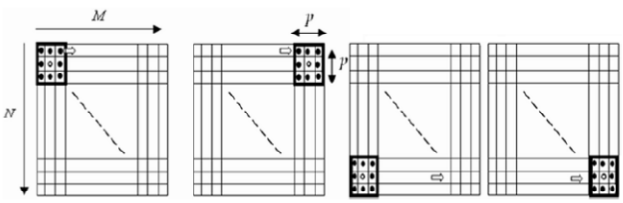


Fig 1. Extraction of statistical features.

So the equation (17) and (18) become

$$\mu_{ik} = \frac{(1-K(H_k,V_i))^{\frac{-1}{m-1}}}{\sum_{j=1}^{c}(1-K(H_k,V_j))^{\frac{-1}{m-1}}} \quad (23)$$

$$V_i = \frac{\sum_{i=1}^{n}\mu_{ik}^m K(H_k,V_i)H_k}{\sum_{k=1}^{n}\mu_{ik}^m K(H_k,V_i)} \quad (24)$$

The steps of the algorithm are summarized in the following diagram:

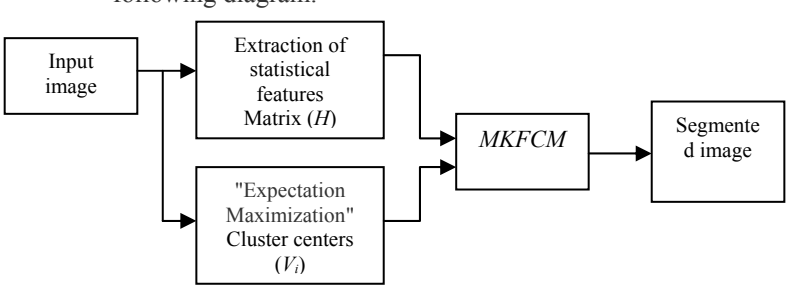


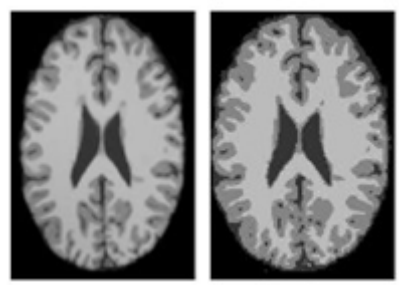
Fig 2. Proposed segmentation process.

3. Results and discussion

To apply the classification approach described above, we use the Brainweb database [10] of the Neurological Institute of McGill University in Montreal. The Brainweb website simulates brain MRI with different levels of noise and inhomogeneities. These simulations are generated from a ground truth which is the classification of a brain image (millimeter resolution) into different classes corresponding to different tissues and structures.

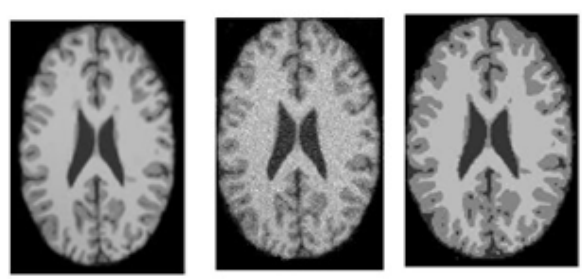
For our evaluations, we have thus made different data sets, varying noise levels. A particular set consists of T1-weighting for each noise level (0%, 5% and 9%). To segment brain tissues in MRI images, we must define the different parameters governing the algorithm, namely the values of m, the number of classes C, and the image pixels vectors form. Our goal is to segment the brain, we set the number of class *i* (c = 4) corresponding to the three brain tissue present in the brain (white matter, grey matter, cerebrospinal fluid and background).

The choice of the vectors form is crucial since their relevance will allow the pixels discrimination. The vector form x_j of a pixel *j* is its grey level.



Original image Segmented image (FCM)

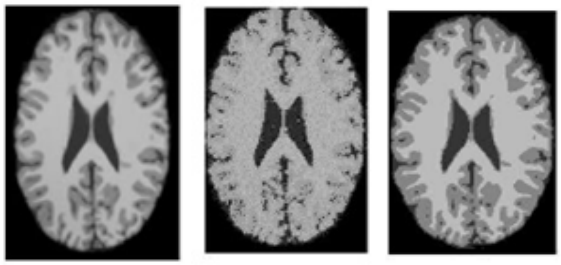
Fig 3. Results for an axial slice T_1 with 0% of noise with classical FCM



Original image FCM MKFCM

Fig 4. Results for an axial slice T_1 with 5% of noise

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.



Original image FCM MKFCM

Fig 5. Results for an axial slice T_1 with 9% of noise

Figure. 4 and 5 show a comparison of segmentation results between FCM and MKFCM, when applied on T1-weighted MR phantom corrupted with noise. From these images, we can see that traditional FCM is unable to correctly segment the images.

MKFCM segmented the image into three classes corresponding to background, grey matter (GM) and white matter (WM) and cerebrospinal fluid (CSF). MKFCM produced better results than FCM due to its ability to cope with noise.

Two types of cluster validity functions, fuzzy partition and feature structure, are often used to evaluate the performance of clustering. The representative functions for the fuzzy partition are partition coefficient *Vpc* [11] and partition entropy *Vpe* [12]. They are defined as:

$$V_{pc} = \frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^2 \quad (25)$$

$$V_{pe} = \frac{-1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}\, log\, u_{ij} \quad (26)$$

The idea of these validity functions is that the partition with less fuzziness means better performance. As a result, the best clustering is achieved when the value *Vpc* is maximum or *Vpe* minimum.

Table 1. Results for the image corrupted by 5% of noise

Algorithm	*Vpc*	*Vpe*
FCM	0.8520	0.2891
MKFCM	0.9693	0.0835

Table 2. Results for the image corrupted by 9% of noise

Algorithm	*Vpc*	*Vpe*
FCM	0.8293	0.2949
MKFCM	0.9484	0.0975

4. Conclusion

The results show that the MKFCM algorithm outperforms the FCM. Although the problem of image segmentation is a problem that is far from being resolved, segmentation result, as defined, is obviously not unique. The choice of a method is related to several factors and parameters adjustment that governs the algorithm deserves special attention. Finally, it seems interesting to consider the integration of other constraints on the pixels spatial arrangement and to combine several classification algorithms working in cooperation. Cooperation strategies should be considered in order to achieve faster convergence and improve the segmentation results.

References

[1] N.A. Mohamed, “Modified fuzzy C-mean algorithm for medical image segmentation,” M.Sc. thesis, Elect. Eng. Dept., Univ. Louisville, KY, 1999.

[2] J. B. MacQueen, "Some Methods for classification and Analysis of Multivariate Observations”, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, 1971.

[3] L.A Zadeh , ”Fuzzy sets, Information and Control”, 8,1965 pp 338-353.

[4] J.P. Stitt, R.L.Tutwiler and A.S.Lewis,”Fuzzy C-Means image segmentation of side scan sonar images”, Proceeding of, International Symposium on Signal and Image Processing, IASTED 2001, vol. 1, pp. 27-32, 2001.

[5] R. Krishnapuram and J.M Keller, ”A possibilistic approach to clustering”. IEEE Transactions on Fuzzy Systems, 1(2), pp.98-110, 1993.

[6] Nikhil R. Pal, Kuhu Pal, James M. Keller, and James C. Bezdek, ”A possibilistic fuzzy c-means clustering algorithm”, IEEE Transactions on Fuzzy Systems, 13(4), pp.517–530, 1997.

[7] J.C. Bezdek, J. Hathaway, M.J. Sabin, and W. T. Tucker, ”Convergence Theory for Fuzzy C-Means: Counterexamples and Repairs”, IEEE Trans, 17, pp. 873-877, Sept./Oct. 1987

[8] D.Q. Zhang, and S.C. Chen, "A novel kernelized fuzzy c-means algorithm with application in medical image segmentation," Artif. Intell. Med, vol.32, pp.37–50, 2004.

[9] Geoffrey McLachlan and Thriyambakam Krishnan. The EM Algorithm and Extensions. John Wiley & Sons, New York, 1996.

[10] BrainWeb: “Simulated Brain Database”, McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill.

[11] J. C. Bezdek. Cluster validity with fuzzy sets. Cybernetics and Systems, 3(3), pp. 58–73, 1973

[12] J. C. Bezdek. Mathematical models for systematic and taxonomy. Proc. of Eight International Conference on Numerical Taxonomy, San Francisco, pp. 143–166, 1975.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Gossip Based Routing Protocol Design for Ad Hoc Networks

Toqeer Mahmood[1], Dr. Tabassam Nawaz[2], Rehan Ashraf[3] and Syed M. Adnan Shah[4]

**[1] Ph.D. Research Scholar, Department of Computer Engineering,
University of Engineering and Technology, Taxila (47050) Pakistan**

**[2] Chairman, Department of Software Engineering,
University of Engineering and Technology, Taxila (47050) Pakistan**

**[3] Ph.D. Research Scholar, Department of Computer Engineering,
University of Engineering and Technology, Taxila (47050) Pakistan**

**[4] Ph.D. Research Scholar, Department of Computer Engineering,
University of Engineering and Technology, Taxila (47050) Pakistan**

Abstract

A spontaneously mannered decentralized network with no formal infrastructure and limited in temporal and spatial extent where each node communicate with each other over a wireless channel and is willing to forward data for other nodes is called as Wireless Ad Hoc network. In this research study, we proposed a routing strategy based on gossip based routing approach that follows the proactive routing with some treatment for wireless Ad Hoc network. The analytical verification of our proposed idea shows that it is a better approach based on gossip routing.

Keywords: *Decentralized, Ad Hoc Network, Gossip, Proactive Routing, Flooding.*

1. Introduction

By late nineties the term "Ad Hoc Network" was adopted by IEEE 802.11 subcommittee. Ad hoc network is a self-organized network and is a spontaneously mannered decentralized multi-hop wireless network and have many potential application in real life [1]. In Ad Hoc wireless networks each node forwards data to other nodes willingly. Each node communicates through wireless transmission of limited range in this kind of network, sometimes requiring the use of intermediate nodes to reach a destination. Usually the nodes are limited in power supply and bandwidth. Thus it complicates the system's mobility Ad hoc networks inherently have snags for instance, inappropriate boundary coverage demarcation, unpredictable time in-varying asymmetric channel, lossy link etc. like the other wireless communication networks have. With all these snags the issues like dynamic change in topology, multi hop network, location awareness, node state of agreement [2], [3], [4] and channel vulnerability were also there. So that wireless protocols simple in nature might not be a better choice for adaptation in Ad hoc networks without customization.

1.1 Example of Ad Hoc Routing protocols

Many researchers have been presented number of Ad Hoc routing protocols like LAR [5], DSR [6], AODV [7], DBF [8], GPSR [9], DREAM [10], ZRP [11], TORA [12] and a lot more.

The protocols like LAR [5], GPSR [9] and DREAM [10] assumes the nodes are provided GPS hardware and their location is known to them while some protocols like DSR [6], AODV [7], ZRP [11] and TORA [12] do not assumes this. The protocols that do not use GPS employ flooding with certain techniques of optimization.

2. What is routing?

The process of path selection and directing packets from a network source node to the destination node is called Routing and is an active area of research in ad hoc networks.

In a network source node sends packet to its neighboring node(s). This packet is further passed on to the node(s) closer to the destination node(s) [13]. This procedure is called forwarding and continues until the packet reaches to the destination. There may exist one-to-many or many-to-one or many-to-many relationship between source and

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

destination nodes. Due to which routing may follow static or dynamic path as well as it may be Unipath or Multipath depending on the designed algorithm and required scenario.

A routing algorithm determines the most appropriate path between the source and destination nodes. In their delivery semantics, routing schemes may differ that are defined as under [Figure 1]:

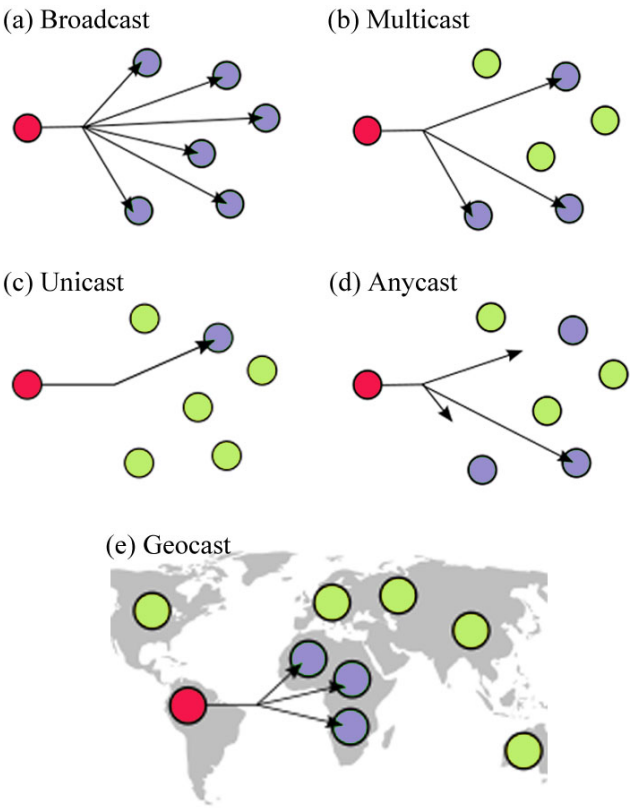


Fig. 1 Routing Schemes for Wireless Ad Hoc Network

2.1 Broadcast

Transmit the packet(s) to all nodes simultaneously [Figure 1(a)].

2.2 Multicast

Transmit the packet(s) to a specified group of interested nodes [Figure 1(b)].

2.3 Unicast

Transmit the packet(s) to a single specified node [Figure 1(c)].

2.4 Anycast

Transmit the packet(s) to any one nearest to the source out of a group of nodes [Figure 1(d)].

2.5 Geocast

Transmit the packet(s) to a group of destination in geographical locations [Figure 1(e)]

3. Flooding Routing Techniques

There exists a saturated container of routing algorithms but it still seems to "ask for more" to satisfy the requirements of a lot of hungry applications. The simplest one is to flood the network by sending an incoming packet to all the neighbors. As long as there does not exist deep network partitioning, the packet is sure to reach the destination.

Many ad hoc routing protocols use some kind of flooding technique to route the packets over the network. Royer et al. [14] classifies the routing protocols as:

3.1 Proactive Routing Protocol

It ensures maintaining the updated routing information and instantly provides the same on need. It is also called as Table-Driven Routing Protocols and is preferable in static network topology. Proactive routing protocol is used in DSDV [15] and OLSR [16].

3.2 Reactive Routing Protocol

It does not maintain updated routing table all the time but the routes are computed on demand. It is also called as On-Demand Routing Protocols. It is preferable in the scenario of frequently changing network topology. Reactive routing protocol is used in, DSR [6], AODV [7] and TORA [12].

This classification is of the time when the Wireless Ad Hoc Network Technology was at its infant stage and was not self-sufficient in the field of protocols.

With flooding every node in the network receives a message and retransmits that message exactly one. In ad hoc networks flooding has several helpful properties such as maximal coverage, distance preservation and redundancy. The refined flooding technique is to sending only the newly received packet and is called as controlled flooding.

A close counterpart of flooding technique is Gossip. Gossip [17] is a probabilistic retransmission algorithm. Gossip is same as flooding but with important distinction. The main objective of gossip protocol is to minimize the number of retransmissions, while maintaining the main benefits of flooding [18]. In gossip, when first time a node receives a packet, it starts a probabilistic process on that received packet to determine whether or not to retransmit. Essentially, gossip protocol retransmits a packet with some probability.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

A generic gossip protocol is given as under:

For every node u do

If *u has a rumor* then

v ← Choice (); /* Choice of the receiver V

Send (*v*); /* "PUSH" the rumor to V

Gossip routing states the forwarding of packets to an arbitrary node on the basis of some calculated probability. Thus here the overhead is reduced as compared to flooding with improved performance [19], [20]. The packet randomly traverses the network in the hope of ultimately finding the destination. This technique is also termed as random walk routing. Bimodal behavior which is a well-known concept in the Percolation theory is the key observation for about all the routing methods following gossiping technique [21], [22], [23].

Probabilistic nature of Gossip protocol enables a node to select its partner node randomly, with which it wants to communicate. Every node sends only a limited number of messages, regardless of the number of nodes in the network making it scalable. A node receives copies of a message from different nodes in a network due to which fault tolerance is achieved. A node does not perform a particular role. If a node fails does not disrupt others from sending messages that is why failure detection or any recovery mechanism is not required.

4. Proposed Idea

Our proposed idea is for the network where deployment of nodes is in a static fashion [Figure 2]. Every node knows about its parent node that is why flooding can be controlled resulting in adapting the gossip routing technique. How much parents a node may have but it only transmit the packet to three of its parents. Although, a node may be a parent of multiple nodes yet it follows some specific strategy to lessen the burden. On receiving data from multiple child nodes, parent node performs some specific calculation on the receiving data and transmits the resulted data to its parent node instead of transmitting data received from its child nodes. This process continues and the resulted data ultimately catches the destination.

The optimal parent selection criterion is that the node which has maximum energy as well as in the coverage area is selected as a parent node in the network. More-over our proposed idea uses the proactive routing with some treatment. Routes are updated in advance but only when energy of the node touches the threshold level.

In Ad Hoc network energy of nodes is a constraint factor; therefore to efficiently manage the energy, updating is on some specific event instead of frequently updating the values. Also each node has information of more than one neighbor nodes. So at the time of failure, best alternate path can be adopted for further transmission of packet to the way of destination.

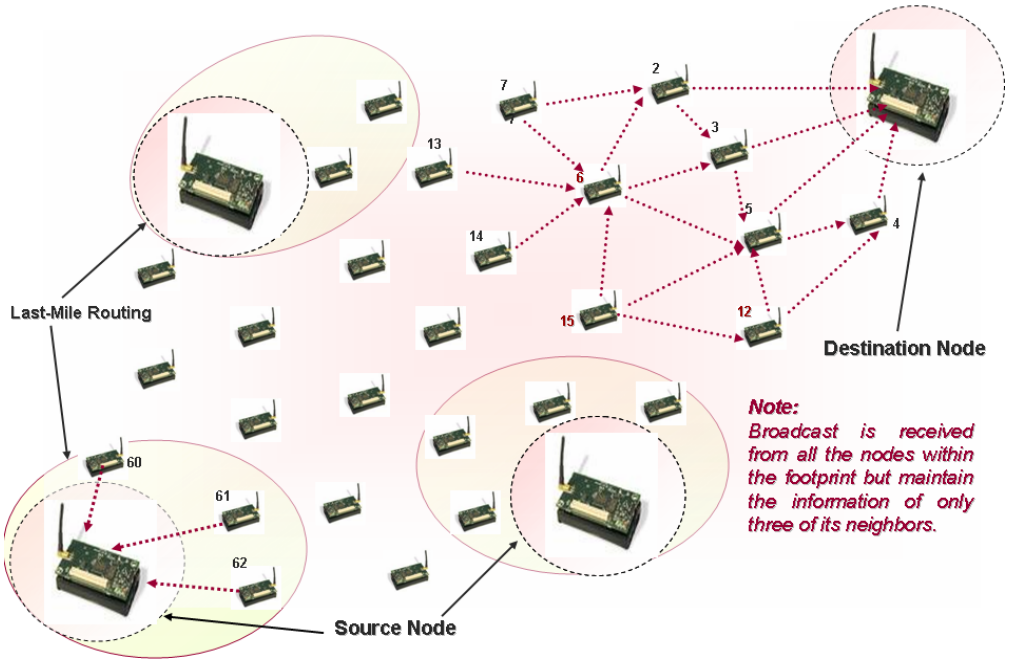


Fig. 2 A Typical Wireless Ad Hoc Sensor Network Static Deployment Scenario

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

5. Analytical Verification of Proposed Idea

Flooding is not only the wastage of precious resources of Ad Hoc network but also over burden the network. Gossip routing treats the above mentioned issue of flooding techniques in a most suitable manner. Our proposed idea presents the solution in a way that it only forwards the packet to the node which has maximum energy among its neighboring nodes. Moreover our idea uses the proactive routing with some treatment although proactive routing normally wastes the network energy due to its time to time updating of network table and others. However, proactive routing is preferable in static network topology while in the scenario of frequently changing network topology; reactive routing is a suitable option. But our proposed algorithm only updates the route in advance only when energy of the node touches the threshold level. This threshold level is defined each time after transmitting specific number of packet. On touching the defined threshold level, another threshold level is defined and on specific number of packets transmission counting that level is touched and network energy updating event occur. As a result another threshold level is defined. This process not only save the network energy but also reduces the burden of the network. A typical calculation for optimum forwarding node selection at node # 6 is given [Figure 3].

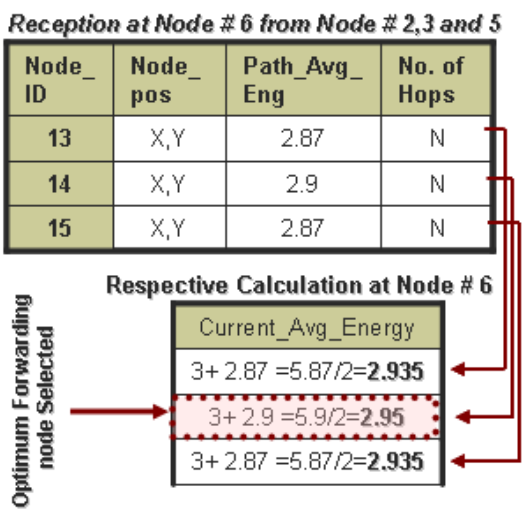


Fig. 3 Calculation for Optimum Forwarding Node

6. Conclusion

In static network topology, routing strategy based on gossip based routing following the proactive routing is a better approach. Flooding can be controlled by adapting the gossip routing technique. It saves the network energy also reduces the burden of the network.

References

[1] M. Gerla, "From battlefields to urban grids: New research challenges in ad hoc wireless networks", Pervasive Mobile Comput. Elsevier 1(1), 2005.

[2] C. C. Moallemi, and B. V. Roy, "Consensus propagation", IEEE Trans. Inf. Theory, Vol. 52, No. 11, 2006, pp. 4753-4766.

[3] W. Ren, and R. Beard, "Consensus Seeking in Multiagent Systems Under Dynamically Changing Interaction Topologies", IEEE Trans. Autom. Control, Vol. 50, No. 5, 2005, pp. 655–661.

[4] R. Olfati-Saber, and R. Murray, "Consensus Problems in Networks of Agents with Switching Topology and Time Delays", IEEE Trans. Autom. Control, Vol. 49, No. 9, 2004, pp. 1520–1533.

[5] Y. B. Ko, and N. H. Vaidya. "Location-aided Routing (LAR) in Mobile Ad Hoc Networks", In Proc. Fourth Annual CM/IEEE International Conference on Mobile Computing and Networking, 1998, pp. 66-75.

[6] D. B. Johnson, and D. A. Maltz, "Dynamic Source Routing in Ad Hoc Wireless Network," T. Imielinkski and H. Korth, Eds., Mobile Computing. Kluwer Academy, 1996, pp. 153–181.

[7] C. E. Perkins, and E.M. Royer, "Ad-hoc on Demand Distance Vector Router," 2nd IEEE Workshop on Mobile Computing System and Applications, New Orleans, Louisiana, USA, 1999.

[8] D. P. Bertsekas, and R. G. Gallager, "Distributed asynchronous bellmanford algorithm", in Data Networks. Prentice Hall, Englewood Cliffs, 1987.

[9] B. Karp, and H. T. Kung, "Greedy Perimeter Stateless Routing (GPSR) for Wireless Networks", In Proc. Sixth Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom), 2000, pp. 243-254.

[10] S. Basagni, I. Chlamtac, V. R. Syrotiuk, and B. A. Woodward, "A distance routing effect algorithm for mobility (DREAM)", In Proc. Fourth Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom), 1998, pp. 76–84.

[11] Z. Haas, and M. Pearlman, "The performance of query control schemes for the zone routing protocol", In Proc. ACM SIGCOMM, 1998, pp. 167–177.

[12] V. Park, and M.S. Corson, "A Highly Adaptive Distributed Routing Algorithm for Mobile Wireless Networks", In Proc. Infocom, 1997, pp. 1405-1413.

[13] Dazhi Chen, and Pramod K. Varshney, "Geographic Routing in Wireless Ad Hoc Networks", Book Chapter "Guide to Wireless Ad Hoc Networks", Computer Communications and Networks, pp. 151-188, 2009.

[14] E. M. Royer, and Chai-Keong Toh, "A Review of Current Routing Protocols for Ad Hoc Mobile Wireless Network", IEEE Personal Communication, 1999.

[15] C. E. Perkins, and P. Bhagwat, "Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for Mobile Computers", Association of Computing Machinery, 1994.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

[16] T. Clausen, and P. Jacquet, "Optimized Link State Routing Protocol (OLSR)", Network Working Group, 2003.
[17] X. Hou, and David Tipper, "Gossip-based Sleep Protocol (GSP) for Energy Efficient Routing in Wireless Ad Hoc Networks", IEEE, 2004.
[18] S. Rajeswari, and Dr. Y. Venkataramani, "An Adaptive Energy Efficient and Reliable Gossip Routing Protocol For Mobile Ad Hoc Networks", IJCTE, 2010, Vol. 2, No. 5 pp. 1793-8201.
[19] Z. J. Haas, J. Y. Halpern, and Li (Erran) Li, "Gossip-Based Ad Hoc Routing", IEEE/ACM Transactions on Networking, Vol. 14, No. 3, 2006, pp. 479-491.
[20] F. Wuhib, M. Dam, and R. Stadler, "A Gossiping Protocol for Detecting Global Threshold Crossings", IEEE Transactions on Network and Service Management, Vol. 7, No. 1, 2010 pp.1-16.
[21] P. Hall, "On continuum percolation", The Annals of Probability, Vol. 13, No. 4, 1985.
[22] R. Meester, and R. Roy, Continuum Percolation, Cambridge University Press, 1996.
[23] T. Ishikawa, and T. Hayakawa, "Gossip Protocol on the Ad Hoc Networks and Its Approximated Saturation", 49th IEEE Conference on Decision and Control, Hilton Atlanta Hotel, Atlanta, GA, USA, 2010, pp. 2638-2643.

Toqeer Mahmood completed his MS Computer Engineering in 2010 from Center for Advanced Studies in Engineering (CASE), Islamabad–Pakistan. He is currently pursuing his Ph.D in Computer Networks from University of Engineering and Technology (UET), Taxila – Pakistan and currently working as Programmer at UET Taxila. His areas of interests are Computer Networks, Mobile Networks, Numerical Techniques and Database Management Systems (DBMS).

Dr. Tabassam Nawaz completed his MS Computer Engineering in 2005 from CASE (Center for Advanced Studies in Engineering), Islamabad, Pakistan and subsequently he completed his Ph.D in 2008. He has published number of research papers in different Journals. His research areas include Software Engineering, Programming Languages, Data Structure, Computer Graphics, Networks and Digital Image Processing. He is the Chairman of Software Engineering Department at University of Engineering and Technology Taxila, Pakistan.

Rehan Ashraf completed his MS Computer Engineering in 2010 from Center for Advanced Studies in Engineering (CASE), Islamabad–Pakistan. He is currently pursuing his Ph.D in Digital Image Processing from University of Engineering and Technology (UET), Taxila. His areas of interests are Digital Image Processing, Computer Networks, Mobile Networks and Numerical Techniques.

Syed Muhammad Adnan Shah completed his MS Computer Engineering in 2009 from Center for Advanced Studies in Engineering (CASE), Islamabad–Pakistan. He is currently pursuing his Ph.D in Computer Networks from University of Engineering and Technology (UET), Taxila–Pakistan and currently working as Web Manager at UET Taxila. His areas of interest are Computer Networks, Mobiles Networks and Web Technologies.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Design and Implementation of Mobile IPv6 Data Communication in Dual Networks

Chaman Singh[1] , Sanjeev Kumar[2], Sanjeev Kumar[3] and K.L.Bansal[4]

[1] Department of Computer Application, Govt. P.G. College, Chamba
Affiliated to H.P.University , Shimla , Himachal Pradesh 176310, India

[2]Department of Journalism and Mass Communication, Govt. P.G. College, Chamba
Affiliated to H.P.University , Shimla , Himachal Pradesh 176310, India

[3]Department of Physics, Govt. P.G. College, Chamba
Affiliated to H.P.University , Shimla , Himachal Pradesh 176310, India

[4] Department of Computer Science, H.P.University, Shimla
Himachal Pradesh 171005, India

Abstract

Dual Stack Mobile IPv6 is an extension of Mobile IPv6 to support mobility of devices irrespective of IPv4 and IPv6 Networks. This is implemented by combining different modules. IPv4 private networks are behind Network Address Translation (NAT) devices. So, to bypass the Binding Update and Binding Acknowledgment by NAT, we need to encapsulate it in User Datagram Protocol (UDP) Packets. So, the Dual Stack Mobile IPv6 should support NAT Traversal and Detection. With the support of NAT Detection and Traversal Module in Dual Stack Mobile IPv6, the mobile node is able to move freely from IPv6 Network to IPv4 Network or vice-versa. The main objective of not breaking the connectivity at the time of switching from one network to other is accomplished by NAT Module in Dual Stack Mobile IPv6.

Keywords: *Dual Stack, Network Address Translation, Detection, Tunnelling, Home Agent, Mobility.*

1. Introduction

IPv6 is introduced generally to resolve the address space issues and also provides several advanced features. Dual Stack Mobile IPv6 is an extension of Mobile IPv6 to support mobility of devices irrespective of IPv4 and IPv6 network. The application interface is required to exchange mobility information with Mobility subsystem [1]. Mobile IPv6 (MIPv6) is a protocol developed as a subset of Internet Protocol version 6 (IPv6) [2] to support mobile connections. MIPv6 [3] allows a mobile node to transparently maintain connections while moving from one subnet to another. The Mobile IPv6 protocol takes care of binding addresses between Home Agent (HA) and Mobile Node (MN). It also ensures that the Mobile Node is always reachable through Home Agent. Each mobile node is always identified by its home address [4], regardless of its current point of attachment to the Internet. While situated away from its home, a mobile node is also associated with a care-of address, which provides information about its current point of attachment to the Internet. The protocol provides for registering the care-of address with a home agent. The home agent sends datagram's destined for the mobile node through a tunnel to the care-of address. After arriving at the end of the tunnel, each datagram is then delivered to the mobile node. Currently, two mobility management protocols are defined for IPv4 and IPv6. Deploying both in a dual stack mobile node introduces a number of problems. This has been improved [5]. Mobile IPv6 uses IP Security (IPSec) to protect signaling between the home agent and the mobile node [6] [22]. Generic Packet Tunneling [7] Specifies a method and generic mechanisms by which a packet is encapsulated and carried as payload within an IPv6 packet. The forwarding path between the source and destination of the tunnel packet is called an IPv6 tunnel. The technique is called IPv6 tunneling. The current Mobile IPv6 [3] and Network Mobility [8] [21] specifications support IPv6 only. These extend those standards to allow the registration of IPv4 addresses and prefixes, respectively, and the transport of both IPv4 and IPv6 packets over the tunnel to the Home Agent. [9] Allows the Mobile Node to roam over both IPv6 and IPv4, including the case where Network Address Translation is present on the path between the mobile node and its home agent. Middle Boxes [10], such as Firewalls and Network

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Address Translators (NAT) [11], are an important component for a majority of Internet Protocol (IP) [5] networks today. Current IP networks are predominantly based on IPv4 [5] technology, and hence various firewalls as well as Network Address Translators have been originally designed for these networks. NAT's are necessary to overcome the IPv4 address space limitations of many network domains.

A network address translator a box that interconnects a local network to the public internet, where the local network runs on a block of private IPv4 addresses [12]. In the original design of the internet architecture, each IP address was defined to be globally unique and globally reachable. In contrast, a private IPv4 address is meaningful only within the scope of the local network behind a NAT and as such, the same private address block can be reused in multiple local networks [20] , as long as those networks do not directly talk to each other. Instead, they communicate with each other and with the rest of internet through NAT boxes. Mobile IPv6 (MIPv6) [15] is a protocol developed as a subset of Internet Protocol veMyon 6 (IPv6) to support mobile connections. The impact to IPv4, which changes IP address semantics, provide ample evidence , since now coming time MIPv6 are in progress so need of network address translation traversal and detection on Dual Stack implementation of mobile IPv6 [16]. NEPL (NEMO Platform for Linux) [17] is a freely available implementation of DSMIPv6 for Linux platform. The original NEPL release was based on MIPL (Mobile IPv6 for Linux) [18]. Without the support of NAT Detection and Traversal module in DSMIPv6, the mobile node will not be able to move freely from IPv6 network to IPv4 network or vice-versa. Connectivity also breaks at the time of switching from one network to other will be accomplished by this research.

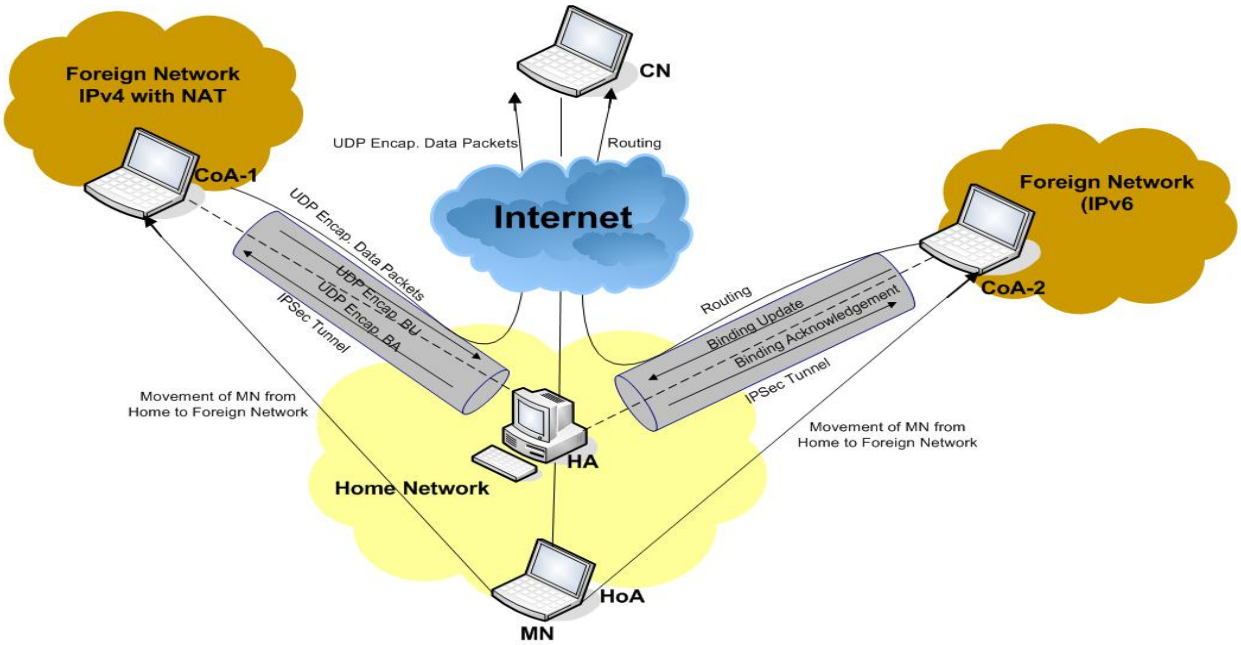


Figure 1:- Basic Function of Dual Stack Mobile IPv6

2. Dual Stack Implementation of Mobile IPv6

First, NEPL (NEMO Platform for Linux) [17] is a freely available implementation of DSMIPv6 for Linux platform. The original NEPL release was based on MIPL (Mobile IPv6 for Linux) [18]. In Figure-1: Basic Operation of DSMIPv6, all Mobile Nodes (MN) has a fixed address, called a Home Address (HoA), assigned by Home Agent. When the Mobile Node moves to other networks, it gets Care-of Address (CoA) from foreign network. Mobile Node sends a Binding Update (BU) message to its Home Agent. Then Home Agent replies to the Mobile Node with a Binding Acknowledgement (BA) message to confirm the request. When Mobile Node is moved to any foreign network all packets sent to the Home Agent will be IPSec encrypted. A bi-directional tunnel is established between the Home Agent and the Care of address of the Mobile Node after the binding information has been successfully

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

exchanged. DSMIPv6 extends the MIPv6 and NEMO [6] Basic Support standards to allow mobile nodes to roam in both IPv6 [23] and IPv4-only networks. The following features are supported by the DSMIPv6 Architecture:-

1. The mobile node can register an IPv6/IPv4 Care of address to its Home Agent and so roam in IPv6-only networks and IPv4-only networks by the use of IPv6 and IPv6-in-IPv4 tunnels between the Node and its Home Agent.
2. A Network Address Translation Detection and Traversal Mechanism allow the Mobile Node to communicate with its Home Agent even though it uses an IPv4 private address as a Care of address. The signaling messages are always UDP encapsulated in IPv4 network. However, when the Mobile Node is located behind a NAT, data traffic is also encapsulated in UDP.
3. Securing the signaling packets between Home Agent and Mobile Node when Mobile Node is moved to foreign network and Session management on movement from one foreign link to another.

3. Solution Description

Solution is an extension to the existing NEPL solution provided by Nautilus [17]. We validated DSMIPv6 functionality as per the requirements provided against the draft, along with other IETF standards. We took the baseline architecture implementation from the Nautilus6 which uses Linux platform. The below mentioned steps are taken by us to achieve the requirements:-

1. Setup DSMIPv6 Test Lab using Kernel 2.6.28.2 and UMIP veMyon 0.4. In order to test the basic functionality between Home Agent and Mobile Node according to [3] the Test Bed has been setup.
2. Code change in MIP6 Daemon and Linux kernel and also applied the open source patches/packages .The Routing Advertisement daemon (radvd), IPSec daemon (strongswan) and Web Server (httpd) daemon has been configured on Home Agent.
3. The Mobile Node is configured with IPSec daemon (strongswan). Mobile Node gets IPv6 address whenever it is moved to any IPv6 foreign network through the radvd server running on the router.
4. When Mobile Node is moved to IPv4 network, it gets configured with IPv4 Care of address from the DHCP server running on IPv4 Router. In IPv4 network, DHCP is configured on the private network behind router. The network behind IPv4 router can be public or private.

4. Modules Representation in Dual Stack Mobile IPv6

MIPL (Mobile IPv6 for Linux) [24] is an open-source implementation of the Mobile IPv6 standard for the GNU/Linux operating system. MIPv6 is a user space for Mobile Node and Home Agent which aims at providing the necessary changes to MIPL in order to run on the latest kernels.Figure-2: Block Diagram of MIPv6 shows the internal data flow between two major components i.e. Home Agent and Mobile Node. Both of these two components consist of several helper modules which are also shown in this figure. This section describes IPv4 address assignment mechanism used by DSMIPv6.DHCP DNA module is used to obtain IPv4 address from the DHCP server running on IPv4 network.

4.1 DNA / DHCP Module

Module describes IPv4 address assignment mechanism used by DSMIPv6. When Mobile Node moves to IPv4 FL (Foreign Link) and its egress interface becomes enabled, Mip6d code in Mobile Node listens for Router Advertisement message, and since it does not receives Router Advertisement message in IPv4 FL, it gets timeout and sends Router Solicitation message (that will request the router to generate the Router Advertisement message immediately rather than at there next scheduled time), and Mobile Node wait for some time interval for Router Advertisement message before repeating the same procedure of sending Router Solicitation message. Meanwhile after sending Router Solicitation message, mip6d daemon will check the presence of DHCP server on the Egress interface link of Mobile Node by sending the DHCP discover message and wait for DHCP offer packet. Since the DHCP server is running on the IPv4 FL, it gets the IPv4 address from DHCP server and then mip6d code maps IPv4 address to IPv6 address, which is further used as Care of address. Mip6d daemon sets the default route on Mobile Node figure 3.

4.2 Movement Detection Modules

Describe Movement Detection in DSMIPv6 implementation. The movement of a mobile node away from its home link is transparent to transport and higher-layer protocols and applications. Describe Movement Detection in DSMIPv6 implementations. The Movement detection uses Neighbor Unreachablility Detection [13] to detect when the default router is no longer bi-directionally reachable, in which case the mobile node must discover a new default router (usually on a new link). However, this

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

detection only occurs when the mobile node has packets to send. When the mobile node detects handover, it expire the previous routers and Care-of-Address (es) and selects a new default router as a consequence of Router Discovery, and then performs Prefix Discovery with that new router to form new care-of address (es). It then registers its new primary care-of address with its home agent. After updating its home registration, the mobile node then updates associated mobility bindings in correspondent nodes. It triggers a movement event and then detects that Mobile Node is in foreign network. Route is modified and Home Address which was previously assigned to the physical interface is now moved to the tunnels. This is handled in 'mn.c'.

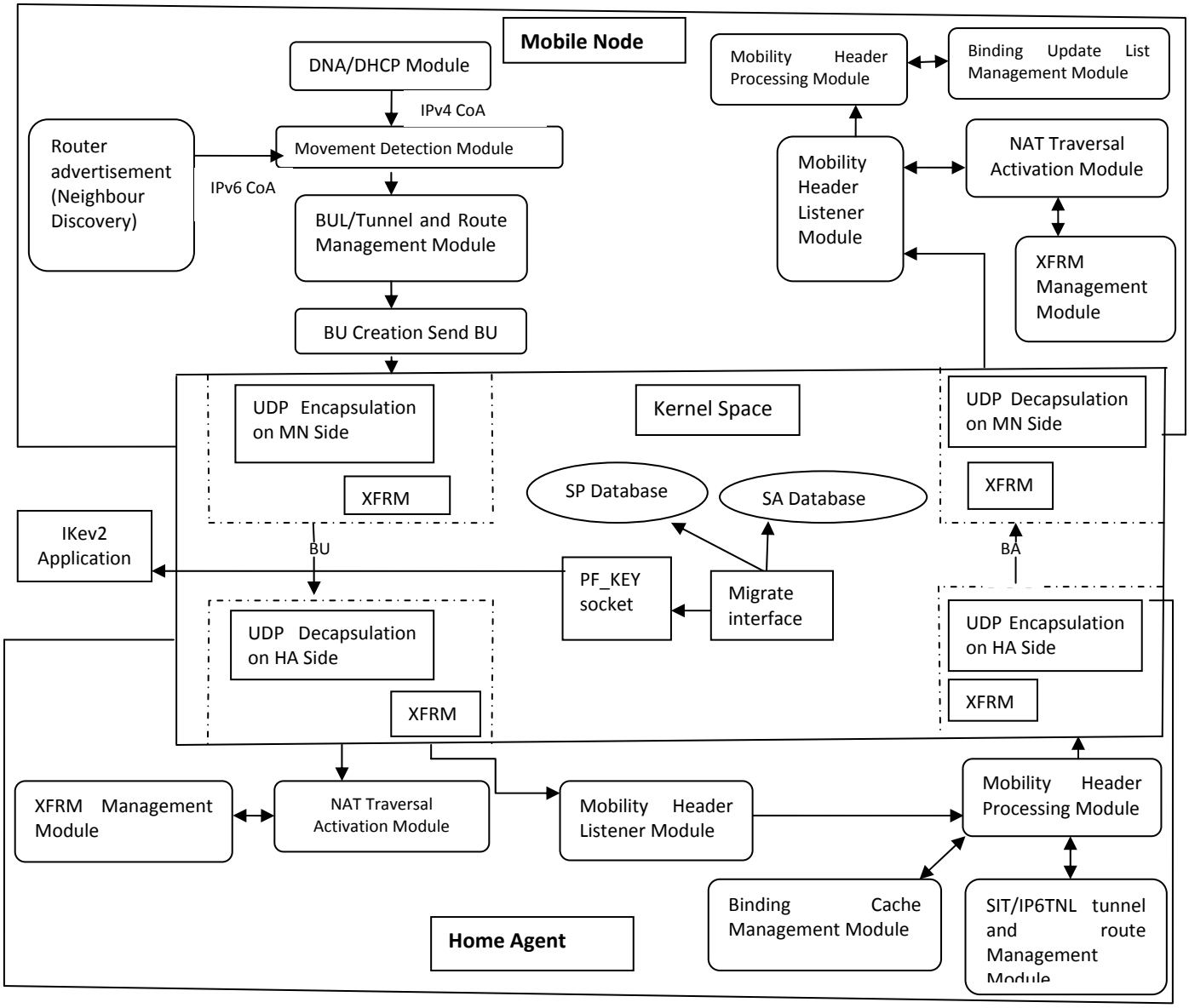


Figure 2: MIPL for Home Agent and Mobile Node

4.3 Binding Management Modules

Describe Binding Management in DSMIPv6 implementation. When a Mobile Node moves between different networks, it is essential that binding update messages are sent to that node's Home Agent and Correspondent Nodes as soon as possible, in order to facilitate a fast handoff. Mobile Nodes therefore cannot rely on the soft state timeout mechanism used in binding caches to refresh stale bindings maintained by Correspondent Nodes (typical binding lifetimes are of the order of minutes).

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

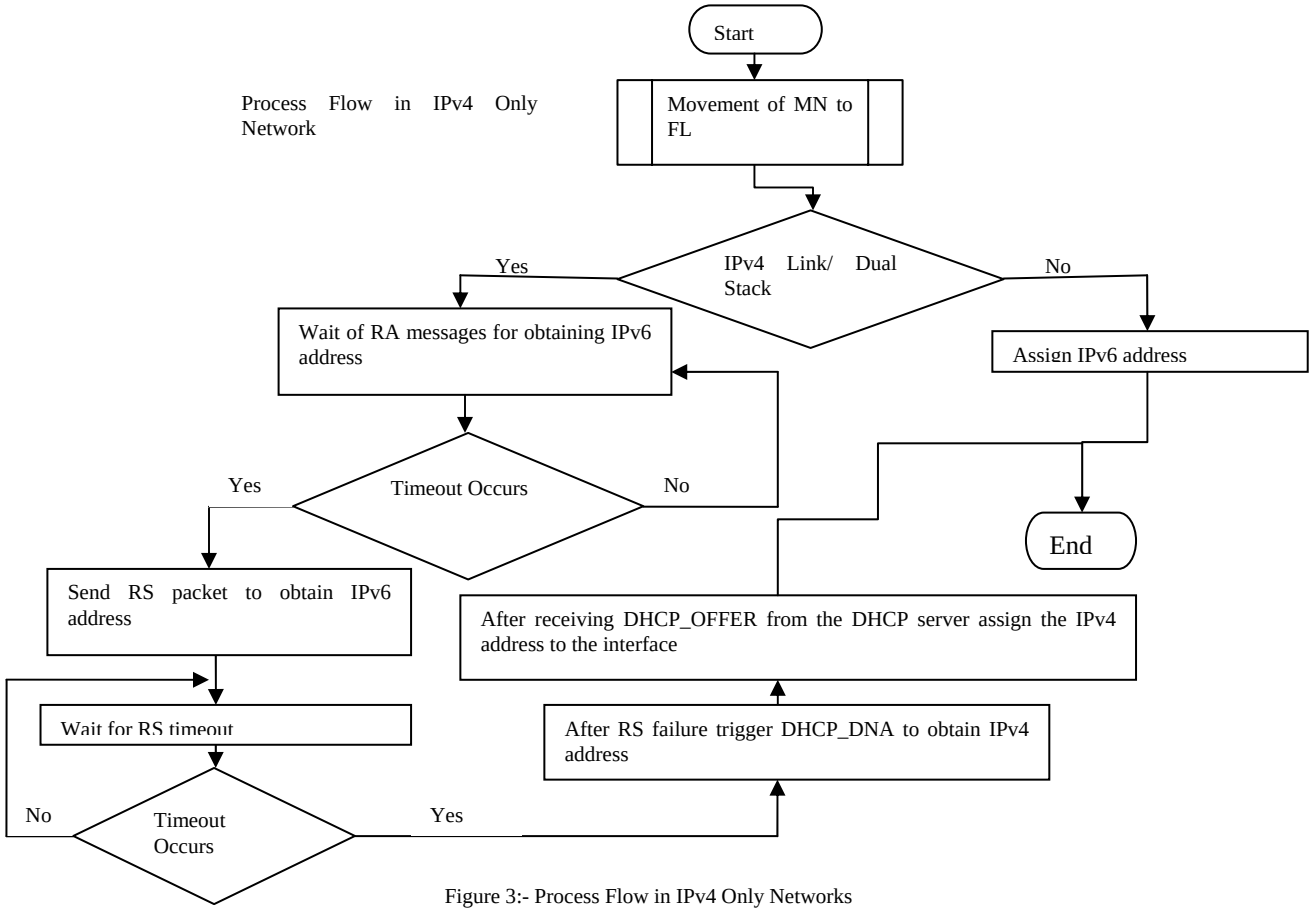


Figure 3:- Process Flow in IPv4 Only Networks

An additional data structure, the binding update list, is therefore kept by Mobile Nodes, which maintains state on any Correspondent Nodes or Home Agents. And in Home Agent mobile IPv6 binding cache maps home addresses to the current care-of addresses for each mobile node. This allows the home agent to tunnel traffic to the mobile node at its current location, and allows a correspondent node to send packets directly to a mobile node at its current location. Binding Management Module is classified as:

1. Binding Update List
2. Binding Cache

Binding Update: When Mobile Node moves to any Foreign Link, Mip6d code in Mobile Node initiate binding update functionality after configuring Care of address either from Router Advertisement message received in IPv6 or Dual stack FL or from DHCP server in IPv4 FL on the egress interface of Mobile Node. After Mobile Node configures Care of address on its egress interface, it creates BUL which consists of information about Mobile Node's Home Address, Care of address, CN address, life time and delay time of this binding message in seconds and set of various flag.

Binding Update List format is as follows:
== BUL_ENTRY ==
Home address 2001: x: x: x: x: x: x: x Care-of addresses 2001: x: x: x: x: x: x: x
CN address 2001: x: x: x: x: x: x: x Lifetime = 32, delay = 1500
Flags: IP6_MH_BU_HOME
IP6_MH_BU_ACK IP6_MH_BU_TLV

When Mip6d code in Mobile Node creates BUL successfully, it send Binding Update message to Home Agent to register the new Care of address of Mobile Node to Home Agent. So that Home Agent gets updated with current Care of address of Mobile Node.

Binding Cache: Mip6d code in Home Agent received Binding Update message sent by Mobile Node when it moves to any Foreign Link or when it directly boots in IPv6/IPv4 FL. Home Agent process BU by performing DAD (dynamic address discover)[14] and latter parse Binding Update message. After paMyng the binding update message Home Agent creates or updates its Binding Cache entry. Binding Cache entry consists of

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Mobile Node's Home Address, Mobile Node's Care of address, its local address, lifetime and sequence no.

Binding Cache entry format is as follows:

== BC_ENTRY ==

HoA 2001: a: b: 0:0:0:0:1 status registered CoA 2001: a: d: 1:20c:29ff:fea0:4026 flags AH-- Local 2001: a: b: 0:0:0:0:1000 Lifetime 23 / 32 sequence 3435 Unreach 0 / 959299 retry -2

After updating its Binding Cache entry, the mip6d code creates devices to tunnel traffic to Mobile Node. Mip6d code in Home Agent creates and send Binding Acknowledgment message to Mobile Node so that Mobile Node gets acknowledged that it's new Care of address gets successfully registers with Home Agent. When Mobile Node receives Binding Acknowledgment message, mip6d code in Mobile Node parses the BA packet and update the BUL entry. It checks for various options set in BA and proceed accordingly. If NAT is detected between Mobile Node and Home Agent, it set xfrm policies/states to UDP Encapsulate IPv6/IPv4 data traffic to bypass NAT. Mip6d daemon sets the callback function to resend the BA, once lifetime of BUL entry is expired. And finally set the binding update timer to decrease the lifetime of BUL entry. When Mobile Node moves back from any FL to Home Link, Mip6d code in Mobile Node sends Binding Update message with lifetime set as zero to Home Agent to indicate that it as returned to Home Link mip6d code in Mobile Node deletes corresponding BUL entry. On Home Agent side Mip6d code receives BU message with lifetime set as zero, it indicate that Mobile Node moved to Home Link, so it deletes corresponding Binding Cache entry and send Binding Acknowledgment back to Mobile Node.

4.4 Route and Tunnel Management Modules

Tunnel and Route Management module figure 4 is mainly responsible for tunneling, when mobile node changes from IPv6 to IPv6, IPv6 to IPv4 and vice versa network. This module configures sit and ip6tnl interface via IOCTL system call which in turns performs the task at kernel level. Route Management handles the return rout-ability with CN. Some data structures are being used between some of the important functions in Tunnel Management module. Some user land data structures used in various routines in tunnel management module. Some data structures used in various routines in tunnel management module.

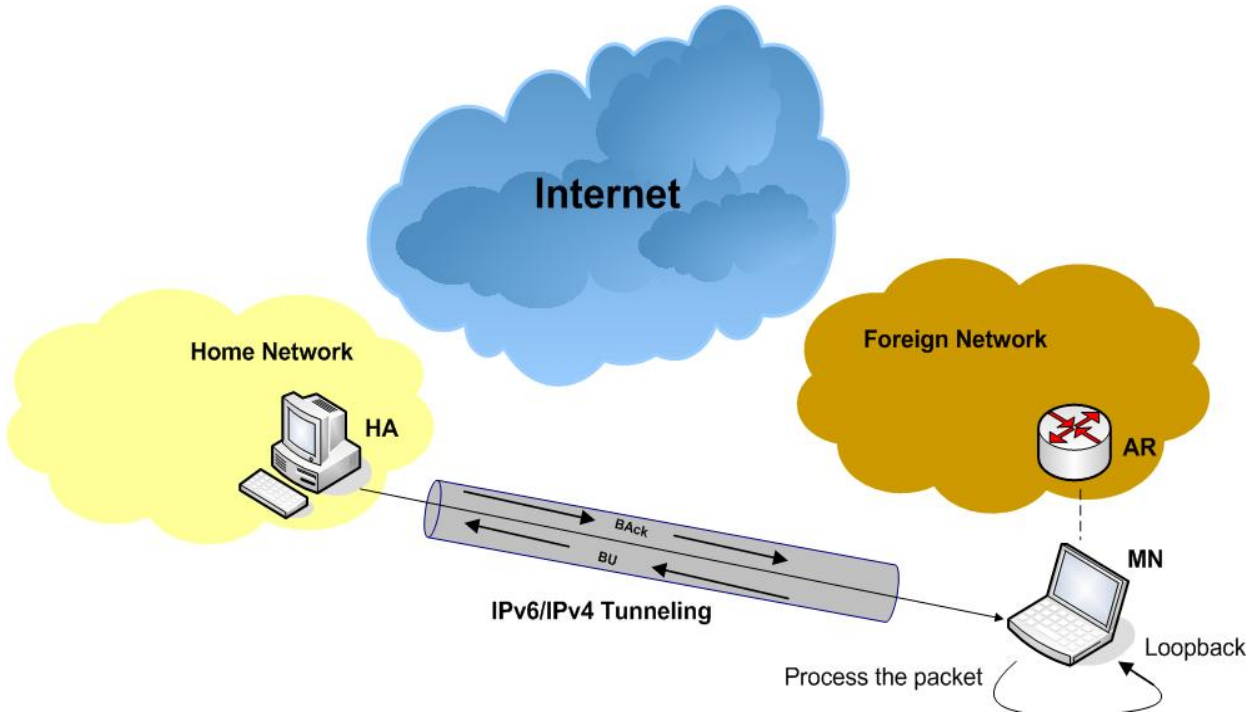


Figure 4 : Tunnel and Route

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

4.5 XFRM and IPSEC Management Modules

XFRM [15] is a packet transformation framework residing in the Linux kernel. It performs operations on IP packets such as inserting, modifying headers, UDP encapsulation and de-capsulation. DSMIPv6 XFRM module will take the advantage of existing IPSEC transformation and defines a simple UDP encapsulation scheme. IPSEC module is responsible for interaction with IKE through MIGRATE messages. IPSec will be used to protect the following traffic between Home Agent and Mobile Node.

1. BU/BA messages.
2. Mobile prefix sollicitation and advertisement messages.
3. Normal traffic between Mobile Node and Home Agent.
4. All tunneled normal traffic between Mobile Node and correspondent Node.

In Mip6d, the Mobile Node and the Home Agent uses IPsec Security Associations in transport mode to protect BU/BA messages, since the MN may change its attachment point to the Internet, it is necessary to update its endpoint address of the IPsec SAs. This indicates that corresponding entry in IPsec databases (Security Policy and SA databases) should be updated when Mobile Node performs movements. IPSec is used to protect the following traffic between Home Agent and Mobile Node. When Mobile Node move in FL a new Care of address is assign to it by FL network. After detecting the movement following steps are taken to create IPSec tunnel.

1. Mip6d issues a PF_KEY MIGRATE message to the PF_KEY socket. The operating system validates the message and checks if corresponding security policy entry exists in SPD. When the message is confirmed to be valid, the target SPD entry is updated according to the MIGRATE message. If there is any target SA found that is also target of the update, those should also be updated.
2. After the MIGRATE message is successfully processed inside the kernel, it will be sent to all open PF_KEY sockets. The IKE daemon receives the MIGRATE message from its PF_KEY socket and updates its SPD and SAD images. The IKE daemon may also update its state to keep the IKE session alive. After that ESP protected BU is send with K–bit set.

Mobile IPv6 specifies a flag named Key Management Mobility Capability bit (K-bit) in Binding Update (BU) and Binding Acknowledgement (BA) messages, which indicates the ability of IKE sessions to survive movement. When both the Mobile Node and Home Agent agree to use this functionality, the IKE daemons dynamically update the IKE session when the Mobile Node moves. The following methods are used.

4.6 NAT Detection and Traversal Modules

NAT (Network Address Translation (figure 5)) is the translation of an Internet Protocol address (IP address) used within one network to a different IP address known within another network. One network is designated the inside network and the other is the outside. In DSMIPv6 the mip6d daemon should bypass NAT, when Mobile Node is behind NAT' ed device in IPv4 FL. NAT detection is done when the initial Binding Update message is sent from the mobile node to the home agent. When located in an IPv4-only foreign link, the mobile node sends the Binding Update message encapsulated in UDP (User Datagram Protocol) and IPv4; this is handled in xfrm.c file. The mip6d daemon adds xfrm policy/state for UDP encapsulation for BU packet. When the home agent receives the encapsulated Binding Update, it compares the IPv4 address of the source address field in the IPv4 header with the IPv4 address included in the IPv4 care-of address option. If the two addresses match, no NAT device is in the path. Otherwise, a NAT is detected in the path and the NAT detection option is included in the Binding Acknowledgement. The Binding Acknowledgement, and all future packets, is then encapsulated in UDP and IPv4. Note that the home agent also stores the port numbers and associates them with the mobile node's tunnel in order to forward future packets. This is handled in ha.c file. The mip6d daemon adds the xfrm polices/states for UDP encapsulation of BA and IPv6/IPv4 data traffic. Upon receiving the Binding Acknowledgement with the NAT detection option, the mobile node sets the tunnel to the home agent for UDP encapsulation. Hence, all future packets to the home agent are tunneled in UDP and IPv4. If no NAT device is detected in the path between the mobile node and the home agent then IPv4/IPv6 data traffic is not UDP encapsulated. A mobile node will always tunnel the Binding Updates in UDP when located in an IPv4-only network. Essentially, this process allows for perpetual NAT detection. Similarly, the home agent will encapsulate Binding Acknowledgements in a UDP header whenever the Binding Update is encapsulated in UDP. This is handled in mn.c and xfrm .c file. The mip6d daemon adds xfrm polices/states for UDP encapsulation of IPv6/IPv4 data traffic, when NAT is detected between Mobile Node and Home Agent.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

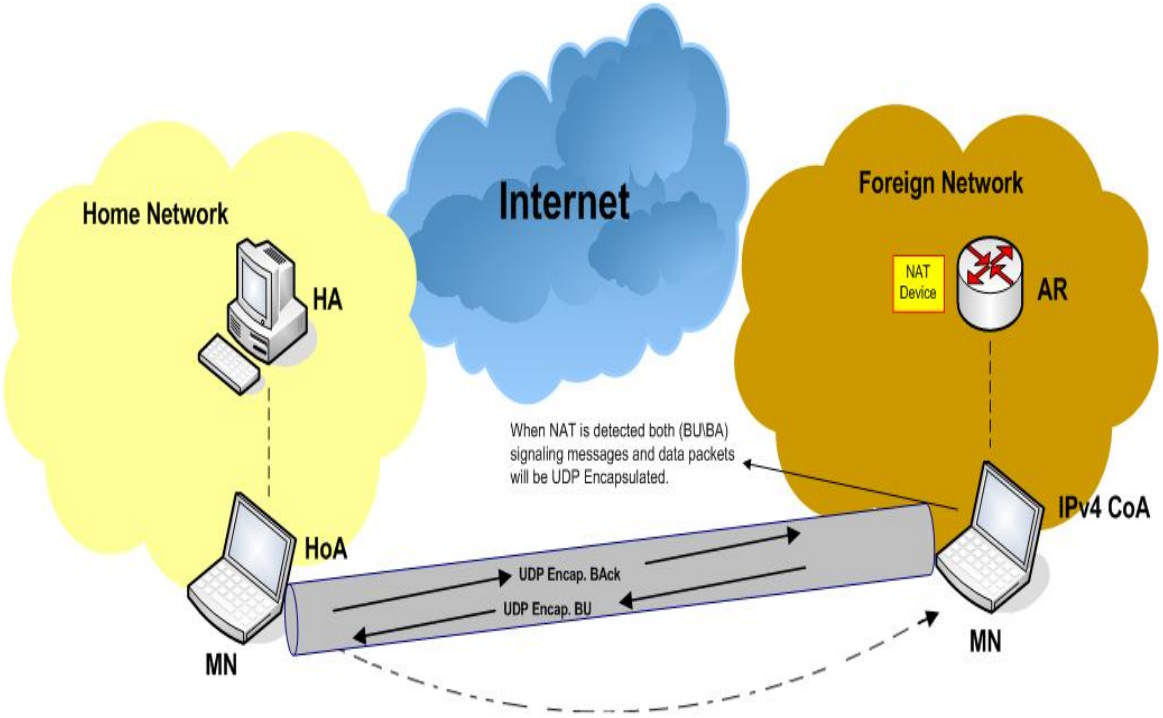


Figure-5: NAT Detection/Traversal

4.7 Mobility Listener Modules

The Mobility Header is an extension header used by mobile nodes, correspondent nodes, and home agents in all messaging related to the creation and management of bindings. The Mobility Header is identified by a Next Header value of 135 in the immediately preceding header. Header is used to carry various messages. A mobile node uses the Home Test Init message to initiate the return rout-ability procedure and request a home keygen token from a correspondent node. The Home Test Init message uses the MH Type value 1. The Home Test message is a response to the Home Test Init message, and is sent from the correspondent node to the mobile node. The Home Test message uses the MH Type value 3.A mobile node uses the Care-of Test Init message to initiate the return rout-ability procedure and request a care-of keygen token from a correspondent node. The Care-of Test Init message uses the MH Type value 2.The Care-of Test message is a response to the Care-of Test Init message, and is sent from the correspondent node to the mobile node. The Care-of Test message uses the MH Type value 4. The Binding Update message is used by a mobile node to notify their nodes of a new care-of address for itself. The Binding Update uses the MH Type value 5.The Binding Acknowledgement is used to acknowledge receipt of a Binding Update. The Binding Acknowledgement has the MH Type value 6.The Binding Refresh Request message requests a mobile node to update its mobility binding. The Binding Refresh Request message uses the MH Type value 0.The Binding Error message is used by the correspondent node to signal an error related to mobility, such as an inappropriate attempt to use the Home Address destination option without an existing binding. The Binding Error message uses the MH Type value 7.

5. Conclusions

This paper is one of the earliest attempts in the community to investigate the problems and impacts when middle boxes, especially NAT devices are placed in Dual stack Mobile IPv6are implemented in computer laboratory. With the support of modules in DSMIPv6, the mobile node is able to move freely from IPv6 network to IPv4 network or vice-versa. It accomplishes the main objective of not breaking the connectivity at the time of switching from one network to other. The transition from IPv4 to IPv6 will be time consuming process, so there will be time, when both IPv4 and IPv6 networks will be there and there will be always being scope for further development.

Acknowledgments

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

The authors would also like to express their sincere thanks to Mr. Manish Jamwal for his worthy Help and support for This Work.

References

[1]. T.Momoseetal, The Internet Engineering Task Force, July 2005 "The application interface to exchange mobility information with Mobility subsystem", Internet Drafts draft-momose-mip6-mipsock-00.
[2]. Vida, R. and L. Costa, Eds., RFC 3810, June 2004. "Multicast Listener Discovery VeMyon 2 (MLDv2) for IPv6".
[3]. Perkins, C., RFC 3344, August 2002. "IP Mobility Support for IPv4".
[4]. Johnson, D., Perkins, C., and J. Arkko, RFC 3775, June 2004. "Mobility Support in IPv6".
[5]. G.Tsirtsis, Qualcomm, H. Soliman, Elevate Technologies, [RFC 4977], August 2007. "Dual Stack Mobility".
[6]. Arkko, J., Devarapalli, V. and F. Dupont, RFC 3776, June 2004. "Using IPsec to Protect Mobile IPv6 Signaling Between Mobile Nodes and Home Agents".
[7]. Conta, A. and S. Deering, RFC 2473, December 1998. "Generic Packet Tunneling in IPv6 Specification".
[8]. F Ralf Spenneberg, ipsec-howto, 2003-08-18.
[9]. Soliman, Ed., Elevate Technologies, November 3, 2008. Mobile IPv6 Support for Dual Stack Hosts and Routers draft-ietf-mext-nemo-v4traversal-06.txt.
[10]. B. Carpenter and S. Brim. Middleboxes: Taxonomy and Issues. RFC 3234, Internet Engineering Task Force, February 2002.
[11]. K. Egevang and P. Francis. The IP Network Address Translator (NAT). RFC 1631, Internet Engineering Task Force, May 1994.
[12]. Y.Rekhter et al, "Address allocation for private Internets" RFC 1918, 1996.
[13]. J. Rosenberg et al., "STUN: Simple Traversal of User Datagram Protocol (UDP) through Network Address Translators (NATs)," RFC 3489, 2003.
[14]. Rosenberg, R. Mahy, and P. Matthews, "Traversal Using Relays around NAT (TURN)," draft-ietf-behave-turn-08, 2008.
[15]. C. Huitema, "Teredo: Tunnelling IPv6 over UDP through Network Address Translations (NATs)," RFC 4380, 2006.
[16]. K.L.Bansal, Chaman Singh, "Dual Stack Implementation of Mobile IPv6 Software Architecture", IJCA- Volume 25, No 9, July 2011.
[17]. NEPL (NEMO Platform for Linux) how to, June 24th, 2009.
[18]. MIPL (Mobile Ipv6 for Linux), how to, 2004-4-20.
[19]. K.L.Bansal, Chaman Singh, "NAT Traversal and Detection on Dual Stack Implementation of Mobile IPv6", IJCA-Volume 29, No 7, September2011.
[20]. L Cao, G J Qi, S F Tsai, M H Tsai, A D Pozo, T S Huang, et al." Multimedia Information Networks in Social Media" in Social Network Data Analytics -2011.
[21]. Mark S. Taylor, Willium Waung and Mohsen Banan, "Imternetwork Mobility – The CDPD Approach",Prentice Hall PTR Upper Saddle River NJ 07458ISBN-0-13-209963-5.
[22]. Christian Huitema, "IPv6- The New Internet Protocol", Prentice Hall PTR Upper Saddle River NJ 07458 ISBN 0-13-241936-X.
[23]. Scott O. Bradner, Allison Mankin ," IPng Internet Protocol Next Generation" Addison-Wesley IPng Series-Massachusetts 01867. ISBN 0-201-63395-7.
[24]. Viktor T.Toth ," Linux:-A Network Solution for Your Office", Sams Publication,USA ISBN 0-672-31628-5.

Chaman Singh is working as Assistant Professor Cum HOD Department of Computer Application, Govt. P.G. College, Chamba Himachal Pradesh University Shimla India. He received Master of Computer Application from Department of Computer Science H.P.University. He also qualified UGC-NET-2006 and Doctor of Philosophy in Computer Science is Under Submission from H.P.University. Having more then 4 years of working experience in Development, Teaching and Networks.

Sanjeev Kumar is working as Assistant Professor Cum HOD Department of Journalism and Mass Communication, Govt. P.G. College, Chamba Himachal Pradesh University Shimla India. He received Master of Philosophy in Journalism and Mass communication from H.P.University. Having more then 3 years of working experience in Teaching and Communication.

Sanjeev Kumar is working as Assistant Professor in Department of Physics Govt. P.G. College, Chamba Himachal Pradesh University Shimla India. He received Master of Philosophy in Physics from Punjabi University Patiala. Having more then 6 years of working experience in Physics and Teaching.

Dr. K.L. Bansal is working as Associate Professor in Department of Computer Science , H.P.University Shimla Himachal Pradesh. Received MCA from P.U. Chandigarh and Ph.D. from H.P.University Shimla Himachal Pradesh India. More then 15 year of experience in Teaching and Guidance.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

A Data Transfer in Wireless Sensor Networks Using AODV Protocol

B.Manimekala[1], M.Kayalvizhi[2]

[1] Assistant Professor, Department of MCA & Software Systems, V.L.B.Janakiammal College of Arts and Science, Coimbatore, TamilNadu, India

[2] Assistant Professor, Department of MCA & Software Systems, V.L.B.Janakiammal College of Arts and Science, Coimbatore, TamilNadu, India

Abstract

Sensor networks present unique opportunities for a broad spectrum of applications such as industrial automation, situation awareness, tactical surveillance for military application and environmental monitoring, chemical/biological detection etc. Sensor Network can monitor ambient condition such as temperature, sound, light and others. Information is collected from many sensor devices for further consumer application in the Sensor Network. For selecting a cluster head, k-means algorithm will be used to find the cluster center. For hop to hop packet forwarding, AODV protocol has been used at network layer. All the simulations of the proposed idea will be simulated on Berkeley's ns2 network simulator and the performance of the proposed scheme has been evaluated.

Keywords: Sensor network, AODV, NS2, Cluster.

1. Introduction

1.1 About Sensor Networks

Sensor networks present unique opportunities for a broad spectrum of applications such as industrial automation, situation awareness, tactical surveillance for military applications, environmental monitoring, chemical/biological detection etc. Microsensor systems enable the reliable monitoring and control of a variety of applications

1.2 Problem Definition

Wireless Sensor Network is one promising application on wireless ad hoc networks. Sensor Network can monitor ambient condition such as temperature, sound, light and others. Information is collected from many sensor devices for further consumer application in the Sensor Network.

The proposed clustering based architecture is similar to the hierarchical fusion architecture. The hierarchical management architecture can be applied to handle numerous sensor nodes. The lower-level nodes are managed and organized by the higher-level nodes by using 20/80 rules. The routing protocols for ad hoc networks can be divided into two categories: table-driven and on-demand routing protocols. The table-driven routing mechanisms discover and maintain routing tables even when no usage of network; in on-demand routing mechanisms, the routes are discovered only when it is need by the source node. These two routing categories have their own merits and demerits. This project a new cluster-based routing protocol that combines the on-demand routing in inter-cluster and the table-driven routing in intra-cluster to be suited for the environment of Sensor Networks.

1.3 About This Research

This project proposes a scheme for Self-Organization Management Protocols of higher-level nodes to contest member nodes with multi-hop form clusters, introduce the "20/80 Rule" for determining the ratio of headers to member nodes. This study implements the proposed management protocols including Clustering Mechanism for constructing cluster headers to solve the problems of clustering and broadcast storm, the suitable protocol to provide low cost communications between clusters.

The Location information of the sensor nodes will be collected from source and the cluster heads were elected/selected based on the locations of the active sensor nodes in network. For selecting a cluster head, k-means algorithm will be used to find the cluster center. The nearby sensor nodes of a cluster-head then will forward their data to sink only via the cluster-head. For hop to hop packet forwarding, AODV or DSDV protocol may be used at Network Layer Protocol. That is to forward a packet from a sensor node to a cluster-head or cluster head to sink; normal routing protocols may be used.

The proposed clustering based application layer

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

protocol reduces the number of management nodes in a large-scale sensor network. All the simulations of the proposed idea will be simulated on ns2 network simulator and the performance of the proposed scheme will be evaluated.

2. Sensor Network

A sensor network comprises of sensor nodes and a base station. Each sensor node is battery powered and equipped with:

1. Integrated sensors
2. Data processing capabilities
3. Short-range radio communications

2.1 Data Fusion in Sensor Network

In sensor networks, thousands of sensor nodes collectively monitor an area. These large sensor networks generate a substantial amount of data, yet the sensor nodes often have limited resources, such as computation power, memory, storage, communication, and most importantly, battery energy. The process wherein, sensors detect an event, and the data related to the event is eventually fused at the sink via multiple levels of fusion en route, is called a "round" of aggregation. Typically, the sink is distant from the area where the sensor nodes reside. The sensor data has to be ultimately relayed to the sink via multiple sensor node relays.

The large scale of sensor networks and the resource constraints make it an important challenge to design and develop efficient information processing and aggregation techniques to make effective use of the data.

Given a query, it may be unnecessary and inefficient to return all raw data collected from each sensor—instead, information should be processed and aggregated within the network and only processed and aggregated information is returned.

In such a setting, certain nodes in the sensor network, called aggregators, collect the raw information from the sensors, process it locally, and reply to the aggregate queries of a remote user.

Each sensor node has to decide on when to begin the process of fusion and how long to wait before the end of the fusion. Intuitively, the longer the time that the sensor node which performs the fusion waits, the larger the number of reports it will receive.

In this setting, we focus on stealthy attacks, where the attacker's goal is to make the Base Station accept false aggregation results, which are significantly different from the true results determined by the measured values.

2.2 Routing in Sensor Networks

There are a few inherent limitations of wireless or wired media such as low bandwidth, error prone transmissions, need for collision free channel access etc. These wireless nodes also have only a limited amount of energy available to them, since they derive energy from a personal battery and not from a constant power supply. Furthermore, since these sensor nodes are deployed in places where it is difficult to replace the nodes or their batteries, it is desirable to increase the longevity of the network. Also preferably all the nodes should die together so that we can replace all the nodes simultaneously in the whole area. Finding individual dead nodes and then replacing those nodes selectively would require pre-planned deployment and eliminate some advantages of these networks. Thus, the protocols designed for these networks must strategically distribute the dissipation of energy, which also increases the average life of the overall system.

2.3 The Proposed Network Architecture

The proposed mechanism reduces the number of management nodes in a large-scale sensor network. One of the best methods is to develop hierarchical architecture and apply the 20/80 rule [1].

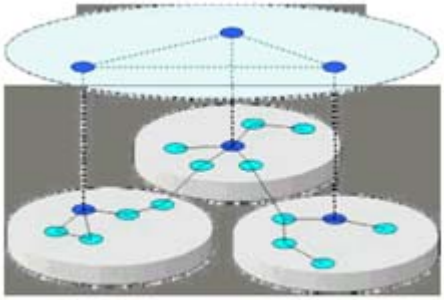


Fig 1: The Architecture of Clustering Network

Construct the intra-cluster routing with table-driven routing mechanisms similar to destination sequence distance vector (DSDV). The source node broadcasts a route query packet to nodes until the border nodes in adjacent clusters around the local cluster receive it. Therefore, the intra-cluster routing limits the broadcast range within one distance hop distance of the cluster to lower the control overhead and reduce the interference of the shared media. Figure 5 shows how the Intra-cluster routing builds.

Also, the proposed protocol acquires the inter-cluster routing by exchanging the relationship between clusters. To perform the inter-cluster route discovery when the route is demanded can reduce the overhead of building and maintaining inter-cluster routing. When the demand for inter- cluster

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

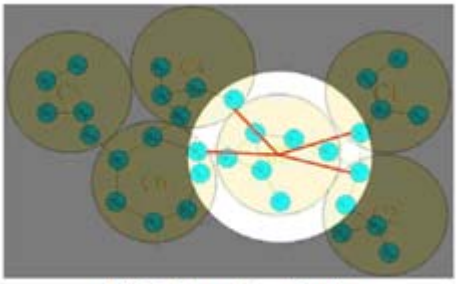


Fig 2: Intra-cluster Routing

route occurs, the source node sends the inter-cluster route request packet (RREQ) in unicast mode to the border nodes to acquire the adjacent cluster's intra-cluster routing information.

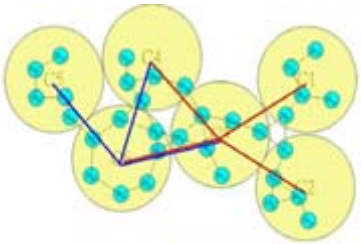


Fig 3: Inter-cluster Routing

After constructing the clustering network, this previous work proposes cluster-based routing algorithm into two aspects: intra-cluster routing and inter-cluster routing [1].

2.4 Classification of Sensor Networks

Sensor networks can be classified into two types based on their mode of operation or functionality and the type of target applications [15] [16].

Proactive Networks: In this scheme the nodes periodically switch on their sensors and transmitters, sense the environment and transmit the data of interest. Thus, they provide a snapshot of the relevant parameters at regular intervals and are well suited for applications that require periodic data monitoring.

Reactive Networks: In this scheme the nodes react immediately to sudden and drastic changes in the value of a sensed attribute and are well suited for time critical applications.

Once we have a network, we have to come up with protocols which efficiently route data from the nodes to the users, preferably using a suitable MAC (Medium Access Control) sub-layer protocol to avoid collisions.

2.5 Proactive Network Protocol

At each cluster change time, once the cluster-heads are decided, the cluster-head broadcasts the following parameters:

Report Time (T_R), is the time period between successive reports sent by a node.

Attributes (A), is a set of physical parameters which the user is interested in obtaining data about.

At every report time, the cluster members sense the parameters specified in the attributes and send the data to the cluster-head. The cluster-head aggregates this data and sends it to the base station or a higher level cluster-head, as the case may be. This ensures that the user has a complete picture of the entire area covered by the network.

Since the nodes switch off their sensors and transmitters at all times except the report times, the energy of the network is conserved. At every cluster change time, T_R and A are transmitted afresh and thus can be changed. By changing A and T_R, the user can decide what parameters to sense and how often to sense them. Also, different clusters can sense different attributes for different T_R. This scheme, however, has an important drawback. Because of the periodicity with which the data is sensed, it is possible that time critical data may reach the user only after the report time making this scheme ineffective for time-critical data sensing applications.

LEACH (Low-Energy Adaptive Clustering Hierarchy) is a family of protocols developed in [14]. LEACH is a good approximation of a proactive network protocol, with some minor differences.

Once the clusters are formed, the cluster heads broadcast a TDMA schedule giving the order in which the cluster members can transmit their data. The total time required to complete this schedule is called the frame time T_F. Every node in the cluster has its own slot in the frame, during which it transmits data to the cluster head. When the last node in the schedule has transmitted its data, the schedule is repeated.

The report time T_R discussed earlier is equivalent to the frame time T_F in LEACH. However T_F is not broadcasted by the cluster head but is derived from the TDMA schedule and hence is not under user control. Also, the attributes are predetermined and not changed after initial installation.

In a new network protocol called TEEN [15] (Threshold sensitive Energy Efficient sensor Network protocol) has been developed that is targeted at reactive networks and is the first protocol developed for reactive networks.

3. Simulation of Proposed Sensor Network

This section gives a brief overview of each simulator. Its aim is to summarize the different implementation approaches of each simulator. The way a new algorithm is integrated can be pretty different from

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

one simulator to another. The various simulators commonly available are: OPNET, NS-2,

OPNET Modeler is a powerful network simulator developed by OPNET. It can simulate all kinds of wired networks, and an 802.11 compliant MAC layer implementation is also provided. Although OPNET is rather intended for companies to diagnose or reorganize their network, it is possible to implement one's own algorithm by reusing a lot of existing components. Most part of the deployment is made through a hierarchical graphic user interface. Basically, the deployment process goes through the following phases.

First step it to choose and configure the node models (i.e. types) that is to be used in the simulations -for example a wireless node, a workstation, a firewall, a router, a web server, etc. Then build and organize the network by connecting the different entities. The last step consists in selecting the statistics needed to collect during the simulations.

The difficulty with OPNET Modeler is to build this state machine for each level of the protocol stack. It can be difficult to abstract such a state machine starting from a pseudo-coded algorithm.

But anyway, state machines are the most practical input for discrete simulators. In summary, it is possible to reuse a lot of existing components (MAC layer, transceivers, links, etc.) improving the deployment process. But on the other hand, any new feature must be described as a finite state machine which can be difficult to debug, extend and validate.

NS-2 is a discrete event network simulator that has begun in 1989 as a variant of the REAL network simulator. Initially intended for wired networks, the Monarch Group at CMU have extended NS-2 to support wireless networking such as MANET and wireless LANs as well. Most MANET routing protocols are available for NS-2, as well as an 802.11 MAC layer implementation.

NS-2's code source is split between C++ for its core engine and OTcl, an object oriented version of TCL for configuration and simulation scripts. The combination of the two languages offers an interesting compromise between performance and ease of use.

Implementation and simulation under NS-2 consists of 4 steps:

- Implementing the protocol by adding a combination of C++ and OTcl code to NS-2's source base;
- Describing the simulation in an OTcl script;
- Running the simulation and
- Analyzing the generated trace files.

Implementing a new protocol in NS-2 typically requires adding C++ code for the protocol's functionality, as well as updating key NS-2 OTcl configuration files in order for NS -2 to recognize the new protocol and its default parameters. The C++ code also describes which parameters and methods are to be made available for OTcl scripting. The NS-2 architecture follows closely the OSI model. We have adapted the implementation of flooding provided in NS-2 in the context of diffusion in sensor networks.

An agent in NS-2 terminology represents an endpoint where network packets are constructed, processed or consumed. Such an Agent was implemented at the Application layer for the broadcast source, and the simulation trace was collected at the MAC layer. Some disadvantages of NS-2 stem from its open source nature. First, documentation is often limited and out of date with the current release of the simulator. Fortunately, most problems may be solved by consulting the highly dynamic newsgroups and browsing the source code. Then code consistency is lacking at times in the code base and across releases. Finally, there is a lack of tools to describe simulation scenarios and analyze or visualize simulation trace files. These tools are often written with scripting languages. The lack of generalized analysis tools may lead to different people measuring different values for the same metric names.

The learning curve for NS-2 is steep and debugging is difficult due to the dual C++/OTcl nature of the simulator. A more troublesome limitation of NS-2 is its large memory footprint and its lack of scalability as soon as simulations of a few hundred to a few thousand of nodes are undertaken.

The following shows specifications of Proposed Simulation Environment

The following hardware is used to simulate the algorithm. The Processor with Pentium IV, 1.7GHz, and Memory: 128 MB and Video card: AGP with 32 Mb RAM. The following software used during developing the system, Operating System: Red Hat Linux 9, Simulator: NS2 version 2.28, Graph: X-Graph, Visualization: NAM (Network animator), Programming Language: TCL Shell Scripting.

3.1 The Working of TCL and NS2

An NS simulation script starts with making a Simulator object instance.

set ns [new Simulator]: generates an NS simulator object instance, and assigns it to variable ns. The line does is the following:

–Initialize the packet format (ignore this for now)

–Create a scheduler (default is calendar scheduler)

–Select the default address format (ignore this for now)

The "Simulator" object has member functions that do the following:

–Create compound objects such as nodes and links (described later)

–Connect network component objects created (ex. attach-agent)

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

–Set network component parameters (mostly for compound objects)
–Create connections between agents (ex. make connection between a "tcp" and "sink")
–Specify NAM display options

Most of member functions are for simulation setup (referred to as plumbing functions in the Overview section) and scheduling, however some of them are for the NAM display. The "Simulator" object member function implementations are located in the "ns-2/tcl/lib/ns-lib.tcl"file.

3.2 Proposed Clustering Based Algorithm

K-Means algorithm is very popular for data clustering. The proposed data transfer protocol K-Means algorithm is used to cluster the nodes with respect to their locations in the MANET and select a central node in each cluster to make it as a cluster head [11].

1. From sink, resolve the locations of all the nodes in the network.
2. For this use a service like GPS or we can resolve them by using the agents in all the nodes which will provide the information on request. For the second approach simple flooding can be used for initial location discovery. (in this simulation, GPS was assumed)
3. Then Select k Center in the problem space (it can be random).
4. Partition the data into k clusters by grouping points that are closest to those k centers.
5. Use the mean of these k clusters to find new centers.
6. Repeat steps 3 and 5 until centers do not change.
7. Find the nearby central nodes from the calculated cluster centers.
8. Make the central nodes as a cluster head.
9. Now, the Cluster heads will announce its presence to all its neighbors.
10. The neighboring nodes in turn will forward the data to sink via the cluster-head.
11. The sink will periodically update the cluster-heads.
12. From sink, resolve the locations of all the nodes in the network.
13. For this we can use a service like GPS or we can resolve them by using the agents in all the nodes which will provide the information on request. For the second approach simple flooding can be used for initial location discovery. (in this simulation, GPS was assumed)
14. Then Select k Center in the problem space (it can be random).
15. Partition the data into k clusters by grouping points that are closest to those k centers.
16. Use the mean of these k clusters to find new centers.
17. Repeat steps 3 and 5 until centers do not change.
18. Find the nearby central nodes from the calculated cluster centers.
19. Make the central nodes as a cluster head.
20. Now, the Cluster heads will announce its presence to all its neighbors.
21. The neighboring nodes in turn will forward the data to sink via the cluster-head.
22. The sink will periodically update the cluster-heads.

The purpose of study is to experiment with various kinds of simulations on NS2 to understand the behaviors of the algorithms under consideration. The following list shows different kinds of simulation parameters used in the proposed data transfer protocol simulation.

3.3 Simulation Results and Analysis

For the purpose of this study, we have experimented with various kinds of simulations on NS2 to understand the behaviors of the algorithms under consideration. The following list shows different kinds of simulation parameters used in the proposed data transfer protocol simulation.

The General Network Topology Parameters are,
Channel Type: Wireless Channel
Radio-Propagation Model: TwoRayGround
Network Interface Type: *WirelessPhy*
MAC type: Mac/802_11
Interface Queue Type: DropTail/PriQueue
Link Layer Type: LL
Antenna Model: Omni Antenna
Queue Length: 50
Network Layer Protocol: AODV
Size of the Topography: 700 X 500
The Constant Simulation Parameters are,
Active Data Senders: 75% Sensor Nodes
Sensor Data Size: 64 Bytes
Fused Data Size: 512
Sensor Data Interval: 1 Data Packet per seconds
Channel Error Rate: 0.15
Total Simulation Time: 10
The Variable Simulation Parameter has Total Sensor Nodes : 50, 75, 100, 125, and 150

To summarize the different implementation approaches of each simulator. The way a new algorithm is integrated can be pretty different from one simulator to another [10].

NS-2 is a discrete event network simulator that has begun in 1989 as a variant of the REAL network simulator. NS-2's code source is split between C++ for its core engine and OTcl, an object oriented version of TCL for configuration and simulation scripts. The combination of the two languages offers an interesting compromise between performance and

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

ease of use. Implementation and simulation under NS-2 consists of 4 steps:

– Implementing the protocol by adding a combination of C++ and OTcl code to NS-2's source base;

– Describing the simulation in an OTcl script;

– Running the simulation and

– Analyzing the generated trace files.

Algorithm I – Normal Sensor Network

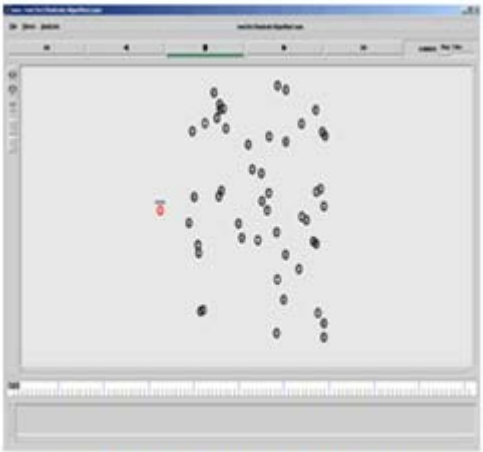


Fig 4: A Sample Sensor Network Topology

All the Sensor Nodes sending Data to the Sink at the periodic intervals.

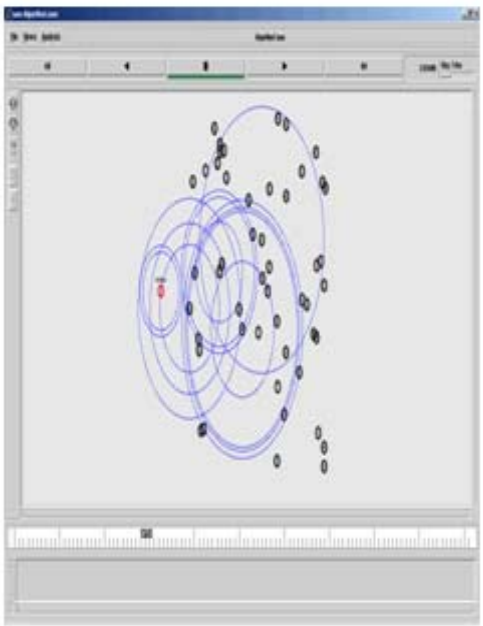


Fig 5: The 50 Node Sensor Network

The Initial Message Broadcast for Route Discovery

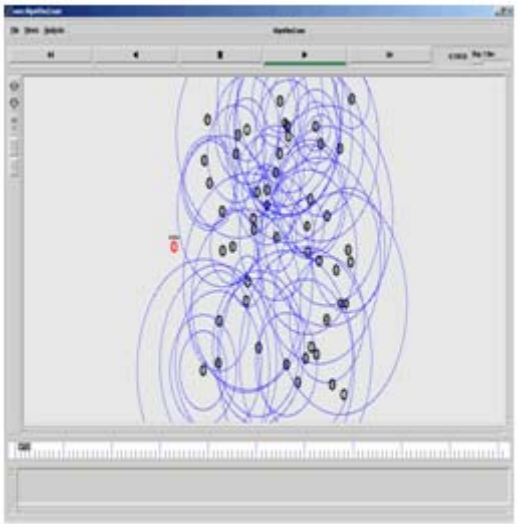


Fig 6: Initial Message Broadcast for Route Discovery

The Clustered Sensor Network, each cluster member is shown in different color. The Sink node (Left-Middle) is shown as red circle.

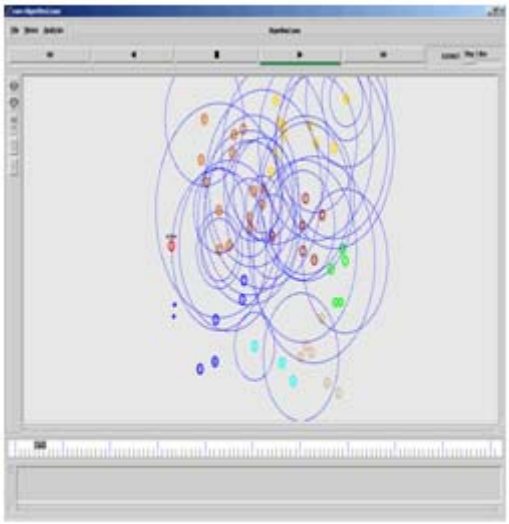


Fig 7: The Clustered Sensor Network

Marking Clusters and Cluster Centres.

The Calculated Cluster Centres

215 174, 413 92, 237 471, 513 383, 567 103, 492 599, 646 282

The Sensor Network with 7 Selected Cluster Heads

The Cluster Heads are shown as marked red Circles (double lines)

The Sink node (Left-Middle) is shown as red circle (single line).

The node Locations nearby the centres

Node 5: (186, 188), Node 39: (434, 93), Node 27: (198, 428), Node 41: (509, 396), Node 11: (543, 114), Node 37: (540, 614), Node 28: (617, 299)

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

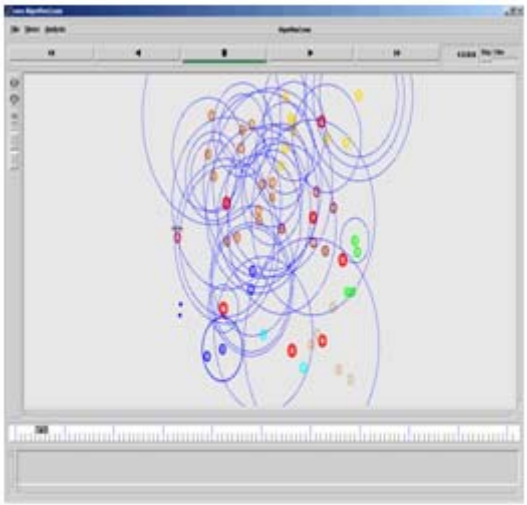


Fig 8: Sensor Network with Selected Cluster Heads

The Cluster Head is forwarding the collected data to sink

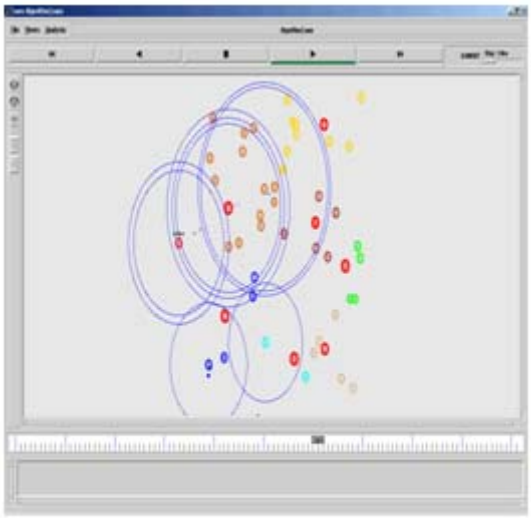


Fig 9: Cluster Head is forwarding the collected Data to Sink

The Final Terminal Window output
Total Sent Bytes: 21056.0
Total Received Bytes: 20928
The Delivery Ratio: 99.392097264437695

3.4 Comparative Results

Metrics considered for evaluation, the following metrics were considered for evaluating the flooding algorithms: Packet Delivery Ratio and Normalized Routing Load.
The following table shows the Packet Delivery Ratio of the two algorithms with respect to different number of nodes.

Sl No	Number of Nodes	Packet Delivery Ratio	
		Normal Method (Algorithm I)	Clustering Based Method (Algorithm II)
1	50	99.25	99.80
2	75	90.87	93.23
3	100	83.98	86.66
4	125	72.67	80.45
5	150	60.32	75.34

Table 1: Number_of_SensorNodes vs. Packet Delivery Ratio

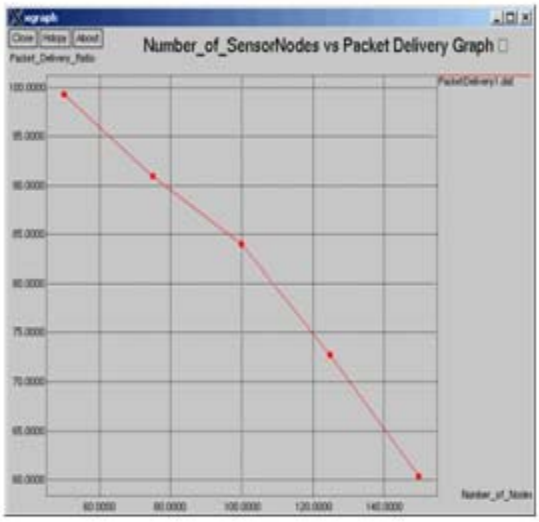


Fig 10: Packet Delivery Ratio of Algorithm I

The Packet Delivery Ratio Graph - Algorithm I

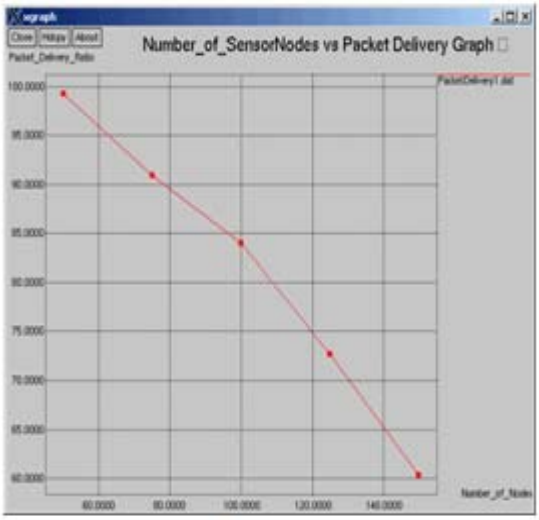


Fig 10: Packet Delivery Ratio of Algorithm I

The Packet Delivery Ratio Graph - Algorithm II

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

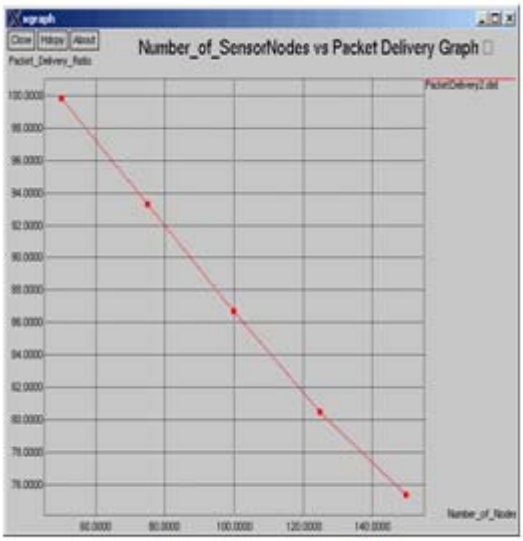


Fig 11: Packet Delivery Ratio of Algorithm II

The following line chart shows the Packet Delivery Ratio of the two algorithms with respect to different number of the nodes.

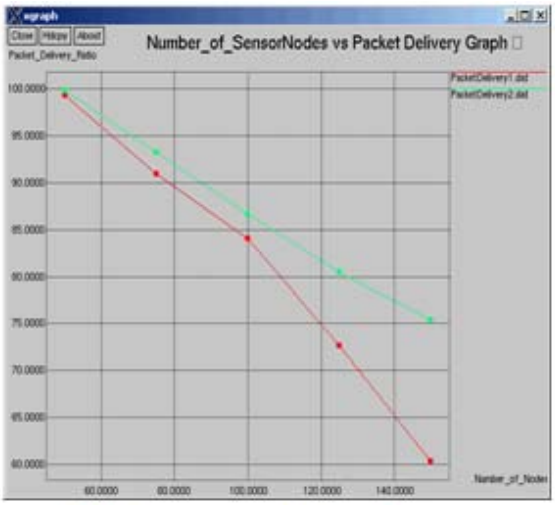


Fig 11: The Combined Packet Delivery Ratio Graph

Normalized Routing Load, the following table shows the normalized routing load of the two algorithms with respect to different number of nodes.

Sl No	Number of Nodes	Routing Load	
		Normal Method (Algorithm I)	Clustering Based Method (Algorithm II)
1	50	0.389	0.263
2	75	0.288	0.198
3	100	0.210	0.132
4	125	0.174	0.090
5	150	0.120	0.075

Table 2: Number_of_SensorNodes vs. Routing_Load

Normalized Routing Load Charts for Number_of_SensorNodes vs. Routing_Load Graph – Algorithm I

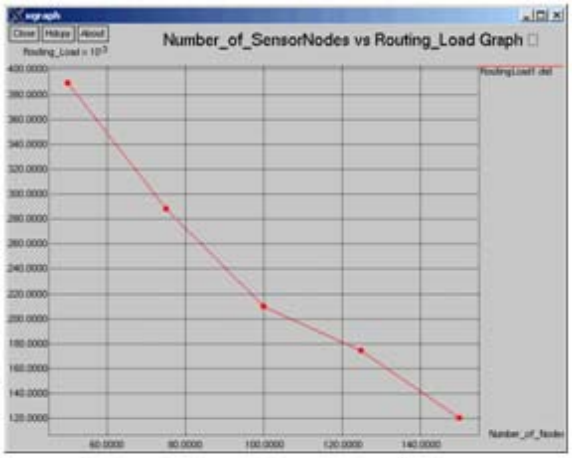


Fig 12: Routing Load of Algorithm I

Number_of_SensorNodes vs. Routing_Load Graph – Algorithm II

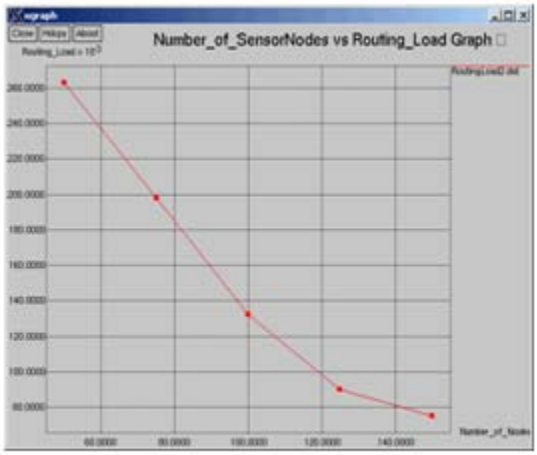


Fig13: Routing Load of Algorithm II

Number_of_SensorNodes vs. Routing_Load - Combined Graph, shows the normalized routing load of the two algorithms with respect to different number of nodes.

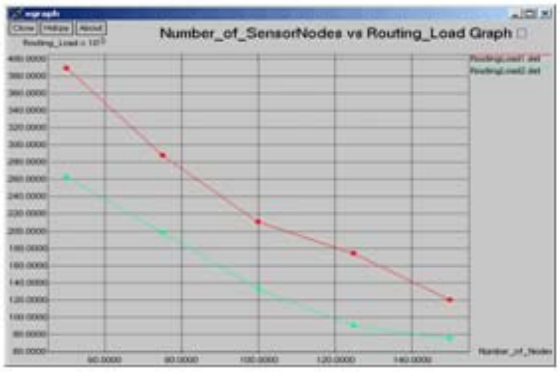


Fig 14: Normalized Routing Load Chart

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

As shown in the previous tables and graphs, the proposed algorithm performed well and provided good packet delivery with minimum routing load.

4. Conclusion and Scope for Further Enhancements

The main scope of this research is to device a new clustering based data transfer protocol for sensor network for efficient data transfer in sensor network. The proposed clustering based data transfer protocol has been successfully implemented and evaluated under Berkeley's Network Simulator ns2.

The performance of the network was tested with different simulation parameters and the simulation was repeated for different number of sensor nodes in the network. According to the trace analysis, the arrived results were significant and more comparable. While comparing the proposed method with a normal method, the proposed data transfer mechanism provided very good packet delivery ratio with very low routing load. The graphs and tables shows the enhancement in performance while using the proposed method.

In this work, the classic k-mean clustering algorithm was used to cluster the mobile nodes. Since the k-means algorithm has some draw backs and produce wrong clusters if there were lot of outliers in the location data, the future works may address more suitable algorithm for clustering the nodes in the mobile scenario.

In this implementation the numbers of clusters were decided with respect to the model scenario at hand. But in future works, one may address the possibility of deciding the number of clusters by some indirect means. In the proposed algorithm, for getting the location information, the same old flooding was used and for further message passing, the proposed clustering based method was used. Future works may address the possibilities of removing the classical flooding phase which is used to discover location information.

The whole experiments were done on Berkeley's Network Simulator. The future works may address the issues for real implementation which may involve real GPS for resolving location information.

Sensor networks present unprecedented opportunities for a broad spectrum of applications such as industrial automation, situation awareness, tactical surveillance for military applications, environmental monitoring, chemical/biological detection etc. So the future works may address the possibilities of enhancements of the proposed data transfer protocol for specific application.

References

[1] Yao-Chung Chang, Zhi-Sheng Lin and JiannLiang Chen, "Cluster Based Self-Organization Management Protocols for Wireless Sensor Networks", IEEE Transactions on Consumer Electronics, FEBRUARY 2006.

[2] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-Efficient Communication Protocols forWireless Microsensor Networks," In Proceedings Hawaaian International Conference on Systems Science, January 2000

[3] J. Hill, R. Szewczyk, A.Woo, S. Hollar, D. Culler, and K. Pister, "System Architecture Directions for Network Sensors," In Proceedings of the 9th International Conference on Architectural Support for Programming Languages and Operating Systems, November 2000.

[4] C. Intanagonwiwat, R. Govindan, and D. Estrin, "Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks," In Proceedings of the 6th Annual ACM/IEEE International Conference on Mobile Computing and Networking (MOBICOM), August 2000.

[5] C. Intanagonwiwat, D. Estrin, R. Govindan, and J. Heidemann, "Impact of network density on data aggregation in wireless sensor networks," Technical Report 01-750, University of Southern California, November 2001.

[6] Intel Research Oregon. Heterogeneous sensor networks. Technical report, Intel Corporation, 2003. WebPage.http://www.intel.com/research/exploratory/heterogeneous.htm.

[7] Brad Karp and H. T. Kung, "GPSR: Greedy perimeter stateless routing for wireless networks," In Proc. ACM/IEEE MobiCom, August 2000.

[8] H. Suzuki and F. A. Tobagi, "Fast Bandwidth Reservation Scheme with Multi link Multi path routing in ATM networks," In Proceedings of IEEE Infocom, 1992.

[9] CITRIS, Center for Information Technology Research in the Interest of Society. Web Page. http://www.citris-uc.org

[10] J. Broch, D. Maltz, D. Johnson, Y. Hu, and J. Jetcheva, "A Performance Comparison of Multi-hop Wireless Ad Hoc Network Routing Protocols". In Proc. Of 4th Annual ACM/IEEE International Conference on Mobile Computing(MOBICOM), ACM, October 1998.

[11] Elizabeth M. Royer, and C.-K. Toh. "A Review of Current Routing Protocols for Ad-Hoc Mobile Wireless Networks," In IEEE Personal Communications Magazine, April 1999.

[12] Jeremy Nelson and Deborah Estrin. "An Address-Free Architecture for Dynamic Sensor Networks," Technical Report 00-724, Computer Science Department, USC, January 2000.

[13] D. Estrin, R. Govindan, J. Heidemann, and S. Kumar, "Next Century Challenges: Scalable Coordination in Wireless Networks," In Proc. Of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking (MOBICOM),1999.

[14] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "uAMPS ns Code Extensions," http: // www-mtl.mit.edu /research/ icsystems /uamps/leach

[15] A. Manjeshwar and D. P. Agrawal, "TEEN: A Routing Protocol for Enhanced Efficiency in Wireless Sensor Networks," In Proc. Of 15th International Parallel and

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Distributed Processing Symposium (IPDPS '01) Workshops, 2001.
[16] A. Manjeshwar and D. P. Agrawal, "APTEEN: A Hybrid Protocol for Efficient Routing and Comprehensive Information Retrieval in Wireless Sensor Networks," In Proc. Of International Parallel and Distributed Processing Symposium (IPDPS '02) Workshops, 2002.
[16] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-Efficient Communication Protocols for Wireless Microsensor Networks," In Proc. Of Hawaiian International Conference on Systems Science, January 2000.
[17] A. Manjeshwar, M.S. Thesis "Energy Efficient Routing Protocols with Comprehensive Information Retrieval for Wireless Sensor Networks", 2001.

B.Manimekala received her B.Sc degree in Computer Science, Master degree (MCA) in Computer Applications and M.Phil in Computer Science in 2003, 2006 and 2010 respectively from Bharathiar University and Periyar University. Previosly she had worked as Assistant Professor inDepartment of Computer Applications, Dr.N.G.P.Arts and Science College, Coimbatore during 2007 to 2010 and she had worked as Assistant professor in Department of Computer Science, Kovai Kalaimagal College Of Arts and Science, Coimbatore during 2010 – 2011. Currently she is working as a Assistant to Professor, Department of Software Systems, V.L.B. Janakiammal College of Arts and Science, Coimbatore. Her area of interest includes Networking, DBMS, and Software Testing. Her career objective is seeking a position to inspire and guide the students to acquire knowledge, develop skills and a positive attitude, which will enhance their personality providing self-confidence to face the competitive world.

M.Kayalvizhi MCA., M.Phil., She is a Research Scholar in Bharathiar University,Coimbatore, Tamilnadu, India. She is working as a Assistant Professor, in VLB Janakiammal Arts and Science College, Coimbatore. she has 5 years of experience in teaching line 4 years of experience in research.she conducted International, National Conference and she presented paper in International, National Conference and Journals. Her Interest areas are Networking Image Processing, Current research project sensor network.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

An Advanced Survey on Cloud Computing and State-of-the-art Research Issues

Mohiuddin Ahmed[1], Abu Sina Md. Raju Chowdhury[2], Mustaq Ahmed[3], Md. Mahmudul Hasan Rafee[4]

[1,2,3,4] Department of Computer Science and Information Technology, Islamic University of Technology Board Bazar, Gazipur-1704, Bangladesh

Abstract

Cloud Computing is considered as one of the emerging arenas of computer science in recent times. It is providing excellent facilities to business entrepreneurs by flexible infrastructure. Although, cloud computing is facilitating the Information Technology industry, the research and development in this arena is yet to be satisfactory. Our contribution in this paper is an advanced survey focusing on cloud computing concept and most advanced research issues. This paper provides a better understanding of the cloud computing and identifies important research issues in this burgeoning area of computer science.

Keywords: *Cloud Computing; Virtualization; Data Center; Server Consolidation.*

1. Introduction

Due to the unprecedented success of internet in last few years, computing resources is now more ubiquitously available. And it enabled the realization of a new computing concept called Cloud Computing. Cloud Computing environment requires the traditional service providers to have two different ways. These are infrastructure and service providers. Infrastructure providers manage cloud platforms and lease resources according to usage. Service providers rent resources from infrastructure providers to serve the end users. Cloud Computing has attracted the giant companies like Google, Microsoft, and Amazon and considered as a great influence in today's Information Technology industry. Business owners are attracted to cloud computing concept because of several features [1, 2].

These are as follows:

- Lower initial investment
- Easier to manage
- Scalability
- Deploy faster
- Location independent
- Device independent
- Reliability
- Security

Although cloud computing has shown considerable opportunities to the IT industry of today's world, but still there are number of challenges that requires to be carefully addressed. In our paper, we present a survey of cloud computing and state-of-the-art research challenges. Our aim is to provide a better understanding of cloud computing and focus on the research ongoing in this tremendously flourishing arena of computer science.

In the section 2, we provide an overview of cloud computing, section 3 contains the state-of-the-art and section 4 will focus on the research issues. We conclude the paper on section 5 along with references.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

2. Cloud Computing Overview

2.1 Definitions

What is Cloud Computing?

Cloud computing is a way of leveraging the Internet to consume software or other IT services on demand. Users share processing power, storage space, bandwidth, memory, and software. With cloud computing, the resources are shared and so are the costs. Users can pay as they go and only use what they need at any given time, keeping cost to the user down. Cloud computing is very much a business model as well. Providers of cloud computing solutions, whether they are software, hardware, platform, or storage providers, deliver their offerings over the Internet. There are no shrink wrapped boxes containing discs or hardware for you to buy and set up yourself. Cloud providers typically charge monthly recurring fees based on your usage. [1]

Understanding Cloud Computing Applications [2]

Cloud computing, at its simplest, is a collection of computing software and services available from a decentralized network of servers. The term “cloud” has long been used as a metaphor for the Internet, and there are many popular services and Web sites which you may already be enjoying, without being aware that they are cloud-based. Social networking sites, Web-based email clients like Yahoo! and Gmail, Wikipedia and YouTube, and even peer-to-peer networks like Skype or Bit Torrent are all applications that run in the cloud. In other words, there is no one centralized location or organization that controls them, and nothing is required to utilize them besides a Web browser and an Internet connection. Enterprise cloud computing is cloud computing for the business world. Instead of purchasing and installing the physical infrastructure necessary to run software programs, a business instead consumes resources on a software-as-a-service (SaaS) basis. Running individual applications such as Microsoft, SAP, or Oracle will require hardware and an extensive infrastructure to support it: office space, power, networks, servers, storage, cooling, and bandwidth, not to mention the experts needed to install and run them. Cloud computing offers a streamlined, simplified solution to this complexity and the capital expenditure it necessitates.

Cloud computing [3] according to Wikipedia

Cloud computing is the delivery of computing as a service rather than a product, whereby shared resources, software, and information are provided to computers and other devices as a utility (like the electricity grid) over a network (typically the Internet).

Cloud computing [4] is a technology that uses the internet and central remote servers to maintain data and applications. Cloud computing allows consumers and businesses to use applications without installation and access their personal files at any computer with internet access. This technology allows for much more inefficient computing by centralizing storage, memory, processing and bandwidth.

A simple example of cloud computing is Yahoo email, Gmail, or Hotmail etc. You don’t need software or a server to use them. All a consumer would need is just an internet connection and you can start sending emails. The server and email management software is all on the cloud (internet) and is totally managed by the cloud service provider Yeah, Google etc. The consumer gets to use the software alone and enjoy the benefits. The analogy is, 'If you need milk, would you buy a cow?' All the users or consumers need is to get the benefits of using the software or hardware of the computer like sending emails etc. Just to get this benefit (milk) why should a consumer buy a (cow) software /hardware?

Cloud computing is broken down into three segments: "application" "storage" and "connectivity." Each segment serves a different purpose and offers different products for businesses and individuals around the world. In June 2011, a study conducted by VersionOne found that 91% of senior IT professionals actually don't know what cloud computing is and two-thirds of senior finance professionals are clear by the concept,[5] highlighting the young nature of the technology. In Sept 2011, an Aberdeen Group study found that disciplined companies achieved on average an 68% increase in their IT expense because cloud computing and only a 10% reduction in data center power costs[6].

Final Version of NIST Cloud Computing Definition[7]

Cloud computing is a relatively new business model in the computing world. According to the official NIST definition, "cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction."

The NIST definition lists five essential characteristics of cloud computing: on-demand self-service, broad network access, resource pooling, rapid elasticity or expansion, and measured service. It also lists three "service models" (software, platform and infrastructure), and four "deployment models" (private, community, public and

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

hybrid) that together categorize ways to deliver cloud services. The definition is intended to serve as a means for broad comparisons of cloud services and deployment strategies, and to provide a baseline for discussion from what is cloud computing to how to best use cloud computing.

2.2 Types of Cloud

Cloud computing is typically classified in two ways [9]:

1. Location of the cloud computing
2. Type of services offered

Location of the cloud
Cloud computing is typically classified in the following ways:

1. **Public cloud**: In Public cloud the computing infrastructure is hosted by the cloud vendor at the vendor's premises. The customer has no visibility and control over where the computing infrastructure is hosted. The computing infrastructure is shared between any organizations.

2. **Private cloud**: The computing infrastructure is dedicated to a particular organization and not shared with other organizations. Some experts consider that private clouds are not real examples of cloud computing. Private clouds are more expensive and more secure when compared to public clouds.

Private clouds are of two types: On-premise private clouds and externally hosted private clouds. Externally hosted private clouds are also exclusively used by one organization, but are hosted by a third party specializing in cloud infrastructure. Externally hosted private clouds are cheaper than On-premise private clouds.

3. **Hybrid cloud** Organizations may host critical applications on private clouds and applications with relatively less security concerns on the public cloud. The usage of both private and public clouds together is called hybrid cloud. A related term is Cloud Bursting. In Cloud bursting organization use their own computing infrastructure for normal usage, but access the cloud for high/peak load requirements. This ensures that a sudden increase in computing requirement is handled gracefully.

4. **Community cloud** involves sharing of computing infrastructure in between organizations of the same community. For example all Government organizations within the state of California may share computing infrastructure on the cloud to manage data related to citizens residing in California.

Classification based upon service provided

Based upon the services offered, clouds are classified in the following ways:

1. Infrastructure as a service (IaaS) involves offering hardware related services using the principles of cloud computing. These could include some kind of storage services (database or disk storage) or virtual servers. Leading vendors that provide Infrastructure as a service are Amazon EC2, Amazon S3, Rackspace Cloud Servers and Flexiscale.

2. Platform as a Service (PaaS) involves offering a development platform on the cloud. Platforms provided by different vendors are typically not compatible. Typical players in PaaS are Google's Application Engine, Microsofts Azure, Salesforce.com's force.com .

3. Software as a service (SaaS) includes a complete software offering on the cloud. Users can access a software application hosted by the cloud vendor on pay-per-use basis. This is a well-established sector. The pioneer in this field has been Salesforce.coms offering in the online Customer Relationship Management (CRM) space. Other examples are online email providers like Googles gmail and Microsofts hotmail, Google docs and Microsofts online version of office called BPOS (Business Productivity Online Standard Suite).

The above classification is well accepted in the industry. David Linthicum describes a more granular classification on the basis of service provided. These are listed below:

1. Storage-as-a-service
2. Database-as-a-service
3. Information-as-a-service
4. Process-as-a-service
5. Application-as-a-service

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

6. Platform-as-a-service
7. Integration-as-a-service
8. Security-as-a-service
9. Management/Governance-as-a-service
10. Testing-as-a-service
11. Infrastructure-as-a-service

2.3 Related Technologies

Cloud computing typically has characteristics of all these technologies [8]:

a. Grid computing
b. Virtualization
c. Utility Computing
d. Autonomic Computing

A quick overview of these technologies is given here.

Grid Computing

Grid Computing involves a network of computers that are utilized together to gain large supercomputing type computing resources. Using this network of computers large and complex computing operations can be performed. In grid computing these network of computers may be present in different locations.
A famous Grid Computing project is Folding@Home. The project involves utilizing unused computing powers of thousands of computers to perform a complex scientific problem. The goal of the project is "to understand protein folding, misfolding, and related diseases".

Virtualization

Virtualization introduces a layer between Hardware and operating system. During the sixties mainframe started supporting many users using virtual machines. These virtual machines simulated behavior of an operating system for each user. VMWare launched a product called VMware Workstation in 1999 that allows multiple operating systems to run on personal computers.
The virtualization forms the foundation of cloud technology. Using virtualization, users can access servers or storage without knowing specific server or storage details. The virtualization layer will execute user request for computing resources by accessing appropriate resources.
Typically server utilization in data centers can be as low as 10%. Virtualization can help in significantly improving server utilization.

Utility Computing

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

1. Billing per user count. As an example if an organization of 100 people uses Google's gmail or Microsoft Live as their internal email system with email residing on servers in the cloud, Google/Microsoft may bill the organization on per user basis.
2. Billing per Gigabyte. If an organization is using Amazon to host their data on the cloud, Amazon may bill the organization on the disk space usage.
3. Billing per hour/day. As an example a user may pay for usage of virtual servers by time utilized in hours.

3. State-of-the-art Technologies

In this section, the state-of-the-art implementations of cloud computing is presented. Technologies used for cloud computing are describes here.

3.1 Architectural Design of Data Centre

A **data center [10]** (or **data centre** or **datacenter** or **datacenter**) is a facility used to house computer systems and associated components, such as telecommunications and storage systems. It generally includes redundant or backup power supplies, redundant data communications connections, environmental controls (e.g., air conditioning, fire suppression) and security devices.

Key Design Areas

- Resilience - ensuring maximum uptime without compromising on performance.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

- Availability - business continuity is especially important.
- Performance - the faster the better; in a predictable manner.
- Security - ensuring data separation and controlled access to all Data centre resources.
- Effective architecture for data separation - a common infrastructure that provides facilities for network-based backup and efficient back-end network access.
- Predictable Failover - for maximum service availability.

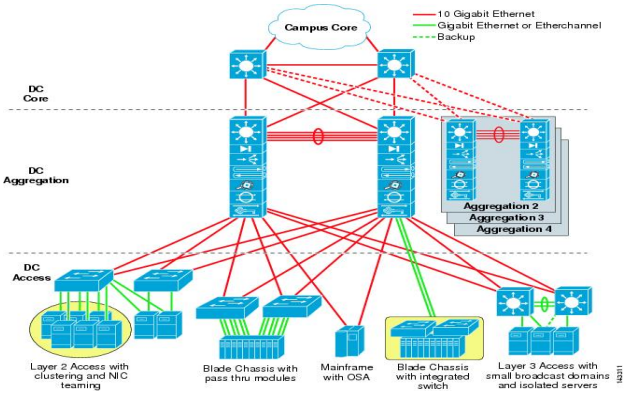


Figure : Data Centre Design

3.2 Distributed File System over Cloud

This In this paper [11], they focused on Google File System (GFS)[12] which is a proprietary distributed file system developed by Google and specially designed to provide efficient, reliable access to data using large clusters of commodity servers. Files are divided into chunks of 64 megabytes and are usually appended to a read and only extremely rarely overwritten or shrunk. Comparing traditional file systems, GFS is designed and optimized to run on data centers to provide extremely high data throughputs, low latency and survive individual server failures. Inspired by GFS, the open source Hadoop Distributed File System (HDFS)[13] stores large files across multiple machines. It achieves reliability by replacing the data across multiple servers.

3.3 Distributed Application Framework over Clouds

MapReduce[14] is a software framework introduced by Google to support distributed computing on large data sets on clusters of computers. HTTP-based applications are usually conform to some web application framework such as Java EE. MapReduce consists of one Master to which client applications submit MapReduce jobs. The Master pushes work out to available task nodes in the data centre striving to keep the tasks as close to the data as possible. The open source Hadoop MapReduce project[15] is inspired by Google's work. Currently, many organizations are using Hadoop MapReduce to run large data intensive computations.

4. Research Issues

Cloud Computing has been widely practiced by IT industry as well as business enterprises in recent times. But , research on it is still at an immature stage. Many existing issues have not been fully addressed, where as newer problems are arising due to its extensive usage. In this section, we summarize some of the crucial research issues in cloud computing.

4.1 Virtual Machine Migration

Some vendors have implemented VM migration in their virtualization solution—a big advantage for application uptime in a data center[16]. What is VM migration? Consider the case of a server with a hypervisor and several VMs, each running an OS and applications. If you need to bring down the server for maintenance (say, adding more memory to the server), you have to shut down the software components and restart them after the maintenance window—significantly affecting application availability. VM migration allows you to move an entire VM (with its contained operating system and applications) from one machine to another and continue operation of the VM on the second machine. This advantage is unique to virtualized environments because you can take down physical servers for maintenance with minimal effect on running applications. You can perform this migration after suspending the VM on the source machine, moving its attendant information to the target machine and starting it on the target machine. To lower the downtime, you can perform this migration while the VM is running (hence the name "live migration") and resuming its operation on the target machine after all the state is migrated. The following are some of the benefits of virtualization in a cloud-computing environment:

- *Elasticity and scalability:* Firing up and shutting down VMs involves less effort as opposed to bringing servers up or down.

- *Workload migration:* Through facilities such as live VM migration, you can carry out workload migration with much less effort as compared to

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

workload migration across physical servers at different locations.

- *Resiliency:* You can isolate physical-server failure from user services through migration of VMs.

It must be clarified that virtualization is not a prerequisite for cloud computing. In fact, there are examples of large cloud service providers using only commodity hardware servers (with no virtualization) to realize their infrastructure. However, virtualization provides a valuable toolkit and enables significant flexibility in cloud-computing deployments.

4.2 Server Consolidation

Server consolidation is an effective approach to maximize resource utilization while minimizing energy consumption in a cloud computing environment. Live VMmigration technology is often used to consolidate VMs residing on multiple under-utilized servers onto a single server, so that the remaining servers can be set to an energy-saving state. The problem of optimally consolidating servers in a data center is often formulated as a variant of the vector bin-packing problem [17], which is an NP-hard optimization problem. Various heuristics have been proposed for this problem [18,19]. Additionally, dependencies among VMs, such as communication requirements, have also been considered recently [20].

4.3 Energy Management

Improving energy efficiency is another major issue in cloud computing. It has been estimated that the cost of powering and cooling accounts for 53% of the total operational expenditure of data centers [21]. In 2006, data centers in the US consumed more than 1.5% of the total energy generated in that year, and the percentage is projected to grow 18% annually [18]. Hence infrastructure providers are under enormous pressure to reduce energy consumption. The goal is not only to cut down energy cost in data centers, but also to meet government regulations and environmental standards. Designing energy-efficient data centers has recently received considerable attention. This problem can be approached from several directions. For example, energy efficient hardware architecture that enables slowing down CPU speeds and turning off partial hardware components has become commonplace. Energy-aware job scheduling and server consolidation are two other ways to reduce power consumption by turning off unused machines. Recent research has also begun to study energy-efficient network protocols and infrastructures . A key challenge in all the above methods is to achieve a good trade-off between energy savings and application performance. In this respect, few researchers have recently started to investigate coordinated solutions for performance and power management in a dynamic cloud environment.

4.4 Information Security

Information security is another important research topic in cloud computing. Since service providers typically do not have access to the physical security system of data centers, they must rely on the infrastructure provider to achieve full data security. Even for a virtual private cloud, the service provider can only specify the security setting remotely, without knowing whether it is fully implemented. The infrastructure provider, in this context, must achieve the following objectives: (1) *confidentiality*, for secure data access and transfer, and (2) *audit ability*, for attesting whether security setting of applications has been tampered or not. Confidentiality is usually achieved using cryptographic protocols, whereas audit ability can be achieved using remote attestation techniques. Remote attestation typically requires a trusted platform module (TPM) to generate non-forgeable system summary (i.e. system state encrypted using TPM's private key) as the proof of system security. However, in a virtualized environment like the clouds, VMs can dynamically migrate from one location to another, hence directly using remote attestation is not sufficient. In this case, it is critical to build trust mechanisms at every architectural layer of the cloud. Firstly, the hardware layer must be trusted using hardware TPM. Secondly, the virtualization platform must be trusted using secure virtual machine monitors. VM migration should only be allowed if both source and destination servers are trusted. Recent work has been devoted to designing efficient protocols for trust establishment and management.

4.5 Novel Cloud Architecture

In recent times, commercial clouds are implemented in mammoth data centers and operated centrally. Although, there are economical pros but it too comes along with limitations like high energy expenses and initial investment for constructing data centers. In this paper[22], small sized data centers are suggested to be more

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

advantageous than large data centers. Small sized data centers eliminates the problems of energy,power,cooling system and, more over economical & geographically better distributed.

Another research trend on using voluntary resources for hosting cloud application. Voluntary resources help to build clouds which are cheaper and more suitable for non-profit applications such as scientific computing. Although it is providing some benefits but management of such heterogeneous resources is a challenge and creating such architecture is a research problem.

5. Conclusion

The advancement of cloud computing is dramatically changing the horizon of information technology and ultimately turns the utility computing into a reality. However, it provides a large array of benefits, but many challenges in this domain, including automatic resource positioning, energy management, information security are only attracted the research community. There are still so many issues to be explored. Opportunities are enough in this arena for some groundbreaking contribution and bring significant development in the industry.

In our paper, we have presented an overview of cloud computing and focused on the state-of-the-art research and future issues to be handled by the research community. Cloud computing is at an early stage of research and development, we believe our paper will provide a better understanding of the cloud computing and different research issues, thereby bolstering further research in this arena.

References

[1] http://www.thesmallbusiness.org/software/benefits-of-cloud-computing.html

[2] http://www.allthingscrm.com/cloud-computing/understanding-cloud-computing-applications.html

[3] http://en.wikipedia.org/wiki/Cloud_computing

[4] http://www.wikinvest.com/concept/Cloud_Computing

[5] Cloud Confusion Amongst IT Professionals. VersionOne (June 6, 2011).

[6] Business Adoption of Cloud Computing. AberdeenGroup (Sept 9, 2011).

[7] http://www.nist.gov/itl/csd/cloud-102511.cfm

[8] http://thecloudtutorial.com/related.html

[9] http://thecloudtutorial.com/cloudtypes.html

[10] http://en.wikipedia.org/wiki/Data_center

[11] Qi Zhang,Lu Cheng, Raouf Boutaba."Cloud Computing: State-of-the-art and research challenges". The Brazilian Computer Society 2010.

[12] GhemawatS, Gobioff H, Leung S-T(2003)
The Google file system. In: Proc of SOSP, October 2003.

[13] Hadoop Distributed File System, hadoop.apache.org/hdfs

[14] Dean J, Ghemawat S(2004) MapReduce:simplified data processing on large clusters. In: Proc of OSDI

[15] Hadoop MapReduce, hadoop.apache.org/mapreduce

[16] http://www.cisco.com/web/about/ac123/ac147/archived_issues/ipj_12-3/123_cloud1.html

[17] Chekuri C, Khanna S (2004) On multi-dimensional packing problems.
SIAM J Comput 33(4):837–851

[18]Li B et al (2009) EnaCloud: an energy-saving application live placement approach for cloud computing environments. In: Proc of international conf on cloud computing

[19] Srikantaiah S et al (2008) Energy aware consolidation for cloud computing. In: Proc of HotPower

[20] Meng X et al (2010) Improving the scalability of data center networks
with traffic-aware virtual machine placement. In: Proc INFOCOM

[21] Hamilton J (2009) Cooperative expendable micro-slice servers (CEMS): low cost, low power servers for Internet-scale services
In: Proc of CIDR

[22] Valancius V, Laoutaris N et al (2009) Greening the Internet with nano data centers. In: Proc of CoNext

Mohiuddin Ahmed has completed Bachelor of Science in Computer Science and Information Technology from Islamic University of Technology, OIC. Research interest includes Human Computer Interaction, Cloud Computing, Artificial Intelligence and Wireless Network.

Abu Sina Md. Raju Chowdhury has completed Bachelor of Science in Computer Science and Information Technology from Islamic University of Technology, OIC. Research interest includes Cloud Computing, Human Computer Interaction, Peer to Peer and Overlay Network, Data Mining and Web Engineering.

Mustaq Ahmed has completed Bachelor of Science in Computer Science and Information Technology from Islamic University of Technology, OIC. Research interest includes Cloud Computing, Human Computer Interaction, Peer to Peer and Overlay Network, Open Source Computing.

Md. Mahmudul Hasan Rafee has completed Bachelor of Science in Computer Science and Information Technology from Islamic University of Technology, OIC. Research interest includes Cloud Computing, Human Computer Interaction, Web Engineering.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

An Empirical Study of Underwater Acoustic Sensor Network and Terrestrial Wireless Network

Mohd. Ehmer Khan[1], Farmeena Khan[2]

[1] Department of Computer Science, Singhania University
Jhunjhunu, Rajasthan, India

[2] Department of Computer Science, EIILM University
Jorethang, Sikkim, India

Abstract

Two third of the earth is covered with water and there has been a growing interest in monitoring underwater medium for scientific exploration and attack protection. Underwater sensor nodes will find application in ocean-graphic data collection, pollution monitoring, disaster prevention and tactical surveillance application. Underwater network consist of a variable number of sensors and vehicles that are deployed to perform collaborative monitoring task over a given area. On the other hand the terrestrial wireless networks are the networks without cable and are very easy to install and offer portability and convenience which other network solutions do not offer. The terrestrial wireless networks utilized radio waves to maintain communication channels between computers. In our paper some of the fundamental key aspects of underwater acoustic communication are investigated and three different UASN architectures are discussed.

Keywords: *Communication Architecture, Terrestrial Wireless Network, Underwater Acoustic Sensor Network, WWAN*

1. Introduction

Underwater Acoustic Sensor Network (UASN): A collection of sensor nodes which communicates among them through the emerging underwater acoustic communication technology is known as UASN. Acoustic communication technology is the best choice when compare to the radio waves and optical waves that is why it was chosen for the communication in underwater. Underwater acoustic network gaining attention due to their importance in underwater applications for military and commercial purpose.

Some of the major applications of UASN are [1]

I. Exploring the unpredictable underwater world.

II. Reports Ocean related disaster like tsunanmi to coastal areas.

III. Exchange information among nodes.

IV. It can perform pollution monitoring in lakes and rivers e.g. Oil leakage, chemical pollution, biological pollution etc.

V. It can also helps in detecting minerals, oil fields etc.

VI. UAN with acoustic and optical sensors can be used to perform rapid environmental assessment and detect mine like objects. [2]

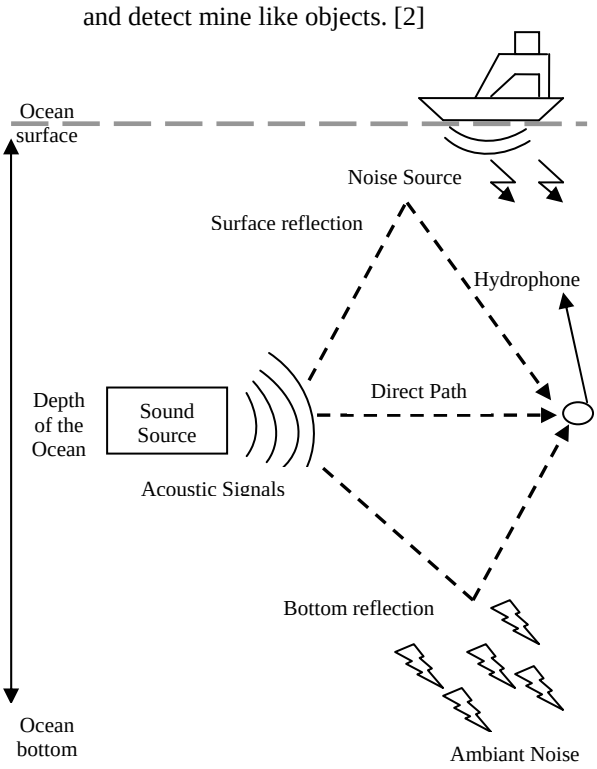


Fig. 1 Represent underwater acoustic environment

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Major challenges in the designing of UASN are [3]

I. Battery power is limited.
II. Limited available bandwidth.
III. Due to corrosions underwater sensors are more prone to failure.
IV. Bit error rates are quite high.
V. Propagation delay.
VI. Channel is severely impaired.

Terrestrial Wireless Network: Wireless network refers to any type of network that is not connected by cables or any kind of wires, instead it uses radio waves. Through wireless networking we can move our computer any where very easily. Wireless network uses air as medium. It provides us more privacy and security than before.

Some of the most important advantages of wireless networks are [4]

- Mobility/Portability
- It is very cost effective.
- Convenient to use.
- Easy deployment
- Expendability

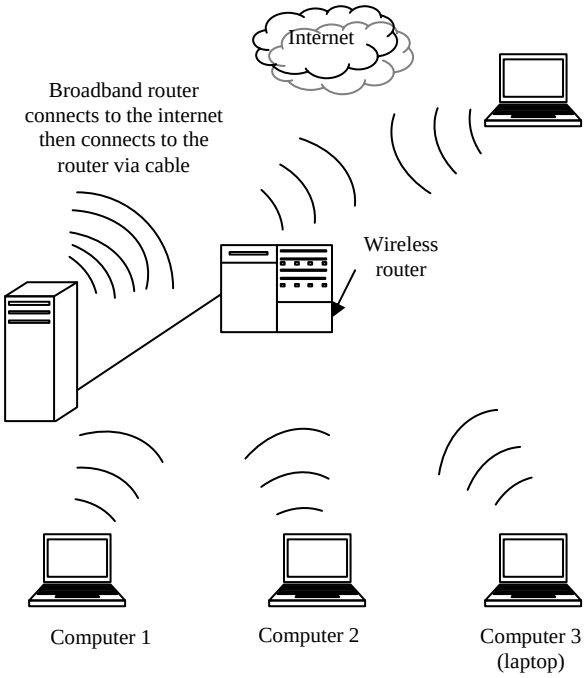


Fig. 2 Represent wireless network

But with all these advantages there are certain disadvantages also such as [5]

- Security: If wireless networks not secure properly it can be hacked easily.
- Speed: The speed of wireless networks is far slower than wire network.
- Reliability: Wireless networking signals are subject to a wide variety of interference.

2. Communication Architectures for Underwater Acoustic Sensor Network

Three types of underwater acoustic sensor network are

2.1 Static-Two Dimensional UASN (for Ocean Bottom Monitoring)

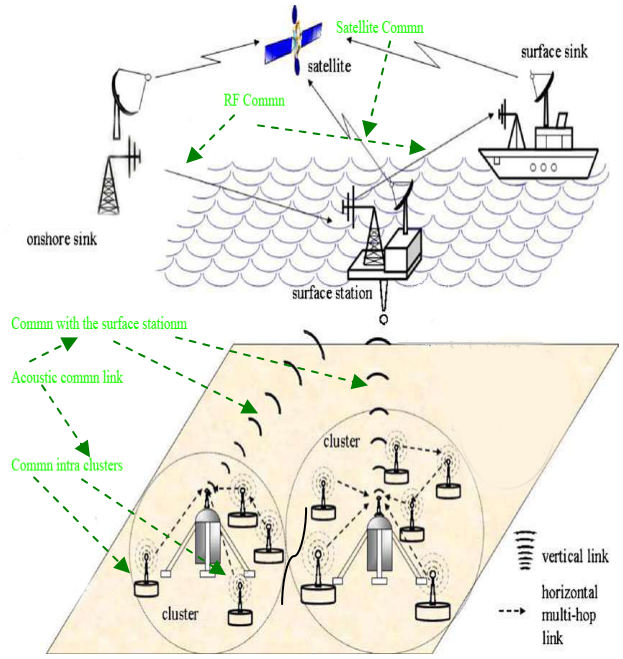


Fig. 3 Represent architecture for static 2D UASN

It is a collection of sensor nodes that are anchored to the bottom of the ocean bed. The sensor nodes are connected to one or more sinks by the means of wireless acoustic links and have vertical and horizontal transceiver. Main application can be environmental monitoring. The major problem presented in this architecture is various multi-hopp paths which can increase signaling overhead over the networks as well as the complexity of routing protocols. [6]

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

2.2 Static-Three Dimensional UASN (for Ocean Column Monitoring)

It is a collection of sensor nodes that are connected to the bottom of the ocean-bed by tethere. This technique can be used for surveillance monitoring of ocean phenomena.

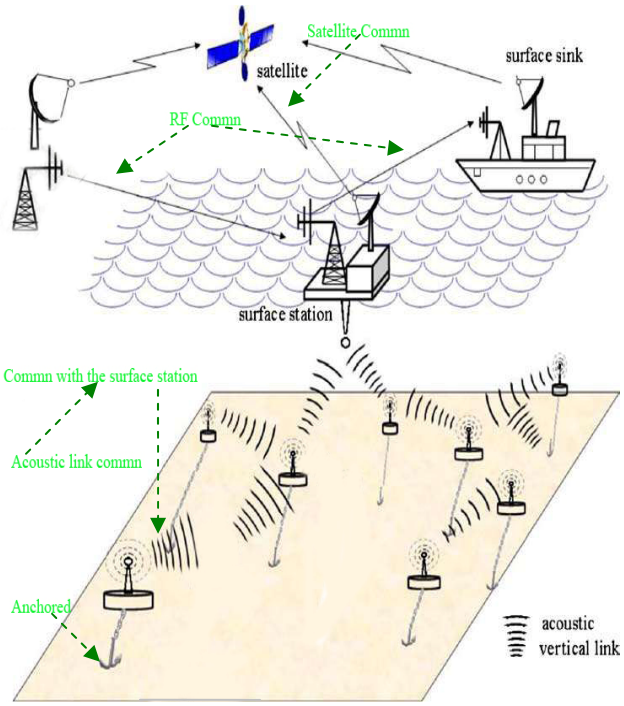


Fig. 4 Represent architecture for static 3D UASN

Some major problems presented in this architecture are [6]

- If the architecture is for military purpose than they can be easily detected by enemies.
- It can obstruct ship navigating on the surface.

2.3 Three Dimensional Networks of Autonomous Underwater Vehicle (AUV)

In this architecture a sensor node is built inside underwater vehicle and is used to enhance the whole sensor network capabilities. It is necessary for this architecture to have algorithm that guarantee the perfect coordination between AUV and the sensors. It also ensures the correct functionality of AUV and sensors in the scenario. [6]

Two distinct characteristics of AUV are

- Adaptive sampling
- Self configuration

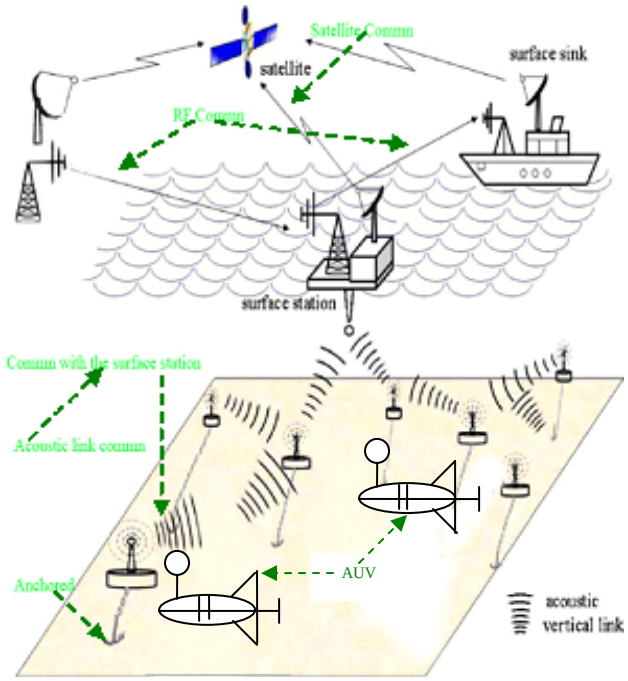


Fig. 5 Represent architecture for 3D AUV

3. Various Terrestrial Wireless Networks

The four different types of terrestrial wireless networks are described below:

3.1 Wireless Personal Area Network (WPAN)

A wireless PAN involves some one wirelessly synchronizing his/her PDA to a laptop or desktop computer; it can also provide wireless connectivity to a printer. “Plugging in” (every device in a WPAN will be able to plug in to any other devices in the same WPAN) is the key concept in the WPAN technology. When two WPAN equipped devices come into close proximity, they can communicate as if they are connected by a cable. It can also prevent the unauthorized access to information as each WPAN equipped device has ability to lock out other device selectively. Two kinds of wireless technology used in WPAN are Bluetooth and infrared. [7]

3.2 Wireless Local Area Network (WLAN)

Using a wireless distribution method i.e. linking of two or more devices over a short distance (within a building) is called WLAN. The working of WLAN depends on three main parts; they are as follows [8]

1. Infrastructure mode
2. Ad-hoc network mode
3. Mixed network mode

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

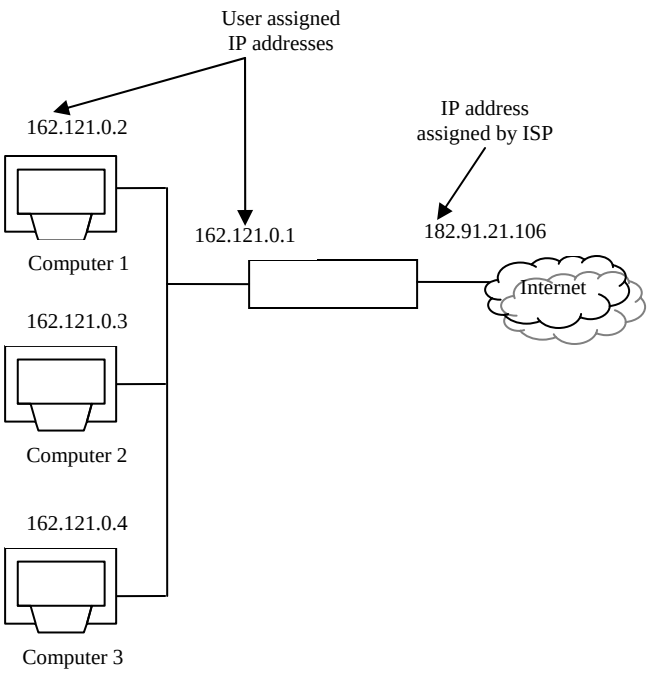


Fig. 6 Represent WLAN architecture

WLAN architecture is a technique of arranging and designing of different components in WLAN networking device in a specific way.

Three important components of WLAN architecture are

1. Access Point
2. Clients
3. Bridge

3.3 Wireless Metropolitan Area Network (WMAN)

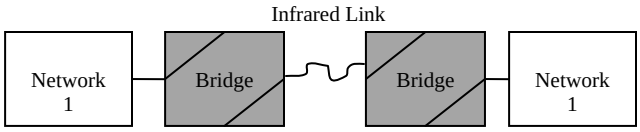


Fig. 7 Represent Wireless MAN

Wireless MANs are the type of wireless networks that connect several wireless LANs for e.g. WiMAX. Wireless MANs encompass areas of about the size of the cities and can span upto 50 kilometers. DQDB is the standard of WMAN which cover upto 48 kilometers with the speed of 34 Mbit/s to 155Mbit/s. WMAN performance varies, and the actual performance depends on the choice from the wide assortment of technologies. The technology of WMAN consists of ATM, FDDI and SMDS. WMAN provide high speed internet access to business and thus opens the door for the success.

3.4 Wireless Wide Area Network (WWAN)

Wireless WANs covers large geographical area, such as country or continent. Data throughput rate for WWAN are a great deal slower, that is upto 170 Kbps and typical rates of about 56 Kbps. [9]

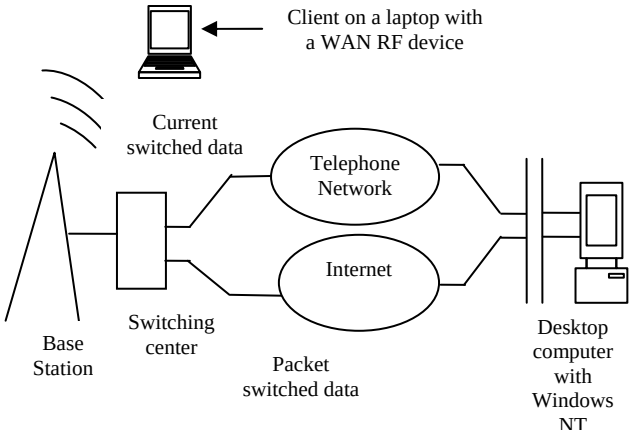


Fig. 8 Represent various components of WWAN

Various advantages of WWAN application involve user accessing internet wherever and whenever needed, stay connected while traveling, enabling users to carry on business and leisure activities form different location. With all these advantages there are certain limitations also, such as

1. Application persistence
2. Poor performance within buildings
3. Security
4. Expensive to install within buildings
5. Off-the-shelf solutions

Table 1: Comparing different terrestrial wireless networks

S.No.	***Types***	***Coverage***	***Standards***
1.	WPAN	Within reach of a person	Bluetooth, IEEE 802.15 and Irda
2.	WLAN	Within a campus or same building	IEEE 802.11, Wifi, Hyper LAN
3.	WMAN	Within a city	IEEE 802.16, WIMAX and proprietary
4.	WWAN	All over the world	CDPD, 2G, 3G and 3.5G

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

4. A Comparison between Underwater Acoustic Sensor Network and Terrestrial Wireless Network

The comparison between UASN and Terrestrial wireless network are described below: [1] [3]

1. The ground based wireless networks are expected to become increasingly expensive but the expenses for the underwater equipment i.e. the cost of manufacturing, deployment, maintenance and recovery are much higher.

2. Due to the fouling and corrosion underwater equipments can be easily damaged as compared to ground based wireless network.

3. Communication protocol used in ground based wireless network works better than underwater acoustic protocols.

4. The power needed for acoustic underwater communication is higher than ground based wireless network.

5. As compared to ground based wireless networks underwater sensors typically do not use RF communication. Acoustic communications are used instead.

6. Underwater sensors are mobile while terrestrial wireless network are fixed.

7. As compared to terrestrial wireless network underwater acoustic network typically use solar power.

5. Conclusion

Use of radio frequency technology or electromagnetic waves to transmit and received data over the air is known as wireless network. In our paper we have described four important types of wireless networks, they are as follows: WPAN, WLAN, WMAN, WWAN [5]

Whereas underwater acoustic network consist of a variable number of sensors and vehicles that are deployed to perform collaborative monitoring over a given area. There are three types of underwater acoustic networks.

1. Static 2D UASN
2. Static 3D UASN
3. 3D network for autonomous underwater vehicle

In our paper we have also explain some of the major differences between UASN and ground based wireless networks.

References:

[1] Ziahan Jiang, "Underwater Acoustic Networks-Issues and Solutions", International Journal of Intelligent and Systems, Vol. 13, No. 3, September 2008, pp. 152-161

[2] Underwater Acoustic Communication available at http://en.wikipedia.org/wiki/Underwater_acoustic_communication

[3] Sharankumar Sivakumar, "Delay Sensitive Routing in Three Dimensional Underwater Acoustic Sensor Networks", M.Sc. Thesis, Faculty of the Graduate College, Oklahoma State University, Oklahoma, USA, May 2006

[4] Wireless Computer Networking available at http://compnetworking.about.com/cs/wireless/f/whatiswireless.htm

[5] Advantages and Disadvantages of Wireless Networks, Published on Saturday, 07 May 2011, available at http://www.the-server-admin.com/networking/7-advantages-and-disadvantages-of-wireless-networks.html

[6] Underwater sensor Networks by Saurabh Kumar Pandey available at http://intranet.daiict.ac.in/~ranjan/isn2006/presentation/saurabh_under_water.pdf

[7] Types of Wireless Networks and Usage available at http://www.tutorial-reports.com/wireless/usage.php

[8] Wireless LAN Architecture available at http://www.wifinotes.com/wlan-architecture.html

[9] Wireless Networks Defined available at http://etutorials.org/Networking/wn/Chapter+1.+The+Wireless+World+An+Introduction+to+Concepts/Wireless+Networks+Defined/

Mohd. Ehmer Khan

I completed my B.Sc in 1997 and M.C.A. in 2001 from Aligarh Muslim University, Aligarh, India, and pursuing Ph.D (Computer Science) from Singhania University, Jhunjhunu, India. I have worked as a lecturer at Aligarh College Engineering & Management, Aligarh, India from 1999 to 2003. From 2003 to 2005 worked as a lecturer at Institute of Foreign Trade & Management, Moradabad, India. From 2006 to present working as a lecturer in the Department of Information Technology, Al Musanna College of Technology, Ministry of Manpower, Sultanate of Oman. I am recipient of PG Merit Scholarship in MCA. My research area is software engineering with special interest in driving and monitoring program executions to find bugs, using various software testing techniques.

Farmeena Khan

I completed my BCA in 2003 from Dr. B. R. Ambedkar University, Agra, India and MBA in 2011 from Janardan Rai Nagar Rajasthan Vidyapeeth University, Rajasthan, India, and pursuing M.Sc (Computer Science) from EIILM University, Jorethang, Sikkim, India.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Application of Decision Tree in Analysis of Intra College Festival Data set

[1]Sumit Srivastava (Dr.)
[1]Department of Information Technology,
Poornima College of Engineering, Jaipur, Rajasthan-302022(India)

[2]Megha Yadav
[2]Department of Computer Science,
Poornima College of Engineering, Jaipur, Rajasthan-302022(India)

[3]Sandeep Devra
[3]Department of Computer Science,
Jagannath Institute of Engineering & Technology, Jaipur, Rajasthan-302022(India)

Abstract - Data mining has been used very frequently to extract hidden information from large databases. The paper suggested the use of data mining technique named decision trees along with its algorithm for continuously monitoring the College fest. In this way the data extracted could be used to teach the students on the real time problem scenario application to be monitored at college level. Also the model can be used for the future planning of such fest at college level. The algorithm used is CHAID (Chi-squared Automatic Interaction Detection) using PASW18 as software.

Keywords- Data Mining, ID3, C4.5, CART, CHAID

1. Introduction

It is the process of extracting patterns from large data sets by combining methods from statistics and artificial intelligence with database management. It is an increasingly important tool by modern business to transform data into business intelligence. It is currently used in a wide range of profiling practices, such as marketing, surveillance, fraud detection, and scientific discovery. In this paper, the effective use of the information available by using the decision trees on a festival data set. The problem is to classify the information available on the basis of year of student participated in the various events, participation on the basis of the gender, particiaption on the basis of type of event(can be clutural, club, sports and tehnical), branch wise information. The algorithm available to be used in Decision trees ID3, C4.5, CART, CHAID to be used with software tools like PASW18 onwards. Also the comparison with the other algorithms is being done.

2. Decision Trees

Decision trees are often used in classification and prediction. It is simple yet a powerful way of knowledge Representation [1]. The models produced by decision trees are represented in the form of tree structure. A leaf node indicates the class of the examples. The instances are classified by sorting them down the tree from the root node to some leaf node.

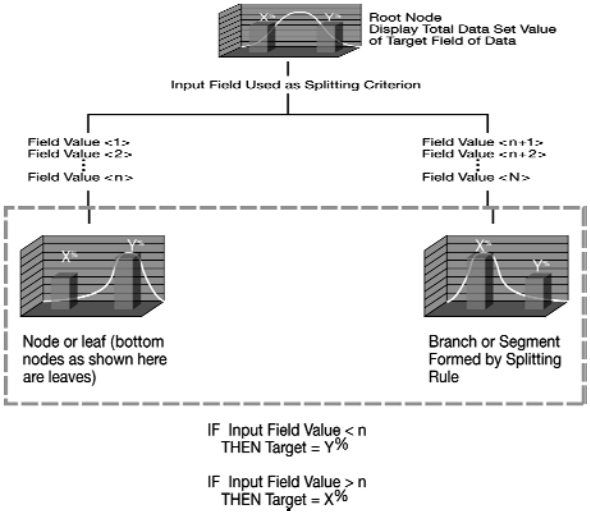


FIG 2.1: Decision tree reflecting both continuous and categorical object of analysis.

The reason of using Decision trees are

1. Decision trees are white boxes means they generate simple, understandable rules
2. Decision trees are non-parametric means no specific data distribution is necessary.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

3. Decision trees handle missing values as easily as any normal value of the variable.
4. In decision trees elegant tweaking is possible.
5. Decision trees identify subgroups. Each terminal or intermediate leave in a decision tree can be seen as a subgroup/segment of your population.
6. Decision trees run fast even with lots of observations and variables
7. Decision trees can be used for supervised and unsupervised learning.
8. Decision trees can easily handle unbalanced datasets.
9. Versatility for a wide variety of data mining tasks, such as classification, regression, clustering and feature selection

while some of the disadvantage lies as following

1. While decision trees classify quickly, the time for building a tree may be higher than another type of classifier.
2. Decision trees suffer from a problem of error propagation
3. Disjunctive descriptions may be required
   a. naturally represent disjunctive expressions

the algorithms being used while decision tree analysis is as follows.

2.1 ID3 Algorithm:- it is a greedy search technique, for which top-down recursive divide and conquer manner. Generally the question relates to classifying the information. It generally use information gain as a quantitative measure of an attribute[3]. The algorithm for attribute selection measure

- Select the attribute with the highest information gain
- Let p_i be the probability that an arbitrary tuple in D belongs to class C_i, estimated by $|C_{i,D}|/|D|$
- Expected information (entropy) needed to classify a tuple in D:

$$Info(D) = -\sum_{i=1}^{m} p_i \log_2(p_i) \quad \ldots\ldots(2.1)$$

- Information needed (after using A to split D into v partitions) to classify D:
  - Information gained by branching on attribute A (If A is discrete valued):

$$Info_A(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} \times I(D_j) \quad \ldots\ldots(2.2)$$

- Information gained by branching on attribute A (If A is discrete valued):

$$Gain(A) = Info(D) - Info_A(D) \quad \ldots\ldots(2.3)$$

2.2 C4.5 Algorithm:- C4.5 (a successor ofID3) uses gain ratio to overcome the problem. Here for Computing the information- gain for continuous value attributes the best spilt-point is being calculated [2][4]. The algorithm for calculating the same is:

- Let attribute A be a continuous-valued attribute
- Must determine the *best split point* for A
- Sort the value A in increasing order
- Typically, the midpoint between each pair of adjacent values is considered as a possible *split point*
- $(a_i+a_{i+1})/2$ is the midpoint between the values of a_i and a_{i+1}
- The point with the *minimum expected information requirement* for A is selected as the split-point for A

and the SPLIT is calculated as

- D1 is the set of tuples in D satisfying A ≤ split-point, and D2 is the set of tuples in D satisfying A > split-point

It uses the kind of normalization to information gain using a "Split Information" value defined analogously with Info(D) as:

$$\boxed{SplitINFO_A(D) = -\sum_{i=1}^{k} \frac{D_i}{D} \log \frac{D_i}{D}} \quad \ldots\ldots(2.4)$$

and the gain ratio defined as

$$\boxed{GainRATIO(A) = \frac{GAIN(A)}{SplitINFO(A)}} \quad \ldots\ldots(2.5)$$

2.3 CART Algorithm: Classification and Regression Trees (CART) is a non-parametric technique that produces either classification or regression trees, depending on whether the dependent variable is categorical or numeric, respectively. The Gini index is used in CART. Gini index measures the impurity of *D, a data partition or set of training tuples, as*

$$gini(D) = 1 - \sum_{j=1}^{n} p_j^2 \quad \ldots(2.6)$$

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

where p_j is the relative frequency of class j in D. The Gini Index consider the binary split of the each of the attribute. If the data set D is split into subset D_1 and D_2, the gini-index can be defined as

$$gini_A(D) = \frac{|D_1|}{|D|} gini(D_1) + \frac{|D_2|}{|D|} gini(D_2) \quad ...(2.7)$$

and Reduction in Impurity (that would be incurred by a binary split on a discrete- or continuous-valued attribute A) is:

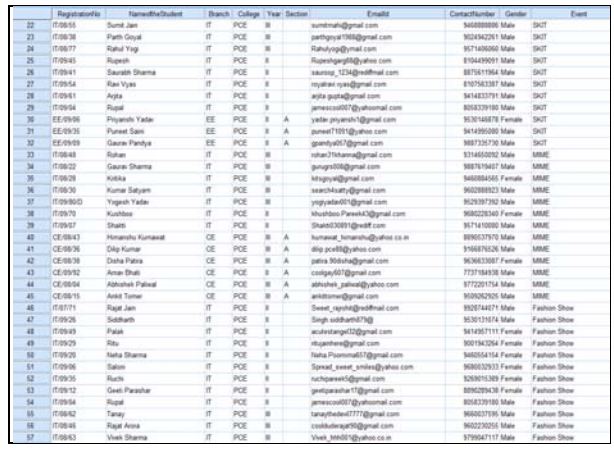
$$\Delta gini(A) = gini(D) - gini_A(D) \quad ....(2.8)$$

2.5 CHAID Algorithm:- Chi-squared Automatic Interaction Detection is a non-binary decision tree algorithm, where split search is designed for categorical values. CHAID uses a Chi-squared test using contingency tables. The advantage for using CHAID as method of evaluation is:

- Used in fields of marketing research and public health
- Can handle a large number of variables
- Designed to identify potentially meaningful patterns in a dataset
- Easy to use and understand

3. Decision Trees

The application data set consists of the cultural fest named "AAROHAN2k10" organized as annual fest in one of the private engineering colleges. It includes nearly 22 technical events, 13 club events, 7-8 sports events, 9-10 cultural events with 5500 approx student participating for the competition. For each student participation in activity involving Technical and club is a must with total of 4 events per student has to be submitted. So with almost 5500*4= 22,000 entries as input the event has been successfully completed. The snapshot of the data taken from AAROHAN2k10 is shown below



	RegistrationNo	NameoftheStudent	Branch	College	Year	Section	EmailId	ContactNumber	Gender	Event
22	IT/08/55	Sumit Jain	IT	PCE	III		sumitmahi@gmail.com	946888886	Male	SKIT
23	IT/09/38	Parth Goyal	IT	PCE	III		parthgoyal1988@gmail.com	9024342261	Male	SKIT
24	IT/08/77	Rahul Yogi	IT	PCE	III		Rahulyogi@gmail.com	9571456060	Male	SKIT
25	IT/09/45	Rupesh	IT	PCE	II		Rupeshjaiswal88@yahoo.com	8104499091	Male	SKIT
26	IT/09/41	Saurabh Sharma	IT	PCE	II		saursop_1234@rediffmail.com	8875611964	Male	SKIT
27	IT/09/54	Ravi Vyas	IT	PCE	II		royalravi vyas@gmail.com	8107563387	Male	SKIT
28	IT/09/61	Arjita	IT	PCE	II		arjita gupta@gmail.com	9414833791	Male	SKIT
29	IT/09/04	Rupal	IT	PCE	II		jamescool007@yahoomail.com	8058339180	Male	SKIT
30	EE/09/06	Priyanshi Yadav	EE	PCE	II	A	yadav.priyanshi1@gmail.com	9530166878	Female	SKIT
31	EE/09/35	Puneet Saini	EE	PCE	II	A	puneet71091@yahoo.com	9414955080	Male	SKIT
32	EE/09/09	Gaurav Pandya	EE	PCE	II	A	gpandya067@gmail.com	9887335730	Male	SKIT
33	IT/08/48	Rohan	IT	PCE	III		rohan31khanna@gmail.com	9314650092	Male	MIME
34	IT/08/22	Gaurav Sharma	IT	PCE	III		gurugrs006@gmail.com	9887619407	Male	MIME
35	IT/08/28	Kritika	IT	PCE	III		kitsgnyal@gmail.com	9460884565	Female	MIME
36	IT/08/30	Kumar Satyam	IT	PCE	III		search4satty@gmail.com	9602888923	Male	MIME
37	IT/09/06/D	Yogesh Yadav	IT	PCE	III		yogyadav001@gmail.com	9529397392	Male	MIME
38	IT/09/70	Kushboo	IT	PCE	II		khushboo.Parwani3@gmail.com	9680228340	Female	MIME
39	IT/09/07	Shakti	IT	PCE	II		Shakti030891@rediff.com	9571410080	Male	MIME
40	CE/08/43	Himanshu Kumawat	CE	PCE	III	A	kumawat_himanshu@yahoo.co.in	8890537570	Male	MIME
41	CE/08/36	Dilip Kumar	CE	PCE	III	A	dilip.pce88@yahoo.com	9166876526	Male	MIME
42	CE/08/38	Disha Patira	CE	PCE	III	A	patira.96disha@gmail.com	9636633087	Female	MIME
43	CE/09/02	Aman Bhati	CE	PCE	II	A	codige661@gmail.com	7737184938	Male	MIME
44	CE/08/04	Abhishek Paliwal	CE	PCE	III	A	abhishek_paliwal@yahoo.com	9772201754	Male	MIME
45	CE/08/15	Ankit Tomer	CE	PCE	III	A	ankittomer@gmail.com	9509262926	Male	MIME
46	IT/07/71	Rajat Jain	IT	PCE	II		Sweet_rajrishi@rediffmail.com	9926744671	Male	Fashion Show
47	IT/09/26	Siddharth	IT	PCE	II		Singh.siddharth879@	9530131674	Male	Fashion Show
48	IT/09/49	Palak	IT	PCE	II		aculvistangel32@gmail.com	9414957111	Female	Fashion Show
49	IT/09/29	Ritu	IT	PCE	II		ritupatihere@gmail.com	9001843264	Female	Fashion Show
50	IT/09/20	Neha Sharma	IT	PCE	II		Neha.Poonmal67@gmail.com	9460554154	Female	Fashion Show
51	IT/09/06	Saloni	IT	PCE	II		Spread_sweet_smiles@yahoo.com	9680032933	Female	Fashion Show
52	IT/09/35	Ruchi	IT	PCE	II		ruchiparek5@gmail.com	9269015369	Female	Fashion Show
53	IT/09/12	Geeti Parashar	IT	PCE	II		geetiparashar17@gmail.com	8890289436	Female	Fashion Show
54	IT/09/04	Rupal	IT	PCE	II		jamescool007@yahoomail.com	8058339180	Male	Fashion Show
55	IT/08/62	Tanay	IT	PCE	III		tanaythedeal7777@gmail.com	9680037195	Male	Fashion Show
56	IT/08/48	Rajat Arora	IT	PCE	III		coolbuddyrajat90@gmail.com	9602230255	Male	Fashion Show
57	IT/08/63	Vivek Sharma	IT	PCE	III		Vivek_tmh001@yahoo.co.in	9799047117	Male	Fashion Show

Fig 3.1: Training Data Set (taken from Aahoran database)

During the result analysis, it was asked to predict the participation in each types of events by each branch of each college with Gender of the students can be as the dependent variable for the analysis. Also if the event is supposed to a group then classification of the information on the basis of college, year, branch, team-name, event-name is requested. For such problem CHAID is selected as the best technique for the solution as information can be branched using non-binary variable. Some of the results snapshots calculated using PASW 18 beta version as software is shown below.

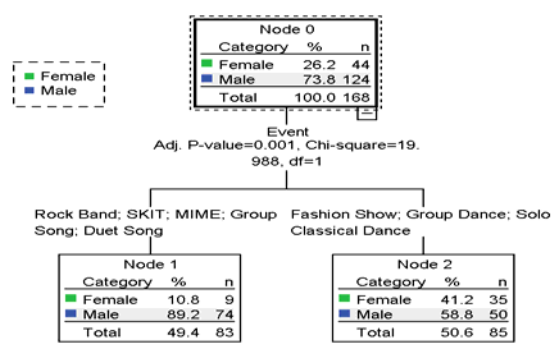


Fig 3.2:- When the dependant variable is taken to be Gender and independent variables are: Event, College,

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Branch)

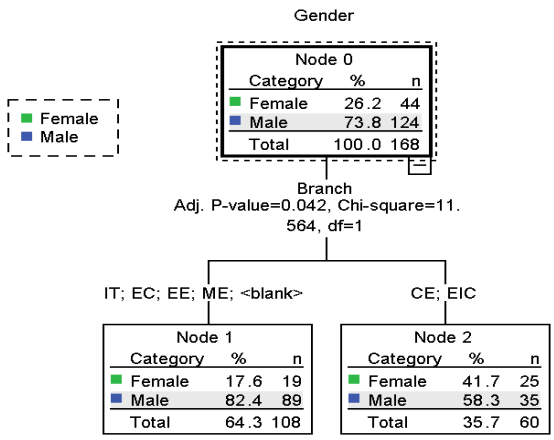


Fig 3.3:- When the dependant variable is taken to be Gender and independent variables are: College, Branch)

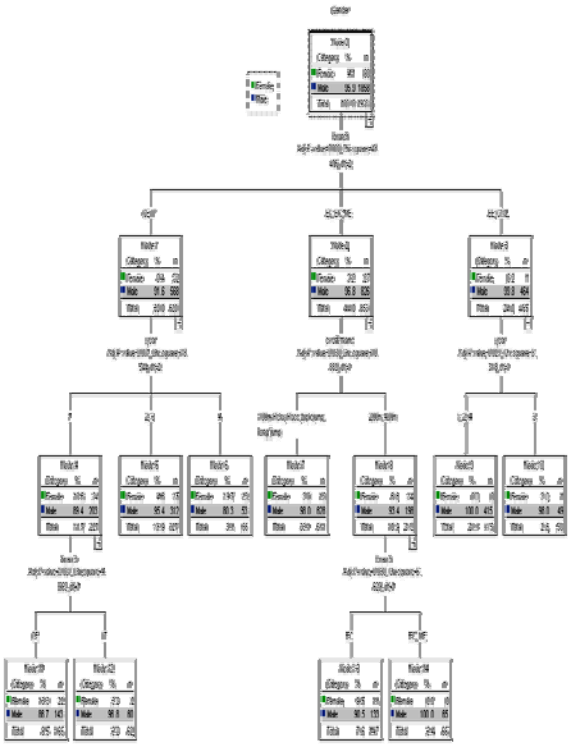


Fig 3.4:- When the dependant variable is taken to be Gender and independent variables are: college, year, branch, team_name, event_name

4. Conclusion

The inference from the various result analyses can be shown as.

- It was predicted that 89.2% out of 73.9% of the males have participated in events like Rock Band, SKIT, MIME, Group Song and Duet Song.
- On the basis of the data studied it can be concluded that 58.8% out of 73.9% of the males have participated in events Cultural and Club Events
- it can be further concluded that 58.3% out of 73.8% of the males are from CE, EIC branches who have participated in cultural events
- On the basis of the data studied it can be concluded that 82.4% out of 73.8% of the males are from IT, ME, EC, EE branches who have participated in cultural events, sports and technical events

With fast analysis of large databases and classification of information using Decision trees, CHAID is founded as the one among the best algorithm for the large data analysis.

5. Acknowledgement

We acknowledgement the Director of the engineering institute for providing us the opportunity to work on their database.

6. Bibliographies and References

[1] T. Mitchell, *"Decision Tree Learning*", in T. Mitchell, Machine
Learning, the McGraw-Hill Companies, Inc., 1997, pp. 52-78.
[2] P. Winston, "*Learning by Building Identification Trees*", in P. Winston, Artificial Intelligence, Addison-Wesley Publishing Company, 1992, pp. 423-442.
[3] Howard J. Hamilton's CS Course: Knowledge Discovery in Databases. Accessed 06/06/12.
[4] http://www.cs.waikato.ac.nz/ml/weka/, accessed 06/05/21.
[5]http://grb.mnsu.edu/grbts/doc/manual/J48_Decision_Trees.html, accessed 06/06/12.
[6] Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kauffman, 1993.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

[1]Dr. Sumit Srivastava
Professor,
Department of Information Technology,
Poornima College of Engineering,
Jaipur-302022(raj)

[2]Mrs. Megha Yadav
M.Tech, Research Scholar,
Department of Computer Science Engineering,
Poornima College of Engineering,
Jaipur-302022(raj)

[3]Mr.Sandeep Devra
Associate Professor,
Department of Computer Engineering,
JNIT, Jaipur-302022(Raj)

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

String Matching and its Applications in Diversified Fields

Vidya SaiKrishna[1], Prof. Akhtar Rasool[2] and Dr. Nilay Khare[3]

[1]Department Of Computer Science And Engineering Maulana Azad National Institute Of Technology Bhopal-462051

[2]Asst.Professor,
[2]Department Of Computer Science And Engineering Maulana Azad National Institute Of Technology Bhopal-462051

[3]Associate Professor, HOD
[3]Department Of Computer Science And Engineering Maulana Azad National Institute Of Technology Bhopal-462051

Abstract

String searching algorithms, sometimes called string matching algorithms, are an important class of string algorithms that try to find a place where one or several strings (also called patterns) are found within a larger string or text.[11] String matching is a classical problem in computer science. In this paper we are trying to explore the various diversified fields where string matching has an eminent role to play and is found as a solution to many problems. Few of the fields exploited are intrusion detection in network, application in bioinformatics, detecting plagiarism, information security, pattern recognition, document matching and text mining. Here we discuss how string matching is found useful in finding solutions to above problems. String matching algorithms can be categorized either as exact string matching algorithms or approximate string matching algorithms . Also depending upon the kind of application, string matching algorithms are designed either to work on single pattern or multiple patterns. Followed by a brief introduction to string matching, the application areas are discussed.

Keywords: *String matching, intrusion detection, bioinformatics, pattern recognition, text mining, digital forensics.*

1. Introduction

String searching algorithms, sometimes called string matching algorithms, are an important class of string algorithms that try to find a place where one or several strings (also called patterns) are found within a larger string or text. Let Σ be an alphabet (finite set). Formally, both the pattern and searched text are vectors of elements of Σ. The Σ may be a usual human alphabet (for example, the letters A through Z in the Latin alphabet). Other applications may use binary alphabet (Σ = {0,1}) or *DNA alphabet* (Σ = A,C,G,T}) in bioinformatics.[11]

We assume that the text is an array T[1..n] of length n and that the pattern is an array of length[1..m] of length m and that m<=n. The character arrays T and P are often called strings of characters.

We say that pattern P occurs with shift s in text T (or equivalently that the pattern P occurs beginning at position s+1 in text T) if 0<=s<=n-m and T[s+1….s+m]=P[1..m]. If P occurs with shift s in T then we calls a valid shift otherwise we call s an invalid shift. The string matching algorithm is the problem of finding all valid shift with which a pattern P occurs in given text T.[1]

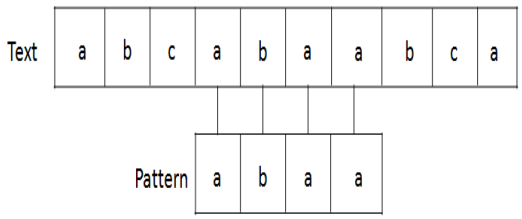


Fig 1

Large number of algorithms is known to exist to solve string matching problem. Based on the number of patterns searched for the algorithms can be classified as single pattern and multiple pattern algorithms. Applications may require exact or approximate string matching.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

2. Exact String Matching Problem

We are given a text string pattern string we want to find all occurrences of *P* in *T*. In Exact string matching problem the pattern is exactly found inside the text.[12]

Consider the following example:

$$T = AGCCTAAGCTCCTAAGTC$$

$$P = CCTA$$

There are two occurrences of *P* in *T* as shown below:

$$AG\underline{CCTA}AGCT\underline{CCTA}AGTC$$

A brute force method for exact string matching algorithm:

T = ACCACTAGA
P = ACTA
ACTA
ACTA
ACTA

If the brute force method is used, many characters which had been matched will be matched again because each time a mismatch occurs, the pattern is moved only one step.

There are many exact string matching algorithms. Nearly all of them are concerned with how to slide the pattern. Few of them are listed below.

2.1 Brute Force Algorithm[12, page 19-20]

- No preprocessing phase.
- Constant extra space needed.
- Always shifts the window by exactly 1 position to the right.
- Comparisons can be done in any order.
- Searching phase in O (m×n) time complexity.
- 2n expected text character comparisons.

2.2 Searching with automation [12, page 25-30]

- Builds the minimal Deterministic Finite Automaton recognizing the language ∑*x.
- Extra space in O(m×σ) if the automaton is stored in a direct access table.
- Preprocessing phase in O(m×σ) time complexity.
- Searching phase in O(n) time complexity.

2.3 Rabin Karp Algorithm[12, page 31-35]

- Uses an hashing function.
- Preprocessing phase in O(m) time complexity and constant space.
- Searching phase in O(m× n) time complexity.
- O(m+n) expected running time.

2.4 Shift OR Algorithm[12, page 37-40]

- Uses bitwise techniques.
- Efficient if the pattern length is no longer than the memory word size of the machine.
- Preprocessing phase in O(m+σ) time and space complexity.
- Searching phase in O(n) time complexity(independent from the alphabet size and the pattern length).
- Adapts easily to approximate string matching.

2.5 Morris Pratt Algorithm[12, page 41-44]

- Performs the comparisons from left to right.
- Preprocessing phase in O(m) space and time complexity.
- Searching phase in O(m+n) time complexity independent from the alphabet size.
- Performs at most 2n -1 text character comparisons during the searching phase.
- Delay bounded by m.

2.6 Knuth- Morris Pratt Algorithm[[12, page 47-50]

- Performs the comparisons from left to right.
- Preprocessing phase in O(m) space and time complexity.
- Searching phase in O(m+n) time complexity independent from the alphabet size.
- Performs at most 2n -1 text character comparisons during the searching phase.
- Delay bounded by $\log_{\Phi}$ (m) where $_{\Phi}$ is the golden ratio(1+√5)/2.

2.7 Colussi Algorithm[12, page 61-67]

- Refinement of the Knuth Morris Pratt algorithm.
- Partitions the set of pattern positions into two disjoint subsets. The positions in the first set are scanned from left to right and when no mismatch occurs the positions of the second subset are scanned from right to left.
- Preprocessing phase in O(m) time and space complexity.
- Searching phase in O(n) time complexity.
- Performs 3/2n text character comparisons in the worst case.

2.8 Forward DAWG Matching algorithm[12, page 87-90]

- Uses the suffix automaton of x.
- O(n) worst case time complexity.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

- Performs exactly n text character inspections.

2.9 Boyer Moore Algorithm[12, page 91-96]

- Performs the comparisons from right to left.
- Preprocessing phase in O(m+σ) time and space complexity.
- Searching phase in O(m×n) time complexity.
- n text character comparisons in the worst case when searching for non periodic pattern.
- O(n/m) best performance.

2.3.0 Quick Search algorithm[12, page 121-123]

- Simplification of the Boyer Moore algorithm.
- Uses only the bad character shift.
- Easy to implement.
- Preprocessing phase in O(m+σ) time and O(σ) space complexity.
- Searching phase in O(m×n) time Complexity.
- Very fast in practice for short patterns and large alphabets.

3. Approximate String Matching Problem

Approximate string matching is a recurrent problem in computer Science which is applied in text searching, computational biology, pattern recognition and signal processing applications[13]. The problem can be stated as follows:

For a length n and pattern of length m , we are supposed to find all the occurrences of pattern in the text whose edit distance to the pattern is at most K. The edit distance between two strings is defined as minimum number of character insertion, deletion and replacements needed to make them equal.

An Example of approximate string matching is shown below[14]:

T ="appropriate_meaning"

P="approximate_matching"

```
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
    T a p p r o p r i a t e _ m ε e a n i n g
      | | | | |     | | |  | |           |  | |
    p: a p p r o x I m a t e _ m a t c h i n g
```

δ(t, p) = 7

Dynamic Programming is a method to solve approximate string matching problem. For a Text of length n and pattern of length m ,the dynamic programming method returns a time complexity of O(nm). Bit Parallelism results in faster approximate string matching algorithm like the the fastest non-filtering algorithms in practice are the O(kn[m/w]) algorithm of Wu & Manber, the O([km/w]n) algorithm of Baeza-Yates & Navarro, and the O([m/w]n) algorithm of Myers, where m is the pattern length, n is the text length, k is the error threshold and w is the computer word size.

The motivation for approximate string matching comes from low quality of text, heterogeneousness of databases, spelling errors in the pattern or text, searching for foreign names and searching with uncertainty [13].

4. Multiple String Matching Problem

In multiple string matching we are given a text T = $t_1t_2 : : : t_n$ and want to search simultaneously for a set of strings P = $\{p^1, p^2, \ldots p^r\}$ Where $p^i=p^i_1\, p^i_2 \ldots\ldots\ldots\ldots p^i_{mi}$ is a string of length mi, for i=1…..r[16]. There are many algorithms used for multipattern searching , which varies in speed measured in terms of time complexity. A few are described below .

4.1 Aho-Corasick string matching algorithm

In computer science, the Aho–Corasick string matching algorithm is a string searching algorithm invented by Alfred V. Aho and Margaret J. Corasick. It is a kind of dictionary-matching algorithm that locates elements of a finite set of strings (the "dictionary") within an input text. It matches all patterns simultaneously. The complexity of the algorithm is linear in the length of the patterns plus the length of the searched text plus the number of output matches.

Informally, the algorithm constructs a finite state machine that resembles a trie with additional links between the various internal nodes. These extra internal links allow fast transitions between failed pattern matches (e.g. a search for cat in a trie that does not contain cat, but contains cart, and thus would fail at the node prefixed by ca), to other branches of the trie that share a common prefix (e.g., in the previous case, a branch for attribute might be the best lateral transition). This allows the automaton to transition between pattern matches without the need for backtracking.[17]

4.2 Rabin Karp String matching Algorithm

In computer science, the Rabin–Karp algorithm is a string searching algorithm created by Michael O. Rabin and Richard M. Karp in 1987 that uses hashing to find any one of a set of pattern strings in a text. For text of length *n* and *p* patterns of combined length *m*, its average and best case running time is $\underline{O}(n+m)$ in space O(*p*), but its worst-case time is O(*nm*). In

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

contrast, the Aho–Corasick string matching algorithm has asymptotic worst-time complexity O(*n*+*m*) in space O(*m*).A practical application of Rabin–Karp is detecting plagiarism. Given source material, Rabin–Karp can rapidly search through a paper for instances of sentences from the source material, ignoring details such as case and punctuation. Because of the abundance of the sought strings, single-string searching algorithms are impractical [18].

4.3 Commentz-Walter algorithm

In computer science, the **Commentz-Walter algorithm** is a string searching algorithm invented by Beate Commentz-Walter. Like the Aho–Corasick string matching algorithm, it can search for multiple patterns at once. It combines ideas from Aho–Corasick with the fast matching of the Boyer–Moore string search algorithm [19].

5. Applications of String Matching

5.1. Intrusion Detection

Security is a big issue for all networks in today's enterprise environment. Hackers and intruders have made many successful attempts to bring down high-profile company networks and web services. Many methods have been developed to secure the network infrastructure and communication over the Internet, among them the use of firewalls, encryption, and virtual private networks. Intrusion detection is a relatively new addition to such techniques. Intrusion detection methods started appearing in the last few years. Using intrusion detection methods, you can collect and use information from known types of attacks and find out if someone is trying to attack your network or particular hosts. The information collected this way can be used to harden your network security, as well as for legal purposes.[2]
Intrusion detection is a set of techniques and methods that are used to detect suspicious activity both at the network and host level. Intrusion detection systems fall into two basic categories: signature-based intrusion detection systems and anomaly detection systems. Intruders have signatures, like computer viruses, that can be detected using software. You try to find data packets that contain any known intrusion-related signatures or anomalies related to Internet protocols. Based upon a set of signatures and rules, the detection system is able to find and log suspicious activity and generate alerts. Anomaly-based intrusion detection usually depends on packet anomalies present in protocol header parts. In some cases these methods produce better results compared to signature-based IDS. Usually an intrusion detection system captures data from the network and applies its rules to that data or detects anomalies in it. Snort is an open source IDS available to the general public. NIDS are intrusion detection systems that capture data packets traveling on the network media (cables, wireless) and match them to a database of signatures. Depending upon whether a packet is matched with an intruder signature, an alert is generated or the packet is logged to a file or database. One major use of Snort is as a NIDS.

String matching in intrusion detection

The earlier intrusion detection systems makes use of the AC Algorithm (Aho- Corasick) .It is an automaton based multiple string matching algorithm which locates all the occurrences of keywords in a string or text. It first builds a finite state machine of all the keywords in a string and then uses the machine to process the payload in a single pass. The AC algorithm is having a deterministic performance which does not depends on specific input and therefore not vulnerable to various attacks, making it attractive to Network intrusion detection systems.[3]
Let us consider an example to understand how the algorithm works. P={at, cat,rat} is a set of keywords and the algorithm searches for the occurrences of keywords in the text ="cratcatar". The AC automaton built on pattern P is shown in figure 2. The Dashed arrows represent failure transitions. The failure function is shown in figure 3 and output function in figure 4.

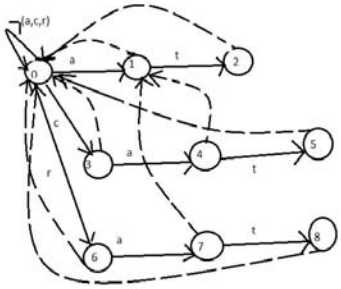


Fig 2: AC Automaton

I	1	2	3	4	5	6	7	8
F(i)	0	0	0	1	0	0	1	0

Fig 3: failure Function

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

i	output(i)
2	{at}
5	{cat,at}
8	{rat,at}

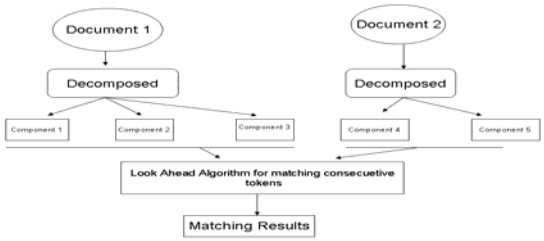
Fig 4: output function

Original AC automaton representations are unsuitable for high performance requirements, which require too much memory, and too much time in the matching process. Recently many works have been presented with the goal of memory reduction for DFAs, by exploiting the intrinsic redundancy in regular expression sets. NIDS not only focus on the header fields but also have to check signatures in the data payload portion of a packet. The most widely used NIDS like Snort and Bro use DFA to describe attack signatures in their rule sets. Because of the large traffic volume and complexity of the process, signature matching can easily become the performance bottleneck in deep packet inspection.[4]

5.2. String matching in detecting plagiarism

Management of large collection of replicated data in centralized or distributed environments is important for many systems that provide data mining, mirroring, storage, and content distribution. In its simplest form, the documents are generated, duplicated and updated by emails and web pages. Although redundancy may increase the reliability at a level, uncontrolled redundancy aggravates the retrieval performance and might be useless if the returned documents are obsolete. Document similarity matching algorithms do not provide the information on the differences of documents, and file synchronization algorithms are usually inefficient and ignore the structural and syntactic organization of documents. For this purpose the *S2S* matching approach is used The *S2S* matching is composed of structural and syntactic phases to compare documents [5]. Firstly, in the structural phase, documents are decomposed into components by its syntax and compared at the coarse level. The structural mapping processes the decomposed documents based on its syntax without actually mapping at the word level. The structural mapping can be applied in a hierarchical way based on the structural organization of a document. Secondly, the syntactic matching algorithm uses a heuristic look-ahead algorithm for matching consecutive tokens with a verification patch. The two-phase *S2S* matching approach provides faster results than currently available string matching algorithms.[4,5]

Figure 5: S2S matching



5.3 String matching in bioinformatics

Bioinformatics is the application of information technology and computer science to biological problems, in particular to issues involving genetic sequences. String algorithms are centrally important in bioinformatics for dealing with sequence information. Modern automated high throughput experimental procedures produce large amounts of data for which machine learning and data mining approaches hold great promise as interpretive means[6].

Approximate matching of a search pattern to a target (called the "text" in string algorithms) is a fundamental tool in molecular biology. The pattern is often called the "query" and the text is called a "sequence database", but we will use "pattern" and "text" consistent with usage in computer science.

While exact string matching is more commonly used in computer science, it is often not useful in biology. One reason for this is that biological sequences are experimentally determined, and may include errors: a single error can render an exact match useless, where approximate matches are less susceptible to errors and other sequence differences. Another, perhaps more important, reason for the importance of approximate matching is that biological sequences change and evolve. Related genes in different organisms, or even similar genes within the same organism, most commonly have similar, but not identical sequences. Determining which sequences of known function are most similar to a new gene of unknown function is often the first step in finding out what the new gene does.

Another application for approximate string matching is predicting the results of hybridization experiments. Since strands may hybridize if they are similar to each other's reverse complements, prediction of which strands will bind to which other strands, and how stable the binding will be, requires approximate, rather than exact, string matching.[6]

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Approximate string matching

```
character[] pattern, text;
integer i,j;
for (i=0;i<length text;i++){
for (j=0; j< length pattern; j++){
if (text[i+j] != pattern[j]) next i;
}
record_match(i);
}
```

Note: the command "next i" exits the current j loop without completing the call to record match.

Biological sequences can be represented as strings. Approximate matching algorithms that can tolerate insertions, deletions, and substitutions are extremely important for biological sequence comparison.

The Shift-AND method uses a bit manipulation approach to accelerate the process of approximate matching. The approach can be explained by comparing it to the naïve exact matching method, in which the pattern is compared character by character at each position along the text. This simple approach is inefficient (its time complexity is O(n*m)) because it contains two nested loops, where the inner loop is executed at each position along the text. Shift-AND uses bit-wise operations in entire registers to perform the inner loop operations on multiple positions in parallel. For patterns that can be contained within the length of a register, it has a time complexity proportionate to the length of the text being searched (O(n)). Shift-AND actually uses a set of four registers (called the "U" registers [Gusfield 1997]) to contain the pattern, with one bit in each register to represent each base of the pattern. Thus, a machine with 32 bit registers can easily represent 32 base patterns for use in shift-AND. Longer registers, such as the 128 bit registers of the PowerPC Alti Vec vector engine, can represent proportionately longer patterns. The algorithm can also be extended to use multiple registers to represent longer patterns, but (depending on the architecture), this would likely be at a cost of increased time complexity.

Of course, shift-AND is more sophisticated than a simple short-circuit of the inner loop in the naïve exact matching approach, in that it can be extended to allow for approximate matches 9
[Wu 1991].

The full algorithm allows for a given number of substitutions, insertions, or deletions in the pattern. This is achieved by maintaining a set of matrixes, which allow different numbers of errors. A bit is set if the current characters match and the prefixes of each string were within the error limits; if the prefixes had not reached the error limits, the bit is set even if there is an error at the current position.

5.4 String matching in Digital Forensics

Textual evidence is important to the vast majority of digital investigations. This is because a great deal of stored digital data is linguistic in nature (e.g. human languages, programming languages, and system and application logging conventions). Some examples of important text-based evidence include: email, Internet browsing history (both logs and the content itself), instant messaging, word processing documents, spreadsheets, presentations, address books, calendar appointments, network activity logs, and system logs.

Digital forensic1 text string searches are designed to search every byte of the digital evidence, at the physical level, to locate specific text strings of interest to the investigation. Given the nature of the data sets typically encountered, text string search results are extremely noisy, which results in inordinately high levels of information retrieval (IR) overhead and information overload (Beebe and Dietrich, 2007). Frequently, investigators are left to wade through hundreds of thousands of search hits for even reasonably small queries (i.e. 10 search strings) on reasonably small devices (i.e. 80 GB) – most of which (i.e. 80–90% of the search hits) are irrelevant to investigative objectives. The investigator becomes inundated with data and wastes valuable investigative time scanning through noisy search results and reviewing irrelevant search hits presumably by some automated means. This can prohibit There are fundamentally two classes of solutions to this problem: (1) decrease the number of irrelevant search hits returned, or (2) present the search hits in a manner which enables the investigator to find the relevant hits more quickly. The first solution class is disconcerting to many investigators and litigators, since it results in information reduction, the investigator from finding important evidence. The second solution class encompasses the basic approach that revolutionized web-based knowledge discovery: search hit ranking. This approach presents hits in priority order based on some determination of similarity and/or relevancy to the query. This approach is much more attractive to investigators and litigators, since it improves the ability to obtain important information, without sacrificing fidelity.

Current digital forensic text string search approaches fail to employ either solution class. They use simple matching and/or indexing algorithms that return all hits. They fail to successfully implement grouping and/or ranking algorithms in a manner that appreciably reduce IR overhead (time spent scanning/reviewing irrelevant search hits). Search hits are commonly grouped by search string and/or

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

''file item3'' and/ or ordered by their physical location on the digital media.
Such grouping and ordering are inadequate; as neither substantially helps investigators get to the relevant hits first (or at least more quickly). New, better approaches in the second solution class are needed.
The Current researches in digital forensics aims at improving IR(information Retrieval) effectiveness of digital forensics text string searches.[7,8]

5.5 Text Mining Research

Text mining includes tasks designed to extract previously unknown information by analyzing large quantities of text, as well as tasks geared toward the retrieval of textual data from a large corpus of documents (Sebastian, 2002; Fan et al., 2006; Sullivan, 2001). Several information processing tasks fall under the umbrella of text mining: information extraction, topic tracking, content summarization, information visualization, question answering, concept linkage, text categorization/ classification, and text clustering (Fan et al., 2006).
These are defined as follows:
1. Information extraction: identifies conceptual relationships, using known syntactic patterns and rules within a language.
2. Topic tracking: facilitates automated information filtering, wherein user interest profiles are defined and fine-tuned based on what documents users read.
3. Content summarization: abstracts and condenses document content.
4. Information visualization: represents textual data graphically (e.g. hierarchical concept maps, social networks, timeline representations).
5. Question answering: automatically extracts key concepts from a submitted question, and
Subsequently extracts relevant text from its data store to answer the question(s).
6. Concept linkage: identifies conceptual relationships between documents based on Transitive relationships between words/concepts in the documents.
7. Text categorization/classification: automatically and probabilistically assigns text documents into predefined thematic categories, using only the textual content (i.e. no metadata).
8. Text clustering: automatically identifies thematic categories and then automatically assigns text documents to those categories, using only textual content (i.e. no metadata).

If applied to digital forensic text string searching, information extraction, content summarization, information visualization, and concept linkage would fit into the first solution class identified earlier (reduction of search results set size).These text mining approaches reduce the search result set size via data abstraction techniques, by and large.

5.6 String matching Based Video Retrieval

String matching can be effectively used to retrieve fast video as it uses the content based video retrieval in contrast with the traditional video retrieval which was slow and time consuming. String based video retrieval method first converts the unstructured video into a curve and marks the feature string of it. Approximate string matching is then used to retrieve video quickly. In this method the characteristic curve of the key frame sequence is first extracted followed by marking the feature string and then approximate string matching is used on the feature string to get fast video retrieval [10].

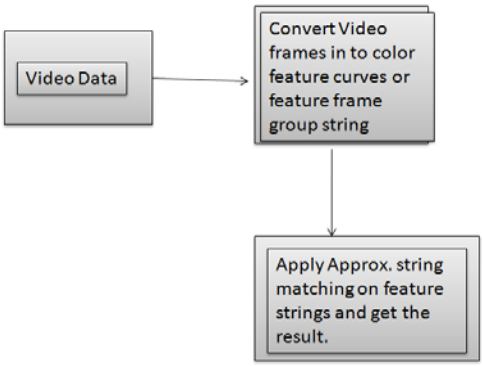


Fig 6: Video Retrieval Process

6. Conclusion

String matching has greatly influenced the field of computer science and will play an important role in future technology. The importance of memory and time efficient string matching algorithm will be increased in computer science. There are many more possible areas in which string matching can play a key role for excelling. Exact and approximate string matching algorithm makes various problems in the solvable state. Innovation and creativity in string matching can play a immense role for getting time efficient performance in various domains of computer science.

References

[1] Thomas H Corman, Charles E. Leiserson, Ronald L. Rivest & Clifford Stein "Introduction to Algorithms-String matching", EEE Edition, 2nd Edition, Page no 906-907.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

[2] Ali Peiravi, "Application of string matching in Internet Security and Reliability", Marsland Press Journal of American Science 2010, 6(1): 25-33
[3] Peifeng Wang , Yue Hu, Li Li, "An Efficient Automaton Based String Matching Algorithm and its application in Intrusion Detection", International Journal of Advancements in Computing Techology(IJACT), Vol 3, Number 9 , October 2011
[4] Pekka Kilpelainen, "Set Matching and Aho-Corasick Algorithm", Biosequence Algorithms, Spring 2005, BSA Lecture 4
[5] Ramazan S. Aygün "structural-to-syntactic matching similar documents", Journal Knowledge and Information Systems archive, Volume 16 Issue 3, August 2008.

[6] Robert M. Horton, Ph.D. "Bioinformatics Algorithm Demonstrations in Microsoft Excel" , 2004 - cybertory.org
[7] Nicole Lang Beebe, Jan Guynes Clark, "Digital forensic text string searching: Improving information retrieval effectiveness by thematically clustering search results", d i g i t a l i n v e s t i g a t ion 4 S (2 0 0 7) S 4 9 – S 5 4
[8] Beebe NL, Dietrich G. "A new process model for text string searching". In: Shenoi S, Craiger P, editors. Research advances in digital forensics III. Norwell: Springer; 2007. p. 73–85.
[9] Rafeeq Ur Rehman , "Intrusion Detection Systems with Snort Advanced IDS Techniques Using Snort Apache, MySQL, PHP, and ACID"
[10]Yin Jian, Yu Xiu ,Dong Meng, " Application of Approximate String Matching in Video Retrieval", 2010 3rd International Conference on Advanced Computer Theory and Engineering(ICACTE),vol 4, page 348-351.
[11]Wikipedia The free Encyclopedia *en.wikipedia.org/wiki/**String**_searching_algorithm*,
[12] Christian Charras,Thierry Lecroq , "Handbook of exact string matching algorithms"
[13] Ricardo Baeza-Yates, Gonzalo Novarro, "Fast Approximate string matching in a Dictionary",□Bulletin of the Technical Committee, 2000
[14] Yoan Pinzon, "Algorithm for approximate string Matching",□*dis.unal.edu.co/~fgonza/courses/2006.../ approx_**string_matching**.pdf* August 2006
[15] Heikki Hyyo, "Bit Parallel approximate string matching algorithm with transposition"□String Processing and Information Retrieval, 2003 – Springer.
[16] D. Huson , "multiple string matching", Comp. Sequence Analysis, Nov 17 , 2004
[17] Aho, Alfred V.; Margaret J. Corasick (June 1975). "Efficient string matching: An aid to bibliographic search". Communications of the ACM **18** (6): 333–340.
[18] http://en.wikipedia.org/wiki/Rabin_karp_string_search_algorithm.
[19] http://en.wikipedia.org/wiki/Commentz_Walter_algorithm

Vidya Saikrishna, B.E in Computer Science From University Institute of Technology, Bhopal in the year 2002, M.Tech in Computer Science From Maulana Azad National Institute of Technology. Worked as Asst. Prof in Truba Institute of Science and Technology, Bhopal, Asst. Prof. in Institute of Technology and Science , Ghaziabad. and Presently working as Asst. Prof. in SAM College of Engineering and Technology.

Akhtar Rasool, B.E in Computer Science from Rajiv Gandhi Technical University in the year 2003,M.Tech in Computer Science from Maulana Azad National Institute of Technology. Presently Working as Asst. Prof in Department of Computer Science in Maulana Azad National Institute Technology, Bhopal.

Dr. Nilay Khare, Associate Prof. and Head in Department of Computer Science in Maulana Azad National Institute of Technology,Bhopal. Reviewer of Journal Elsevier

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

A Study on the Behavior of a Neural Network for Grouping the Data

Suneetha Chittineni [1] and Dr.Raveendra Babu Bhogapathi [2]

[1]Department of Computer Applications,
R.V.R. & J.C. College of Engineering,
Guntur, India.

[2]Director-Operations
DELTA Technology & Management Services Pvt. Ltd.,
Hyderabad, India.

Abstract

One of the frequently stated advantages of neural networks is that they can work effectively with non-normally distributed data. But optimal results are possible with normalized data.In this paper, how normality of the input affects the behaviour of a K-means fast learning artificial neural network(KFLANN) for grouping the data is presented. Basically, the grouping of high dimensional input data is controlled by additional neural network input parameters namely vigilance and tolerance. Neural networks learn faster and give better performance if the input variables are pre-processed before being fed to the input units of the neural network. A common way of dealing with data that is not normally distributed is to perform some form of mathematical transformation on the data that shifts it towards a normal distribution.In a neural network, data preprocessing transforms the data into a format that will be more easily and effectively processed for the purpose of the user. Among various methods, Normalization is one which organizes data for more efficient access. Experimental results on several artificial and synthetic data sets indicate that the groups formed in the data vary with non-normally distributed data and normalized data and also depends on the normalization method used.

Keywords: Data Preprocessing, Normalization, Fast Learning Artificial Neural Network.

1. Introduction

Data preprocessing describes any type of processing on atomic data that has not been processed further to preparing it for another processing procedure. Different data preprocessing techniques like cleaning method, outlier detection, data integration and transformation can be carried out before clustering process to achieve successful analysis. Normalization is an important preprocessing step in data mining to standardize the values of all variables from dynamic range into specific range.

Typical objective functions in clustering formalize the goal of attaining high intra-cluster similarity (Patterns within a cluster are similar) and low inter-cluster similarity (Patterns from different clusters are dissimilar). This is an internal criterion for the quality of a clustering.

In unsupervised learning, an output unit is trained to respond to clusters of pattern within the input. In this paradigm, the system is supposed to discover statistically salient features of the input population. Unlike the supervised learning paradigm, there is no a priori set of categories into which the patterns are to be classified; rather, the system must develop its own representation of the input stimuli.

Supervised networks like Simple Perceptrons, Back Propagation (BP), and Radial Basis Function (RBF) networks need a teacher to tell the network what the desired output should be. Unsupervised nets include Kohonen Self Organizing maps (SOM), Adaptive Resonance theory (ART) etc. The major applications of unsupervised nets include clustering data, reducing the dimensionality of the data. In data clustering, exactly one of small number of output units comes on in response to an input. In dimension reduction, large number of input units is compressed into a small number of output units.

Kohonen clustering algorithm takes high-dimensional input, clusters it, and retaining some topological ordering of the output [7]. Adaptive Resonance Theory (ART) was initially introduced by Grossberg (1976) as a theory of human information processing [3]. ART neural networks are extensively used for supervised, unsupervised classification tasks and functional approximation. There exist many different variations of ART networks today (Carpenter and Grossberg, 1998). ART1 performs unsupervised learning for binary input patterns, ART2 is modified to handle both analog and binary input patterns, and ART3 performs parallel searches of distributed recognition codes in a multilevel network hierarchy. Fuzzy ARTMAP represents a synthesis of elements from neural networks, expert systems, and fuzzy logic.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Many variations of fast learning artificial neural network algorithms have been proposed. A fast learning artificial neural network (FLANN) models was first developed by Tay and Evans [17] to solve a set of problems in the area of pattern classification. FLANN [3] [4] was designed with concepts found in ART but imposed the Winner Take All (WTA) property within the algorithm. Further improvement was done to take in numerical continuous value in FLANN II [11]. The original FLANN II was restricted by its sensitivity to the pattern sequence. This was later overcome by the inclusion of k-means calculations, which served to remove inconsistent cluster formations [13]. The KFLANN utilizes the Leader-type algorithm first addressed by Hartigan [9] and also draws some parallel similarities established in the Adaptive Resonance Theories developed by Grossberg [7] and later ART algorithms by Carpenter et al [2].The later improvement on KFLANN [15] includes data point reshuffling which resolves the data sequence sensitivity that creates stable clusters. Clusters are said to be stable if the cluster formation is complete after some iterations and the cluster centers remain consistent.

The organization of this paper is as follows. In section II, an overview of the data normalization methods are described. Section III gives the details of neural network input parameters. Section IV presents various formulas for computing tolerance values of the attributes. In Section V the K-means Fast Learning Artificial Neural Network algorithm is presented. Section VI presents the experimental analysis of the results and conclusions follow in section VII.

2. Data normalization

1) Z-score normalization

The *Z*-score is called as standardized unit. It indicates how far and in what direction a variable deviates from its distribution's mean, expressed in units of its distribution's standard deviation. The *Z*-scores are especially informative when the distribution to which they refer is normal. If the value of *Z* is positive, it means the variable *X* is above its mean and if Z is negative, then *X* is below its mean (Cryer & Miller, 1994). It can also measure the outliers of a dataset. If the value of a *Z*-score is greater than 3, it indicates that the data distribution has outliers (Tamhane & Dunlop, 2000). It can be defined as follows:

$$z - score = \frac{\left(x - \bar{x} \right)}{\sigma} \tag{1}$$

2) Min-Max normalization

Min-max normalization performs a linear transformation on the original data. Suppose that min_a and max_a are the minimum and the maximum values for attribute A. Min-max normalization maps a value v of A to v' in the range [new-min_a,new-$_{maxa}$] by computing

$$v' = \left(new_max_a - new_min_a \right) * \left(\frac{\left(v - min_a \right)}{\left(max_a - min_a \right)} \right) + new_min_a \tag{2}$$

3. Neural Network Input Parameters

KFLANN is an atomic module, suitable for creating scalable networks which performs similar processing evident within biological neural systems. The original FLANN and KFLANN share the exact same network parameters. Both algorithms require an initialization of two network parameters, namely the Vigilance factor (ρ) and the Tolerance setting (δ). These network parameters are the key to obtain an optimal clustering outcome.

3.1 Vigilance

The Vigilance (ρ) is a parameter that originated from [2]. It was designed as a means to influence the matching degree between the current exemplar and long term memory trace. The higher the ρ value, the stricter the match, while for a smaller ρ value, a more relaxed matching criteria is set .The ρ value in the KFLANN is similar and it is used to determine the number of the attributes in the current exemplar that is similar to the selected output node. For example, if a pattern consists of 12 attributes and a clustering criteria was set such that a similarity of 4 attributes was needed for consideration into the same cluster, then the ρ should be held at 3.The Vigilance formulation is given by equation (3)

$$\rho = \frac{f_{match}}{f_{total}} \tag{3}$$

If the vigilance value is high more number of clusters were formed than when it was set lower.

3.2 Tolerance

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Tolerance parameter provides the localized control, affects individual input features where as the Global variation in the input features is governed by vigilance parameter. Tolerance setting, δ, is the measurement of a particular feature consistency that measures the maximum range that the specified feature is allowed to fluctuate. The tolerance setting can be performed through various methodologies. Three methods are described in the sections that follow. The tolerance setting of the exemplar attributes is the measurement of attributes dispersion, and thus computation is performed for every feature of the training exemplar at the initial stage.

1. The presence of Domain knowledge is perhaps the most helpful form of tolerance setting. With the existence of knowledge from the domain expert, the tolerance of pattern attributes may be determined. This approach is able to produce an acceptable clustering result since the domain knowledge is usually superior. However, domain knowledge is not always available. Two other approaches have been included to harness the clustering capabilities of K-FLANN even in the absence of domain knowledge.

2. In the mathematical context, Standard deviation is the measurement of dispersion
3. of a particular variable from the mean value [1]. Standard deviation method used as the tolerance setting of each feature. The tolerance computation formula is given in equation (4). From the equation, it indicates that it is susceptible to outlier's values, which may increase the δ value. This method is appropriate only if the variation of each feature in exemplar is uniform. There should be a minimal presence of the extreme points lying in the outer regions of the main cluster body. However in most pattern classification problems, the data sets are usually attach with the outliers points.

1.

1.

$$\sigma_d = \partial_d = \sqrt{\frac{\sum_{k=1}^{n}\left(x_{dk} - \mu_d\right)^2}{n}} \qquad (4)$$

Where x_{dk} is the d_{th} feature of the k_{th} pattern and μ_d is the respective mean value of the attribute.

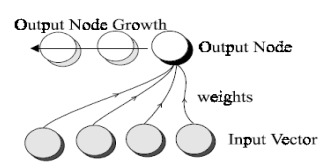
3. This method uses of the equation shown in Eq. (5) (Max-Min method)

$$\frac{\max\ diff\ + \min\ diff}{2} \qquad (5)$$

4. KFLANN Algorithm

4.1 KFLANN Architecture

The basic architecture of KFLANN is delineated in Figure 1. It consists of two layers, similar to the basic Kohonen network model [2] and the ART1 model [4].Connecting the output layer and the input layer is a set of weight vectors. Each output node has weights connected to each element of the input vector. The weight assignment system is unlike that of the Kohonen network or ART1. If a novel pattern is to be stored, it performs a direct copy of the input vector into the weight vector. No other forms of equation are needed for evaluating the weight vectors. The output layer can be viewed as a single dimensional layer which grows dynamically as more novel patterns are encountered. As these novel patterns are encountered, the output system will allocate memory space for the pattern. This implies that the design stages of the network would not require a detailed plan of the output layer, but only a correct configuration of the input layer. The output layer will determine its own size, as the system is trained. Classification results are eventually obtained directly from the winning output node.



4.2 Algorithm

Step 1 Initialize the network parameters.

Step 2 Present the pattern to the input layer. If there is no output node, GOTO step 6

Step 3 Determine all possible matches output node using Eq (6).

$$\frac{\sum_{i=0}^{n}\left[\delta_i^2 - \left(w_{ij} - x_i\right)^2\right]}{n} \geq \rho \qquad (6)$$

Step 4 Determine the winner from all matches output nodes using Eq.(7)

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

$$\text{Winner} = \min\left[\sum_{i=0}^{n}\left(w_{ij} - x_i\right)^2\right] \quad (7)$$

Step 5 Match node is found. Assign the pattern to the match output node. GOTO Step 2

Step 6 Create new output node. Perform direct mapping of the input vector into weight vectors.

Step 7 If complete a single epoch compute clusters centroid. If centroid points of all clusters unchanged Terminate

Else

GO TO Step 2.

Step 8 Find closest pattern to the centroid and re-shuffle it to the top of the dataset list, GOTO Step 2.

Note: ρ is the Vigilance Value, δ_i is the tolerance for i_{th} feature of the input space , W_{ij} used to denote the weight connection of jth output to the ith input, X_i represent the ith feature.

4.3 Tolerance Tuning

Using the domain knowledge it is possible to know number of clusters in the given data set. After completing a single epoch, if desired number of clusters are not formed then KFLANN requires tolerance tuning .It is given as follows:

i. Initialize $\delta_i = (\delta_{imax} + \delta_{imin})/2$ (8)

ii. While number of clusters formed is not appropriate

iii. run KFLANN algorithm without step 7 based on current δ, values

iv. if number of clusters is less than expected

$\delta_i = (\delta_i + \delta_{imin}) / 2$ (9)

v. else if number of clusters is more than expected

$\delta_i = (\delta_i + \delta_{imax}) / 2$ (10)

vi. end while

δ_i - Tolerance value for attribute i. δimin - Minimum difference in attribute values for attribute ***i.*** This is the difference between the smallest and the second smallest values of the attribute. δ_{imax} - **Maximum** difference in attribute values for attribute ***i.*** This is the difference between the smallest and largest values of the attribute.

5. Experimental Results

5.1 Data Sets

5.1.1 Artificial Data Sets Used

The data sets that are used to test the KFLANN algorithm are obtained from the site http://kdd.ics.uci.edu/)

Table 1: Description of Artificial Data Sets

Data Set	# of patterns	# of features	# of clusters
Iris	150	4	3(class 1-50,class 2-50,class 3-50)
Wine	178	13	3(class 1-59,class 2-70,class 3-49)
Glass	214	9	7(class 1-70,class 2-76,class 3-17,class 4-0,class 5-13,class 6-9,class 7-29)
Haberman	306	3	2(class 1-227,class 2-79)
New Thyroid	215	5	3(class 1-150,class 2-35,class 3-30)
Image Segmentation	210	19	7(30 per class)
Pima-Indian diabetes	768	8	2(class 1-500,class 2-268)
Ionosphere	351	34	2(class 1-225,class 2-126)

5.1.2 Synthetic Data Sets Used

Table 2: Description of Synthetic Data Sets

Data Set	# of patterns	#of features	# of clusters
Synthetic Data Set 1(Well separated)	1000	2	2(class 1-500,class 2-500)
Synthetic Data Set 2(Half separated)	1000	2	2(class 1-500,class 2-500)
Synthetic Data Set 3(Not separated)	1000	2	2(class 1-500,class 2-500)
Synthetic Data Set 4	500	8	3(class 1-250,class 2-150,class 3-100)
Synthetic Data Set 5	400	8	3(class 1-150,class 2-150,class 3-100)
Synthetic Data Set 6	350	8	3(class 1-100,class 2-150,class 3-100)

5.2 Setting the Vigilance Parameter (ρ) and Tolerance Parameter (δ) is fixed for each attribute.

The following plots shows the number of clusters formed with the given vigilance value. The groups formed in the data grow with the increase in the vigilance value.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

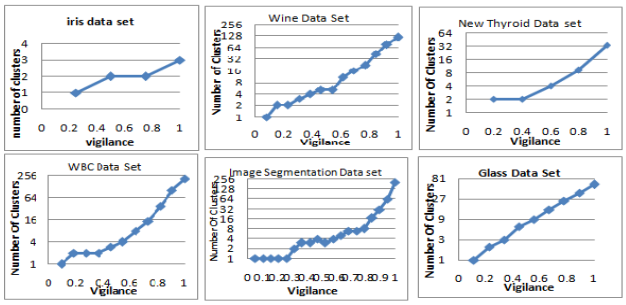


Fig. 1 Plots of Rho Vs Vigilance

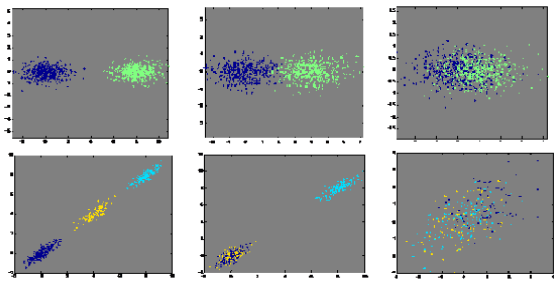


Fig. 2 Plots of Synthetic Data Sets

5.3 Non-normalized data

The Experiments are conducted with non-normalized input data with the given neural network input parameters namely vigilance and tolerance. The following table shows the misclassification rate when Eq. (5) was used for computing tolerance setting for each attribute.

Table 3: Results with non-normalized data and tolerance with Eq.5

data set	Vigilance (ρ)	#of clusters	With Eq. (5) Error Rate (%)
Iris	1	3	7.3333
New Thyroid	1	4(3 is the actual value, (tolerance tuning is required))	20.4651
New Thyroid	0.8	2(3 is the actual value, (tolerance tuning is required))	24.6512
Ionosphere	0.5294	2	29.6296
Pima Indian diabetes	0.6250	2	33.4635
Wine	0.7692	3	35.3933
Glass	0.8889	7	54.6729
Haberman	0.3333	6(2 is the actual value, tolerance tuning is required)	70.2614
Image Segmentation	0.8947,0.9474	4,8(7 is the actual value, tolerance tuning is required)	79.0476, 64.4286

The following table shows the misclassification rate when (standard deviation i.e. Eq. (4) was used for computing tolerance setting for each attribute.

Table 4: Results with non-normalized data and tolerance With Eq.4

data set	Vigilance (ρ)	#of clusters	With Eq. (4) Error Rate (%)
Iris	0.5	3	10
New Thyroid	0.4	2(3 is the actual value, tolerance tuning is required)	30.2326
Ionosphere	0.2647	2	31.3390
Pima Indian diabetes	0.2500	2	34.2448
wine	0.3077	3	38.2022
New Thyroid	0.6	4(3 is the actual value ,tolerance tuning is required)	45.5814
Haberman	0.3333	2	48.6928
Image Segmentation	0.6842,0.7368	7	64.2857,64.7619
Glass	0.4444	7	67.2897

The above results show that the error rate was low when Eq. (5) was used for computing tolerance than Eq. (4) was used .KFLANN performs well only for iris data. For New Thyroid data set and Image Segmentation, actual number of clusters are not formed with the given vigilance value, so tolerance tuning was required. So Normalization was required in order to classify samples with lower misclassification.

5.4 Normalized data

5.4.1 Z-score normalization

Table 5: Results with Z-Score and tolerance with Eq.5

Data Set	Vigilance (ρ)	#of Clusters	With Eq. (5) Error Rate (%)
Iris	0.75	3	5.3333
New Thyroid	0.8	4(actual value is 3)	16.7442
Haberman	0.6667	2	25.4092
wine	0.9231	3	32.5843
Ionosphere	0.8235	2	42.4501
Pima Indian diabetes	0.75	2	48.4375
Glass	0.8889	7	52.8037
Image Segmentation	0.9,1	6 ,10(actual value is 7)	80.25, 82.30

Table 6: Results with Z-Score and tolerance with Eq.4

data set	Vigilance (ρ)	#of clusters	With Eq. (4) Error Rate (%)
Iris	0.5	3	5.3333
New Thyroid	0.2	3	13.4884
Ionosphere	0.3235	2	40.1709

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

wine	0.4615	3	43.2584
Pima Indian diabetes	0.1250	2	48.3073
Haberman	0.3333	2	53.2680
Glass	0.4444	7	56.0748
Image Segmentation	0.6842	6 (actual value is 7,tolerance tuning is required)	87.6190

5.4.2 Max-Min Normalization

Table 7: Results with Max-Min and tolerance with Eq.5

Data Set	Vigilance (ρ)	#of clusters	With Eq.(5)Error Rate (%)
Pima Indian diabetes	0.6250	2	0
Iris	1	3	10
wine	0.7692	3	14.6067
New Thyroid	0.8,1	2,4(actual value is 3,tolerance tuning is required)	21.8605,12.5581
Ionosphere	0.5294	2	29.6296
Glass	0.8889	7	49.5327
Image Segmentation	0.9474	8 (tolerance tuning is required)	59.5238
Haberman	0.6667,1	3,6(actual value is 2,tolerance tuning is required)	60.7843,80.7190

Table 8: Results with Max-Min and tolerance with Eq.4

data set	Vigilance (ρ)	#of clusters	With Eq.(4) Error Rate (%)
Pima Indian diabetes	0.2500	2	0.1302
Iris	0.5	3	12
New Thyroid	0.4,1	2,4(tolerance tuning is required)	30.2326,12.5581
Ionosphere	0.2647	2	31.3390
Haberman	0.3333	2	54.5752
Image Segmentation	0.7368	7	57.1429
Wine	0.3077	3	64.6067
Glass	0.4444	7	65.8879

With the above results, it was observed that accuracy is high with normalized data. KFLANN performs well for Pima Indian Diabetes data set where the error rate is zero. Using Z-score normalization for the Iris data set the error rate was same when either of the method (stated above) was used for computing tolerance. It was found that Z-score Normalization was well suited for Haberman survival data and Max-Min normalization was well suited for Wine, Ionosphere and Glass (tolerance setting with Max-Min) data sets as the error rate was low when compared with Z-Score Normalization.

5.5 Why tolerance tuning?

From the above results, it was found that desired number of clusters is not formed with maximum vigilance value (1) and also with the next maximum value. So there is no chance of getting stated number of clusters because if the vigilance value is reduced number of clusters formed is less than the previous value. Getting number of clusters stated in the domain knowledge is possible only with tolerance tuning when max-min method was used for setting tolerance parameter for the attributes. For some of the artificial data sets tolerance tuning is required as expected number of clusters is not formed. For example consider New Thyroid Data set .The number of clusters are 3.but with the given vigilance values 2 and 4 are formed.

5.5.1 Synthetic Data Sets

On all the Synthetic data sets, KFLANN requires tolerance tuning. The misclassification rate was as follows:

Table 9: Results with Synthetic Data sets

Data Set	Vigilance (ρ)	#of clusters	Eq. (4) or Eq. (5) Error Rate (%)
Synthetic Data Set 1	1	2	0
Synthetic Data Set 2	1	2	1.8
Synthetic Data Set 3	1	2	29.2
Synthetic Data Set 4	1	3	0
Synthetic Data Set 5	1	3	0
Synthetic Data Set 6	1	3	54.2857

6. Conclusions

This paper was intended to observe the behavior of a KFLANN algorithm used for unsupervised learning. The above results show that the performance of the algorithm and the quality of the clustering was based on the input data presented (normalized or non-normalized). The result of the KFLANN was varied between non-normalized and normalized data and was not same for all types of data normalization. For a given input data, one of the normalization method stated above was well suited for a given input because the efficiency of the KFLANN was improved. So normalization is one of the most important factors in clustering. The operation of the KFLANN also depends on above mentioned two parameters namely vigilance and tolerance. It was observed that the misclassification rate is low when Max-Min formula was used for tolerance and also independent of normalization method. For all synthetic data sets the misclassification rate is same with the vigilance setting and with different formulas used for

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

tolerance setting.A comparative analysis showed that the clustering results depended on the normalization method used and the noisiness of the data. In particular, the selection of the vigilance and tolerance setting values for the KFLANN algorithm was sensitive to the normalization method used for datasets with large variations across samples. The fine behavior obtained with neural network for the clustering problems studied here can be used with other difficult optimization problems.

References

[1] Han, J. and M. Kamber, 2001. Data Mining: Concepts and Techniques, Morgan Kaufmann, USA.
[2] Carpenter, G. A. and Gross berg, S. (1987). A Massively Parallel Architecture for a Self-Organizing Neural Pattern Recognition Machine, Computer vision, Graphics, and Image Processing, Vol. 37, pp.54-115.
[3] Adaptive Resonance Theory (ART): An Introduction by L.G. Heins & D.R. Tauritz May/June 1995.
[4] Kohonen, T., "Self-organized formation of topologically correct feature maps", Biological Cybernetics, Vol. 43, 1982, pp. 49-59.
[5] J. Hartigan, *Clustering Algorithms*. New York: Wiley, 1975.
[6] Fukushima K., A hierarchical Neural Network Model for Pattern Recognition. Advanced Neural Computers. Pp 263 - 270. (1990).
[7] Jacek M. Zurada, Introduction to Artificial Neural Systems, West Publishing, (1994).
[8] Grossberg, S (1976). Adaptive pattern classification and universal recoding, Parallel development and coding of neural feature detector, biological Cybernet, Vol. 23, pp.121-134.
[9] Kohonen, T., Self Organization and Associative Memory, **3rd** ed. Berlin: Springer-Verlag, (1989).
[10] Hartigan, J. (1975). Clustering algorithms, New York: John Wiley & Sons.
[11] Chao, P Y, Hwang, Y D. An Improved Neural Network Model for the Prediction of Cutting Tool Life. *Journal of Intelligent Manufacturing*, 8,107-115, 1997.
[12] Hartigan, J. A. and Wong, M. A. (1979). A k-means clustering algorithm. *Applied Statistics 28*, 100-108.
[13] D.J.Evans, L.P. Tay, "Fast Learning Artificial Neural Network for continuous Input Applications", Kybernetes.24 (3), 1995.
[14] Z. S. H. Chan and N. Kasabov, "Fast neural network ensemble learning via negative-correlation data correction," *IEEE Trans. Neural Netw.*,vol. 16, no. 6, pp. 1707–1710, Nov. 2005.
[15] Alex T.L.P., Sandeep Prakash, "K- Means Fast Learning Artificial Neural Network, An Alternate Network For Classification", ICANN, 2002.
[16] L. P. Wong and L. P. Tay, "Centroid stability with K-means fast learning artificial neural networks," in *Proc. Int. Joint Conf. Neural Netw.*, 2003, pp. 1517–1522.
[17] L. P. Wong, L. P. Tay, and J. Xu, "Hierarchical fast learning artificial neural network," in *Proc. Int. Joint Conf. Neural Netw.*, 2005, pp.3300–3305.
[18] L. P. Tay, L. P. Tan, C. A. Bastion, and K. Zhang, "Biologically inspired KFLANN place fields for robot localization," in *Proc. Int. Joint Conf. Neural Netw.*, 2006, pp. 8317–8324.
[19] Merz, C. J., Murphy, P. M. (1998). UCI Repository of Machine learning databases [http://www.ics.uci.edu/~mlearn/MLRepository.html]. Irvine, CA: University of California, Department of Information and Computer Science.
[20] Wong, L. P., Alex, T. L. P. and Xu, J. (2005). The K-means fast learning artificial neural network (KFLANN), an atomic model for hierarchical network architectures, 9th International Conference on Cognitive and Neural System.
[21] Tay, L. P. and Evans, D. J. (1994). Fast Learning Artificial Neural Network Using Nearest Neighbors Recall, Neural Parallel and Scientific Computations, Vol. 2 (1).
[22] Wong L. P., Xu J., Alex T. L. P., Liquid Drop Photonic signal using Fast Learning Artificial Neural Network, ICICS, (2003).
[23] L. P. Tay, J. M. Zurada, L. P. Wong, and J. Xu, "The Hierarchical Fast Learning Artificial Neural Network (HieFLANN) - an autonomous platform for hierarchical neural network construction," IEEE Transactions on Neural Networks, vol. 18, pp. 1645-1657, November 2007.19)
[24] R. C. K. Ho, A. L. P. Tay, and X. Zhang, "Progressive Learning Paradigms using the Parallel K-Iterations Fast Learning Artificial Neural Network," in Proceeding of
[25] International Joint Conference on Neural Networks IJCNN 2007, FL, USA, pp. 552-557, August 2007.
[26] L. P. Tay, L. P. Tan, C. A. Bastion, and K. Zhang, "Biologically Inspired KFLANN Place Fields for Robot Localization," Neural Networks, 2006. IJCNN06. International Joint Conference on, pp. 4201-4208, 2006.

[1]**Suneetha Chittineni**, Associate Professor in the department of Computer Applications, R.V.R.& J.C College of Engineering,Chowdavarm,Guntur. She has 12 years of teaching experience. Currently She is persuing Ph.D from Acharya Nagarjuna University, Guntur. Her research interests include Artificial Intelligence, Machine learning, Pattern Recognition.

[2]**Dr B. Raveendra Babu** obtained his Masters in Computer Science and Engineering from Anna University, Chennai. He received his Ph.D. in Applied Mathematics at S.V University, Tirupati. He is currently leading a Team as Director (Operations), M/s. Delta Technologies (P) Ltd., Madhapur, Hyderabad. He has 26 years of teaching experience. He has more than 25 International & National publications to his credit. His research areas of interest include VLDB, Image Processing, Pattern Analysis and Information Security.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Performance Evaluation of Bluetooth and Zigbee Using Monte Carlo Simulation

U.R.Shoaib , Obaid Rehman ,Zeeshan Akbar and Javed Iqbal

Department of Electrical Engineering, Sarhad University of Science and IT, Peshawar, Pakistan

Abstract

Wireless networking technologies are being implemented in quite a few devices including phones, mobiles, mp3 players etc. The performance of these technologies varies from one to another. The aim of this paper is to analyze the performances of two different low power consumption technologies: Bluetooth and Zigbee. The performance parameters include data coding efficiency and transmission rate. For the analysis we used Monte Carlo simulation and conclusion is drawn on the basis of its results to find the optimum performance among Bluetooth and zigbee. A brief analysis of these graphs is presented in the paper and from this analysis. We propose which of these two technologies Bluetooth or Zigbee gives better performance for data transfer. This analysis will help engineers to select proper technology for their applications.

The results are carried out by using deterministic and stochastic models for evaluating the performance of Bluetooth and Zigbee in terms of transmission time and data coding efficiency. This paper draws a conclusion that Zigbee is superior at the low rate payload size than the misconception of Bluetooth in practical applications for data transfer, whereas Bluetooth is observed to perform efficiently at the higher end payload size. The simulation results verify that the stochastic model can effectively provide real time traffic with ideal performance. The analysis with random noise and jamming can be challenging future work.

Keywords: *Bluetooth, Zigbee, Wireless personal area networks.*

1. Introduction

The evolution of new technologies has brought a drastic change in the field of networking and mobile communication systems. With the introduction of wireless networks many new protocols have come into existence. Bluetooth and Zigbee are the two most prominent wireless technologies because of their low power consumption, easy implementation, low cost, etc[1]. These two technologies operate under an ISM (Industrial, scientific and medical) band in 2.4 GHz range and are mainly used for short range data communications where participating devices may have low battery power [2]. The performance of a network plays an important role in transferring the data. It also helps the users to have secure connections between the participating devices. The performance can be measured by using such parameters as coding efficiency, data transfer rate. The analysis of these performance parameters help to select appropriate technology for the data transfer in wireless personal area networks.

2. Review of State of the art

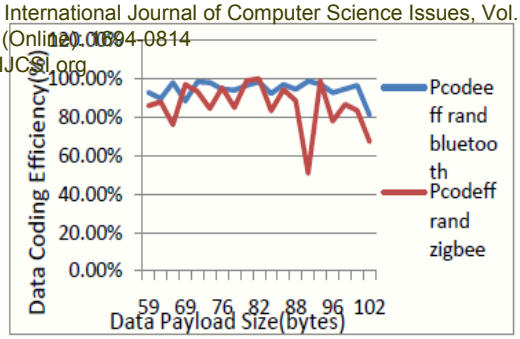
So far the research in analyzing the performance of these two technologies is based on analysis of coding efficiency, power consumption, complexity and data transfer. This is done based on deterministic models [1]. When these two technologies are built using deterministic model, the performance is stable and is also defined by standards [2]. But we feel that randomness is very important in real time performance evaluation and stochastic models should also be applied.

3. Problem Statement and Main Contribution

The performances of Bluetooth and Zigbee have been evaluated previously by using deterministic models. A deterministic model cannot provide the future prediction of performance. For real-time short range data transfer the exact output has to be given, this problem statement is related to our research question: How to evaluate the performance of the two technologies and to judge which one of these technologies is better for data transfer?

The performance of Bluetooth and Zigbee are defined by the data rate parameters like data coding efficiency and transmission rate. We use Monte Carlo simulation for different data sizes and data rate times which involves changing a deterministic model to a stochastic model to

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.



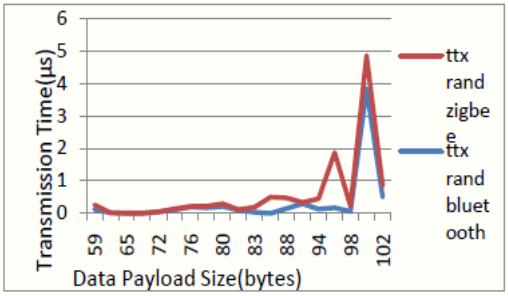
analyze the performance and from the stochastic model we try to deduce which of the two technologies is best suited for short range data transfer.
The main contribution in this paper is that, the performance parameters are given in two different models. The stochastic model is implemented on Excel by giving random inputs in place of expected inputs. The graphical analysis of the stochastic model provides the accurate results of the data rate for both the technologies.

4. Problem Solution

The problem solution involves the application of Monte Carlo simulation for the two data rate parameters we have taken to compare, in which a deterministic model is changed to a stochastic model and the results are analyzed from the stochastic model.

4.1 Application of Deterministic Model

The deterministic system is built by evaluating the transmission time and data coding efficiency. The transmission time is calculated from the data rate, message size and the distance between the nodes. The fo rmula for transmission time can be described as [1].
The data coding efficiency is the ratio of data size and the message size. The formula for calculating data coding efficiency is given as [1].
The performance is calculated using the formulae and the graphs deduced are shown for transmission time and data coding efficiency in “Fig. 1” and “Fig. 2”



respectively.

Fig.1. Comparison of transmission of Bluetooth and Zigbee in the deterministic model

“Fig. 1” indicates that the transmission time of Zigbee increases at a much higher rate than Bluetooth as the data payload size is increased.

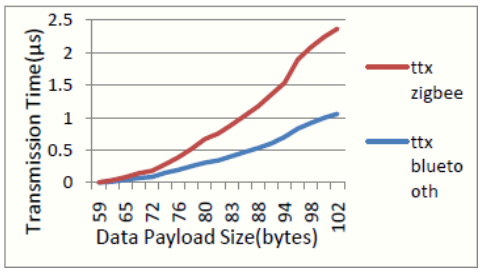


Fig.2. Comparison of data coding efficiency of Bluetooth and Zigbee in the deterministic model.

“Fig. 2” indicates the data coding efficiency of Bluetooth is higher when compared to Zigbee with the increase in data payload size.

4.2 Application of Stochastic Model

The stochastic model is built from the deterministic model by generating different sets of random numbers for each input distribution. The random simulations obtained are taken in a number of 400 and averages of these simulations are analyzed to obtain the output of the stochastic model.
“Fig. 3” and “Fig. 4” show the output for the transmission time and data coding efficiency respectively in the stochastic model.

4.3. Analysis of the Simulation Result

This section includes the analysis of performance by using two parameters transmission time and data coding efficiency. In comparison with the two models, the stochastic model shows increasing fluctuations for Zigbee beyond the payload size of 83 bytes in the transmission time when compared to Bluetooth. The data coding efficiency in Zigbee shows a consistency up to 83 bytes of data payload and decreased later.

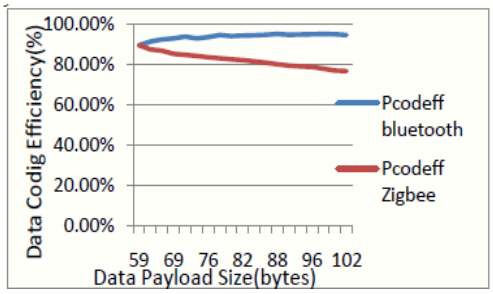


Fig.3. Comparison of transmission time of Bluetooth and Zigbee in the stochastic model.

Fig.4. Comparison of data coding efficiency of Bluetooth and Zigbee in the deterministic model.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

5. Conclusion

This paper presents the results of using deterministic and stochastic models for evaluating the performance of Bluetooth and Zigbee in terms of transmission time and data coding efficiency. This paper draws a conclusion that Zigbee is superior at the low rate payload size than the misconception of Bluetooth in practical applications for data transfer, whereas Bluetooth is observed to perform efficiently at the higher end payload size. The simulation results verify that the stochastic model can effectively provide real time traffic with ideal performance. The analysis with random noise and jamming can be challenging future works.

References

[1] Jin-shyan Lee, Yu-Wei su, Chung-Chou Shen,"A Comparative Study of Wireless Protcols: Bluetooth, UWB, Zigbee and Wi-Fi", IEEE IECON 2007, 5-8 nov 20 07,pp.46-51.
[2] Ferro, E.; Potorti, F, "Bluetooth and Wi-Fi Wireless Protocols: A Survey and a Comparison", *IEEE Wireless Communications*,Volume 12, Issue 1, Feb. 2005, pp.12 – 26.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Cloud Load Balancing Techniques : A Step Towards Green Computing

Nidhi Jain Kansal[1], Inderveer Chana[2]

[1] **Computer Science and Engineering Department, Thapar University**
Patiala-147004, Punjab, India

[2] **Computer Science and Engineering Department, Thapar University**
Patiala-147004, Punjab, India

Abstract

Cloud computing is emerging as a new paradigm of large-scale distributed computing. It is a framework for enabling convenient, on-demand network access to a shared pool of computing resources. Load balancing is one of the main challenges in cloud computing which is required to distribute the dynamic workload across multiple nodes to ensure that no single node is overwhelmed. It helps in optimal utilization of resources and hence in enhancing the performance of the system. The goal of load balancing is to minimize the resource consumption which will further reduce energy consumption and carbon emission rate that is the dire need of cloud computing. This determines the need of new metrics, energy consumption and carbon emission for energy-efficient load balancing in cloud computing. This paper discusses the existing load balancing techniques in cloud computing and further compares them based on various parameters like performance, scalability, associated overhead etc. that are considered in different techniques. It further discusses these techniques from energy consumption and carbon emission perspective.

Keywords: *Cloud computing, Load balancing, Energy efficiency, Green computing.*

1. Introduction

Cloud computing is emerging as a new paradigm of large-scale distributed computing. It has moved computing and data away from desktop and portable PCs, into large data centers [1]. It has the capability to harness the power of Internet and wide area network (WAN) to use resources that are available remotely, thereby providing cost-effective solution to most of the real life requirements [26]. It provides the scalable IT resources such as applications and services, as well as the infrastructure on which they operate, over the Internet, on pay-per-use basis to adjust the capacity quickly and easily. It helps to accommodate changes in demand and helps any organization in avoiding the capital costs of software and hardware [2] [3]. Thus, cloud computing is a framework for enabling a suitable, on-demand network access to a shared pool of computing resources (e.g. networks, servers, storage, applications, and services). These resources can be provisioned and de-provisioned quickly with minimal management effort or service provider interaction. This further helps in promoting availability [4]. Due to the exponential growth of cloud computing, it has been widely adopted by the industry and there is a rapid expansion in data-centers. This expansion has caused the dramatic increase in energy use and its impact on the environment in terms of carbon footprints. The link between energy consumption and carbon emission has given rise to an energy management issue which is to improve energy-efficiency [19] in cloud computing to achieve Green computing [6]. Besides this, there are many other existing issues like Load Balancing, Virtual Machine Migration, Server Consolidation etc. that have not been fully addressed [5] [19]. Virtual Machine Migration enabled by virtualization can help in balancing load, enabling highly responsive provisioning and avoiding hot-spots in data centers thereby reducing power consumption. Server Consolidation helps in improving resource utilization by consolidating various VMs residing on multiple under-utilized servers onto a single server, thereby turning off unused servers, hence reducing energy consumption in a cloud computing environment [19]. Load balancing can help in reducing energy consumption by evenly distributing the load and minimizing the resource consumption. This paper focuses on the prevalent load balancing techniques in cloud computing environment.

Load balancing is one of the central issues in cloud computing [5]. It is a mechanism that distributes the dynamic local workload evenly across all the nodes in the whole cloud to avoid a situation where some nodes are heavily loaded while others are idle or doing little work. It helps to achieve a high user satisfaction and resource utilization ratio, hence improving the overall performance

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

and resource utility of the system. It also ensures that every computing resource is distributed efficiently and fairly [13]. It further prevents bottlenecks of the system which may occur due to load imbalance. When one or more components of any service fail, load balancing helps in continuation of the service by implementing fair-over, i.e. in provisioning and de-provisioning of instances of applications without fail. It also ensures that every computing resource is distributed efficiently and fairly [5] [9]. Consumption of resources and conservation of energy are not always a prime focus of discussion in cloud computing. However, resource consumption can be kept to a minimum with proper load balancing which not only helps in reducing costs but making enterprises greener [6] [13]. Scalability which is one of the very important features of cloud computing is also enabled by load balancing. Hence, improving resource utility and the performance of a distributed system in such a way will reduce the energy consumption and carbon footprints to achieve Green computing [6] [10] [12].

The motivation of the survey of existing load balancing techniques in cloud computing is to encourage the amateur researcher to contribute in developing more efficient load balancing algorithms. This will benefit interested researchers to carry out further work in this research area. This paper is organized as follows: Section II discusses Green computing in clouds, Section III presents the need of load balancing in clouds, Section IV shows the study and analysis of the existing load balancing techniques in cloud computing, Section V identifies the metrics considered in the existing load balancing techniques and carries out the comparison between them based on those identified metrics and Section VI concludes the paper. To the best of our knowledge, none of the techniques has focused on energy consumption and carbon emission factors that is a dire need of cloud computing.

2. Green Computing in Clouds

Green Computing [6], or Green IT, is the practice of implementing policies and procedures that improve the efficiency of computing resources in such a way as to reduce the energy consumption and environmental impact of their utilization [10] [12].

As High Performance Computing (HPC) is becoming popular in commercial and consumer IT applications, it needs the ability to gain rapid and scalable access to high-end computing capabilities. This computing infrastructure is provided by cloud computing by making use of data-centers. It helps the HPC users in an on-demand and payable access to their applications and data, anywhere from a cloud [7]. Cloud computing data-centers have been enabled by high-speed computer networks that allow applications to run more efficiently on these remote, broadband computer networks, compared to local personal computers. These data-centers cost less for application hosting and operation than individual application software licenses running on clusters of on-site computer clusters [11]. However, the explosion of cloud computing networks and the growing demand drastically increases the energy consumption of data-centers, which has become a critical issue and a major concern for both industry and society [8]. This increase in energy consumption not only increases energy cost but also increases carbon-emission. High energy cost results in reducing cloud providers' profit margin and high carbon emission is not good for the environment [7]. Hence, energy-efficient solutions that can address the high energy consumption, both from the perspective of the cloud provider and the environment are required. This is a dire need of cloud computing to achieve Green computing. This whole scenario is depicted in Fig. 1. Load balancing can be one such energy-saving solution in cloud computing environment.

3. Need of Load Balancing in Cloud Computing

Load balancing in clouds is a mechanism that distributes the excess dynamic local workload evenly across all the nodes. It is used to achieve a high user satisfaction and resource utilization ratio [15], making sure that no single node is overwhelmed, hence improving the overall performance of the system. Proper load balancing can help in utilizing the available resources optimally, thereby minimizing the resource consumption. It also helps in implementing fail-over, enabling scalability, avoiding bottlenecks and over-provisioning, reducing response time etc.

Apart from the above-mentioned factors, load balancing is also required to achieve Green computing in clouds which can be done with the help of the following two factors:

• *Reducing Energy Consumption* - Load balancing helps in avoiding overheating by balancing the workload across all the nodes of a cloud, hence reducing the amount of energy consumed.

• *Reducing Carbon Emission* - Energy consumption and carbon emission go hand in hand. The more the energy consumed, higher is the carbon footprint. As the energy consumption is reduced with the help of Load balancing, so is the carbon emission helping in achieving Green computing.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

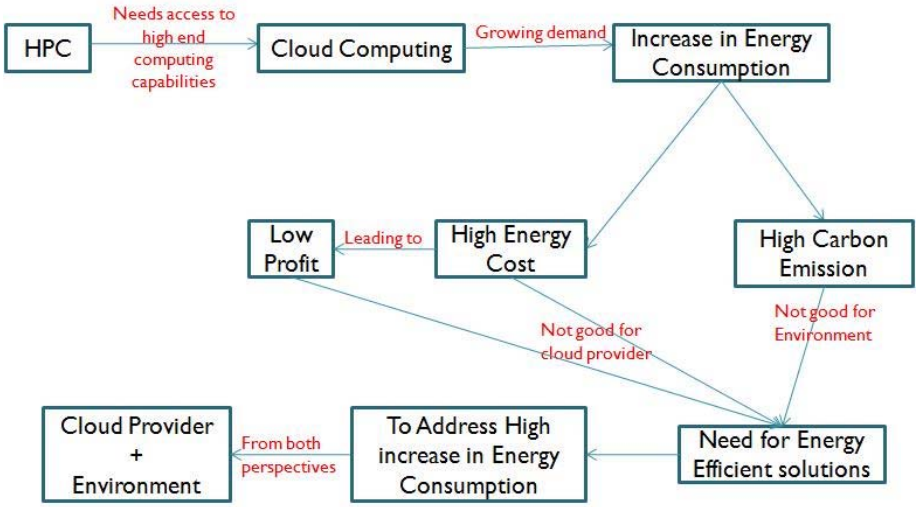


Fig. 1: Green Computing in Clouds

4. Existing Load Balancing Techniques in Cloud Computing

Following load balancing techniques are currently prevalent in clouds:-

• *Decentralized content aware load balancing* - H. Mehta et al. [21] proposed a new content aware load balancing policy named as workload and client aware policy (WCAP). It uses a unique and special property (USP) to specify the unique and special property of the requests as well as computing nodes. USP helps the scheduler to decide the best suitable node for the processing the requests. This strategy is implemented in a decentralized manner with low overhead. By using the content information to narrow down the search, this technique improves the searching performance and hence overall performance of the system. It also helps in reducing the idle time of the computing nodes hence improving their utilization.

• *Server-based load balancing for Internet distributed services* - A. M. Nakai et al. [22] proposed a new server-based load balancing policy for web servers which are distributed all over the world. It helps in reducing the service response times by using a protocol that limits the redirection of requests to the closest remote servers without overloading them. A middleware is described to implement this protocol. It also uses a heuristic to help web servers to endure overloads.

• *Join-Idle-Queue* - Y. Lua et al. [23] proposed a Join-Idle-Queue load balancing algorithm for dynamically scalable web services. This algorithm provides large-scale load balancing with distributed dispatchers by, first load balancing idle processors across dispatchers for the availability of idle processors at each dispatcher and then, assigning jobs to processors to reduce average queue length at each processor. By removing the load balancing work from the critical path of request processing, it effectively reduces the system load, incurs no communication overhead at job arrivals and does not increase actual response time.

• *A Lock-free multiprocessing solution for LB* - X. Liu et al. [24] proposed a lock-free multiprocessing load balancing solution that avoids the use of shared memory in contrast to other multiprocessing load balancing solutions which use shared memory and lock to maintain a user session. It is achieved by modifying Linux kernel. This solution helps in improving the overall performance of load balancer in a multi-core environment by running multiple load-balancing processes in one load balancer.

• *Scheduling strategy on load balancing of virtual machine resources* - J. Hu et al. [25] proposed a scheduling strategy on load balancing of VM resources that uses historical data and current state of the system. This strategy achieves the best load balancing and reduced dynamic migration by using a genetic algorithm. It helps in resolving the issue of load-imbalance and high cost of migration thus achieving better resource utilization.

• *Central load balancing policy for virtual machines* - A. Bhadani et al. [26] proposed a Central Load Balancing Policy for Virtual Machines (CLBVM) that balances the load evenly in a distributed virtual machine/cloud

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

computing environment. This policy improves the overall performance of the system but does not consider the
systems that are fault-tolerant.

• *LBVS: Load Balancing strategy for Virtual Storage* - H. Liu et al. [27] proposed a load balancing virtual storage strategy (LBVS) that provides a large scale net data storage model and Storage as a Service model based on Cloud Storage. Storage virtualization is achieved using an architecture that is three-layered and load balancing is achieved using two load balancing modules. It helps in improving the efficiency of concurrent access by using replica balancing further reducing the response time and enhancing the capacity of disaster recovery. This strategy also helps in improving the use rate of storage resource, flexibility and robustness of the system.

• *A Task Scheduling Algorithm Based on Load Balancing* - Y. Fang et al. [28] discussed a two-level task scheduling mechanism based on load balancing to meet dynamic requirements of users and obtain a high resource utilization. It achieves load balancing by first mapping tasks to virtual machines and then virtual machines to host resources thereby improving the task response time, resource utilization and overall performance of the cloud computing environment.

• *Honeybee Foraging Behavior* - M. Randles et al. [14] investigated a decentralized honeybee-based load balancing technique that is a nature-inspired algorithm for self-organization. It achieves global load balancing through local server actions. Performance of the system is enhanced with increased system diversity but throughput is not increased with an increase in system size. It is best suited for the conditions where the diverse population of service types is required.

• *Biased Random Sampling* - M. Randles et al. [14] investigated a distributed and scalable load balancing approach that uses random sampling of the system domain to achieve self-organization thus balancing the load across all nodes of the system. The performance of the system is improved with high and similar population of resources thus resulting in an increased throughput by effectively utilizing the increased system resources. It is degraded with an increase in population diversity.

• *Active Clustering* - M. Randles et al. [14] investigated a self-aggregation load balancing technique that is a self-aggregation algorithm to optimize job assignments by connecting similar services using local re-wiring. The performance of the system is enhanced with high resources thereby increasing the throughput by using these resources effectively. It is degraded with an increase in system diversity.

• *ACCLB (Load Balancing mechanism based on ant colony and complex network theory)* - Z. Zhang et al. [15] proposed a load balancing mechanism based on ant colony and complex network theory in an open cloud computing federation. It uses small-world and scale-free characteristics of a complex network to achieve better load balancing. This technique overcomes heterogeneity, is adaptive to dynamic environments, is excellent in fault tolerance and has good scalability hence helps in improving the performance of the system.

• *Two-phase load balancing algorithm (OLB + LBMM)* - S.-C. Wang et al. [18] proposed a two- phase scheduling algorithm that combines OLB (Opportunistic Load Balancing) and LBMM (Load Balance Min-Min) scheduling algorithms to utilize better executing efficiency and maintain the load balancing of the system. OLB scheduling algorithm, keeps every node in working state to achieve the goal of load balance and LBMM scheduling algorithm is utilized to minimize the execution time of each task on the node thereby minimizing the overall completion time. This combined approach hence helps in an efficient utilization of resources and enhances the work efficiency.

• *Event-driven* - V. Nae et al. [20] presented an event-driven load balancing algorithm for real-time Massively Multiplayer Online Games (MMOG). This algorithm after receiving capacity events as input, analyzes its components in context of the resources and the global state of the game session, thereby generating the game session load balancing actions. It is capable of scaling up and down a game session on multiple resources according to the variable user load but has occasional QoS breaches.

• *CARTON* - R. Stanojevic et al. [17] proposed a mechanism for cloud control named as CARTON that unifies the use of LB and DRL. LB (Load Balancing) is used to equally distribute the jobs to different servers so that the associated costs can be minimized and DRL (Distributed Rate Limiting) is used to make sure that the resources are distributed in a way to keep a fair resource allocation. DRL also adapts to server capacities for the dynamic workloads so that performance levels at all servers are equal. With very low computation and communication overhead, this algorithm is simple and easy to implement.

• *COMPARE AND BALANCE* - Y. Zhao et al. [16] addressed the problem of intra-cloud load balancing

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

amongst physical hosts by adaptive live migration of virtual machines. A load balancing model is designed and implemented to reduce virtual machines' migration time

Table 1: Existing Load Balancing techniques

Techniques	***Environment***	***Description***	***Findings***
Decentralized content aware [21]	Distributed computing	1. Uses a unique and special property(USP) of requests and computing nodes to help scheduler to decide the best node for processing the requests 2. Uses the content information to narrow down the search	1. Improves the searching performance hence increasing overall performance 2. Reduces idle time of the nodes
LB for Internet distributed services [22]	Distributed web servers	1. Uses a protocol to limit redirection rates to avoid remote servers overloading 2. Uses a middleware to support this protocol 3. Uses a heuristic to tolerate abrupt load changes	1. Reduces service response times by redirecting requests to the closest servers without overloading them 2. Mean response time is 29% smaller than RR(Round Robin) and 31% smaller than SL(Smallest Latency)
Join-Idle-Queue [23]	Cloud data centers	1. First assigns idle processors to dispatchers for the availability of the idle processors at each dispatcher 2. Then assigns jobs to processors to reduce average queue length of jobs at each processor	1. Effectively reduces the system load 2. Incurs no communication overhead at job arrivals 3. Does not increase actual response times
Lock-free multiprocessing[24]	Multi-core	1. Runs multiple load-balancing processes in one load balancer	1. Improves overall performance of load balancer
Scheduling strategy on LB of VM resources [25]	Cloud Computing	1. Uses Genetic algorithm, historical data and current state of system to achieve best load balancing and to reduce dynamic migration	1. Solves the problems of load imbalance and high migration cost
Central LB policy for VMs [26]	Cloud Computing	1. Uses global state information to make load balancing decisions	1. Balances the load evenly to improve overall performance 2. Up to 20% improvement in performance 3. Does not consider fault tolerance
LBVS: LB strategy for Virtual Storage [27]	Cloud Storage	1. Uses Fair-Share Replication strategy to achieve Replica Load balancing module which in turn controls the access load balancing 2. Uses writing balancing algorithm to control data writing load balancing	1. Enhances flexibility and robustness 2. Provides large scale net data storage and storage as a service
Task Scheduling Based on LB [28]	Cloud Computing	1. First maps tasks to virtual machines and then virtual machines to host resources	1. Improves task response time 2. Improves resource utilization
Honeybee Foraging Behavior [14]	Large scale Cloud Systems	1. Achieves global load balancing through local serve actions	1. Performs well as system diversity increases 2. Does not increase throughput as system size increases
Biased Random Sampling [14]	Large scale Cloud Systems	1. Achieves load balancing across all system nodes using random sampling of the system domain	1. Performs better with high and similar population of resources 2. Degrades as population diversity increases
Active Clustering [14]	Large scale Cloud systems	1. Optimizes job assignment by connecting similar services by local re-wiring	1. Performs better with high resources 2. Utilizes the increased system resources to increase throughput 3. Degrades as system diversity increases
ACCLB (Ant Colony and Complex Network Theory) [15]	Open Cloud Computing Federation	1. Uses small-world and scale-free characteristics of complex network to achieve better load balancing	1. Overcomes heterogeneity 2. Adaptive to dynamic environments 3. Excellent in fault tolerance 4. Good scalability
Two-phase scheduling(OLB + LBMM) [18]	Three-level Cloud Computing Network	1. Uses OLB (Opportunistic Load Balancing) to keep each node busy and uses LBMM(Load Balance Min-Min) to achieve the minimum execution time of each task	1. Efficient utilization of resources 2. Enhances work efficiency

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Event-driven [20]	Massively Multiplayer Online Games	1. Uses complete capacity event as input, analyzes its components and generates the game session load balancing actions	1. Capable of scaling up and down a game session on multiple resources according to the variable user load 2. Occasional QoS breaches as low as 0.66%
Carton(LB + DRL) [17]	Unifying framework for cloud control	1. Uses Load Balancing to minimize the associated cost and uses Distributed Rate Limiting for fair allocation of resources	1. Simple 2. Easy to implement 3.Very low computation and communication overhead
Compare and Balance [16]	Intra-Cloud	1. Based on sampling 2. Uses adaptive live migration of virtual machines	1. Balances load amongst servers 2. Reaches equilibrium fast 3. Assures migration of VMs from high-cost physical hosts to low-cost host 4. Assumption of having enough memory with each physical host
VectorDot [29]	Datacenters with integrated server and storage virtualization	1. Uses dot product to distinguish node based on the item requirement	1. Handles hierarchical and multidimensional resource constraints 2. Removes overloads on server, switch and storage

by shared storage, to balance load amongst servers according to their processor or IO usage, etc. and to keep virtual machines' zero-downtime in the process. A distributed load balancing algorithm COMPARE AND BALANCE is also proposed that is based on sampling and reaches an equilibrium very fast. This algorithm assures that the migration of VMs is always from high-cost physical hosts to low-cost host but assumes that each physical host has enough memory which is a weak assumption.

• *VectorDot* - A. Singh et al. [29] proposed a novel load balancing algorithm called VectorDot. It handles the hierarchical complexity of the data-center and multi-dimensionality of resource loads across servers, network switches, and storage in an agile data center that has integrated server and storage virtualization technologies. VectorDot uses dot product to distinguish nodes based on the item requirements and helps in removing overloads on servers, switches and storage nodes.

These existing techniques have been summarized in Table 1.

5. Metrics for Load Balancing in Clouds

The existing load balancing techniques in clouds, consider various parameters like performance, response time, scalability, throughput, resource utilization, fault tolerance, migration time and associated overhead. But, for an energy-efficient load balancing, metrics like energy consumption and carbon emission should also be considered.

• *Overhead Associated* - determines the amount of overhead involved while implementing a load-balancing algorithm. It is composed of overhead due to movement of tasks, inter-processor and inter-process communication. This should be minimized so that a load balancing technique can work efficiently.

• *Throughput* - is used to calculate the no. of tasks whose execution has been completed. It should be high to improve the performance of the system.

• *Performance* – is used to check the efficiency of the system. It has to be improved at a reasonable cost e.g. reduce response time while keeping acceptable delays.
• *Resource Utilization* - is used to check the utilization of resources. It should be optimized for an efficient load balancing.

• *Scalability* - is the ability of an algorithm to perform load balancing for a system with any finite number of nodes. This metric should be improved.

• *Response Time* - is the amount of time taken to respond by a particular load balancing algorithm in a distributed system. This parameter should be minimized.

• *Fault Tolerance* - is the ability of an algorithm to perform uniform load balancing in spite of arbitrary node or link failure. The load balancing should be a good fault-tolerant technique.

• *Migration time* - is the time to migrate the jobs or resources from one node to other. It should be minimized in order to enhance the performance of the system.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

• *Energy Consumption (EC)* - determines the energy consumption of all the resources in the system. Load balancing helps in avoiding overheating by balancing the workload across all the nodes of a Cloud, hence reducing energy consumption.

Table 2: Metrics considered by existing Load Balancing techniques

Techniq--ues	*Perform--ance*	*Response Time*	*Scalability*	*Over-head*	*Through-put*	*Resource utilization*	*Fault tolerance*	*Migration time*	*EC*	*CE*
T1 [21]	✓	✓	✓	✓	×	✓	×	×	×	×
T2 [22]	✓	✓	×	×	×	×	×	×	×	×
T3 [23]	✓	✓	×	✓	×	×	×	×	×	×
T4 [24]	✓	×	×	×	✓	×	×	×	×	×
T5 [25]	×	×	×	✓	×	✓	×	×	×	×
T6 [26]	✓	✓	×	×	✓	✓	×	×	×	×
T7 [27]	✓	✓	✓	×	×	×	✓	×	×	×
T8 [28]	✓	✓	×	×	×	✓	×	×	×	×
T9 [14]	✓	×	✓	×	✓	×	×	×	×	×
T10 [14]	✓	×	✓	×	✓	×	×	×	×	×
T11 [14]	✓	×	✓	×	✓	×	×	×	×	×
T12 [15]	✓	×	✓	×	×	✓	✓	×	×	×
T13 [18]	✓	×	×	×	×	✓	×	×	×	×
T14 [20]	×	×	✓	×	×	✓	×	×	×	×
T15 [17]	✓	×	×	✓	×	✓	×	×	×	×
T16 [16]	×	×	×	✓	×	✓	×	✓	×	×
T17 [29]	×	×	×	×	×	✓	×	×	×	×

• *Carbon Emission (CE)* - calculates the carbon emission of all the resources in the system. As energy consumption and carbon emission go hand in hand, the more the energy consumed, higher is the carbon footprint. So, for an energy-efficient load balancing solution, it should be reduced.

Based on the above metrics, the existing load balancing techniques have been compared in Table 2.

6. Conclusions

Cloud Computing has widely been adopted by the industry, though there are many existing issues like Load Balancing, Virtual Machine Migration, Server Consolidation, Energy Management, etc. which have not been fully addressed. Central to these issues is the issue of load balancing, that is required to distribute the excess dynamic local workload evenly to all the nodes in the whole Cloud to achieve a high user satisfaction and resource utilization ratio. It also ensures that every computing resource is distributed efficiently and fairly. Existing Load Balancing techniques that have been studied, mainly focus on reducing overhead, service response time and improving performance etc., but none of the techniques has considered the energy consumption and carbon emission factors. Therefore, there is a need to develop an Energy-efficient load balancing technique that can improve the performance of cloud computing along with maximum resource utilization, in turn reducing energy consumption as well as carbon emission to an extent that will help achieve Green Computing.

References

[1] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, “Above the Clouds: A Berkeley View of Cloud Computing”, EECS Department, University of California, Berkeley, Technical Report No., UCB/EECS-2009-28, pages 1-23, February 2009.

[2] R. W. Lucky, “Cloud computing”, IEEE Journal of Spectrum, Vol. 46, No. 5, May 2009, pages 27-45.

[3] M. D. Dikaiakos, G. Pallis, D. Katsa, P. Mehra, and A. Vakali, “Cloud Computing: Distributed Internet Computing for IT and Scientific Research”, IEEE Journal of Internet Computing, Vol. 13, No. 5, September/October 2009, pages 10-13.

[4] G. Pallis, “Cloud Computing: The New Frontier of Internet Computing”, IEEE Journal of Internet Computing, Vol. 14, No. 5, September/October 2010, pages 70-73.

[5] B. P. Rima, E. Choi, and I. Lumb, “A Taxonomy and Survey of Cloud Computing Systems”, Proceedings of 5th IEEE International Joint Conference on INC, IMS and IDC, Seoul, Korea, August 2009, pages 44-51.

[6] R. Mata-Toledo, and P. Gupta, “Green data center: how green can we perform”, Journal of Technology Research, Academic and Business Research Institute, Vol. 2, No. 1, May 2010, pages 1-8.

[7] S. K. Garg, C. S. Yeob, A. Anandasivamc, and R. Buyya, “Environment-conscious scheduling of HPC applications on

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

distributed Cloud-oriented data centers", Journal of Parallel and Distributed Computing, Elsevier, Vol. 70, No. 6, May 2010, pages 1-18.

[8] K. M. Nagothu, B. Kelley, J. Prevost, and M. Jamshidi, "Ultra low energy cloud computing using adaptive load prediction", Proceedings of IEEE World Automation Congress(WAC) , Kobe, September 2010, pages 1-7.

[9] B. P. Rimal, E. Choi, and I. Lumb, "A Taxonomy, Survey, and Issues of Cloud Computing Ecosystems, Cloud Computing: Principles, Systems and Applications", Computer Communications and Networks, Chapter 2 , pages 21-46, DOI 10.1007/978-1-84996-241-42, Springer – V erlagLondonLimited, 2010.

[10] S. Kabiraj, V. Topka, and R. C. Walke, "Going Green: A Holistic Approach to Transform Business", International Journal of Managing Information Technology (IJMIT), Vol. 2, No. 3, August 2010, pages 22-31.

[11] K. T. Rao, P. S. Kiran, and L. S. S. Reddy, "Energy Efficiency in Datacenters through Virtualization: A Case Study", Global Journal of Computer Science and Technology, Vol. 10, No. 3, April 2010, pages 2-6.

[12] J. Baliga, R. W. A. Ayre, K. Hinton, and R. S. Tucker, "Green Cloud Computing: Balancing Energy in Processing, Storage, and Transport", Proceedings of the IEEE, Vol. 99, No. 1, January 2011, pages 149-167.

[13] A. M. Alakeel, "A Guide to dynamic Load balancing in Distributed Computer Systems", International Journal of Computer Science and Network Security (IJCSNS), Vol. 10, No. 6, June 2010, pages 153-160.

[14] M. Randles, D. Lamb, and A. Taleb-Bendiab, "A Comparative Study into Distributed Load Balancing Algorithms for Cloud Computing", Proceedings of 24th IEEE International Conference on Advanced Information Networking and Applications Workshops, Perth, Australia, April 2010, pages 551-556.

[15] Z. Zhang, and X. Zhang, "A Load Balancing Mechanism Based on Ant Colony and Complex Network Theory in Open Cloud Computing Federation", Proceedings of 2nd International Conference on Industrial Mechatronics and Automation (ICIMA), Wuhan, China, May 2010, pages 240-243.

[16] Y. Zhao, and W. Huang, "Adaptive Distributed Load Balancing Algorithm based on Live Migration of Virtual Machines in Cloud", Proceedings of 5th IEEE International Joint Conference on INC, IMS and IDC, Seoul, Republic of Korea, August 2009, pages 170-175.

[17] R. Stanojevic, and R. Shorten, "Load balancing vs. distributed rate limiting: a unifying framework for cloud control", Proceedings of IEEE ICC, Dresden, Germany, August 2009, pages 1-6,.

[18] S. Wang, K. Yan, W. Liao, and S. Wang, "Towards a Load Balancing in a Three-level Cloud Computing Network", Proceedings of the 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT), Chengdu, China, September 2010, pages 108-113.

[19] Q. Zhang, L. Cheng, and R. Boutaba, "Cloud computing: state-of-the-art and research challenges", Journal of Internet Services and Applications, Vol. 1, No. 1, April 2010, pages 7-18.

[20] V. Nae, R. Prodan, and T. Fahringer, "Cost-Efficient Hosting and Load Balancing of Massively Multiplayer Online Games", Proceedings of the 11th IEEE/ACM International Conference on Grid Computing (Grid), IEEE Computer Society, October 2010, pages 9-17.

[21] H. Mehta, P. Kanungo, and M. Chandwani, "Decentralized content aware load balancing algorithm for distributed computing environments", Proceedings of the International Conference Workshop on Emerging Trends in Technology (ICWET), February 2011, pages 370-375.

[22] A. M. Nakai, E. Madeira, and L. E. Buzato, "Load Balancing for Internet Distributed Services Using Limited Redirection Rates", 5th IEEE Latin-American Symposium on Dependable Computing (LADC), 2011, pages 156-165.

[23] Y. Lua, Q. Xiea, G. Kliotb, A. Gellerb, J. R. Larusb, and A. Greenber, "Join-Idle-Queue: A novel load balancing algorithm for dynamically scalable web services", An international Journal on Performance evaluation, In Press, Accepted Manuscript, Available online 3 August 2011.

[24] Xi. Liu, Lei. Pan, Chong-Jun. Wang, and Jun-Yuan. Xie, "A Lock-Free Solution for Load Balancing in Multi-Core Environment", 3rd IEEE International Workshop on Intelligent Systems and Applications (ISA), 2011, pages 1-4.

[25] J. Hu, J. Gu, G. Sun, and T. Zhao, "A Scheduling Strategy on Load Balancing of Virtual Machine Resources in Cloud Computing Environment", Third International Symposium on Parallel Architectures, Algorithms and Programming (PAAP), 2010, pages 89-96.

[26] A. Bhadani, and S. Chaudhary, "Performance evaluation of web servers using central load balancing policy over virtual machines on cloud", Proceedings of the Third Annual ACM Bangalore Conference (COMPUTE), January 2010.

[27] H. Liu, S. Liu, X. Meng, C. Yang, and Y. Zhang, "LBVS: A Load Balancing Strategy for Virtual Storage", International Conference on Service Sciences (ICSS), IEEE, 2010, pages 257-262.

[28] Y. Fang, F. Wang, and J. Ge, "A Task Scheduling Algorithm Based on Load Balancing in Cloud Computing", Web Information Systems and Mining, Lecture Notes in Computer Science, Vol. 6318, 2010, pages 271-277.

[29] A. Singh, M. Korupolu, and D. Mohapatra, "Server-storage virtualization: integration and load balancing in data centers", Proceedings of the ACM/IEEE conference on Supercomputing (SC), November 2008.

Nidhi Jain Kansal received her Bachelor's Degree (Electronics) from Kurukshetra University, Haryana, India in 1997, and completed her MCA in 2000 from Guru Jambheshwar University, Hisar, Haryana, India. At present, she is a Ph.D. Candidate in Computer Science & Engineering Department at Thapar University. Her research interests lie in Cloud Computing, Green Computing, Load balancing and Resource utilization techniques.

Dr. Inderveer Chana completed her B.Tech in Computer Science (1997) and M.E. in Software Engineering from TIET (2002) and Ph.D. in Resource Management in Grid Computing from Thapar University, Patiala (2009) and has over thirteen years of teaching and research experience. She is working as Associate Professor in Computer Science and Engineering Department of Thapar University, Patiala. Her research interests include Grid computing, Cloud Computing and resource

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

management challenges in Grids and Clouds. She has over 30 publications in International Journals and Conferences of repute. More than 20 Masters have been completed so far under her supervision and is currently supervising 9 Doctoral candidates in the area of Grid and Cloud Computing.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Wind Turbine Micrositing: Comparison of Finite Difference Method and Computational Fluid Dynamics

Samina Rajper[1] and Imran J. Amin[2]

[1] Department of Computing, SZABIST
Karachi, Sindh, Pakistan

[2] Department of Computing, SZABIST
Karachi, Sindh, Pakistan

Abstract

For smooth and optimal operation of wind turbines the location of wind turbines in wind farm is critical. Parameters that need to be considered for micrositing of wind turbines are topographic effect and wind effect. The location under consideration for wind farm is Gharo, Sindh, Pakistan. Several techniques are being researched for finding the most optimal location for wind turbines. These techniques are based on linear and nonlinear mathematical models. In this paper wind pressure distribution and its effect on wind turbine on the wind farm are considered. This study is conducted to compare the mathematical model; Finite Difference Method with a computational fluid dynamics software results. Finally the results of two techniques are compared for micrositing of wind turbines and found that finite difference method is not applicable for wind turbine micrositing.

Keywords: *Wind farm, Finite Difference Method, Wind turbine micrositing, wind pressure, computational fluid dynamics.*

1. Introduction

Arrhenius gave an idea about a century ago that emission of carbon dioxide from the fossil fuels will be the reason of Earth warming [1]. At present, the globe is looking for the alternate solutions instead of expensive and polluted fossil fuel or any other medium that discharge CO2. Unfortunately, Pakistan depends upon fossil fuels for producing electricity, which is costly and polluted source of energy. However, fossil fuel is one of the major imports of Pakistan to meet the energy requirements of country, as 64.12% of country's total electricity production is obtained using fossil fuels [2]. Wind energy is used as the main source of producing electricity in many countries, such as Germany, Spain, United States, China, India [3].

At present, Wind turbine technology is the most mature and pollution free renewable technology available [4, 5]; because wind exists everywhere on the earth and it has been used as a source of mechanical energy for many decades, furthermore there is a technological capacity also [5] .This study is conducted to optimize wind turbine micrositing. For this purpose, a research is being conducted to study wind pressure distribution and its effect on wind turbines, comparison between two techniques, i.e. computational fluid dynamics software VIZIFLOW 2.0 and discretized form of Laplace equation is made. The main objective of wind turbine micrositing is to gain the maximum result (net revenue) and to minimize the overall cost for energy. Systematical approach to wind turbine placement is a technical issue and was first been proposed by Moseti et al. and gradually worked and been improved by Grady et al.; genetic algorithm was used to find the optimal solution of the problem [6, 7]. Whereas, Computational fluid Dynamics techniques based on linear and nonlinear equations were used for the optimal placement of wind turbines in complex terrain [8]. The present study deals with the analysis of Gharo wind corridor which is the main site for installation of wind farm to meet the requirements of energy in Pakistan. Gharo, is a small town near the coast of Sindh, Pakistan. It is 1,046 Km long coastline, average wind speed is more than 7.4 m/s in Gharo and the estimated wind potential from this site is 50,000 MW [9, 10]. The site is shown in figure 1.

2. Methodology

The methodology that is used to conduct this study is based on the comparison of two models to analyze the results of wind pressure distribution and its effects on the wind turbine on wind farm. The experiment Design can be viewed thoroughly using figure 2. The purpose of adopting this methodology is to study the two models and analyze the results of the proposed mathematical model that whether it can be applicable for the problem or not. Suppose the terrain is a square grid and is supposed to be

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

divided into possible locations. To examine the concept only two Wind Turbines are supposed to be placed. The Diameter of wind turbine is supposed as 200 m. Therefore, the total area can be subdivided into 2kmx2km as shown in figure 3.

As the width supposed for each square grid/cell is good enough, the wake of a column of wind turbine should not affect another. As done in [11]. It is observed that 5D square grid size satisfies the Rule of Thumb spacing requirements for all vertical and horizontal directions [11]. Only one case is considered if the wind speed is constant at 10m/s and the direction of wind is from left to right, for the examining of Laplace equation. Same case is also examined using computational fluid dynamics software VIZIFLOW 2.0.

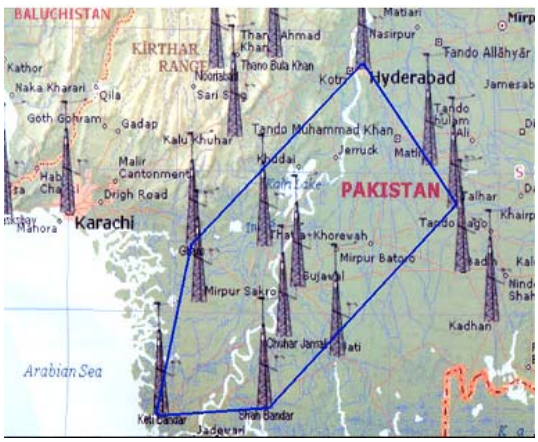


Fig. 1 Depicts the under consideration site Gharo, Thatta, Sindh.

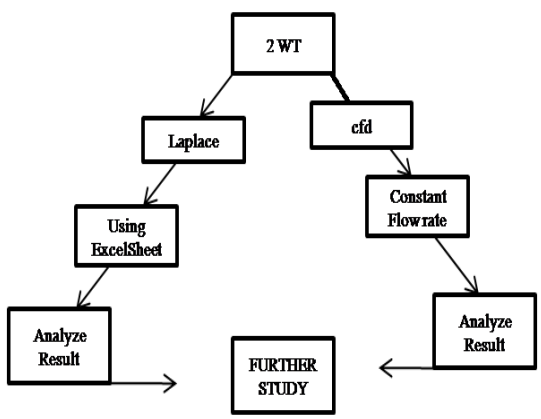


Fig. 2 Represent the Methodology to compare two Techniques.

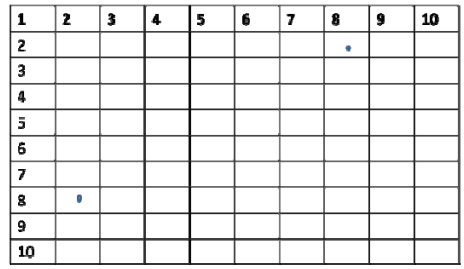


Fig. 3 200 m each cell for a square grid of 2kmx2km

3. Discretized form of Laplace Equation

The Laplace Equation is a second order partial differential equation (PDE). It is used in different areas of science and engineering for example in electricity, fluid mechanics etc. which is

$$\nabla^2 \phi = \frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = 0 \qquad (1)$$

When the selection of a mathematical model is made, the next step is to select a suitable discretization method, i.e., a method of approximating the differential equations by a system of algebraic equations for the variables at some set of discrete locations in space and time. In a finite difference method, approximations for the derivatives at the grid points must be selected. Taylor series expansion is used to obtain approximations to the first and second derivatives of the variables with respect to the coordinates. When necessary, these methods are also used to obtain variable values at locations other than grid points. Here the Laplace equation is derived until and unless to the set of Linear i-e simultaneous equations [12]. A formal basis for developing finite difference approximation to derivative is by the use of Taylor series expansion. Because Taylor series expansion of a function f(x) is about a point x in the forward (positive x) and backward(negative x) directions given, similarly, for a point y upward (positive) and downward (negative) directions.

Assume that there are three points on the X-axis separated by a distance let h. Consider that I, i-1 and i+1 are center, left and right sides of the distance h. Similarly, for Y-axis j, j-1 and j+1 are center, downward and upward points. By applying Taylor expansions [12].

$$\phi_{(i-1)} = \phi_i - \frac{\partial \phi}{\partial x} \Big|_{ih} + \frac{\partial^2 \phi}{\partial x^2} \Big|_{i} \frac{h^2}{2!} - \frac{\partial^3 \phi}{\partial x^3} \Big|_{i} \frac{h^3}{3!} + \frac{\partial^4 \phi}{\partial x^4} \Big|_{i} \frac{h^4}{4!} + O(h^5) \qquad (2)$$

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

$$\phi_{(i+1)} = \phi_i + \frac{\partial\phi}{\partial x}\,|\,ih + \frac{\partial^2\phi}{\partial x^2}\,|\,i\frac{h^2}{2!} + \frac{\partial^3\phi}{\partial x^3}\,|\,\frac{h^3}{3!} + \frac{\partial^4\phi}{\partial x^4}\,|\,i\frac{h^4}{4!} + O(h^5)$$

(3)

∴ |i shows that the derivative is computed at the point i.

$$\phi_{(i-1)} + \phi_{(i+1)} = 2\phi_i + \frac{\partial^2\phi}{\partial x^2}\,|\,ih^2 + \frac{\partial^4\phi}{\partial x^4}\,|\,i\frac{h^4}{12} + O(h^5) \quad (4)$$

$$\frac{\partial^2\phi}{\partial x^2}\,|\,i = \frac{\phi_{(i+1)} - 2\phi_i + \phi_{(i+1)}}{h^2} + O(h^2) \quad (5)$$

Subtracting equation (1) from equation (2), the resultant equation will be:

$$\phi_{i+1} - \phi_{i-1} = 2\frac{\partial\phi}{\partial x}\,|\,ih + \frac{\partial^3\phi}{\partial x^3}\,|\,i\frac{h^3}{3} + o(h^5) \quad (6)$$

Or

$$\frac{\partial\phi}{\partial x}\,|\,ih = \frac{\phi_{(i+1)} - \phi_{(i-1)}}{2h} + O(h^2) \quad (7)$$

Applying same procedure to Y-axis, finally another equation will be:

$$\frac{\partial^2\phi}{\partial y^2}\,|\,j = \frac{\phi_{(j+1)} - 2\phi_j + \phi_{(j+1)}}{h^2} + O(h^2) \quad (8)$$

$$\frac{\partial\phi}{\partial y}\,|\,jh = \frac{\phi_{(j+1)} - \phi_{(j-1)}}{2h} + O(h^2) \quad (9)$$

By combining equations (7) and (9) the resultant equation for x and y will be:

$$\frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2})\,|\,i,j = \frac{\phi_{(i+1),j} - 2\phi_{i,j} + \phi_{(i-1),j}}{h^2} + \frac{\phi_{i,(j+1)} - 2\phi_{i,j} + \phi_{i,(j-1)}}{h^2} \quad (10)$$

After substitution of equation (9) in Laplace equation the resultant equation will be:

$$\phi i,j = \frac{1}{4}\left(\phi_{(i+1),j} + \phi_{(i-1),j} + \phi_{i,(j+1)} + \phi_{i,(j-1)}\right) \quad (11)$$

Or

$$U_{i,j} = \frac{\left(U_{(i+1),j} + U_{(i-1),j} + U_{i,(j+1)} + U_{i,(j-1)}\right)}{4} \quad (12)$$

Assume that the x-axis and y-axis points i-e (i, j) are superposed then the resultant will be a 5- point pattern shown in figure 4.

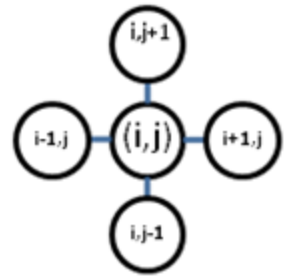


Fig. 4 shows the position of points at x-axis and y-axis

The Discretized form of Laplace equation is used to study the wind pressure distribution and its effect on the wind turbine on the wind farm . Here, the equation (12) shows that there are four variables represent the wind pressure on wind turbine and at different points as well to study the effects of wind pressure distribution. Based on equation 12 [12], air pressure distribution and its effects on wind turbine on the wind farm are considered. Initially, the effects are proven on MS Excel using Laplace equation then the Laplace equation's discretized form is used for finite solution. The computed results based on discretized form of Laplace equation using MS Excel are depicted in table 1 and table 2. The air flow rate is constantly assumed as 10 m/s and using the discretized form of Laplace equation, the results are studied.

It is represented in table 1 and table 2 that cells D5, C6,D6,E6,D7 and M10,L11,M11,N11,M12 are supposed as 0 because the concept is to be examined that if the wind flows at constant speed and in the same direction on the wind turbine, what are wind pressure distribution effects. The mentioned cells of Excel sheet are the representation of wind turbine virtual existence. Suppose that cells D5,C6,D6,E6,D7 are wind turbine called "a" and cells M10,L11,M11,N11,M12 are named as "b".It is depicted that results are satisfactory at wind turbine a but the results are not satisfactory when the same wind speed is applied to wind turbine b. As the wind pressure is gradually distributed and be slowed down if a wind turbine is there. The wind pressure at cells O11, N12 and N10 cannot be raised after it gets down to 0. Figure 7 is the MS-Excel Sheet results which are depicted by Table 1 and 2.

Figure 5 is the graphical representation of the results depicted earlier and represented by table 1 and table 2. There are two wind turbines are represented in figure 5.
The wind pressure is assumed to be from left to right as shown in figure 5. It is shown in the figure 5 that the peaks are the high pressure points in the wind turbine. Table 3 depicts the computed results of wind turbine a and b. Later the effect of air pressure in mathematical model using finite difference method is compared with computational fluid dynamics software discussed in section 5.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

4. Computational fluid Dynamics

A computational fluid dynamics software is used to compare the results of mathematical model with the results of computational fluid dynamics. The gauge theory of this software functions on the principles of two basic equations of fluid flow.

a) The continuity equation

$$A1v1=A2v2 \quad (13)$$

Which means that the volume flow rate is constant, so that a smaller area for the fluid to flow will result in a higher velocity.

b) Bernoulli's equation

This equation relates the pressure, flow velocity, fluid density and fluid height.

$$P1 + \rho gy1 + \tfrac{1}{2}\rho v1^2 = P2 + \rho gy2 + \tfrac{1}{2}\rho v2^2 \quad (14)$$

The second term on each side of the equation represents the potential energy due to the height of the fluid, but since the height is constant in this cfd, these terms cancel and the equation reduces to:

$$P1 + \tfrac{1}{2}\rho v1^2 = P2 + \tfrac{1}{2}\rho v2^2 \quad (15)$$

P is the pressure, v the velocity and ρ the density of the fluid. To find the pressure P2 the equation is rearranged:

$$P2=\tfrac{1}{2}\rho(v2^2-v1^2)-P1 \quad (16)$$

The wind flow is supposed as from left to right again. Table 4 is depicting the tested values for different attributes used for computational fluid dynamics software. Table 5 depicts wind pressure distribution effect is satisfactory at wind turbine a if the reading is taken at all points i.e. up, down, left and right of the turbine and the same procedure of taking readings is applied at wind turbine b. The computed results are not much different from wind turbine a and wind turbine b. The graphical representation of computed results using computational fluid dynamics software based on the discussed equation above as equation (13) and (16) is depicted using figure 6. The figure 6 also represents the same scenario with two wind turbines a and b, the blue color around the circles (virtual representation of wind turbines a and b) shows the high wind pressure.

Table 1: Results of wind speed pressure using Discretized form of Laplace equation on MS Excel.

	A	***B***	***C***	***D***	***E***	***F***	***G***
1		0	0	0	0	0	0
2	10	4.60	2.29	1.16	0.62	0.35	0.23
3	10	6.10	3.39	1.75	0.95	0.57	0.39
4	10	6.40	3.42	1.50	0.84	0.60	0.48
5	10	6.10	2.38	-	0.33	0.49	0.52
6	10	5.61	-	-	-	0.51	0.64
7	10	6.35	2.70	-	0.69	0.91	0.90
8	10	7.10	4.44	2.52	1.87	1.52	1.25
9	10	7.61	5.45	3.78	2.74	2.07	1.58
10	10	7.89	5.97	4.41	3.25	2.41	1.80
11	10	7.98	6.15	4.62	3.43	2.54	1.87
12	10	7.88	6.02	4.50	3.32	2.44	1.79
13	10	7.54	5.54	4.02	2.92	2.12	1.53
14	10	6.75	4.56	3.15	2.21	1.58	1.13

A-Initial Wind Speed 10m/sec

Table 2: Depicts the continued results of wind speed pressure using Discretized form of Laplace Equation on MS Excel.

	H	***I***	***J***	***K***	***L***	***M***	***N***	***O***
1	0	0	0	0	0	0	0	0
2	0.23	0.16	0.12	0.10	0.08	0.06	0.04	0.03
3	0.39	0.30	0.24	0.19	0.15	0.11	0.08	0.05
4	0.48	0.40	0.33	0.27	0.22	0.16	0.12	0.07
5	0.52	0.49	0.43	0.35	0.28	0.21	0.15	0.09
6	0.64	0.61	0.54	0.44	0.34	0.25	0.17	0.11
7	0.90	0.80	0.66	0.52	0.39	0.27	0.18	0.11
8	1.25	1.01	0.80	0.60	0.42	0.28	0.17	0.10
9	1.58	1.21	0.91	0.65	0.43	0.25	0.12	0.07
10	1.80	1.34	0.97	0.67	0.41	0.16	-	0.02
11	1.87	1.37	0.98	0.65	0.35	-	-	-
12	1.79	1.29	0.91	0.61	0.35	0.13	-	0.01

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

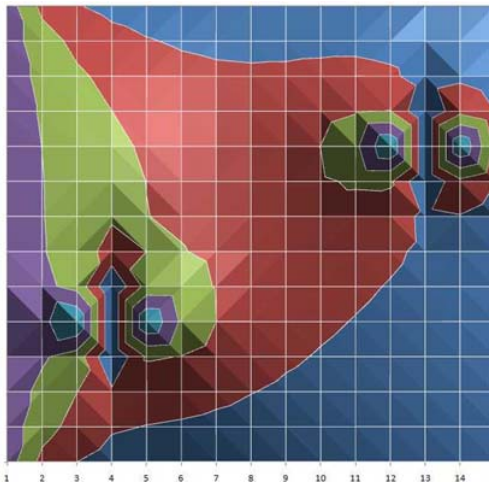


Fig. 5 Wind turbines effect on Air pressure using Finite Difference Method

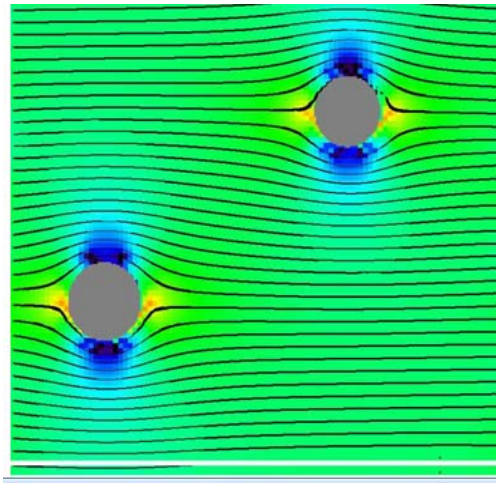


Fig. 6 Wind turbines effect on Air pressure using Computational fluid dynamics model (Two wind Turbines- a, b, Blue color shows high wind pressure)

Table 3.Shows the Computed Results using Discretized form of Laplace Equation.

Wind-Turbine	***Position***	***Flow Rate***
a	UP	1.5
	Down	2.51
	Left	5.61
	Right	0.49
b	UP	0.11
	Down	0.06
	Left	0.34
	Right	0.00

Table 4 . Data for computational fluid Dynamics Test

WT	**Flow Rate**	**Pressure**	**Density**	**Diameter**	**Dist b/w WT**
2	10 m/s	1	1 kg/ m^3	20,20m,m	10 m

WT- Wind turbines

Table 5.Shows the Computed Results using computational fluid Dynamics.

Wind-Turbine	***Position***	***Flow Rate***
a	UP	8.44
	Down	6.42
	Left	4.13
	Right	4.91
b	UP	6.79
	Down	6.04
	Left	2.65
	Right	3.52

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

	A	B	C	D	E	F	G	H	I	J	K
2	10	4.60	2.39	1.16	0.61	0.35	0.22	0.15	0.11	0.09	0.07
3	10	6.10	3.39	1.74	0.94	0.56	0.38	0.28	0.21	0.17	0.13
4	10	6.40	3.42	1.50	0.83	0.58	0.45	0.37	0.30	0.24	0.19
5	10	6.10	2.30	-	0.33	0.47	0.49	0.46	0.39	0.32	0.25
6	10	5.61	-	-	-	0.49	0.60	0.57	0.49	0.40	0.30
7	10	6.35	2.70	-	0.68	0.87	0.86	0.75	0.62	0.48	0.35
8	10	7.10	4.43	2.51	1.84	1.48	1.21	0.97	0.75	0.56	0.39
9	10	7.60	5.43	3.75	2.70	2.02	1.53	1.16	0.86	0.61	0.40
10	10	7.88	5.85	4.37	3.21	2.37	1.75	1.29	0.93	0.64	0.38
11	10	7.97	6.13	4.59	3.39	2.50	1.83	1.33	0.94	0.62	0.34
12	10	7.87	6.00	4.47	3.29	2.40	1.75	1.26	0.88	0.59	0.34
13	10	7.53	5.52	4.00	2.89	2.09	1.50	1.07	0.75	0.50	0.31
14	10	6.74	4.55	3.13	2.20	1.56	1.11	0.79	0.55	0.37	0.23
15	10	4.89	2.81	1.79	1.20	0.84	0.59	0.42	0.29	0.20	0.13

Fig. 7 resulted Values on MS-EXCEL SHEET after using Laplace equation (Data is already given in Tables 1 and 2)

5. Discussion

In this paper, the study of wind turbine micro-siting using Finite Difference Method is represented. In this study wind pressure distribution and its effect on wind turbine on the wind farm are considered. Initially Laplace equation's discretized form is used for finite solution. First the concept is proven on MS. Excel. Later the effect of air pressure in mathematical model using finite difference method is compared with computational fluid dynamics software. The computational fluid Dynamics software is based on another equation. The purpose was to study the wind air pressure distribution effects using two techniques. While using Technique 1; the result is depicted that if the wind speed is constant at 10m/s and direction is from left to right, the wind pressure distribution effect is computed at wind turbine a and wind turbine b are totally different from the results computed using the Technique 2.

The concept that the wind pressure will be less or even zero at the wind turbines and it gradually be raised to high pressure as it goes farther from wind turbines. This concept is proven positively at computational fluid dynamics however, this concept cannot be proved at Laplace equation. Using the Laplace equation, it is observed that when the wind pressure goes to zero it cannot be raised at the points where the wind pressure raised using computational fluid dynamics. Therefore, it is concluded that this model is not successful in comparison of computational fluid dynamics software (VIZIFLOW 2.0)based on the continuity equation and Bernoullis equation

Furthermore, Wind turbine placement is a discrete problem, therefore, classical methods are more complicated for optimization of this problem [13]. This study depicts that Mathematical model based on Laplace equation in this study was initially tested using Excel sheet so it is suggested that it should be tested using Genetic algorithm to see the results. Because the Genetic algorithm technique operate on the coding of the parameter set rather than the parameter itself.

References

[1] I. Martin, et al , 'Advanced Technology Paths to Global Climate Stability: Energy for a Greenhouse Planet' a review in SCIENCE 298, 1, 2002

[2] Electrical Energy Sector Overview, Alternate Energy Development Board, Government of Pakistan, 2010 http://www.aedb.org/re_sector.php

[3] M. A. Ahmed, F. Ahmed and M. W. Akhtar, 'wind characteristics and wind power potential for southern coasts of sindh, pakistan', Journal of Basic and Applied Sciences, 6 (2), 163-168, 2010

[4] A. Houri, 'Effects of Funds Diversion from Fuel Purchasing to Wind Energy Installations: A Case Study'

[5] J. F. Manwell, et al., 'Wind Energy Explained : Theory , design and Application' Second edition , Willey, published in 2009

[6] C. Wan, et al., 'Optimal Micro-Siting of Wind Turbines by Genetic Algorithms Based on Improved Wind and Turbine Models', Joint 48th IEEE Conference on Decision and Control

[7] S.A. Grady et al., "Placement of wind turbines using genetic algorithms", Renewable Energy 30 (2005) 259–270

[8] J.M.L.M. Palma, 'Linear and nonlinear models in wind resource assessment and wind turbine micro-siting in complex terrain', Journal of Wind Engineering and Industrial Aerodynamics, Volume 96, Issue 12, December 2008, Pages 2308-2326

[9] Saifullah, and A. Allauddin, ' Opportunities & Challenges toScaling up wind power in Pakistan', Presentation in Quantum Leap in Wind Power in Asia Structured Consultation , June 21,22 2010

[10] Ch. Qamar-uz-Zaman et al., 'Establishment of commercial wind power plant of 18 MW at Gharo', Technical Report No. PMD-11/2004, pakistan meteorological department sector h-8/2, p. O. Box 1214, Islamabad

[11] G. Marmidis, et al., 'Optimal placement of wind turbines in a wind park usingMonte Carlo simulation', Renewable Energy 33 (2008) 1455–1460

[12] K. Ambar, 'Finite Difference Method for the Solution of Laplace Equation', Department of Aerospace Engineering Iowa State University

[13] A. Emami , P. Noghreh, 'New approach on optimization in placement of wind turbines withinwind farm by genetic algorithms', Renewable Energy 35 (2010) 1559–1564

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Comparative Study between Various Aspects of the Academic and Nonacademic Institutions Websites

Khalid S. Husain, Ph.D
Information Technology Deanship, King Abdulaziz University
Jeddah, Kingdom of Saudi Arabia

ABSTRACT
Designing a good Web site for an academic institution has now become a highly specialized field as more and more educational institutions understanding the magnitude of having an online presence. Web site designing for educational institutions involves creating a web site that is only functional but is attractive, informational and dynamic as well. An educational institution's web site is now more than just an informational site, it works to advertise and promote the particular school, college or university's image. On other hand, with regards to the nonacademic and commercial organizations, we found that the degree of importance of website in running a business depends on the type of business. So, the Websites play an important role in advertising and maintaining business at an international level. This paper will attempt to shed some light on the main difference between the criteria that differentiate between academic and nonacademic institutions specially that is related to design, template, and software available to be guided with as we go on evaluating the performance of both Webpages of educational and commercial purposes.

1. INTRODUCTION

Web site designing for educational institutions is now becoming increasingly specialized in today's world. An educational institutions web site design should be clear, precise, easy to understand, interesting and informative all at the same time [1]. On the other hand if we investigate the rules and criteria that are used in designing the website of the commercial purposes, we will find that many companies today have benefited significantly from the use of corporate websites. The use of websites in export marketing boosts sales through ease of communications and transactions, and publicly traded companies with corporate websites have seen steady increases in stock activity due to their ability to post up- to-date information [2]. In other words, corporate websites increase visibility and convenience for customers. Visibility of a company or brand—how easily it can be viewed—contributes immensely to brand recognition among consumers. More and more, companies are losing customers to those businesses with websites. This is because they are more visible, as websites enable people around the globe to visit the business, and more easily recognized. This problem could easily be resolved by setting up a corporate website; not only will it enable the company to release time-sensitive information, but it is an effective way of increasing visibility in every part of the world [2].

Web site designing for education institutions involves employing a system that empowers the non-technical faculty, staff and students to create, update, manage, and maintain massive amounts of content within the site itself [1]. When deciding to design a web site for educational institutions, the following points should be kept in mind: Make sure you plan the site and consider all the options and map them out; Make a blueprint of how your site will be and how it can further expand with time; When building a website, do so gradually making sure all the relevant information is there; Keep it simple and accessible for all users including those with special needs; Use functionalities that are unique to the web making it easier to navigate around the website and seem interesting at the same time, like use hyperlinks, graphics or small animations; Always remember your target audience; Make sure you invite regular

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

feedback on how others view your web site if it can be more user-friendly; Update your site regularly; Constantly review your site and ensure that the information you provide is regularly updated; As an educational institution the web site should enable current or prospective parents and students to [1]:

- Get an overview of your institution, its philosophy and workings
- Get an overview of the programs offered and/or the curriculum
- View the work done by those in the institution
- Read your institution's newsletter
- Find out the date and time of an event which you are running such as an athletic meet or a student pep rally
- Send email to teachers and/or lecturers within the institution
- Check homework schedules
- Get involved more in the school/college community
- Understand the student life
- Have online communities that showcase your institution's strengths
- Have a link with the outside world in general, such as options for home-school links and overseas collaborations

A well-organized site will help users find key information quickly. It also will make the experience of using the website more coherent, intuitive and satisfying. Additionally, intuitively organized information architecture will help to ensure that all of the phases of the website development run smoothly and efficiently [3]. In fact, it can prevent time-consuming and costly alterations to the visual design and site development by identifying required features, the number and location of navigational links and the placement of content early in the process. If the site is being built in a content management system that allows editors to add their own pages, the site may grow quickly, sometimes in confusing ways. However, if we have planned our organizational structure well, we will be able to easily identify the most appropriate/intuitive location for this additional information [3]. Existing research has investigated factors responsible for website effectiveness. In this context, studies describe different characteristics related to websites such as information quality, website loyalty, consumer shopping experience and usability. Websites serve different purposes-shopping, information gathering, entertainment, research, and others. Hence there are different "categories" of websites. These categories include among others, retailing websites, news websites and search websites. Much of the research in this area has been on websites belonging to one particular category and has focused on a few parameters that influence website success [3]. Although there are not many studies that analyze and compare characteristics of websites belonging to different domains, academic literature and business sources provide preliminary evidence that customers expect websites of different categories to be designed differently from one another, depending on the task and purpose associated with the website. Therefore the relative importance of a particular website characteristic varies, depending on the domain to which the website belongs.

Some parts of the typical Web checkout process are particularly annoying examples of what we call “technology steps”—actions that are not part of the user’s agenda but are present because of the failure to overcome or work around limitations in the technology. The “recalculate total” or “update order” button is a common example. Recalculating the total is not the user’s problem; it should happen automatically anytime the form has been changed. There are dozens of ways to achieve this behavior on the Web. Perhaps the worse sin in the form of a technology step is allowing user entered data to vanish through use of the back button [13]. Never lose, discard, or corrupt user input. Requiring the user to re-enter the information in a long form is one of the surest ways to lose a sale and possibly a customer. Another technology step that mars the final stages of the purchasing process is the need to change the quantity to zero to remove an item from the shopping cart. Best practice provides a button or glyph to remove an item.

2. RELATED WORKS AND LITERATURE REVIEW

2.1. Characteristics of a Good Websites

A number of characteristics can be used to describe websites. *The first* of these is related to the content of the information on the website. Researchers have suggested that

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

the information should be relevant to the purpose of the website, easy to read and comprehend, and useful. It should also be of adequate scope and depth and current. *Second*, the navigation characteristics of the website are also important. Navigability depends on the manner in which the information is organized and arranged, in terms of design, layout and sequencing. Information organization also includes aspects such as the number and effectiveness of hyperlinks, and the overall organization of the information. These characteristics affect the ease with which the website can be navigated [4]. *Third*, the website should be easy to use. It should be designed so as to facilitate effective completion of the tasks associated with it. This property has been variously referred to as usability and ease of use. The extent to which a website is fun to use and visually appealing is seen to be an indicator of how easy it is to use it. The use of graphics, multimedia and other interactive elements also makes the website interesting and easy to use. Features such as search engines and shopping carts make it easy to interact with websites. Overall the ease with which the website can be used enables users to accomplish their purpose and goals at the website with effectiveness, efficiency and satisfaction.

Fourth, personalization characteristics and the capability to cater to customized information requirements of specific groups of customers are important aspects of website design, and make it easy for customers to use the website [4]. For instance customization capabilities such as the "one click" feature or "My Yahoo" features help customers to easily return to specific information and make it easy for them to interact with the website. Customization capabilities also include response related features such as search capabilities and shopping carts. *Finally*, the technical properties of the website are also seen to be important characteristics. Security features are determined by provisions for user authentication and secure transactions. The access speed is characterized by how fast the website can download its pages and how fast it can display successive pages. A third technical property is the accessibility or availability of the website. Accessibility is important because if the website is not available on a sustained basis, browsers are not likely to return to it. Accessibility and access speed depend on the overall technical reliability of the website and are dependent on the underlying IS infrastructural platform.

2.2. Differences in Website Functionality

Recent studies have suggested that users who visit websites do so with a variety of goals, predispositions and purposes. Hence websites in various categories should have different combinations of properties, depending on user requirements and the purpose of the website. In other words, the different website characteristics should be emphasized to varying extents. In a study of website usability characteristics across four different kinds of websites (airlines, bookstores, auto manufacturers and car rentals), Agarwal and Venkatesh [5] found that each domain differed in the extent to which various usability criterion are valued by users. For instance, they found that customers did not think that the extent to which the website engaged the customer emotionally was equally important, for all the four varieties of websites. Similarly, it was found also that in the case of retail websites, users valued website security over other characteristics such as information content and privacy. Other studies on retail websites suggest that information content, ease of use, and trust and security are characteristics that are valued by customers to different extents. Chung and Tan [6] in a study of news and information search websites suggest that perceived playfulness, as given by the "extent to which the individual perceives that his or her attention is focused on the interaction with the World Wide Web; is curious during the interaction; and finds the interaction intrinsically enjoyable or interesting," is an important indicator of the extent to which they would use these websites. They further suggest that perceived playfulness is influenced by navigation properties, ease of use, feedback and website content. Hence there are indications that all characteristics are not equally important for all categories of websites.

3. AN OVERVIEW OF WEBSITE DESIGN PHASES

Fig.1 illustrates an overview of the Website Design (WSDM) method [17]. As shown from this fig., the

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

first step is to define the Mission Statement. The Mission Statement should express the purpose and subject of the web site, and declares the target audience. Based on the Mission Statement, the Audience Modeling is performed, in two steps: Audience Classification and audience class Characterization. During the Audience Classification phase, the different types of users are identified, while during Audience class Characterization, characteristics of the different Audience Classes are given. Next, Conceptual Design, in three steps: Information, Functional and Navigational Modeling. During Information Modeling we observe what kind of information is needed, while in the Functional Modeling phase we observe the functionality. The Navigation Design phase, we consider the global navigation through the information and the functionality. The next phase, Implementation Design, performs Page, Presentation and Logical Database Design. During these phases, grouping in pages, specifying the look and feel, and designing the database is included. The logical data base schema can be derived from the Business information model, which is made during the Conceptual Design [17]. The last phase, Implementation, is the actual realization of the website using the preferred implementation environment.

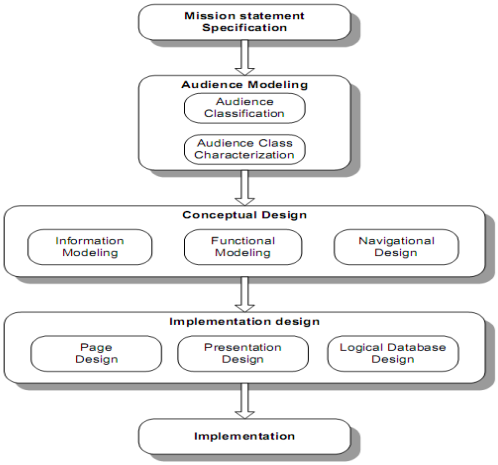


Fig.1 an Overview of Website Design phases [17]

With regards to the university (academic institution) site structure, we should figure out what features and capabilities that we need and knowing how we will use them before we design the layout, place appliances and plugs, and select tiles, curtains and countertops. Good web design requires solid site architecture based on the site's goals and target audience established in the project brief. The deliverables from this phase are:

1. Content Outline
2. Site Diagram
3. Page Description Diagrams
4. Wireframes

These four deliverables are dependent on each other and need to be completed sequentially.

- **Content Outline**

Working closely with our clients, create a list of all existing content. Brainstorm content that needs to be created for the site. Review the list of content, trimming anything that does not match the goals or audience needs as stated in the project brief. Take time to think about the future and how the site content might need to grow. Make sure we leave room for growth. Next group our content into categories. As we categorize our content, considering getting user feedback through a card sort. Once our categories are established, create an outline of our content and review it with our clients for accuracy.

- **Site Diagram**

Take our final content outline and create a sitemap or site diagram. A site diagram is just a visual representation of our content outline and site structure. We can use Excel, Visio or Omni raffle to create our site diagram as shown in Fig.2.

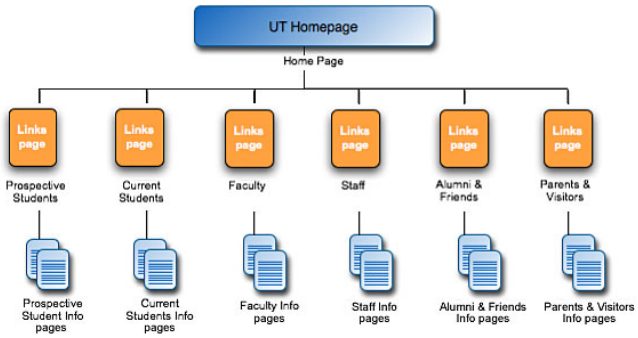


Fig.2 Sitemap Structure of the Academic Institution

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

- **Page Description Diagrams**

Many Information Architects are now recommending the use of Page Description Diagrams (PDD) as a step before wireframes or even as a replacement to wireframes. The focus of a PDD is two-fold:

1. What content belongs on this page
2. What is the priority of each chunk of content

A common layout for the PDD is to use the horizontal access for priority. For example, a PDD for any give page might have three columns. The first column would list the high priority content. The second column would list the medium priority content. The third column would list the lowest priority content. The advantages of the PDD over the wireframe include:

1- Clarifies all content for a given page
2- Clarifies the priority of each chunk of content
3- Completely removes visual design (color, font, and placement) from this stage of the conversation.

- **Wireframes**

A wireframe is a non-graphical layout of a web page. It is a simple drawing of the chunks of information and functionality for each page in our site. We should create a wireframe for the home page, each unique second level page and any other significantly different page on the site. Wireframes include the containers for all the major elements of the page. Elements include navigation, images, content, functional elements (like search) and footer.

4. STANDARDS THAT RECOGNIZE THE ACADEMIC WEB SITES

Academic Web sites communicate the brand and mission of the various institutions. Mainly, they are used by prospective students to learn about an institution and then used later to apply [7]. Current students use them to pay tuition bills, register for classes, access course materials, participate in class discussions, take tests, get grades, and more. Online learning and course-management software programs, such as Blackboard, continue to increase the use of Web sites. They are now an important learning tool for the entire campus community and the primary communication tool for current students, parents, alumni, the community, donors, and funding organizations. Web site standards have developed since the 1990s. Usability and accessibility are now important tenets for Web site designers, especially for educational institutions. As a result, campus Web designers or outside consultants are often responsible for designing large parts of the academic Web site. As Web sites have grown, ongoing maintenance is an important workload issue. Databases and other technologies are used to simplify daily updates and changes to Web sites. This is where the academic template fits in.

An academic template can be defined as a common or shared template used to control the formatting of Web pages in different departments on a campus. Generally, administrators will mandate the use of a specific template or group of templates. This mandate includes guidelines for such things as layout, design, color, font, graphics, and navigation links to be used on all Web pages. Often, the templates are administered using content management systems (CMSs) or Web development software such as Macromedia's Contribute. These programs give different levels of editing rights to individuals, thus keeping tight control over particular Web pages or even parts of Web pages. Academic templates give the Web site administrator the ability to change the template and update all pages with a single keystroke. With regards to the education sites, we see that the on-line course delivery is a major request. Blackboard seems to be the dominate software for on-line delivery. A quick Google for other on-line software shows WebCT, Digital Think, AuthorWare and GeoLearning. Another quick browse through college sites shows an overwhelming use of Blackboard. Blackboard was used at many colleges. An example of an html distance class site is shown in Fig.2. Blackboard includes common links such as syllabus, assignments, etc. It also allows us to create our own links. Students log into Blackboard and can view course material, post

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

to discussion boards or participate in on-line chats. The education site is too interested in satisfying the following [7]:-

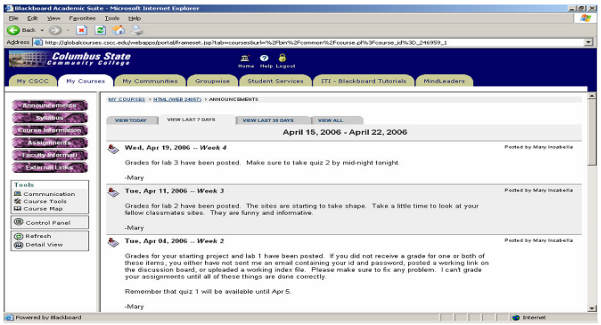


Fig.2 An html distance class site [7]

a- *Test generators;* Creating quizzes and exams to be taken on-line can be done through Blackboard's tools or through third party software tools. Such tools include *Respondus*, Articulate, *PrimeExam and QuizMaker* to name a few. These tools allow us to take tests created in a word processor and upload into a web delivery format.

b- *Screen capturers;* many on-line course sites offer video and/or audio instruction modules to enhance the web experience. Some capturing tools include *Camtasia, Illuminate, Captivate and Screencast*. These video presentations may record a lecture or demonstrate using a computer or calculator. While this kind of enhancement works well for users with digital connections, dial-up users find it takes quite a while to download the files.

c- *Accessibility;* Accessibility issues are another issue that must be addressed when creating education web sites. Federal guidelines such as Section 508 Standards are mandatory for federal and state websites. Voluntary standards such as Web Content Accessibility Guidelines.

The following actions are encouraged to be exist in the educational website:-

 # Check the HTML1 of all new web pages. Make sure that accessible elements are used, including alt tags, long descriptions, and captions, as needed.

 # If images are used, including photos, graphics, scanned images, or image maps, make sure to include alt tags and/or long descriptions for each.

 # If you use online forms and tables, make those elements accessible.

 # when posting documents on the website, always provide them in HTML or a text-based format (even if you are also providing them in another format, such as Portable Document Format (PDF)).

In view point of designing well website for the academic institution, it is recommended to follow up the following criteria to fulfill well the user requests and achieve best utilities:-

I- Ease of Use; In order to let the website more attractive to the users, so the usability of it should be easy as the accessibility of users to the requested information should be smoothly and too easily without any complexity or delay since time of the user is too expensive and the internet service provider is working as the user computer is connected across the network.

II- Targeted Sets of users; The process of knowing the characteristics of the sets of users whom will use the website specially whom having special types of skills will assist in determining the nature of the designed website to be compatible with their qualifications. Also, preparing the computer sets that user will use it from accessories or software either on homes or the institution.

III- Design Pattern; following up certain specified design pattern of all website pages will guarantee the familiarity between these pages and user, accordingly encouraging him to use it effectively with high efficiency for all contents of it.

IV- Approaches of presenting the information; although using various shapes to present information (images, figures, text, and video) can offer it easily on the network websites; however excessive use of it may causes barrier in accessing the information and may deny some users from accessing especially if they have not the programs that run these information files.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

V- Readiness to meet the requirements of use; in order to let all users utilize efficiently all contents of the website in spite of diversity that exist in their computers and programs, the design team should study the technical aspects of the user equipment's as explorers and the type as well as speed of connecting computers with the Internet.

VI- Make sure the public consideration of the pages and their harmonization; the design team should utilize from the opinions of others especially whom use the website concerning the page shape and template and this will be applied on the color compatibility, background used in the pages, font sizes and types of the used fonts.

Regarding the academic institution website structure, we see that the University website is focused around the user and their needs and purpose in visiting the site, regardless of which School or department serves that need.

• Audience-focused

The top-level sections of the website are based on the different audiences using the site whom cover:-

- Prospective students
- Business and community
- Alumni
- Current students (students do not need to log in to see information aimed at current students)
- Staff (there is no staff 'intranet' - all information for staff is available on the website with small pockets of individual pages being password-protected)

In addition there are three topic-based areas which serve external audiences:

- Research (to showcase the University's research)
- About us (general information about the University)
- Social enterprise

All pages on the site will ultimately fall within one of these top-level sections and these form the site's global navigation (found across the top of every page).

• Content development

Content can be in the form of copy, images, video and audio clips, FAQs, and downloads.
The website is built using a content management system (CMS), which allows content to be re-used rather than duplicated if that content is required in more than one place.

1. Categories

Content is added to the website by creating documents, FAQs or downloads. Each item of content can be assigned one or more categories from the taxonomy. The category applied to an item of content will determine where the information appears on the website.

2. Content owners

Each document or download in the CMS has one owner and that owner is responsible for keeping the information current. A document may have several sub-pages, so the owner will also be responsible for the content on all the sub-pages of a document. Content will only be generated by whoever 'owns' that content, for example, the Timetabling manager will create and maintain content about student and staff timetables. Other content owners may add contextual links to Timetabling information, secure in the knowledge that the Timetabling manager will be keeping the Timetabling information up-to-date.

3. Tone and style

Marketing have overall responsibility for the tone and style of content on the website and for the quality and deployment of images across the site.

• Navigation

The left-hand navigation for the site is created from the category names, document titles and subpage titles. Page titles should be brief, but informative. We should also aim to use vocabulary which is meaningful to the intended audience. This can be verified by ongoing usability testing and by regular interrogation of web analytics and search terms.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

• Contextual links

When adding contextual links in a page, we should be careful that the page we are linking to is in the correct area of the site. If multiple categories have been applied to a document, then it would be possible to be linking to the document in a different audience area and this would result in the visitor becoming lost and possibly unable to find the information they were looking for.

• Related links

If you wish to link to another area of the site which may be related to the content we are producing but does not fall within the same section, there is the option to add a related link. This will be shown on the right-hand side of the page. Related links are maintained by the web team.

We can summarize the Website structure of the universities as follows: the information architecture of websites includes an organization scheme (content categories) and a structure (usually hierarchical).

• Organization schemes

Information can be structured according to a variety of organization schemes as:

- **Alphabetical:** phone directories, dictionaries
- **Chronological:** diaries, event calendars, publication archives
- **Geographical:** travel, news and weather
- **Topical:** yellow pages, newspapers, academic courses
- **Task oriented:** menu systems on desktop applications, e.g. file, edit, format menus
- **Audience specific:** students, staff, visitors
- **Metaphor driven:** desktop computers use several, e.g. file and folder system, recycle bin

On the web, hybrids of two or more of these schemes are usually used. Organization schemes work best when users understand them. Work with target audience groups when creating or refining the site's organization scheme. Card sorting and content agreement analysis are user-centered methods for designing an organization scheme. Usability testing can confirm that the resulting scheme is workable. Avoid designing an organization scheme that mirrors the structure of the organization. Users who are unfamiliar with the organizational structure may find it hard to locate resources. Also, organizational changes will require changes to the organization scheme.

• Organization structures

There are two main approaches to organization structures on the web: hierarchical or top down and bottom up as in database models. Some sites also use a linear structure, though this is generally unsuitable for anything other than very small sites.

• Hierarchies (top-down approach)

Hierarchies are simple and effective ways of organizing information and everyone is familiar with how they work. Three important things to remember about designing hierarchies are:-

1. Hierarchical categories are usually mutually exclusive , however, there may be good reason for cross-listing between categories
2. Aim for a balance between breadth and depth; if a hierarchy is too broad, users may have to choose between too many options; if it is too deep, they may have to click too many times to reach the content
3. Don't restrict yourself to the hierarchical model; some content may be better organized using the database model.

• Database models (bottom-up approach)

The database model is useful with relatively homogenous content. Course information and staff directory content are good examples. In the database

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

model, metadata is the key. Here, the metadata relates to structural elements of content objects. For example unit code, unit name, unit description, prerequisites and year offered are elements of the unit content object.

- **Linear structure**

Use of a linear structure assumes that users will move through the content starting from point A and progress through an ordered series of page to point Z. This may be useful for material that requires a linear approach, e.g. some instructional material, but this approach is generally unsuitable for websites with more than a few pages.

5. RULES AND CHARACTERISTICS OF NON ACADEMIC WEBSITES

For organizations engaged in electronic business, the corporate website has emerged as the single most important interface through which transactions are carried out [7]. This being so, appropriate design characteristics are required to make websites effective. Customers expect websites to be designed differently, depending on the task and purpose associated with the website. There are six critical characteristics of websites and the relative importance of these characteristics varies across categories. For organizations using the Internet to conduct commercial and non- commercial transactions, the website has emerged as their "window to the world." It is the interface through which employees and customers carry out their business with the company. Indeed the number of tasks that are accomplished on the Internet is increasing every day. This being so, appropriate design characteristics are required to make websites effective. Therefore the study of website characteristics and their influence on website performance forms an important aspect of framing strategies for website design. Most of websites we visit would be considered commercial in nature. Like education sites, there is a need for clarity and usability in these sites. Web site conversion refers to the likelihood that someone coming to a commercial site will actually purchase something. According to w-edge Design (http://www.w-edge.com) 98% of site visitors do not buy anything. Even people, who come to a site ready to buy, only complete the transaction 30% of the time. "Most failures in getting a customer to convert (to a sale, lead, subscription, registration, etc.) stem from a perception of a lack of value, trust, confidence, security, or relevance," writes Eisenberg Bryan in Persuasive Online Copywriting.

When creating commercial websites, it is important to understand how people scan the site. The Stanford Poynter Project discovered that when people view websites, their eyes to the text first, particularly to captions and summaries. They scan graphics afterward, so copy is deemed more important than graphics. They also discovered that the center of the screen is the prime area to sell a product. The left side of the page functions well for navigation and the right side for action. Users generally ignore the bottom of the screen. Good Site Design Practices (http://goodpractices.com) recommends simple pages that are platform-independent. These pages download faster. Customers who have to wait for a site to download will go on to another. The site also recommends using a validator to test site's compliance with common HTML specification (w3c.com has a validation service). Using color wisely can make the difference between someone staying with our site or leaving it abruptly. Dark backgrounds are hard on the eyes and are harder to print. Fig.3 illustrates a screen shot of a site which allows us to experiment with color combinations, as well as giving the correct html color code [7]. Designing the nonacademic websites verifies the following activities:-

A. Searching

Since many people find websites through the use of a search engine, it's important to know how they work and how to get our site to the top of the list. Search engines vary in the way their spider works, how their database is designed and the algorithms used to give relevant and timely responses to user's queries. A search engine works by sending an automated "crawling" process, called a spider. The spider searches through internet sites and "ranks" sites based on the following content (www.searchengines.com):

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

- Location of search terms in the document
- Relational clustering
- The site's design
- Link popularity
- Click popularity
- "Sector" popularity
- Pay-for-placement rankings"

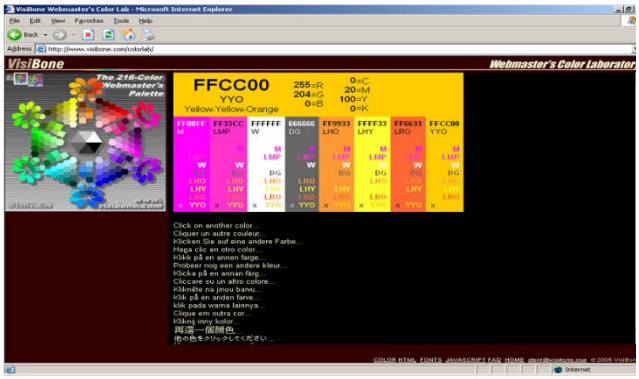


Fig.3 Screen shot of a site that give the correct html color code [7]

B. Metadata

The use of metadata in searching is becoming standardized in the development of Resource Description Framework (RDF), which is a declarative language which uses XML to represent metadata. According to the w3.org site:

RDF provides a framework in which independent communities can develop vocabularies that suit their specific needs and share vocabularies with other communities."

Standardizing vocabularies makes searching a database a much more precise endeavor. People using the internet for research will find their responses to be more directly on-point. The aim is to reduce the "hits" and have the hits be more on-target to the searcher's needs.

C. *Error Messages*

One of the most frustrating things about working with computers, in general, is dealing with error messages. Whether they are generated by your computer, network, system or website, they can be irritating up to infuriating. On the web, error search engine works by sending an automated "crawling" process, called a spider. The spider searches through internet sites and "ranks" sites based on the following content (www.searchengines.com):

- Location of search terms in the document
- Relational clustering
- The site's design
- Link popularity
- Click popularity
- "Sector" popularity
- Pay-for-placement rankings"

Jakob Nielsen's Alertbox (www.useit.com) recommends the following guidelines for effective error messages:

- Explicit indication that something has gone wrong.
- Human-readable language
- Polite
- Precise
- Constructive messages can be hidden on an already overwhelming page.

6. MAIN BENEFITS RETURNED FROM GOOD WEBSITE TO NON-ACADEMIC INSTITUTIONS

There are some important goals can be achieved if we by develop good website [18], below some of these goals:-

a. Far Cheaper and Much More Flexible Than Print Advertising; the Internet is extremely different from print advertising in that space is cheap, our advertisement is accessible for a longer period of time, the content can be changed without having to ask someone to do it for you (if we use a content management system) and we can potentially reach a wider audience. This is not to say that we should not use other forms of advertising at all We can use it to entice people to visit our website and find out about our institution and potentially open two-way communication between the potential customer and a sales person (in case of non-academic institution case) .

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

b. Market Expansion; The Internet has allowed businesses to break through the geographical barriers and become accessible, virtually, from any country in the world by a potential customer that has Internet access .

c. Diversify Revenue Streams; A website is not just a medium for representation of our company, it is a form of media from which everybody can acquire information. We can use this media to sell advertising space to other businesses. A recent trend has risen where businesses feature their very own directory of complimentary services, where the visitor can search for information on a business that will enhance the use of your service. The business sells complimentary businesses a listing in their directory. A good example is a catering company featuring a directory with businesses such as event co-coordinators, electronic equipment rental companies, etc .

d. 24 7 365; No more turning customers away when its time to close shop, putting up a note saying closed for public holiday , or leaving an irritating message on your answering service specifying your trading hours tell them to visit your website for information they are looking for .

e. Offer Convenience; It is far more convenient for a person to research a product on the Internet than it is to get in a car, drive somewhere and look for or ask someone for information on a product. Also, a potential customer won't have to judge a call center agent to determine whether he/she has their best interests in mind, or just wants to make a sale.
The potential customer can visit our website whenever they like in their own privacy and comfort, without the stresses and distractions that exist in the real world.
Our website is a self-service medium for example, instead of having to wait in a long cue to pay our TV License; we can now do it electronically through the TV License website.

f. Add Value and Satisfaction; by offering convenience, a point of reference and that touch of individualized customer service, we ultimately add value to our offering and our customers experience a higher level of satisfaction. Our website can add value in other ways too, by featuring tips, advice and general interest content we can entertain our customers. This will also help them remember us better .

g. Standardize Sales Performance; by looking at which approached / pitches have worked in the past and those which have not, we can produce the ultimate pitch and use it with our website, so that we use it on every customer. No more training of sales people and waiting for them to get a feel for your line of trade .

h. Improve credibility; A website gives us the opportunity to tell potential customers what we are about and why we deserve their trust and confidence. In fact, many people use the internet for pre-purchase research so that they can determine for themselves whether a particular supplier or brand is worthy of their patronage, and won't take them for a ride. The Internet also allows for Viral Marketing where our website visitors spread positive word-of-mouth about our business - our customers do our marketing!

i. Growth Opportunity; A website serves as a great place to refer potential investors to, to show them what our company is about, what it has achieved and what it can achieve in future.

j. Two-Way Communicative Marketing; Customers can quickly and easily give feedback on our product and/or marketing approach .

k. Cheap Market Research; We can use features on our website such as visitor polls, online surveys and our website statistics to find out what our customers like more and how they feel about certain aspects of our business to determine how we can improve our product and the way we do business.

7. CONCLUSION

There are various web design methodologies and strategies that use different mechanisms

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

and algorithms. However, none of them deals with integrating business rules during the website development process. In view point of designing a nonacademic websites we see that business rules which have been around since ancient times covers many human needs and their daily life business and therefore should be considered in parallel with other design issues. From our vision, regarding the nonacademic institutions, we see that it is too important to satisfy some basic elements in its website including [19]: A clear, concise mission statement, Proofread, thoughtful and professional content throughout; A business blog/news section to attract links and demonstrate credibility; Attractive, modern and standout design; A call to action; Design which showcases the content, not the page; Use effects, bright colors; Some empty space placed strategically; Pages that are easily consumable and scan-able; Clear hierarchy of information; Large, easily readable navigation buttons and finally Continuity throughout site design.

On other hand, regarding the academic websites that covers universities and high schools, the use of academic templates is only going to increase as institutional Web sites grow in complexity and importance. Libraries are an important part of institutions both physically—on campus—and virtually—as part of the campus Web site. Academic templates are part of a unified design scheme for colleges and universities. Web site designers to merge the best of the academic template to the best of the library site while not sacrificing users' needs. The result will be highly used, highly usable library Web sites that attract students and keep them coming back to access the fantastic world of information available in today's academic libraries. Also, in order to determine the most beautiful university website among a number of website, we would definitely rate the website best that has a charming appearance, a good way of handling my doubts and manage all of our needs at the same time. The more they fulfill these forms and factors they would be better for us. Same here is the issue with parents when they surf a University Website and they want to know how good or bad is a college or school and how they functionally represent their working. And the better they can represent themselves the better they are considered. There's a saying "what one sees is what one buys." And this completely is true.

REFERENCES

1. "**Web site designing for Educational Institutions**", "http://www.smallbusinessbible.org/websitedesigning_educationalinstitutions.html
2. **"The importance of a business website"**, http://www.helium.com/items/1760241-why-corporate-websites-are-essential
3. **"Guide to Creating Website Information Architecture and Content"** www.princeton.edu/communications/.../IAguide2.pd
4. Tarafder, "**Analysis of Critical Website Characteristics Across Category Study of Successful Websites**" http://www.allbusiness.com/technology/computer-software-customer-relation/1056608-1.html
5. Agarwal, R. and V. Venkatesh. "**Assessing a Firm's Web Presence: A Heuristic Evaluation Procedure for the Measurement of Usability,**" Information Systems Research, 13:2, June 2002, pp. 168-186.
6. Chung, J.F. and Tan. "**Antecedents of Perceived Playfulness: An Exploratory Study on User Acceptance of General Information Searching Websites**," Information and Management, 41, 2004, pp. 869-881.
7. Mary Insabella, "**Designing Web Pages for Educational and Commercial Purposes**", Proceedings of the 2006 ASCUE Conference, June 11-15, 2006, Myrtle Beach, South Carolina
8. **How do Search Engines Differ from one another**?" SearchEngines.com, 2000-2003. http://www.searchengines.com
9. "**Haiku Error Messages**," http://strangeplaces.net/weirdthings/haiku.html
10. "**Error Message Guidelines**," Jakob Nielsen, Jakob Nielsen's Alertbox, June 24, 2001 http://www.useit.com
11. "**Accessibility of State and Local Government Websites to People with Disabilities**," U.S. Department of Justice, Civil Rights Division, June 19, 2003 http://www.usdoj.gov/crt/ada/
12. "**Metadata Activity Statement**," w3c Technology and Society domain, Aug 23, 2002 http://www.w3.org/Metadata/

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

13. "**Web Site Development Information**," kodehost, March 25, 2001 http://goodpractices.com
14. "**The Secrets of Web Site Conversion**," W-edge design, January, 2006 http://www.w-edge.com
15. Kate Peterson, "**Academic Web Site Design and Academic Templates: Where Does the Library Fit In**?", Information Technology and Libraries, December 2006
16. Larry L. Constantine, "**Devilish Details: Best Practices in Web Design**", Constantine & Lockwood, Ltd.
17. Jehad Najjar, "**Business Rules & WSDM Methodology of Web Design**", M.Sc Dissertation, 2002
18. http://www.webworldindex.com/articles/12-Benefits-Of-Having-A-Website.html
19. http://thefuturebuzz.com/2008/03/05/what-defines-a-good-small-business-website/

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Image Processing and Analysis for DTMRI

Dr varaPrasad .s.Kondapalli [1], manike.chiranjeevi [2] **kundan kumar mishra[3]**
And kdbs tanuja4

[1] principal,RGPV,
Indore,m.p,india

2 Associate professor cse,department,RGPV,BM college
Indore,m.p,india

3Associate professor cse,department,RGPV,BM college
Indore,m.p,india

4 final year student of Ntr university,kaminenimedical college
Hyderabad,A.p,india

Abstract.

This paper describes image processing techniques for Diffusion Tensor Magnetic Resonance. In Diffusion Tensor MRI, a tensor describing local water diffusion is acquired for each voxel. The geometric nature of the diffusion tensors can quantitatively characterize the local structure in tissues such as bone, muscles, and white matter of the brain. The close relationship between local image structure and apparent diffusion makes this image modality very interesting for medical image analysis.

We present a decomposition of the diffusion tensor based on its symmetry properties
resulting in useful measures describing the geometry of the diffusion ellipsoid. A simple anisotropy measure follows naturally from this analysis. We describe how the geometry, or shape, of the tensor can be visualized using a coloring scheme based on the derived shape measures. We show how filtering of the tensor data of a human brain can provide a description of macro structural diffusion which can be used for measures of fiber-tract organization. We also describe how tracking of white matter tracts can be implemented using the introducedmethods. These methods offers unique tools for the *in vivo* demonstration of neural connectivity in healthy and diseased brain tissue.

1 Introduction

Diffusion Tensor Magnetic Resonance Imaging (DT-MRI) is a recent MR imaging modality. In Diffusion Tensor MRI, a tensor describing local water diffusion is acquired for each voxel. Diffusion in tissue can be anisotropic depending on the characteristics of the tissue. For example in the white matter fiber tracts the diffusion is mainly in the direction of the fibers. In areas with fluid, such in the CSF filled ventricles, the diffusion is spherical, i.e. isotropic. The advent of robust diffusion tensor imaging techniques has prompted the development of quantitative measures for describing the diffusion anisotropy. A good review by Basser and Pierpaoli can be found in [1].
Since MRI methods in general always obtain a macroscopic measure of a microscopic quantity which necessarily entails intravoxel averaging, the voxel dimensions influence the measured diffusion tensor at any particular location in the brain.
Factors which would affect the shape of the apparent diffusion tensor (shape of the diffusion ellipsoid) in the white matter include the density of fibers, the degree of myelination, the average fiber diameter and the directional similarity of the fibers in the voxel. The geometric nature of the measured diffusion tensor within a voxel is thus a meaningful measure of fiber tract organization.
With current conventional proton magnetic resonance imaging (MRI) techniques, the white matter of the brain appears to be a remarkably homogeneous tissue without any suggestion of the complex arrangement of fiber tracts. Although the individual axons and the surrounding myelin sheaths cannot be revealed with the limited spatial resolution of *in vivo* imaging, distinct bands of white matter fibers with parallel orientation may be distinguished from others running in different directions if MRI techniques are sensitized to water diffusion and the preferred direction of diffusion is determined.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Water diffusion in tissue due to Brownian motion is random but some structural characteristics of tissues may limit diffusion. In the white matter, the mobility of the water is restricted in the directions perpendicular to the axons which are oriented along the fiber tracts. This anisotropic diffusion is due to the presence of tightly packed multiple myelin membranes encompassing the axon. Myelination is not essential for diffusion anisotropy of nerves as shown in studies of nonmyelinated garfish olfactory nerves [3] and anisotropy exists in brains of neonates before the histological appearance of myelin [16] but myelin is widely assumed to be the major barrier to diffusion in myelinated fiber tracts. Therefore the demonstration of anisotropic diffusion in brain by magnetic resonance has opened the way to explore noninvasively the structural anatomy
of the white matter *in vivo* [8, 4, 1, 10].

2 Materials and Methods

In this work we applied a modified version of the recently proposed Line Scan Diffusion Imaging (LSDI) technique [7]. This method, like the commonly used diffusion sensitized,
ultrafast, echo-planar imaging (EPI) technique [12] is relatively insensitive to bulk motion and physiologic pulsations of vascular origin. But unlike EPI, LSDI exhibits minimal image distortion, does not require cardiac gating, head restraints or post-processing image correction, and can be implemented without specialized hardware on all standard MRI scanners.

Here, we present a quantitative characterization of the geometric nature of the diffusion tensors, a method for characterization of marcostructural diffusion properties, and a display method for showing clear and detailed *in vivo* images of human white matter tracts. The orientation and distribution of most of the known major fiber tracts can be identified using these methods.

2.1 Imaging Parameters

Our data were acquired at the Brigham andWomen's Hospital on a GE Signa 1.5 Tesla Horizon Echospeed 5.6 system with standard 2.2 Gauss/cm field gradients. The time required for acquisition of the diffusion tensor data for one slice was 1 min; no averaging was performed. Imaging parameters were: effective TR=2.4 s, TE=65 ms, b_{high}=750 s/mm^2, b_{low}=5 s/mm^2, field of view 22 cm, effective voxel size 4.8_1.6_1.5 mm^2, 6 kHz readout bandwidth, acquisition matrix 128_128.

The gradient cycle in the LSDI interleaving scheme was modified to provide acquisition of more gradient directions and to allow elimination of the crusher gradients. Instead of alternating merely between high and low gradient strengths, the modified sequence cycled through eight configurations of the diffusion gradients. In all other respects it was identical to the sequence described in [7].

2.2 Calculation of Tensors

For each slice, eight images are collected with different diffusion weightings and noncollinear
gradient directions. If S0 represents the signal intensity in the absence of a diffusion-sensitizing field gradient and S the signal intensity in the presence of gradient g = (gx; gy; gz)T , the equation for the loss in signal intensity due to diffusion is given by the Stejskal-Tanner formula:

$$\ln(S)=\ln(S_0)-\gamma^2\delta^2(\Delta-\delta/3)g^T Dg,$$

where γ is the gyromagnetic ratio of hydrogen 1H (protons), δ is the duration of the diffusion sensitizing gradient pulses and Δ is the time between the centers of the two gradient pulses. The eight images provide eight equations for ***S*** in each voxel which are solved in a least-squares sense for the 6+1 unknowns: the six independent components of the symmetric diffusion tensor, D, and S_0. In the LSDI sequence, it is easy to show that cross terms between the slice select gradient for the 180^0 pulse and the diffusion sensitizing gradients account for less than 0.1% of the diffusion weighting, and have therefore been neglected here. Diffusion attenuation due to imaging gradients is already factored into S_0, as is T_2 weighting.

2.3 Geometrical Measures of Diffusion

In order to relate the measure of diffusion anisotropy to the structural geometry of the tissue a mathematical description of diffusion tensors and their quantification is necessary [1]. First, a complete diffusion tensor, D, is calculated (Equation 1) for each voxel. Using the symmetry properties of the diffusion ellipsoid we decomposed the diffusion tensor, and from the tensor basis assigned scalar measures, describing the linearity and the anisotropy, to each voxel [15].

The diffusion tensor can be visualized using an ellipsoid where the principal axes correspond to the directions of the eigenvector system. Let $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$ be the eigenvalues of the symmetric tensor D, and let $\hat{e}_i$ be the normalized eigenvector corresponding to λ_i.

$$D = \lambda_1\hat{e}_1\hat{e}_1^T + \lambda_2\hat{e}_2\hat{e}_2^T + \lambda_3\hat{e}_3\hat{e}_3^T$$

Diffusion can be divided into three basic cases depending on the rank, of the representation tensor:Linear case ($\lambda_1 \gg \lambda_2 \simeq \lambda_3$): diffusion is mainly in the direction cor

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

1) Linear case ($\lambda_1 \gg \lambda_2 \simeq \lambda_3$): diffusion is mainly in the direction corresponding to the largest eigenvalue,

$$D \simeq \lambda_1 D_t = \lambda_1 \hat{e}_1 \hat{e}_1^T$$

2) Planar case ($\lambda_1 \simeq \lambda_2 \gg \lambda_3$): diffusion is restricted to a plane spanned by the two eigenvectors corresponding to the two largest eigenvalues,

$$D \simeq 2\lambda_1 D_p = \lambda_1 (\hat{e}_1 \hat{e}_1^T + \hat{e}_2 \hat{e}_2^T)$$

3) Spherical case ($\lambda_1 \simeq \lambda_2 \simeq \lambda_3$): isotropic diffusion,

$$D \simeq 3\lambda_1 D_8 = \lambda_1 (\hat{e}_1 \hat{e}_1^T + \hat{e}_2 \hat{e}_2^T + \hat{e}_3 \hat{e}_3^T).$$

In general, the diffusion tensor D will be a combination of these cases. Expanding the diffusion tensor using these cases as a basis gives:

$$D = \lambda_1 \hat{e}_1 \hat{e}_1^T + \lambda_2 \hat{e}_2 \hat{e}_2^T + \lambda_3 \hat{e}_3 \hat{e}_3^T$$

$$= (\lambda_1 - \lambda_2)\hat{e}_1 \hat{e}_1^T + (\lambda_2 - \lambda_3)(\hat{e}_1 \hat{e}_1^T + \hat{e}_2 \hat{e}_2^T)$$

$$+ \lambda_3 (\hat{e}_1 \hat{e}_1^T + \hat{e}_2 \hat{e}_2^T + \hat{e}_3 \hat{e}_3^T)$$

$$= (\lambda_1 - \lambda_2) D_t + (\lambda_2 - \lambda_3) D_p + \lambda_3 D_8$$

where $(\lambda_1 - \lambda_2)$, $(\lambda_2 - \lambda_3)$ and λ_3 are the coordinates of D in the tensor basis $\{ D_t, D_p, D_8 \}$.
A similar tensor shape analysis has proven to be useful in a number of computer vision applications [13, 14, 6].
As described, the relationships between the eigenvalues of the diffusion tensor can be used for classification of the diffusion tensor according to geometrically meaningful criteria. By using the coordinates of the tensor in our new basis measures are obtained of how close the diffusion tensor is to the generic cases of line, plane and sphere. The generic shape of a tensor is obtained by normalizing with a magnitude measure of the diffusion. Here we define this magnitude as the largest eigenvalues of the tensor. This gives for the linear, planar and spherical measures:

$$C_l = \frac{(\lambda_1 - \lambda_2)}{\lambda_1}$$
$$C_p = \frac{(\lambda_2 - \lambda_3)}{\lambda_1}$$
$$C_s = \frac{\lambda_3}{\lambda_1}$$
$$C_l + C_p + C_s = 1$$

An anisotropy measure describing the deviation from the spherical case is achieved as follows:

$$C_a + C_l + C_p = 1 - C_s = 1 - \frac{\lambda_3}{\lambda_1}$$

2.4 Relaxation of Data

For many medical image analysis procedures preprocessing of the data in order to reduce the image noise level is important [5]. For diffusion data, regularization based on a Markov model was used to achieve more stable directionality information for tracking white matter fiber tracts [11].
In this paper we argue for staying in the tensor domain when processing simplifies cleaning up of the data. Simple averaging of the tensor components has proven to be a robust relaxation method. One reason for the usefulness of this approach is that tensors have more degrees of freedom compared to scalars and vectors. This is illustrated in Figure 1. Adding two vectors (a) and (b) results in a new vector (c). The data (a) and (b) and the result (c) are of the same order. However, adding two rank 1 tensors (d) and (e), e.g. diffusion tensors from two differently oriented white matter tracts, results in a rank 2 tensor (f), i.e. the output has more degrees of freedom than the input tensors and describes the plane in which diffusion is present. Averaging of tensors are different from averaging a vector field:

- ✓ The average of a set of vectors gives the "mean event"
- ✓ The average of a set of tensors gives the "mean event" and the "range of the present events"

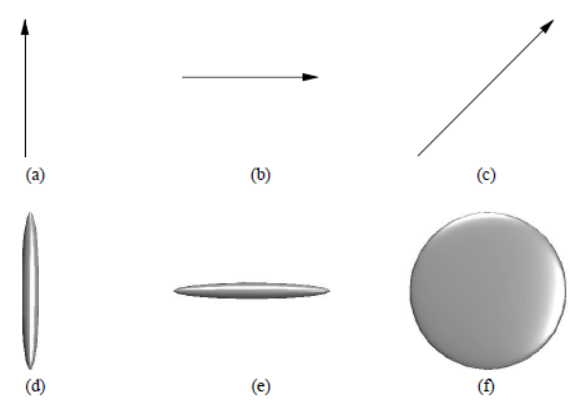


Fig. 1. Vector and tensor summation. Two vectors, (a) and (b), and their sum (c). Two diffusion tensors, (d) and (e), of rank close to 1 visualized as ellipsoids with eigenvectors forming principal axes. The summation of the two tensors gives a rank 2 tensor (f).
Figure 2 shows two stylized 2D examples illustrating the effect of relaxation of a diffusion tensor field. (a) Relaxation of a field that contains inconsistent data (left) gives a result of almost round ellipses (right). In (b) relaxation of a field containing data with a clear bias in

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

one direction resulting in a more stable estimate of the directionality of the field.

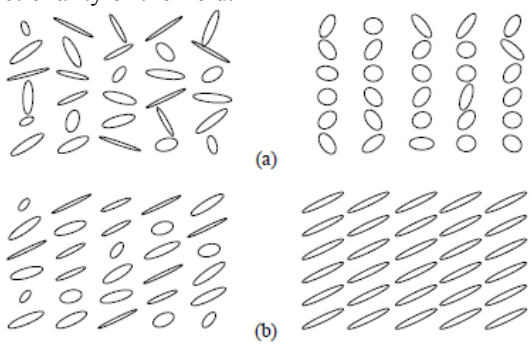


Fig. 2. Stylized 2D examples illustrating the effect of relaxation of a diffusion tensor field. (a)
Relaxation of a field that contains inconsistent data (left) gives a result of almost round ellipses
(right). (b) Relaxation of a field containing data with a clear bias in one direction.

2.5 Macrostructural Diffusive Similarities

In the previous section we characterized diffusion isotropy and anisotropy within a voxel. Here, we will introduce a new method for examining the pattern or distribution of diffusion within an image volume.
Basser and Pierpaoli proposed a scalar measure for macrostructural diffusive similarity based on summing tensor inner products between the center voxel tensor and its neighbors [2]. As in a vector case, the inner product between two tensors measures their degree of similarity.
We will use this idea, and show that averaging the diffusion tensor field has a close relation to this approach. A weighted linear sum of tensor inner products operation over an image volume can be expressed as

$$\sum_k \langle \mathrm{D}(\mathrm{x}), \mathrm{a}(\mathrm{x_k})\mathrm{D}(\mathrm{x_k})\rangle = \langle \mathrm{D}(x_k), \sum_k a(x_k)\mathrm{D}(\mathrm{x_k})\rangle = \langle \mathrm{D}(x), \mathrm{D_a}(\mathrm{x})\rangle$$

where the brackets denote inner product, $a(x_k)$ is a spatial mask defining the local image volume of interest around x, k is an index ordering the voxel in this volume.Da is a weighted average of the tensors in the neighborhood under the mask a, $\sum_k a(x_k)D(x_k)$. Or, in other words, the filtered output from applying the filter $a(x_k)$ to the data $D(x_k)$. Note that averaging of a diffusion tensor field and then deriving a scalar measure from the averaged field is *not the same* as averaging a scalar field derived from the original field.

Equation 11 shows that the scalar macrostructural measure is an inner product between a diffusion tensor with a blurred version of itself. This can be seen as inner products between tensors in a scale pyramid where the tensors with highest spatial resolution, and tensors at a level in the scale pyramid corresponding to the size of interest (scale obtained by blurring with mask a). Figure 4 (rightmost column) shows three level of such a pyramid.

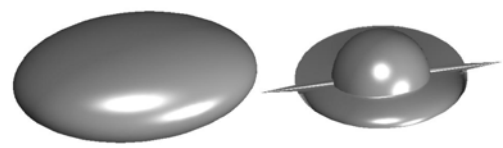


Fig. 3. Comparison of an ellipsoid and a composite shape depicting the same tensor with eigenvalues $\lambda_1 = 1$, $\lambda_2 = 0{:}7$, and $\lambda_3 = 0{:}4$.

The rank of the average tensor D_a describes the complexity of the macroscopic diffusion structure. If the rank is close to one, the structure is highly linear which will be the case in regions of bundles of fibers having the same direction. If the rank is two, fibers are crossing in a plane, or the underlying diffusitivity is planar. Instead of carrying out the inner product in Equation 11, to get a scalar value for similarity, we use the tensor D_a as a description of macrostructural diffusion.

2.6 Visualization of Diffusion Tensors

A 3D diffusion tensor can be visualized using an ellipsoid where the principal axes correspond to the tensor's eigenvector system. However, it is difficult to distinguish between an edge-on, flat ellipsoid and an oblong one using the surface shading information. Similar ambiguity exists between a face-on, flat ellipsoid and a sphere. We propose two techniques for the visualization of tensor fields that overcome the problems with ellipsoids. Figure 3 compares the ellipsoidal representation of a tensor with a composite shape whose linear, planar, and spherical components are scaled according to c_l, c_p, and c_s.
Additionally, coloring based on the shape measures c_l, c_p, and c_s can be used for visualization of shape. Figure 6 shows a coloring scheme where the color is interpolated between the blue linear case, the yellow planar case, and the red spherical case.

2.7 Eigenvalue Transformations

Smoothing a diffusion tensor field removes high frequency noise and as discussed above stabilizes the field. In this section we discuss methods that changes the eigenvalues of the tensors in a way that further

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

can stabilize the data. Let $D = \sum_{k=1}^{3} \lambda_k \hat{e}_k \hat{e}_k^T$. The first method is based on direct manipulation of the eigenvalues λ_k

$$T = \sum_{k=1}^{3} f_k(\lambda_k) \hat{e}_k \hat{e}_k^T$$

where f_k are scalar functions. The function f can for example be a thresholding operator which results in decreased rank of the tensors when the eigenvalues are smaller than the specified threshold. Another method is a "max shape" operator,

$$T = \begin{cases} c_l D_l & \text{if } c_l > c_p, c_s \\ c_p D_p & c_p > c_l, c_s \\ c_s D_s & c_s > c_l, c_p \end{cases}$$

This operation forces the tensor into the closest of the three generic shapes line, plane, and sphere.

2.8 TrackingWhite Matter Tracts

This section gives a sketch of a tracking algorithm that uses the diffusion tensors as operators. Let x_0 be the initial seed point, and v_0 be the seed direction, e.g. the eigenvector corresponding to the largest eigenvalue. A tracking sequence $\{x_0, x_1, \ldots.\}$ can then be obtained by the following iteration formula:

$$x_{k+1} = x_k + \alpha v_k$$
$$v_{k+1} = T(x_{k+1}) v_k$$

where $T(X_{k+1})$ is the tensor in spatial position X_{k+1}. The idea here is based on that V_k will turn towards the largest eigenvector of the tensor $T(X_{k+1})$. When the tensor has lower than full rank, it will act as a projection operator. When rank one, the vector V_k will be projected onto (e1)k+1. Further, when the tensor is spherical, vk will not turn since all vectors are eigenvectors.

3 Results

When applied to white matter, the linear measure, c_l, reflects the uniformity of tract direction within a voxel because it will be high only if the diffusion is restricted in two orthogonal directions. The anisotropy measure, c_a, indicates the relative restriction of the diffusion in the most restricted direction and will emphasize white matter tracts which within a voxel exhibit at least one direction of relatively restricted diffusion. Figure 4 shows the three geometrical measures and the anisotropy index. Major white matter tracts such as the corpus callosum show high linearity. In Figure 5 a diffusion tensor field from an axial slice of the brain is shown (left) and the averaged tensor field (right). The window, a, used was a $5 \times 5 \times 3$ Gaussian window with standard deviation equal to 1, defined in the in-plane resolution of the image. Since the out-of-plane resolution is slightly less than half the in-plane resolution, there is almost no smoothing performed between the slices. The original field (left) and the averaged field (right) have been weighted with their linear diffusion measure, c_l (Equation 8), respectively. This procedure can be used to measure fiber tract organization quantitatively. The size of the filter should be chosen in relation to the size of the fiber tracts of interest.

Figure 7 shows the effect of the two eigenvalue operations described in section 2.7. The left image shows the input data, a simulated tensor field of crossing white matter tracts. Due to partial voluming effects, the tensors in the area where the fibers are crossing have planar shape. The middle image shows the effect of smoothing the field with a 5×5 Gaussian kernel followed by a remapping the eigenvalues when f is a thresholding operator (Equation 12). The threshold was set to 20% of the largest eigenvalue, $0{:}2\lambda_1$. The right image shows the effect of the max shape operator on the same input data.

Figure 8 shows the result of tracking three crossing fiber tracts. A $5 \times 5 \times 5$ Gaussian operator followed by the max shape operator was used to stabilize the data. Each trace line is a composition of the results fromboth seed directions e1 and $-e_1$ joined together.

4 Conclusions

We have proposed measures classifying diffusion tensors into three generic cases based on a tensor basis expansion. When applied to white matter the linear index shows uniformity of tract direction within a voxel while the anisotropic index quantifies the deviation from spatial homogeneity. The non-orthogonal tensor basis chosen is intuitively appealing since it is based on three simple, yet descriptive, geometrically meaningful cases.

We have described how tensor diffusion data can be processed without reverting to the use of only scalar measures of the tensor data. By staying in the tensor domain, cleaning up of the data can be done meaningfully with simple methods such as smoothing. We discuss addition of tensors geometrically and argue that adding tensors and vectors are different in that tensor summation gives more than the "mean" event due to more degrees of freedom. By using the geometric diffusion measures on locally averaged tensors local directionality consistency can be determined (e.g. existence of larger fiber tracts). We have proposed that this averaging approach can be used to derive a tensor field that can be used to describe macrostructural features in the tensor diffusion data. The linear measure c_l derived from the averaged tensor field can for example be used for quantitative evaluation of fiber tract organization.We also have described how non-linear operations can be used to remap the eigenvalues of the diffusion tensors and

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

given a sketch of how this can be used for tracking white matter tracts.

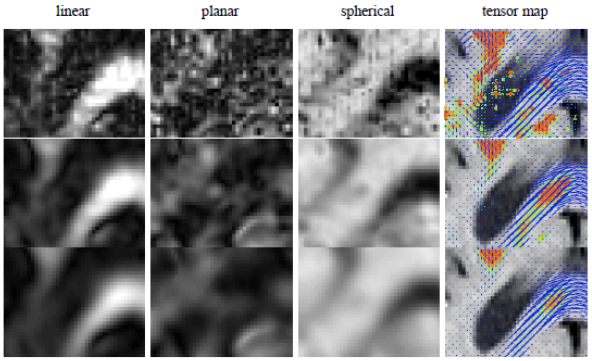


Fig. 4. Axial brain images showing the three geometrical measures and diffusion tensor maps with three different smoothing parameters. **Top:** shows the geometrical measures and the tensor map derived from the original data. **Middle:** shows the same measures derived from data that has been averaged with 9x9x3 Gaussian kernel. **Bottom:** from data averaged with a 15x15x5 Gaussian kernel. The rightmost column shows the tensors. The blue headless arrows represent the in-plane components of $c_l\hat{e}_1$. The out-of-plane components of $c_l\hat{e}_1$ are shown in colors ranging from green through yellow to red, with red indicating the highest value for this component. Display technique from [9]

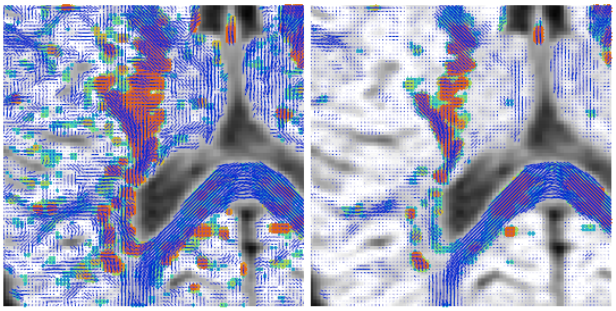


Fig. 5. Left: Diffusion tensors, weighted with their linear measure cl, from an axial slice of a human brain. **Right**: Averaged diffusion tensors using a 5x5x3 Gaussian kernel weighted with their linear measure c_l .

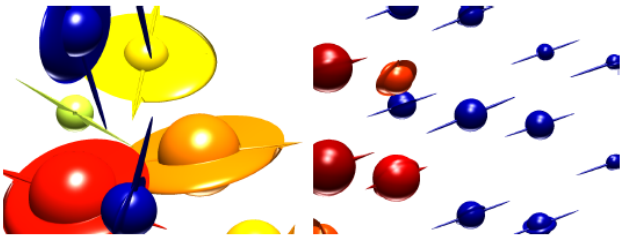


Fig. 6. Visualization of diffusion tensors. The tensors are color coded according to the shape: linear case is blue, planar case is yellow, and spherical case is red. The radius of the sphere is the smallest eigenvalue of the diffusion tensor, the radius of the disk is second largest and the length of the rod is twice the largest eigenvalue.

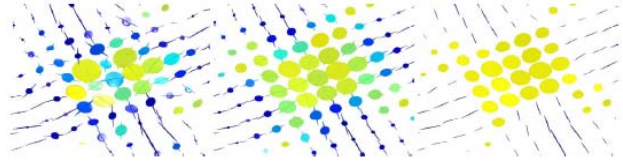


Fig. 7. Left: Simulated diffusion tensor data of two crossing white matter tracts. **Middle** The effect of thresholding the eigenvalues at 20 % of λ_1. **Right**: The effect of the max shape operator.

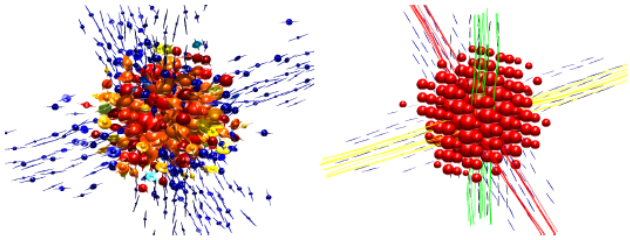


Fig. 8. Left: Original tensor data. **Right:** Tracking fibers after applying the max shape operator. 10 points were randomly seeded at one end of each of the three branches. Note how the stream lines "tunnel" trough the area in the center where the information of directionality is uncertain.

References

1. P.J. Basser. Inferring microstructural features and the physiological state of tissues fromdiffusion-weighted images. *NMR in Biomedicine*, 8:333–344, 1995.
2. P.J. Basser and C. Pierpaoli. Microstructural and physiological features of tissues elucidatedby quantitative-diffusion-tensor MRI. *J. Magn. Reson. Ser. B*, 111:209–219, 1996.
3. C. Beaulieu and P.S. Allen. Determinants of anisotropic water diffusion in nerves. *Magn.Reson. Med.*, 31:394–400, 1994.
4. T. L. Chenevert, J. A Brunberg, and J. G. Pipe. Anisotropic diffusion in human white matter:Demonstration with MR techniques in vivo. *Radiology*, 177:401–405, 1990.
5. G. Gerig, O. K¨ubler, R. Kikinis, and F.A. Jolesz. Non-linear anisotropic filtering of MRIdata. *IEEE Transaction on Medical Imaging*, 11(2):221–232, June 1992.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

6. G. H. Granlund and H. Knutsson. *Signal Processing for Computer Vision*. Kluwer AcademicPublishers, 1995. ISBN 0-7923-9530-1.
7. H. Gudbjartsson, S. E. Maier, R. V. Mulkern, I. A´ . Mo´rocz, S. Patz, and F. A. Jolesz. Linescan diffusion imaging. *Magn. Reson. Med.*, 36:509–519, 1996.
8. M. E. Moseley, Y. Cohen, J. Kucharczyk, J. Mintorovitch, H. S. Asgari, M. F. Wendland,J. Tsuruda, and D. Norman. Diffusion-weighted MR imaging of anisotropic water diffusionin the central nervous system. *Radiology*, 176:439–445, 1990.
9. S. Peled, H. Gudbjartsson, C-F. Westin, R. Kikinis, and F.A. Jolesz. Magnetic ResonanceImaging shows Orientation and Asymmetry of White Matter Tracts. *Brain Research*,780(1):27–33, January 1998.
10. C. Pierpaoli, P. Jezzard, P. J. Basser, A. Barnett, and G. Di Chiro. Diffusion tensor MRimaging of the human brain. *Radiology*, 201:637, 1996.
11. C. Poupon, J.-F. Mangin, F. Frouin, J. R´egis, F. Poupon, M. Pachot-Clouard, D. Le Bihan,and I. Bloch. Regularization of mr diffusion tensor maps for tracking brain white matter bundles.In *Proceedings of MICCAI'98*, number ISSN 0302-9743 in Lecture Notes in Computer Science 1496. Springer Verlag, 1998.
12. R. Turner, D. le Bihan, J.Maier, R. Vavrek, L. K. Hedges, and J. Pekar. Echo planar imagingof intravoxel incoherent motions. *Radiology*, 177:407–414, 1990.
13. C-F. Westin and H. Knutsson. Extraction of local symmetries using tensor field filtering. In*Proceedings of 2nd Singapore International Conference on Image Processing*. IEEE SingaporeSection, September 1992.
14. C-F.Westin and H. Knutsson. Estimation of Motion Vector Fields using Tensor Field Filtering.In *Proceedings of IEEE International Conference on Image Processing*, Austin, Texas,November 1994. IEEE.
15. C-F. Westin, S. Peled, H. Gudbjartsson, R.Kikinis, and F.A Jolesz. Geometrical diffusionmeasures for MRI from tensor basis analysis. In *ISMRM '97*, Vancouver, Canada, April1997.
16. D. M.Wimberger, T. P. Roberts, A. J. Barkovich, L. M. Prayer,M. E.Moseley, and J. Kucharczyk.Identification of "premyelination" by diffusion-weighted MRI. *J. Comp. Assist. Tomogr.*,19(1):28–33, 1995.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Sensitivity Analysis of EADARP Multicast Protocol

Dina Darwish

Mutlimedia and Internet Department, International Academy for Engineering and Media Science
6th October city, Egypt

Abstract

Energy Adaptable Distance Aware Routing Protocol (EADARP) designed for wireless mobile ad hoc networks. EADARP is on-demand mesh-based multicast protocol, depends on choosing the most efficient path in terms of distance and energy and makes some adjustments of nodes batteries power levels when required. It is convenient for networks having high channel capacity mobile hosts, frequent topology changes and constrained power. In this paper, we propose and apply a simulation methodology to perform sensitivity analysis for protocol performance due to changes in network parameters. Eighteen parameters are considered allowing to assess the impact of the changes in such parameters on the performance of the protocol. The results of application are analyzed and evaluated.

Keywords: *Wireless Ad hoc networks, EADARP protocol, Sensitivity analysis, Multicast Network Parameters.*

1. Introduction

Wireless communication use diverse communication techniques, and has two types; fixed or mobile, depending on the network structure. The fixed wireless communication is often called cellular networks, in which communication is achieved through a fixed number of base stations whose locations are known. Mobile wireless communication; also called mobile ad hoc networks[1 -3]; does not have a fixed infrastructure or centralized administration. Each host in the mobile ad hoc network communicates with the other hosts via packet radios to form a temporary network its infrastructure varies according to the hosts' mobility.

Wireless ad hoc networks have several different applications ranging from military applications such as: battlefield communications, law enforcement, disaster recovery, emergency search and rescue and lately to civilian applications such as; electronic classrooms, convention centers, construction sites, and special events.

The objectives of the applications is to determine if a communication session should be unicast (one-to-one), multicast (one-to-many), broadcast (one-to-all) or group communication (many-to-many). The rise in the number of mobile users has led to a wide variety of applications to become available. Some of these new applications depend on multicast communication to perform their operation. Multicasting has been implemented to the wireless ad hoc networks to make benefit from the dynamically reconfigurable nature of these wireless ad hoc networks and to adapt to topological changes in the network.

Multicast ad hoc networks [4 -10] are more complex than cellular wireless networks where all mobiles in a cell can be reached in a single hop, not like multicast mobile ad hoc networks where routes are "multihop". Multicast ad hoc networks are semantically identical to wired networks, but have a number of different characteristics and constraints such as: limited power, limited bandwidth, and high error rates. Multicast ad hoc networks is the focus in this paper. Multicast protocols are divided according to their network structure and formation to three main types; Tree-based multicast protocols, Mesh-based multicast protocols and Overlay multicast protocols.

In this paper, we gave a detailed analysis of the performance of the proposed ad hoc multicast EADARP[13,14] protocol, which is a mesh-based protocol employing the forwarding group concept, when the network parameters mentioned in the simulation environment section are varied, and the results obtained were analyzed.
The remainder of this paper is organized as follows. Section 2 gives a general description of the proposed technique (EADARP). Section 3 describes the simulation environment. In Section 4, the simulation experimental results are investigated and analyzed. Finally, Section 5 gives conclusions for this paper.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

2. EADARP General Description

2.1 The EADARP protocol mesh creation

Group membership and multicast routes are established and updated by the source "on demand", a request phase and a reply phase comprise the protocol as shown in figure 1, while a multicast source has packets to send, it floods a member advertising packet with data payload piggybacked. This packet, called JOIN QUERY, is periodically broadcasted to the entire network to refresh the membership information and update the routes. Once, the JOIN QUERY packet reaches a multicast receiver, the receiver creates and broadcasts a JOIN REPLY to its neighbors. The JOIN REPLY is thus propagated by each forwarding group member until it reaches the multicast source via the route selection method.

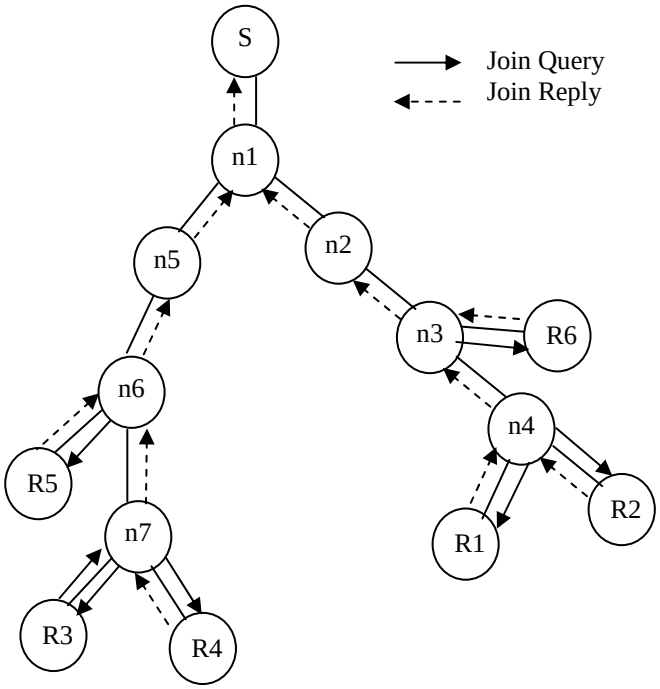


Fig. 1 On-demand procedure for membership and maintenance.

2.2 The EADARP protocol operation

In EADARP route selection, a multicast receiver selects the most stable route having the largest remaining energy, in other words, selecting the route with the highest lifetime. We find the route having the highest lifetime by measuring each route's nodes lifetimes, and choosing the node with the least lifetime in each route, and then selecting the route having the node with the highest lifetime among the least energy remaining nodes.

Then eliminate the nodes having energy below the level required (energy threshold level) in the selected route, for the purpose of avoiding route breakage if these nodes fail. But the eliminated nodes do not include neither the source node nor the destination node, to preserve the original route. After that, some nodes are eliminated between the source node and the destination node to make the selected route shorter, in other terms, reducing the path length leads to decreasing the power consumption during the transmission.

Then, EADARP performs adjustments of nodes batteries power levels when required, and also it increases the network bandwidth when there is a congestion in traffic and decreases it when there is no traffic. To select a route, a multicast receiver must wait for an appropriate amount of time after receiving the first JOIN QUERY so that all possible routes and their lifetimes will be known. The receiver then chooses the most stable route and broadcasts a JOIN REPLY.

3. Simulation Environment

We are going to describe the simulation environment in which we simulated the multicast protocols, and compared their results to conclude the properties of each protocol.

3.1 Simulation model Description

For evaluating EADARP and other multicast routing protocols, a simulation environment was implemented within the GloMoSim library [11]. The GloMoSim library is a scalable simulation environment for wireless network systems using the parallel discrete-event simulation capability provided by PARSEC [12]. The simulation is based on modeling a network of 30 mobile hosts placed uniformly within a 1000 m×1000 m area. Radio transmission power of 15 dBm was used for each node during experiments. The used channel capacity is 10 Mbps when comparing the protocols with each others. Each simulation between any 2 nodes is executed for 100 sec of simulation time. The network traffic load used was 1500 packets/sec.

3.2 Traffic pattern

To study the impact of data traffic load on multicast protocols, the traffic loads were used varied on the network. The network traffic load is

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

1500 packets/sec. A traffic generator was developed to simulate constant bit rate (CBR) sources. The size of the data payload (packet) was 2048 bytes when comparing the protocols with each others. We have 3 multicast source nodes, and each source node transmits to 3 receiver nodes, then, we have 9 receivers during the simulation.

3.3 Mobility Model

For simulation, the random-waypoint model was used for mobility, so there are some packets dropped due to nodes' mobility which stands at 10 m/sec in maximum, unlike the cases when no mobility is used, the packet drops are only caused by buffer overflow, collision and congestion. With mobility random-waypoint pause of 30 sec, and the mobility random-waypoint minimum speed stands at 0 m/sec; the nodes' placement during their mobility is uniform.

3.4 Network parameters used during the simulation

The constant network parameters used during the simulation are:

Simulation-Time: 150 seconds
Seed: 1
Terrain-Dimensions: 1000 x 1000 meters
Number-of-Nodes: 30
Node-Placement: Uniform
Mobility: RANDOM_WAYPOINT
Mobility-WP-Pause: 30 seconds
Mobility-WP-Min-Speed: 0 m/sec
Mobility-WP-Max-Speed: 10 m/sec
Mobility-Position-Granularity: 0.5
Propagation-Limit: -95 dBm
Propagation-Pathloss: FREE-SPACE
Noise-Figure: 10
Temperature: 290 Kelvin
Radio-Type: RADIO-ACCNOISE
Radio-Frequency: 2.4e9 hertz
Radio-Bandwidth: 10000000 bits per second
Radio-RX-Type: SNR-BOUNDED
Radio-RX-SNR-Threshold: 10 dB
Radio-TX-Power: 15 dBm
Radio-Antenna-Gain: 2.0 dB
Radio-RX-Sensitivity: -51 dBm
Radio-RX-Threshold: -81 dBm
Mac-Protocol: 802.11
Mac-Propagation-Delay: 1000 Nanoseconds
Promiscuous-Mode: NO
Network-Protocol: IP
Network-Output-Queue-Size-Per-Priority: 100
Red-Min-Queue-Threshold: 150
Red-Max-Queue-Threshold: 200
Red-Max-Marking-Probability: 0.1
Red-Queue-Weight: 0.0001
Red-Typical-Packet-Transmission-Time: 64000 Nanoseconds
App-Jitter: 100 milliseconds

The network parameters that were varied during consequent simulations are:

(1) Number of Nodes
(2) Terrain Dimensions
(3) Radio Transmission Power
(4) Radio Frequency
(5) Temperature
(6) MAC Protocols
(7) Mobility Models
(8) Propagation Models
(9) MAC Propagation Delay
(10) Node Placement
(11) Noise Figure
(12) Radio Antenna Gain
(13) Radio Receiver Type
(14) Radio Receiver SNR Threshold
(15) Radio Type
(16) Propagation Limit
(17) Radio Receiver Sensitivity
(18) Radio Receiver Threshold

3.5 Performance metrics

3.5.1 Error Loss Percentage

Is measured as the ratio between the data packets lost during transmission between the source node and the receiver node and the total number of data packets sent by the source node.

3.5.2 Average Number of Collisions

Is measured as the total number of collisions that occurred during the simulation over the total number of nodes.

3.5.3 Average Number of Control Packets

Is measured as the total number of routing control packets transmitted during the simulation time over the total number of nodes.

3.5.4 Average Power Consumption

Is the average power consumed by the system, and is measured as the sum of the power consumed at each node at the radio channel during the simulation time over the total number of nodes.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

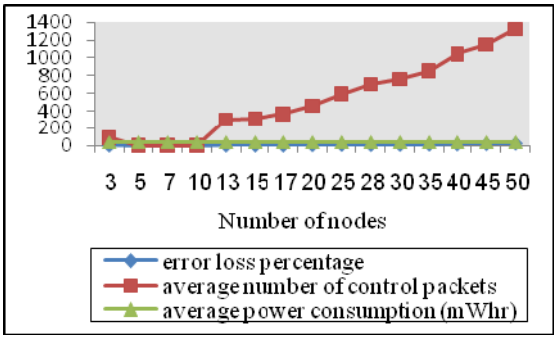
4. Simulation Experimental Results

We simulated the proposed mesh-based EADARP protocol, and the protocol was tested under some conditions, and the results obtained were compared and evaluated to see its performance.

4.1 Performance of EADARP protocol

4.1.1 Changing Number of Nodes

As it appears from figures 2, 3, the error loss percentage varies when changing the number of nodes. The error loss percentage increases when increasing the number of nodes using the EADARP protocol, when the number of nodes is 20 till 50, the error loss percentage of the EADARP protocol is higher in this range. And, when the number of nodes is 3 to 10, the EADARP has lower error loss percentage. Also, the average number of collisions varies when changing the number of nodes. The average number of collisions increases when the number of nodes increases for the EADARP protocol. And, the average number of collisions of the EADARP protocol is at its highest at 50 nodes. Also, the average number of control packets varies when changing the number of nodes. The average number of control packets increases gradually, not in constant rate, when increasing the number of nodes using the EADARP protocol. Also, the average power consumption varies a little bit when changing the number of nodes. And the average power consumption of the EADARP protocol, measured in mWhr, is at its lowest between 5 and 10 nodes. And these results stem from the nature of the protocol, and that the number of nodes affects the protocol operation, hence, has an effect on the performance metrics; error loss percentage, average number of collisions, average number of control packets, average power consumption.



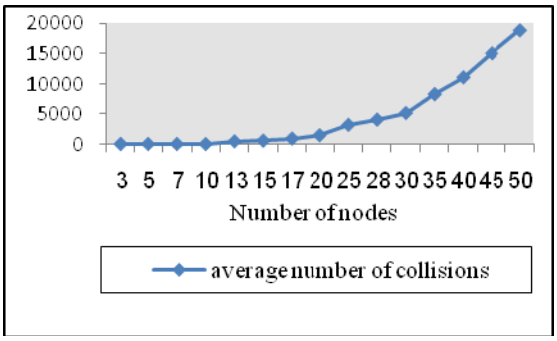
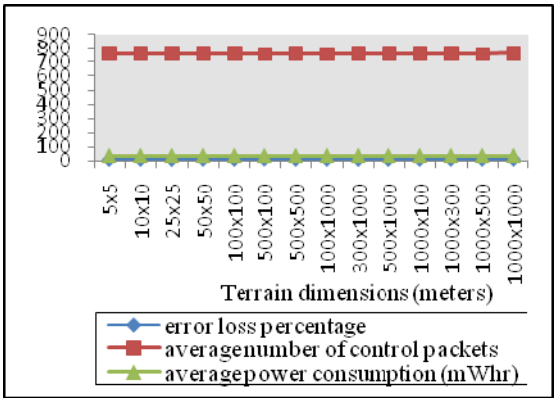


Fig. 2,3 Effect of number of nodes.

4.1.2 Changing Terrain Dimensions

As it appears from figures 4, 5, the error loss percentage does not vary too much when increasing the Terrain Dimensions from 5x5 meters to 1000x1000 meters. Also, the average number of collisions varies with a little margin of values when changing the Terrain Dimensions. Also, the average number of collisions increases gradually, not in constant rate, with a little margin of values, in the case of the EADARP protocol. Also, the average number of control packets does not vary when increasing the Terrain Dimensions from 5x5 meters to 1000x1000 meters. Also, the average power consumption does not vary when increasing the Terrain Dimensions from 5x5 meters to 1000x1000 meters. And these results stem from the nature of the protocol, and that the Terrain Dimensions affects a little bit the protocol operation, hence, has a little effect on the performance metrics; error loss performance, average number of collisions, average number of control packets, average power consumption.



Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

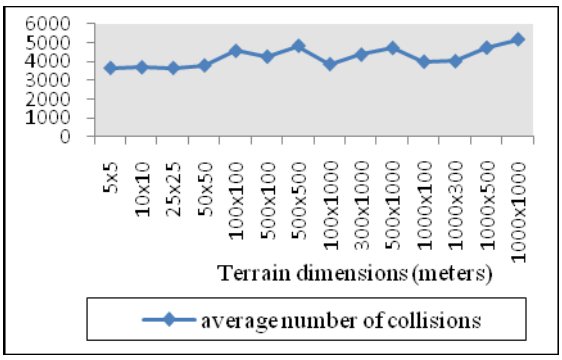
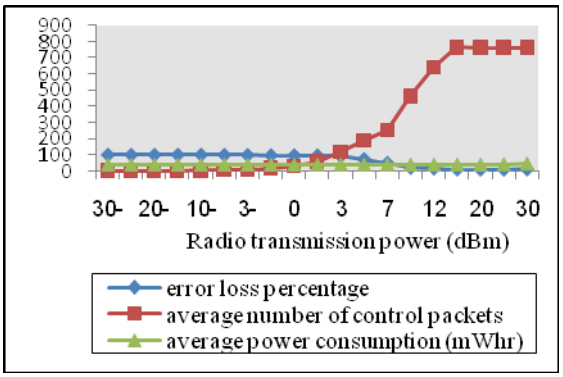


Fig. 4, 5 Effect of Terrain Dimensions.

4.1.3 Changing Radio-TX-Power

As it appears from figures 6, 7, the error loss percentage decreases when increasing the Radio-Tx-Power, measured in dBm. The protocol makes an error loss percentage around 100 % when the Radio-Tx-Power is between -30 dBm and 0 dBm, and, the error loss percentage continues to decrease until the Radio-Tx-Power reaches 30 dBm. The EADARP protocol has an error loss percentage standing between 97% to 11%, when the Radio-Tx-Power is in the range of 1 dBm to 30 dBm, during changing the values of the Radio-Tx-Power. Also, the average number of collisions increases gradually hen increasing the Radio-Tx-Power, then decreases again. The graph below shows that the protocol has a very little constant average number of collisions between -30 dBm and 0 dBm, the average number of collisions continues to increase from 1 dBm till 12 dBm, then, decreases from 12 dBm to 30 dBm. Also, the average number of control packets increases when increasing the Radio-Tx-Power, measured in dBm, using the EADARP protocol, when the range of a Radio-Tx-Power is from 1 dBm to 30 dBm. Also, the average power consumption, measured in mWhr, does not vary too much when increasing the Radio-Tx-Power, in other words, the average power consumption of the protocol, stays constant between -30 dBm and 20 dBm, standing around 38 mWhr, and increases a little bit between 20 dBm and 30 dBm. And these results stem from the nature of the protocol, and that the Radio-Tx-Power affects the protocol operation, hence, has an effect on the performance metrics; error loss percentage, average number of collisions, average number of control packets, average power consumption.



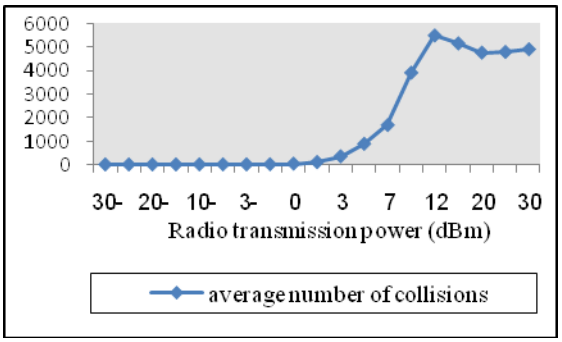


Fig. 6, 7 Effect of Radio-Tx-Power.

4.1.4 Changing Radio Frequency

As it appears from figures 8, 9, the error loss percentage does not vary too much when increasing the Radio Frequency, measured in hertz, in other words, the error loss percentage stays constant from radio frequency -6e9 hertz to radio frequency 9e6 hertz, and increases when the radio frequency becomes 2.4 e9 till the end of the graph. Also, the average number of collisions does not vary too much when increasing the Radio Frequency, measured in hertz, in other words, the average number of collisions stays constant from -6e9 hertz frequency till 5e3 hertz frequency, and then, begin to vary down then up gradually till the end of the graph, using the EADARP protocol. Also, the average number of control packets does not vary too much when increasing the Radio Frequency, measured in hertz. Since the Radio Frequency stays the same from the beginning of the graph till 3e9 hertz, then it decreases gradually till the end of the graph. Also, the average power consumption does not vary when increasing the Radio Frequency, measured in hertz, the Radio Frequency, in other words, when using the EADARP protocol, it stays constant from the beginning of the graph from -6e9 hertz radio frequency to 3e9 hertz, then, it decreases till reaching the end of the graph. And these results

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

stem from the nature of the protocol, and that the Radio Frequency does not affect too much the protocol operation, hence, has a moderate effect on the performance metrics; error loss percentage, average number of collisions, average number of control packets, average power consumption.

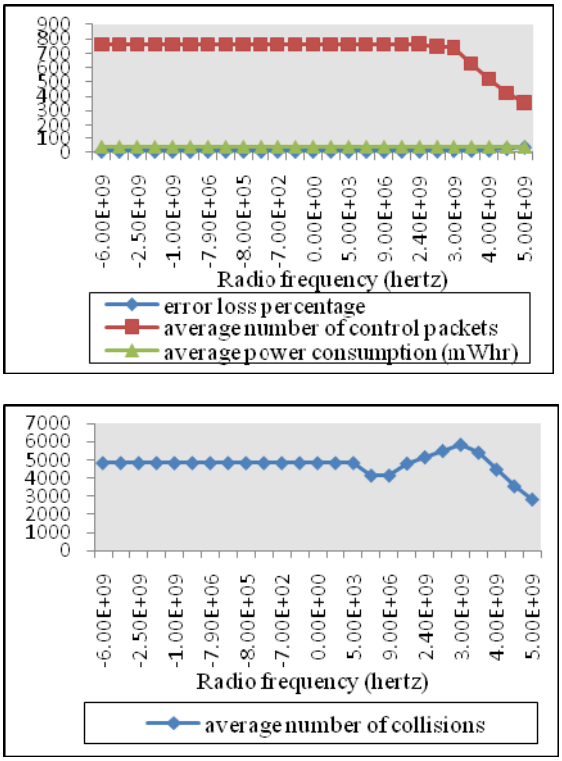


Fig. 8,9 Effect of Radio Frequency.

4.1.5 Changing Temperature

As it appears from figures 10, 11, the error loss percentage does not vary when increasing the values of the Temperature from 0 Kelvin to above 500 Kelvin. The error loss percentage of the EADARP protocol is standing around 11%. Also, the average number of collisions does not vary when increasing the values of the Temperature from 0 Kelvin to above 500 Kelvin. The average number of collisions of the EADARP protocol is standing at 5100. Also, the average number of control packets does not vary when increasing the values of the Temperature from 0 Kelvin to above 500 Kelvin. The average number of control packets of the EADARP protocol is standing at 790. Also, the average power consumption does not vary when increasing the values of the Temperature from 0 Kelvin to above 500 Kelvin. And these results stem from the nature of the protocol, and that the Temperature does not affect the protocol operation, hence, has no effect on the performance metrics; error loss percentage, average number of collisions, average number of control packets, average power consumption.

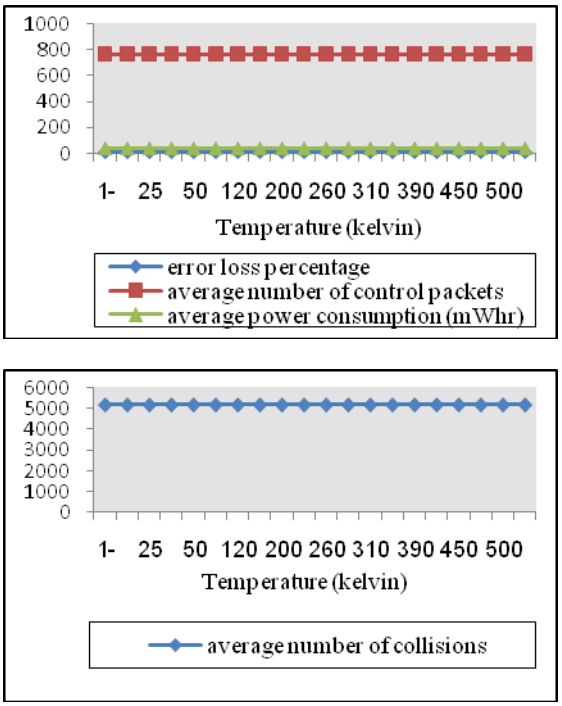


Fig. 10, 11 Effect of Temperature.

4.1.6 Changing MAC Protocols

As it appears from 12, 13, the error loss percentage varies a little bit when changing the MAC protocols. Also, the average number of collisions varies down and up consecutively when changing the MAC protocols. Also, the average number of control packets varies down and up consecutively when changing the MAC protocols. Also, the average power consumption does not vary too much when changing the MAC protocols.

From the graphs, we notice that the MACA protocol with or without promiscuous mode provides the lowest average number of collisions and the lowest average number of control packets, but keeps at the same time, the same error loss percentage and the same average power consumption, compared to other protocols; 802.11 and CSMA protocols. And these results stem from the nature of the protocol, and that the MAC protocol affects the protocol operation, hence, has a moderate effect on the performance metrics; error loss percentage, average number of collisions, average number of control packets, average power consumption.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

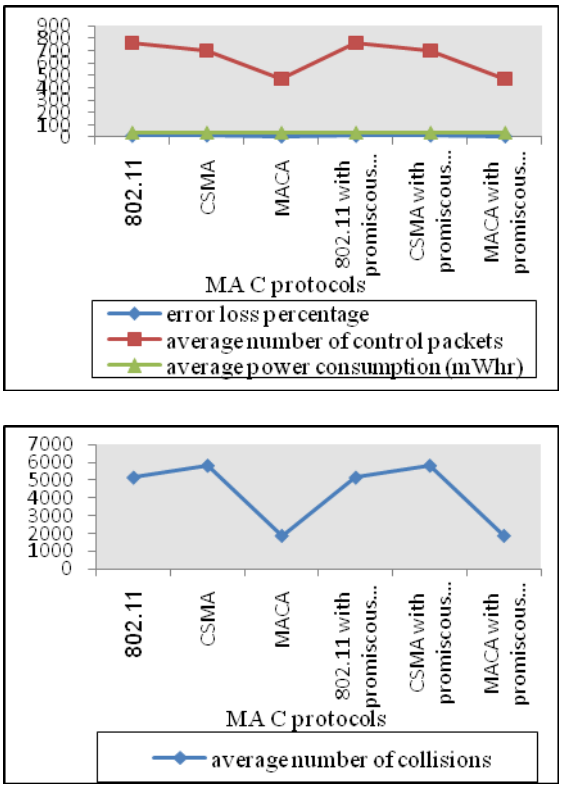


Fig. 12, 13 Effect of MAC Protocols.

4.1.7 Changing Mobility Models

As it appears from figures 14, 15, the error loss percentage does not vary too much when changing mobility models from NONE to RANDOM_WAYPOINT, and stays at 11% for both mobility models. Also, the average number of collisions decreases from 5800 to 5100 when changing the mobility model from NONE to RANDOM_WAYPOINT. Also, the average number of control packets increases from 650 to 800 when changing the mobility model from NONE to RANDOM_WAYPOINT. Also, the average power consumption, measured in mWhr, does not vary too much when changing the mobility model from NONE to RANDOM_WAYPOINT, except with a very small parts of fractions. And these results stem from the nature of the protocol, and that the mobility model does not affect too much the protocol operation, hence, has no big effect on the performance metrics; error loss percentage, average number of collisions, average number of control packets, average power consumption.

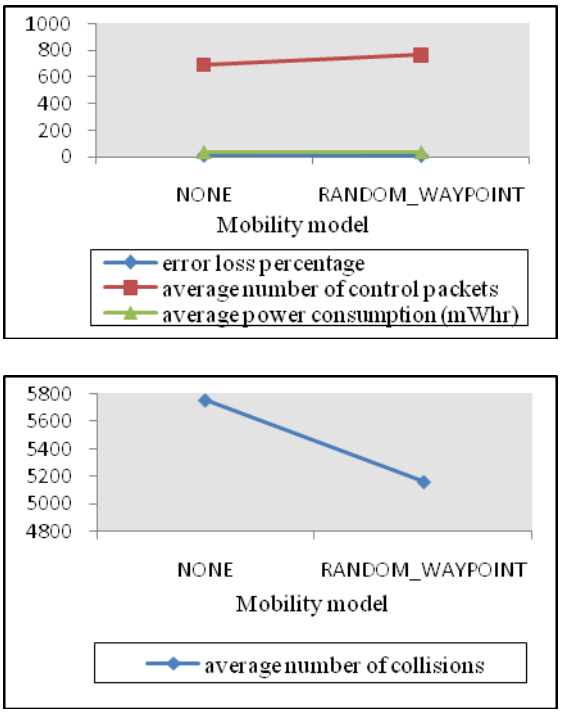
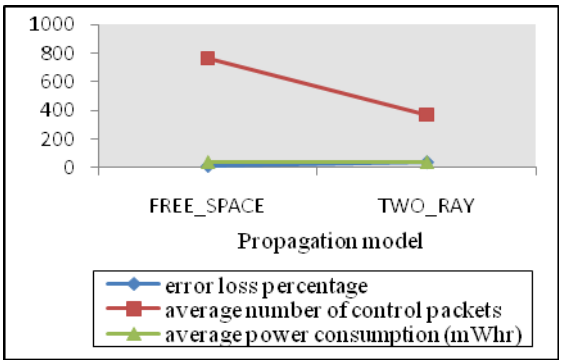


Fig. 14, 15 Effect of Mobility Models.

4.1.8 Changing Propagation Models

As it appears from figures 16,17, the error loss percentage varies a little bit when changing the propagation model from FREE_SPACE to TWO_RAY. Also, the average number of collisions decreases from 5100 to 2200 when changing the propagation model from FREE_SPACE to TWO_RAY. Also, the average number of control packets decreases from 790 to 400 when changing the propagation model from FREE_SPACE to TWO_RAY. Also, the average power consumption, measured in mWhr, does not vary too much when changing the propagation model from FREE_SPACE to TWO_RAY. And these results stem from the nature of the protocol, and that the propagation model affects the protocols operation, hence, has an effect on the performance metrics; error loss percentage, average number of collisions, average number of control packets, average power consumption.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.



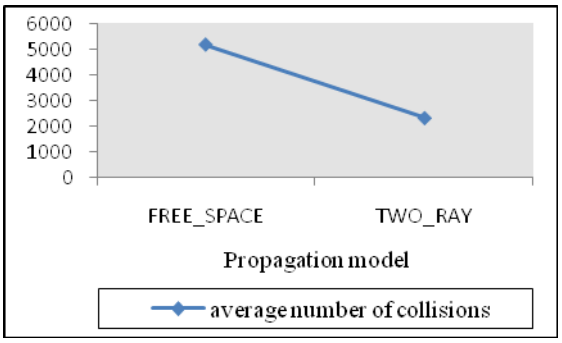
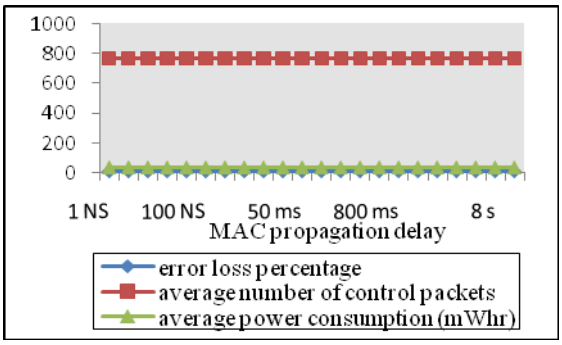


Fig. 16, 17 Effect of Propagation Models.

4.1.9 Changing MAC Propagation Delay

As it appears from figures 18, 19, the error loss percentage does not vary when changing the MAC propagation delay from 1 Nanoseconds to 8 seconds. The EADARP protocol error loss percentage is at 11% along the graph. Also, the average number of collisions does not vary when changing the MAC propagation delay. The average number of collisions of the EADARP protocol is at 5100 along the graph. Also, the average number of control packets does not vary when changing the MAC propagation delay from 1 Nanoseconds to 8 seconds. The average number of control packets of the EADARP protocol is 790 along the graph. Also, the average power consumption, measured in mWhr, does not vary when changing the MAC propagation delay from 1 Nanoseconds to 8 seconds. The average power consumption of the EADARP protocol is 37.7 mWhr along the graph.
And these results stem from the nature of the protocol, and that the MAC propagation delay does not affect the protocol operation, hence, has no effect on the performance metrics; error loss percentage, average number of collisions, average number of control packets, average power consumption.



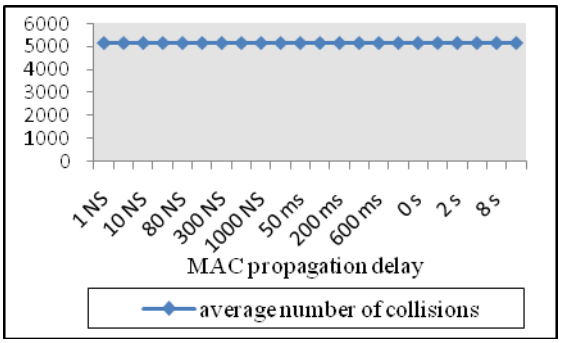


Fig. 18, 19 Effect of MAC Propagation Delay.

4.1.10 Changing Node Placement

As it appears from figures 20, 21, the error loss percentage does not vary when changing the Node placement method from: Random with NONE Mobility, Random with RANDOM_WAYPOINT, Uniform with NONE Mobility, Uniform with RANDOM_WAYPOINT. Also, the average number of collisions varies when changing the Node placement method from: Random with NONE Mobility, Random with RANDOM_WAYPOINT, Uniform with NONE Mobility, Uniform with RANDOM_WAYPOINT. The EADARP protocol shows higher average number of collisions at node placement methods: Random with NONE Mobility and Uniform with NONE Mobility. Also, the average number of control packets varies when changing the Node placement method from: Random with NONE Mobility, Random with RANDOM_WAYPOINT, Uniform with NONE Mobility, Uniform with RANDOM_WAYPOINT. The average power consumption of the EADARP protocol is a little lower using the Random with NONE Mobility and the Uniform with NONE Mobility node placement methods. Also, the average power consumption, measured in mWhr, does not vary when changing the Node placement method from: Random with NONE Mobility, Random with RANDOM_WAYPOINT, Uniform with NONE

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Mobility, Uniform with RANDOM_WAYPOINT. The average power consumed of the EADARP protocol is standing at 37.7 mWhr. And these results stem from the nature of the protocol, and that the Node placement method does not affect too much the protocol operation, hence, has a moderate effect on the performance metrics; error loss percentage, average number of collisions, average number of control packets, average power consumption.

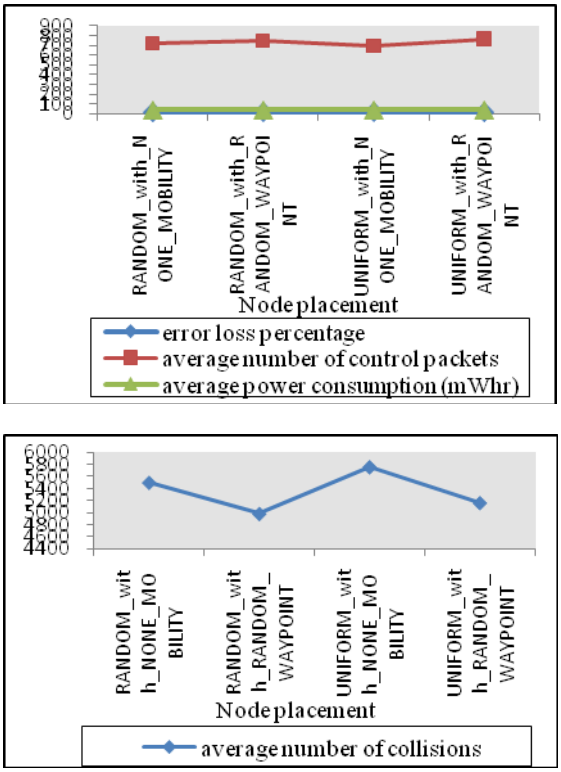


Fig. 20, 21 Effect of Node Placement.

4.1.11 Changing Noise Figure

As it appears from figures 22, 23, the error loss percentage does not vary too much when increasing the Noise Figure from -200 to 20, in other words, the EADARP protocol has a changeable error loss percentage with a little margin of values. The error loss percentage of the EADARP protocol is varying between 10.5% and 12%. Also, the average number of collisions varies when increasing the Noise Figure from -200 to 20 along the graph. The average number of collisions of the EADARP protocol increases and decreases consecutively till reaching the end of the graph. Also, the average number of control packets does not vary when increasing the Noise Figure from -200 to 20. The number of control packets of the EADARP protocol is 790 along the graph. Also, the average power consumption does not vary when increasing the Noise Figure from -200 to 20. The average power consumed of the EADARP protocol is at 37.7 mWhr. And these results stem from the nature of the protocol, and that the Noise Figure does not affect too much the protocol operation, hence, has a small effect in general on the performance metrics; error loss percentage, average number of collisions, average number of control packets, average power consumption.

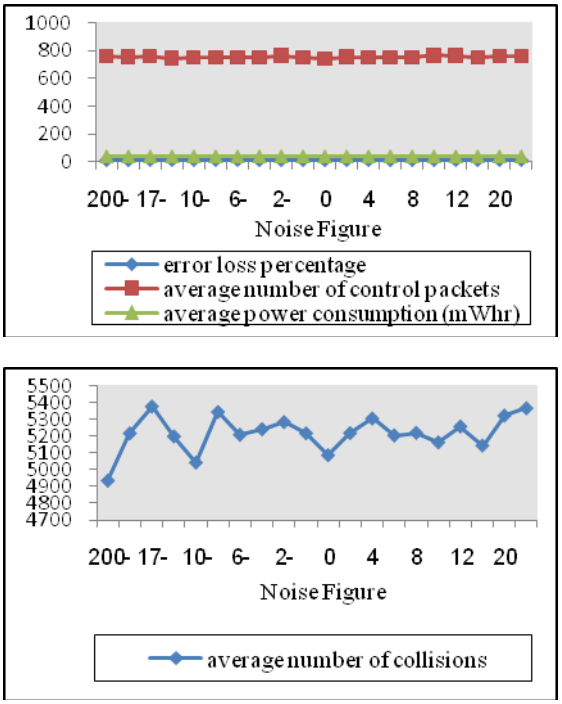


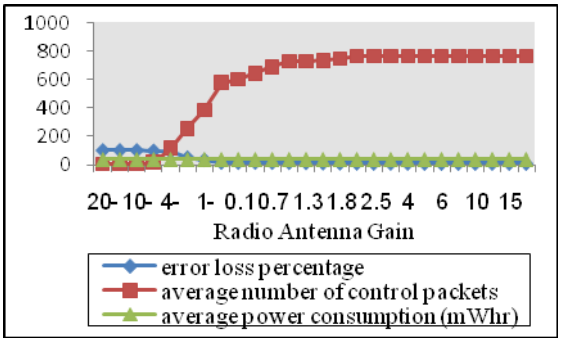
Fig. 22, 23 Effect of Noise Figure.

4.1.12 Changing Radio Antenna Gain

As it appears from figures 24,25, the error loss percentage of the EADARP protocol decreases from 100% to below 12% when increasing the Radio Antenna Gain from -20 db to 0.1 db and stays constant at 11% till reaching the end of the graph. Also, the average number of collisions increases at a Radio Antenna Gain of -4 db, then decreases at a Radio Antenna Gain of 1.8 db before reaching the end of the graph when increasing the Radio Antenna Gain from -20 db to 15 db. Also, the average number of control packets varies when increasing the Radio Antenna Gain from -20 db to 15 db, in other words, the Radio Antenna Gain of the EADARP protocol increases from a Radio Antenna Gain of -4 db to 1.8 db, and stays

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

constant from a Radio Antenna Gain of 2.5 db till the end of the graph. Also, the average power consumption, measured in mWhr, does not vary when increasing the Radio Antenna Gain from -20 db to 15 db. And these results stem from the nature of the protocol, and that the Radio Antenna Gain affects the protocol operation, hence, has an effect on the performance metrics; error loss percentage, average number of collisions, average number of control packets, average power consumption.



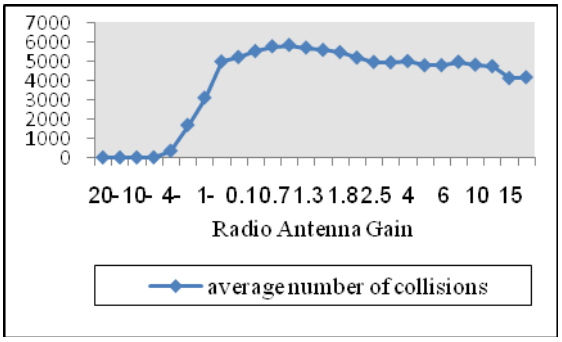
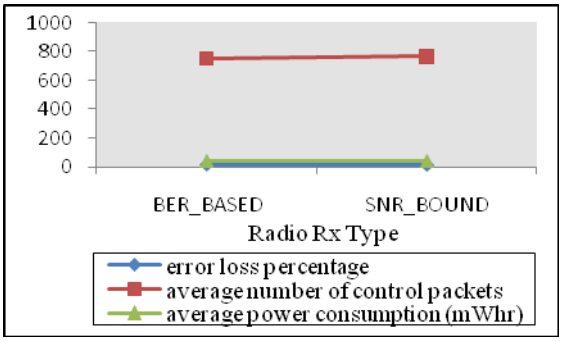


Fig. 24, 25 Effect of Radio Antenna Gain.

4.1.13 Changing Radio-RX-Type

As it appears from figures 26, 27, the error loss percentage decreases a little bit when changing the Radio-RX-Type from BER_BASED to SNR_BOUND using the EADARP protocol. The error loss percentage of the EADARP protocol is ranging from 14% to 11% consecutively. Also, the average number of collisions increases when changing the Radio-RX-Type from BER_BASED to SNR_BOUND using the EADARP protocol. When using the BER_BASED Radio-RX-Type, the EADARP protocol has an average number of collisions equal to zero. When using the SNR_BOUND Radio-RX-Type, the average number of collisions of the EADARP protocol is 5300. Also, the average number of control packets does not vary when changing the Radio-RX-Type from BER_BASED to SNR_BOUND. The average number of control packets of the EADARP protocol stands at 800 for both cases. Also, the average power consumption, measured in mWhr, does not vary when changing the Radio-RX-Type from BER_BASED to SNR_BOUND. The average power consumed of the EADARP protocol is standing at 37.7 mWhr. And these results stem from the nature of the protocol, and that the Radio-RX-Type affects a little the protocol operation, hence, has a small effect on the performance metrics; error loss percentage, average number of collisions, average number of control packets, average power consumption.



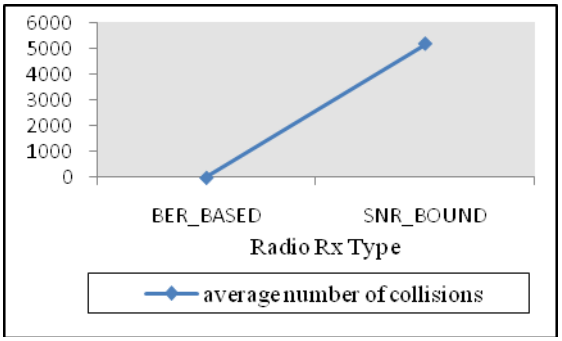


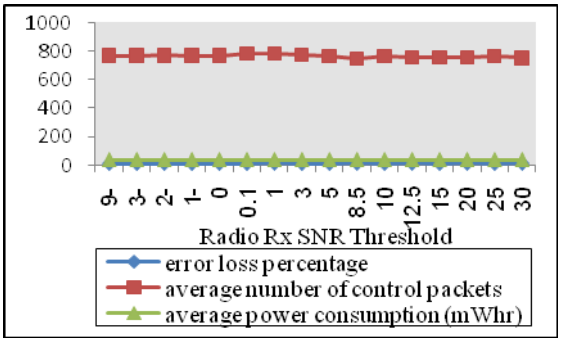
Fig. 26, 27 Effect of Radio-RX-Type.

4.1.14 Changing Radio-RX-SNR-Threshold

As it appears from figures 28, 29, the error loss percentage varies with small fractions when increasing the Radio-RX-SNR-Threshold from -9 dB to 30 dB. Also, the average number of collisions increases when increasing the Radio-RX-SNR-Threshold from -9 dB to 30 dB. When the Radio-RX-SNR-Threshold is from -9 dB to 0.1 dB, the average number of collisions is equal to 1000. When the Radio-RX-SNR-Threshold is from 0.1 dB till 30 dB, the average number of collisions increases gradually till the end of the graph. Also, the average number of control packets does not vary when increasing the Radio-RX-SNR-Threshold from -9 dB to 30 dB, and is staying at

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

790. Also, the average power consumption, measured in mWhr, does not vary when increasing the Radio-RX-SNR-Threshold from -9 dB to 30 dB. And these results stem from the nature of the protocol, and that the Radio-RX-SNR-Threshold affects a little the protocol operation, hence, has a small effect on the performance metrics; error loss percentage, average number of collisions, average number of control packets, average power consumption.



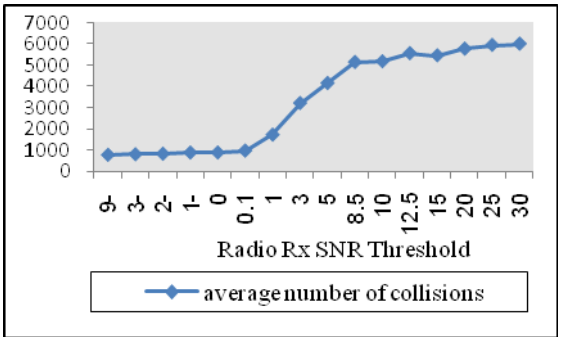
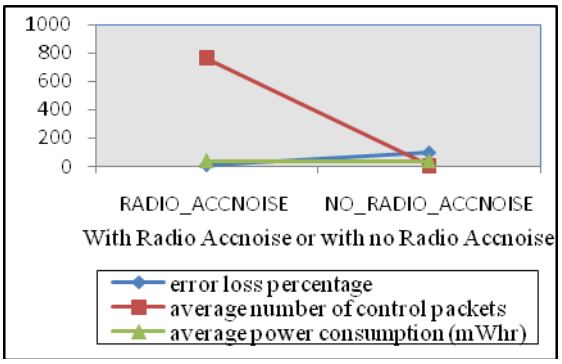


Figure 28, 29: Effect of Radio-RX-SNR-Threshold.

4.1.15 Changing Radio Type

As it appears from figures 30, 31, the error loss percentages of the EADARP protocol increases from below 12% to 100%, when changing the Radio Type from RADIO_ACCNOISE to NO_RADIO_ACCNOISE. Using the NO_RADIO_ACCNOISE Radio Type, the EADARP protocol has an error loss percentage of 100%. When using the RADIO_ACCNOISE Radio Type, the error loss percentage of the EADARP protocol, is standing at 11%. Also, the average number of collisions of the EADARP protocol decreases to 0 when changing the Radio Type from RADIO_ACCNOISE to NO_RADIO_ACCNOISE. Using the NO_RADIO_ACCNOISE Radio Type, the EADARP protocol has an average number of collisions of 0. When using the RADIO_ACCNOISE Radio Type, the average number of collisions of the EADARP protocol is 5000. Also, the average number of control packets of the EADARP protocol varies when changing the Radio Type from NO_RADIO_ACCNOISE to RADIO_ACCNOISE. Using the NO_RADIO_ACCNOISE Radio Type, the EADARP protocol has an average number of control packets equal to 0. Using the RADIO_ACCNOISE Radio Type, the average number of control packets of the EADARP protocol is 800. Also, the average power consumption of the EADARP protocol does not vary when changing the Radio Type from RADIO_ACCNOISE to NO_RADIO_ACCNOISE. Using the NO_RADIO_ACCNOISE Radio Type, the EADARP protocol has an average power consumption, equal to 37.5 mWhr. When the RADIO_ACCNOISE Radio Type is used, the average power consumption of the EADARP protocol is the same. And these results stem from the nature of the protocol, and that the Radio Type affects the protocol operation, hence, has an effect on the performance metrics; error loss percentage, average number of collisions, average number of control packets, average power consumption.



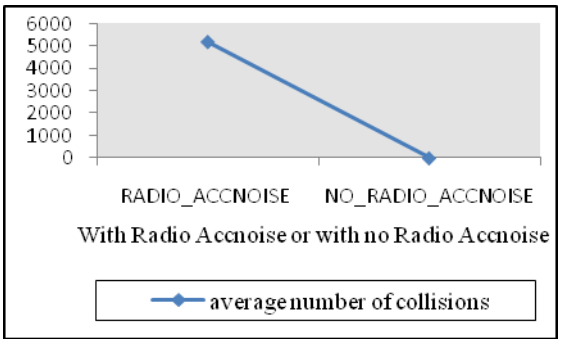
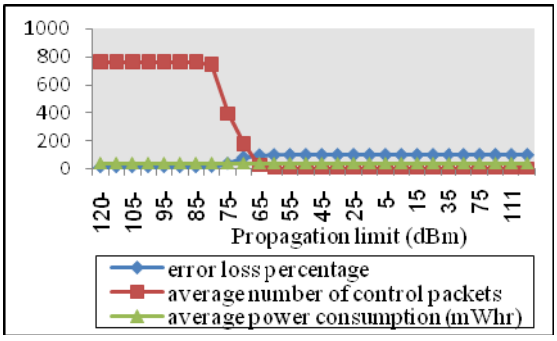


Fig. 30, 31 Effect of Radio Type.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

4.1.16 Changing the Propagation Limit

As it appears from figures 32, 33, the error loss percentage increases when increasing the Propagation Limit from -120 dBm to 111 dBm. When the Propagation Limit is between -120 dBm and -80 dBm, the error loss percentage of the EADARP protocol is standing at 11%. When the Propagation Limit is between -80 dBm to -65 dBm, the EADARP protocol increases its error loss percentages to 100%, and from -65 dBm till reaching the end of the graph, the protocol keeps its error loss percentages at 100%. Also, the average number of collisions decreases when increasing the Propagation Limit from -120 dBm to 111 dBm. When the Propagation Limit is between -120 dBm to -80 dBm, the average number of collisions of the EADARP protocol is around 5000. When the Propagation Limit is between -80 dBm and -65 dBm, the average number of collisions decreases from around 5000 to 0. The EADARP protocol has an average number of collisions equal to 0, when the Propagation Limit is -65 dBm till the end of the graph. Also, the average number of control packets of the EADARP protocol is constant and equal to 0, when increasing the Propagation Limit from -65 dBm to 111 dBm. Taking the range when the Propagation Limit is between -120 dBm to -80 dBm, the EADARP protocol has an average number of control packets equal to 800. When the Propagation Limit is between -80 dBm and -65 dBm, the average number of control packets decreases from 800 to 0. Also, the average power consumption, measured in mWhr, of the EADARP protocol decreases from a higher power to a power of 37.5 mWhr between a Propagation Limit of -65 dBm till reaching the end of the graph, when increasing the Propagation Limit from -120 dBm to 111 dBm. And these results stem from the nature of the protocol, and that the Propagation Limit affects the protocol operation, hence, has an effect on the performance metrics; error loss percentage, average number of collisions, average number of control packets, average power consumption.



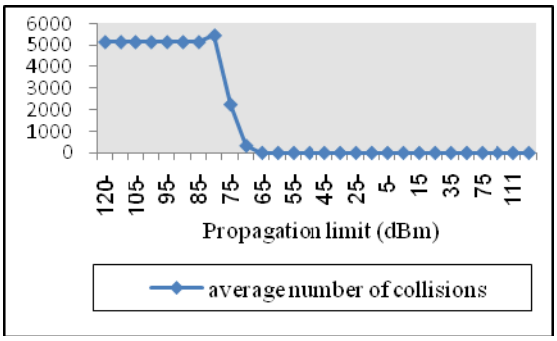


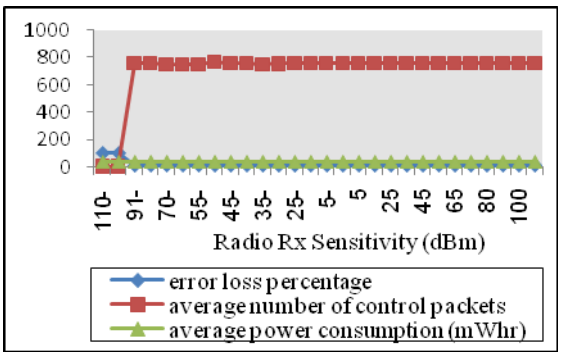
Fig. 32, 33 Effect of Propagation Limit.

4.1.17 Changing the Radio-RX-Sensitivity

As it appears from figures 34, 35, the error loss percentage varies when increasing the Radio-RX-Sensitivity from -110 dBm to 100 dBm. When the Radio-RX-Sensitivity is between -110 dBm to -91 dBm, the EADARP protocol decreases its error loss percentages from 100% to below 12%. But when the Radio-RX-Sensitivity is between -91 dBm till reaching 100 dBm, the error loss percentage of the EADARP protocol is 11%. Also, the average numbers of collisions of the EADARP protocol increases when increasing the Radio-RX-Sensitivity from -110 dBm to 100 dBm. When the Radio-RX-Sensitivity is between -110 dBm to -100 dBm, the EADARP protocol has an average number of collisions of 0. But when the Radio-RX-Sensitivity is between -100 dBm to -70 dBm, the average number of collisions of the EADARP protocol increases from 0 to 5300. When the Radio-RX-Sensitivity is between -70 dBm till 100 dBm, the EADARP protocol keeps its average number of collisions at 5300. Also, the average number of control packets of the EADARP protocol varies along the graph. When the Radio-RX-Sensitivity is between -110 dBm to -100 dBm, the EADARP protocol has an average number of control packets equal to 0. When the Radio-RX-Sensitivity is between -100 dBm to -91 dBm, the EADARP protocol increases its average number of control packets from 0 to 800. When the Radio-RX-Sensitivity is from -91 dBm till 100 dBm, the average number of control packets of the EADARP protocol is equal to 800. Also, the average power consumption of the EADARP protocol increases at -100 dBm from a power of 37.5 mWhr to a higher power when increasing the Radio-RX-Sensitivity from -110 dBm to 100 dBm. And these results stem from the nature of the protocol, and that the Radio-RX-Sensitivity affects the protocol operation, hence, has an effect on the performance metrics; error loss percentage, average number of

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

collisions, average number of control packets, average power consumption.



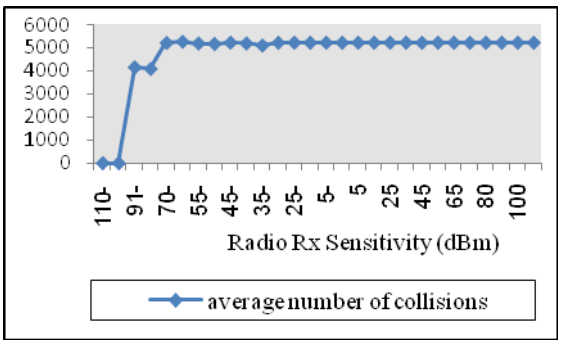
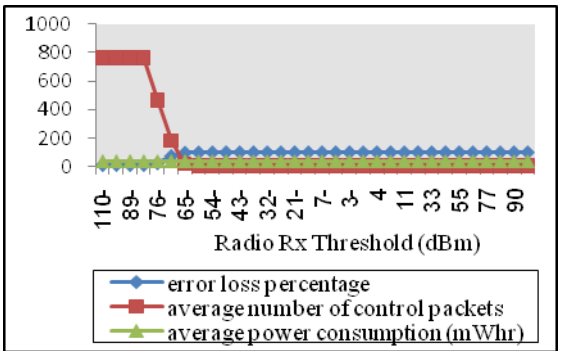


Fig. 34, 35 Effect of Radio-Rx-Sensitivity.

4.1.18 Changing the Radio-RX-Threshold

As it appears from figures 36, 37, the error loss percentage of the EADARP protocol increases from below 12% to 100% when increasing the Radio-RX-Threshold from -110 dBm to 90 dBm. When the Radio-RX-Threshold is between -110 dBm to -76 dBm, the error loss percentage of the EADARP protocol is standing at 11%. When the Radio-RX-Threshold is between -76 dBm to -65 dBm, the EADARP protocol increases its error loss percentage to 100% and this error loss percentage stays constant till the end of the graph. Also, the average number of collisions of the EADARP protocol decreases from higher average number of collisions to 0 at a Radio-RX-Threshold of -65 dBm, when increasing the Radio-RX-Threshold from -110 dBm to 90 dBm. When the Radio-RX-Threshold is between -65 dBm to 90 dBm, the EADARP protocol has an average number of collisions equal to 0. Also, the average number of control packets of the EADARP protocol stays equal to 0, when increasing the Radio-RX-Threshold from -65 dBm to 90 dBm. When the Radio-RX-Threshold is between -65 dBm to -89 dBm, the average number of control packets of the EADARP protocol increases from 0 to 800. When the Radio-RX-Threshold is between -89 dBm to -110 dBm, the EADARP protocol has an average number of control packets equal to 800. Also, the average power consumption, measured in mWhr, of the EADARP protocol decreases from a higher power to a power of 37.5 mWhr at a Radio-RX-Threshold of -60 dBm, when increasing the Radio-RX-Threshold from -110 dBm to -60 dBm. When the Radio-RX-Threshold is between -60 dBm to 90 dBm, the EADARP protocol keeps the same average power consumption, standing at 37.5 mWhr. And these results stem from the nature of the protocol, and that the Radio-RX-Threshold affects the protocol operation, hence, has an effect on the performance metrics; error loss percentage, average number of collisions, average number of control packets, average power consumption.



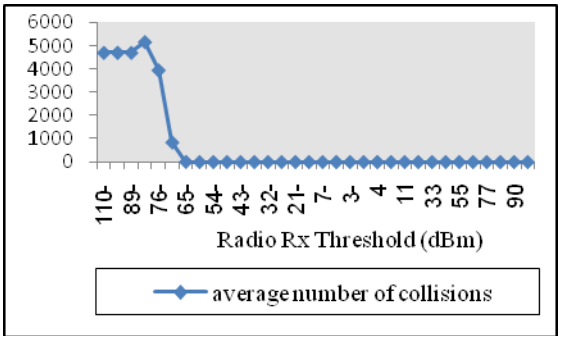


Fig. 36, 37 Effect of Radio-Rx-Threshold.

5. Conclusions

In [13,14], the EADARP protocol for multicast wireless ad hoc networks was proposed. EADARP provides multiple paths by the formation of mesh configuration making the protocol robust to mobility. Alternate routes secure data delivery when there is mobility and link breaks while the primary route is being reconstructed. The key properties of EADARP are; simplicity, low channel

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

and storage overhead, usage of up-to-date shortest routes, reliable construction of routes and forwarding group, robustness to host mobility, maintenance and utilization of multiple paths.

In this paper, we continued to investigate and study sensitivity of the performance of the protocol due the effect of changing the network parameters, Eighteen parameters are considered allowing to assess the impact of the changes in such parameters on the performance of the protocol and analyzing the results obtained from the simulation, to give conclusions.

From the results obtained, the parameters that does not affect the EADARP protocol operation are the following: Temperature, MAC propagation delay. The parameters that have a very limited effect in some cases on the performance of the EADARP protocol performance, are: Terrain Dimensions, MAC Propagation Delay, Noise Figure.

The performance of the EADARP protocol enhances in general when increasing the following parameters: RADIO_TX_POWER, Using MACA MAC protocol with or without promiscuous mode is best for the protocol, Using the RANDOM_WAYPOINT is best for the protocol, the FREE_SPACE propagation model is best, the RANDOM_with_NONE_MOBILITY is the best node placement method, Radio Antenna Gain, the BER_BASED Radio_RX_TYPE is better, the RADIO_ACCNOISE is better, Radio-RX-Sensitivity.

The performance of the EADARP protocol deteriorates in general when increasing the following parameters: number of nodes, Radio Frequency, Radio-RX-SNR-Threshold, the Propagation Limit, Radio-RX-Threshold.

There are other remarks that are concluded from the simulation results as follows:

- Using MACA MAC protocol with or without promiscuous mode is better for the protocol.
- UNIFORM_WITH_NONE_MOBILITY is the worst node placement for the protocol.
- Using a RADIO_TX_POWER below 0 dBm, makes the error loss percentage of the protocol, equal to 100%, the average number of collisions equal to 0, and the average number of control packets equal to 0.
- Increasing the Radio Frequency above 2.4e9 hertz, makes the error loss percentage of the protocol increase.
- Increasing the Radio Antenna Gain leads to decreasing the error loss percentage of the protocol.
- Using a NO_ACCNOISE_RADIO leads to an 100% error loss percentage, a zero average number of collisions, a zero average number of control packets, and an average power consumption equal to 37.5 mWhr for the protocol.
- Increasing the Propagation Limit above -65 dBm leads to an 100% error loss percentage, a zero average number of collisions, a zero average number of control packets, and an average power consumption equal to 37.5 mWhr for the protocol.
- Decreasing the Radio-RX-Sensitivity below the -91 dBm makes an 100% error loss percentage, a zero average number of collisions, a zero average number of control packets, and an average power consumption equal to 37.5 mWhr for the protocol.
- Increasing the Radio-RX-Threshold to -65 dBm and above makes an 100% error loss percentage, a zero average number of collisions, a zero average number of control packets, and an average power consumption equal to 37.5 mWhr for the protocol.

References

[1] B.Biskupski ,J.Dowling and J.Sacha, "Properties and mechanisms of self organizing MANET and P2P Systems", ACM Transactions on Autonomous and Adaptive systems, vol.2, no.1, March 2007, pp:1-34.

[2] Jun Liu, Jiejun Kong, Xiaoyan Hong and Mario Gerla, "Performance Evaluation of Anonymous Routing Protocols in MANETs", IEEE Wireless Communications and Networking Conference 2006, Las Vegas, April, 2006.

[3] Michail and A.Ephremides, "Energy Efficient routing for connection oriented traffic in wireless ad-hoc networks", Mobile Networks and Applications, vol. 8, 2003, pp: 517-533.

[4] C.-C. Chiang, M. Gerla and L. Zhang , "Forwarding Group Multicast Protocol (FGMP) for multihop, mobile wireless networks", Cluster Computing, vol.1, No.2 , 1998 , pp. 187–196.

[5] Kwan-Wu Chin, "A New Link Scheduling Algorithm for Concurrent Tx/Rx Wireless Mesh Networks", IEEE International Conference on Communications (IEEE ICC'08), Beijing, China, May 19-23, 2008.

[6] M.S. Corson and S.G. Batsell, "A Reservation-Based Multicast (RBM) routing protocol for mobile networks: Initial route construction phase", Wireless Networks, vol.1, No.4, December 1995, pp. 427–450.

[7] M. Gerla, C.-C. Chiang and L. Zhang, "Tree multicast strategies in mobile, multihop wireless networks", Mobile Networks and Applications, vol. 4, No.3, October 1999, pp. 193–207.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

[8] T. Ozaki, J.B. Kim and T. Suda, "Bandwidth-efficient multicast routing protocol for ad-hoc networks", Proceedings of IEEE ICCCN'99, Boston, MA ,October 1999, pp. 10–17.

[9] Ravindra Vaishampayan, J. J. Garcia-Luna-Aceves and Katia Obraczka, "An adaptive redundancy protocol for mesh based multicasting", Computer Communications, Volume 30 Issue 5, March 2007.

[10] Jun Tao, Xiaoyan Hong and Phillip G. Bradford, "Single-hop and multi-hop delay Characteristics of Ad-hoc 802.11b wireless network", IASTED International Conference on Parallel and Distributed Computing and Systems (PDCS 2007), Cambridge, MA. USA, Nov 19-21, 2007.

[11] Glomosim User Manual, http://pcl.cs.ucla.edu/projects/glomosim

[12] R. Meyer and R. Bagrodia, "PARSEC User Manual Release 1.1", January 1999, http://pcl.cs.ucla.edu/.

[13] Dina Darwish, Imame Saroit and Abdelwahab Hassan, "Energy Adaptable Distance Aware Routing Protocol for Multicast Wireless ad hoc Networks", International journal of Intelligent Computing and Information science, vol. 8, no. 2, July 2008, pp:186-204.

[14] Dina Darwish, Imame Saroit and Abdelwahab Hassan, "Effect of Packet Size and Channel Capacity on the Performance of EADARP for Multicast Wireless Ad hoc Networks", Third International Joint IEEE - CISSE Conference on Telecommunications and Networking, Bridgeport-USA, 5-13 December, 2008.

Dina Darwish received the B.Sc. in 2004 and the M.Sc. in 2006 with honors degrees in computer engineering from Arab Academy for Science and Technology, Egypt. She received the Ph.D. degree with honors degrees in computer engineering in 2009 from Cairo University, Egypt. Her main interests include communications systems, computer networks, internet technology, and multimedia systems.
She is assistant professor of communications and computer Engineering International Academy of Engineering and Media Science, Egypt. Since September 2009.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Fault Tolerance- Challenges, Techniques and Implementation in Cloud Computing

Anju Bala[1], Inderveer Chana[2]

[1] **Computer Science and Engineering Department, Thapar University**
Patiala-147004, Punjab, India

[2] **Computer Science and Engineering Department, Thapar University**
Patiala-147004, Punjab, India

Abstract

Fault tolerance is a major concern to guarantee availability and reliability of critical services as well as application execution. In order to minimize failure impact on the system and application execution, failures should be anticipated and proactively handled. Fault tolerance techniques are used to predict these failures and take an appropriate action before failures actually occur. This paper discusses the existing fault tolerance techniques in cloud computing based on their policies, tools used and research challenges. Cloud virtualized system architecture has been proposed. In the proposed system autonomic fault tolerance has been implemented. The experimental results demonstrate that the proposed system can deal with various software faults for server applications in a cloud virtualized environment.

Keywords: *Cloud Computing; Virtual Machine; Fault Tolerance; Replication*

1. Introduction

Cloud computing is a style of computing where service is provided across the Internet using different models and layers of abstraction [4]. It refers to the applications delivered as services [5] to the mass, ranging from the end-users hosting their personal documents on the Internet to enterprises outsourcing their entire IT infrastructure to external data centers. A simple example of cloud computing service is Yahoo email or Gmail etc.

Although cloud computing has been widely adopted by the industry, still there are many research issues to be fully addressed like fault tolerance, workflow scheduling, workflow management, security etc.Fault tolerance is one of the key issues amongst all. It is concerned with all the techniques necessary to enable a system to tolerate software faults remaining in the system after its development. When a fault occurs, these techniques provide mechanisms to the software system to prevent system failure occurrence [18]. The main benefits of implementing fault tolerance in cloud computing include failure recovery, lower cost, improved performance metrics etc. This paper aims to provide a better understanding of fault tolerance challenges and identifies various tools and techniques used for fault tolerance. When multiple instances of an application are running on several virtual machines and one of the server goes down, there is a need to implement an autonomic fault tolerance technique that can handle these types of faults. To address this issue, cloud virtualized system architecture has been proposed and implemented using HAProxy.The proposed architecture also has been validated through experimental results.

The rest of the paper is organized as follows. Section 2 discusses fault tolerance techniques based on their policies. Section 3 presents challenges of implementing fault tolerance in cloud computing .Section 4 identifies the comparison between various tools used for implementing fault tolerance techniques with their comparison table. Section 5 presents proposed cloud virtualized architecture and implementation with experimental results. Section 6 finally concludes the paper.

2. Background

There are various faults which can occur in cloud computing .Based on fault tolerance policies various fault tolerance techniques can be used that can either be task level or workflow level .

2.1 Reactive fault tolerance

Reactive fault tolerance policies reduce the effect of failures on application execution when the failure effectively occurs. There are various techniques

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

which are based on these policies like Checkpoint/Restart, Replay and Retry and so on.

- Check pointing/ Restart - When a task fails, it is allowed to be restarted from the recently checked pointed state rather than from the beginning. It is an efficient task level fault tolerance technique for long running applications [2].

- Replication-Various task replicas are run on different resources, for the execution to succeed till the entire replicated task is not crashed. It can be implemented using tools like HAProxy, Hadoop and AmazonEc2 etc.

- Job Migration-During failure of any task, it can be migrated to another machine. This technique can be implemented by using HAProxy.

- SGuard- It is less disruptive to normal stream processing and makes more resources available. SGuard is based on rollback recovery [18] and can be implemented in HADOOP, Amazon EC2.

- Retry-It is the simplest task level technique that retries the failed task on the same cloud resource [20].

- Task Resubmission-It is the most widely used fault tolerance technique in current scientific workflow systems [25]. Whenever a failed task is detected, it is resubmitted either to the same or to a different resource at runtime.

- User defined exception handling-In this user specifies the particular treatment of a task failure for workflows.

- Rescue workflow-This technique [20] allows the workflow to continue even if the task fails until it becomes impossible to move forward without catering the failed task.

2.2 Proactive Fault Tolerance

The principle of proactive fault tolerance policies is to avoid recovery from faults, errors and failures by predicting them and proactively replace the suspected components other working components. Some of the techniques which are based on these policies are Preemptive migration, Software Rejuvenation etc.

- Software Rejuvenation-It is a technique that designs the system for periodic reboots. It restarts the system with clean state [5].

- Proactive Fault Tolerance using Self-Healing- When multiple instances of an application are running on multiple virtual machines, it automatically handles failure of application instances.

- Proactive Fault Tolerance using Preemptive Migration- Preemptive Migration relies on a feedback-loop control mechanism where application is constantly monitored and analyzed.

3. Challenges of Implementing Fault Tolerance in Cloud Computing

Providing fault tolerance requires careful consideration and analysis because of their complexity, inter-dependability and the following reasons.

- There is a need to implement autonomic fault tolerance technique for multiple instances of an application running on several virtual machines [12].

- Different technologies from competing vendors of cloud infrastructure need to be integrated for establishing a reliable system [15].

- The new approach needs to be developed that integrate these fault tolerance techniques with existing workflow scheduling algorithms [14].

- A benchmark based method can be developed in cloud environment for evaluating the performances of fault tolerance component in comparison with similar ones [21].

- To ensure high reliability and availability multiple clouds computing providers with independent software stacks should be used [22] [23].

- Autonomic fault tolerance must react to synchronization among various clouds [15].

4. Tools Used For Implementing Fault Tolerance

Fault tolerance challenges and techniques have been implemented using various tools. Table 1 compares these tools based on their programming framework, environment and application type along with different fault tolerance techniques. HAProxy is used for server failover in the cloud [13]. *S*Help [12] is a lightweight runtime system that can survive software failures in the framework of virtual machines. It may

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

also work in cloud environment for implementing check pointing. ASSURE [9] introduces rescue points for handling programmer anticipated failures. Hadoop [7] is used for data intensive applications but can also be used to implement fault tolerance techniques in cloud enviorment.Amazon Elastic Compute Cloud (EC2) [8] provides a virtual computing environment to run Linux-based applications for fault tolerance.

Table 1: Tools Used To Implement Existing Fault Tolerance Techniques

Fault Tolerance Techniques	*Policies*	*System*	*Programming Framework*	*Environment*	*Fault Detected*	*Application Type*
Self Healing, Job Migration, Replication	Reactive/ Proactive	HAProxy[13]	Java	Virtual Machine	Process/node failures	Load balancing Fault Tolerance
Check pointing	Reactive	SHelp[12]	SQL,JAVA	Virtual Machine	Application Failure	Fault tolerance
Check pointing, Retry, Self Healing	Reactive/ Proactive	Assure[9]	JAVA	Virtual Machine	Host, Network Failure	Fault tolerance
Job Migration,Replication,Sguard,Resc	Reactive/ Proactive	Hadoop[7]	Java,HTML,CSS	Cloud Environment	Application/node failures	Data intensive
Replication, Sguard,Task Resubmission	Reactive/ Proactive	AmazonEC2[8]	Amazon Machine Image, Amazon Map	Cloud Environment	Application/node failures	Load balancing , fault tolerance

5. Proposed Cloud Virtualized System Architecture And Implementation

5.1 Cloud Virtualized System Architecture

A few techniques currently exist for autonomic fault tolerance in cloud enviorment.Shelp can survive the software faults for server applications running in virtual machine environment [12].There is a need to implement autonomic fault tolerance in cloud enviorment.If any one of the servers breaks down, system should automatically redirect user requests to the backup server. So, the cloud virtualized system architecture has been proposed and implemented using HAProxy.The application availability and reliability can be maintained by using the proposed cloud virtualized system architecture as shown in Figure 1. The server virtualized system consists of VMs (server 1 and server 2) on which an Ubuntu 10.04 OS and database application are running. Server 2 is a backup sever in case of failure. HAProxy is configured on the third virtual machine to be used for fault tolerance. The availability of the servers is continuously monitored by HAProxy statistics tool on a fault tolerant server. HAProxy is running on web server to handle requests from web. When one of the servers goes down unexpectedly, connection will automatically be redirected to the other server.

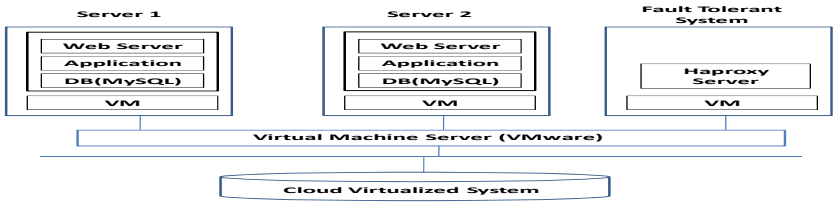


Figure 1: Cloud Virtualized System Architecture

5.2 Implementation and Experimental Results

Fault tolerant system has been implemented using HAProxy and MySQL. HAProxy is used to handle server failures in fault tolerant cloud environment. It provides a web interface for statistics known as HAProxy statistics. Implementation includes two virtual machines as web servers, server 1 and server 2

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

hosting Apache Tomcat 6.0.32. HAProxy software version 1.3.15.2 is configured on third virtual machine. A simple database application written in Java is installed on the web servers. Xampp for Linux Windows XP (SP2) is used to install MySQL. Data in MySQL is replicated using replication technique for local backup. Replication enables data from one MySQL database server (the master) to be replicated to one or more MySQL database servers. Application can be accessed on any of the web server. Data consistency is also maintained through MySQL replication. The experimental results show that HAProxy can assist server applications to recover from server failures in just a few milliseconds with minimum performance overhead.

Case 1: Figure 2 shows the stats when both the servers are working and green line indicates that the servers are up.

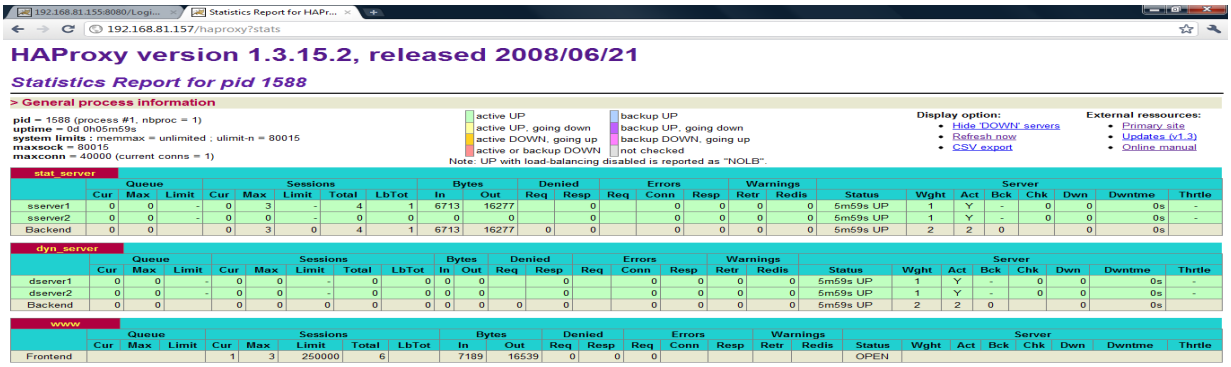


Figure 2: HAProxy Statistics

Case 2: Figure 3 shows the stats when server 1 goes down and server 2 is still up. Red line in this figure indicates that the server is down.

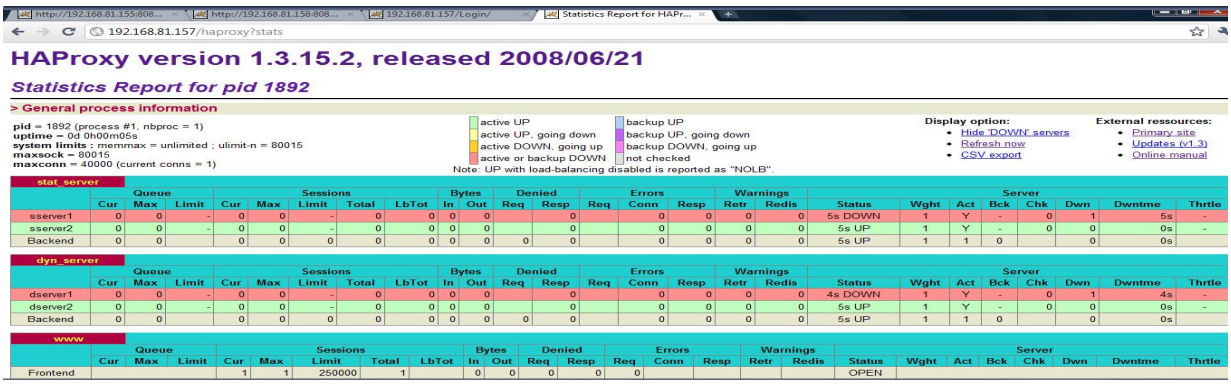


Figure3: Server 1 is down

Case 3: Figure 4 shows the stats when server 2 goes down and server 1 is still up.

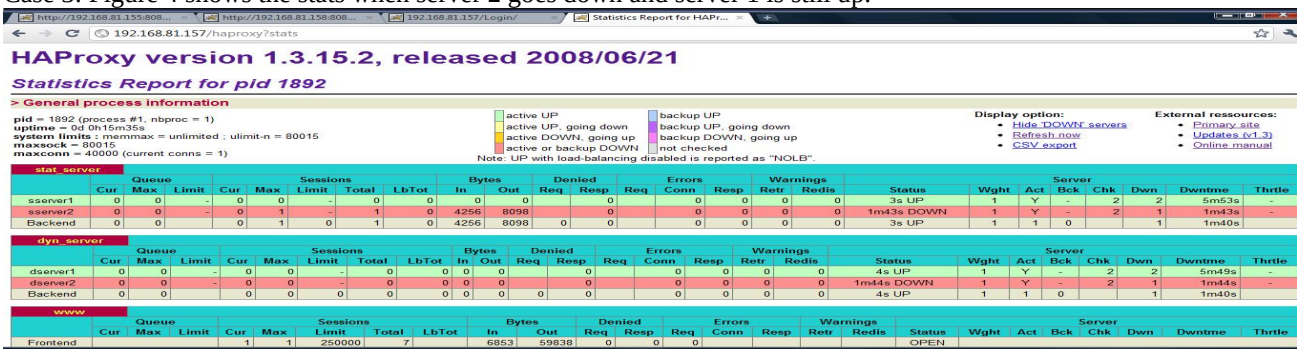


Figure 4: Server 2 is down

Case 4: Replication enables data from one MySQL database server to be replicated to one or more MySQL database servers (the slaves). Figure 5 shows the database table of server 1 using SQLyog and the connection established with server 2 slave.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

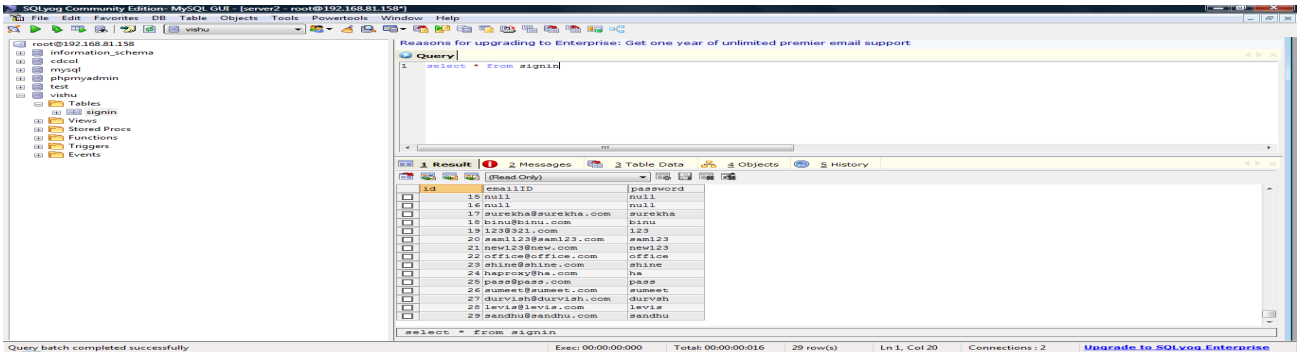


Figure 5: Database entries of sever 1

Case 5: Figure 6 shows the replicated database table of server 2.When server 1 fails data is replicated to server 2.

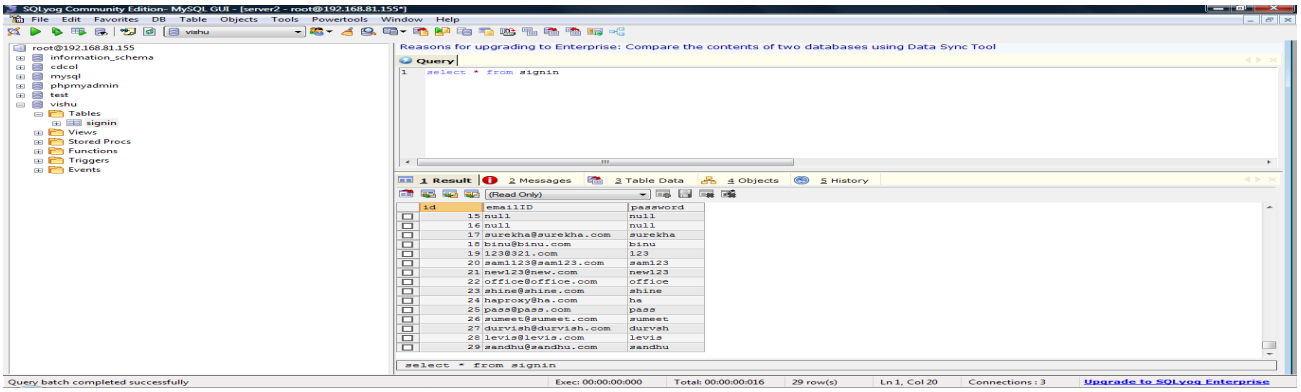


Figure 6: Replicated Database entries of sever 2

6. Conclusion

Fault tolerance is concerned with all the techniques necessary to enable a system to tolerate software faults remaining in the system after its development. This paper discussed the fault tolerance techniques covering its research challenges, tools used for implementing fault tolerance techniques in cloud computing. Cloud virtualized system architecture is also proposed based on HAProxy. Autonomic fault tolerance is implemented dealing with various software faults for server applications in a cloud virtualized environment. When one of the servers goes down unexpectedly, connection will automatically be redirected to the other server. Data replication technique is implemented on virtual machine environment. The experimental results are obtained, that validate the system fault tolerance.

References

[1]Antonina Litvinova, Christian Engelmann and Stephen L. Scott," A Proactive Fault Tolerance Framework for High Performance Computing", 2009.

[2]Golam Moktader Nayeem, Mohammad Jahangir Alam," Analysis of Different Software Fault Tolerance Techniques", 2006.

[3]Steven Y. Ko, Imranul Hoque, Brian Cho and Indranil Gupta, "On Availability of Intermediate Data in Cloud Computations", 2010.

[4]L. M. Vaquero, L. Rodero-Merino, J. Caceres and M. Lindner, "A break in the clouds: towards a cloud definition," SIGCOMM Computer Communication Review,vol. 39, pp. 50–55, December 2008.

[5] M.Armbrust, A.Fox, R. Griffit,et al., "A view of cloud computing", Communications of the ACM, vol. 53, no. 4, pp. 50–58, 2010.

[6]R.Buyya, S.Pandey and C.Vecchiola, "Cloudbus toolkit for market-oriented cloud computing", In Proceeding of the 1st International Conference on Cloud Computing (CloudCom2009), Beijing, China, December 2009.

[7]HadoopMapReduceTutorial.http://hadoop.apache.org/core/docs/current/mapred tutorial.html.

[8]AmazonElasticComputeCloud(EC2) http://www.amazon.com/ec2/

[9]S. Sidiroglou, O. Laadan, C. Perez, N. Viennot, J. Nieh, and A. D. Keromytis, "ASSURE: Automatic Software Self-healing Using REscue points", Proceedings of the 14th International Conference on Architectural Support for Programming Languages and Operating Systems

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

(ASPLOS'09), ACM Press, March 7-11, 2009, Washington, DC, USA, pp.37-48.

[10]B. Buck and J. K. Hollingsworth, "An API For Runtime Code Patching", International Journal of High Performance Computing Applications, Vol.14, No.4, November 2000, pp.317-329.

[11]S. Osman, D. Subhraveti, G. Su, and J. Nieh, "The Design and Implementation of Zap: A System For Migrating Computing Environments", Proceedings of the 5th Symposium on Operating Systems Design and Implementation (OSDI'02), USENIX Association, December 9-11, 2002, Boston, Massachusetts, USA, pp.361-376.

[12]Gang Chen, Hai Jin, Deqing Zou, Bing Bing Zhou, Weizhong Qiang, Gang Hu, "SHelp: Automatic Self-healing for Multiple Application Instances in a Virtual Machine Environment", IEEE International Conference on Cluster Computing, 2010.

[13]http://haproxy.1wt.eu/download/1.3/doc/configuration.txt.

[14]Yang Zhang1, Anirban Mandal2, Charles Koelbel1 and Keith Cooper," Combined Fault Tolerance and Scheduling Techniques for Workflow Applications on Computational Grids "in 9th IEEE/ACM international symposium on clustering and grid, 2010.

[15]Imad M. Abbadi, "Self-Managed Services Conceptual Model in Trustworthy Clouds' Infrastructure", 2010.

[16]Manish Pokharel and Jong Sou Park, "Increasing System Fault Tolerance with Software Rejuvenation in E-government System", IJCSNS International Journal of Computer Science and Network Security, VOL.10 No.5, May 2010.

[17]Zaipeng Xie, Hongyu Sun and Kewal Saluja, "A Survey of Software Fault Tolerance Techniques".

[18]Geoffroy Vallee, Kulathep Charoenpornwattana, Christian Engelmann, Anand Tikotekar, Stephen L. Scott," A Framework for Proactive Fault Tolerance".

[19]Anglano C, Canonico M," Fault-tolerant scheduling for bag-of-tasks grid applications", In: Advances in grid computing—EGC 2005. Lecture notes in computer science, vol 3470/2005. Springer,Berlin/Heidelberg. ISSN: 0302-9743 Print. doi:10.1007/b137919, ISBN: 978-3-540-26918-2,pp 630–639

[20]Elvin Sindrilaru,,Alexandru Costan,, Valentin Cristea," Fault Tolerance and Recovery in Grid Workflow Management Systems", 2010 International Conference on Complex, Intelligent and Software Intensive Systems.

[21]S. Hwang, C. Kesselman, "Grid Workflow: A Flexible Failure Handling Framework for the Grid",12th IEEE international Symposium on Zigh Performance Distributed Computing (HPDC'03), Seattle, Washington,USA., IEEE CS Press, Los Alamitos, CA, USA, June 22 - 24, 2003.

[22]Michael Armbrust, Armando Fox,Rean Griffith, " Above the Clouds: A Berkeley View of Cloud Computing", Electrical Engineering and Computer Sciences University of California at Berkeley, 2009.

[23]Wenbing Zhao, P. M. Melliar-Smith and L. E. Moser," Fault Tolerance Middleware for Cloud Computing", 2010 IEEE 3rd International Conference on Cloud Computing.

[24]Kassian Plankensteiner, Radu Prodan, Thomas Fahringer,"A New Fault Tolerance Heuristic for Scientific Workflows in Highly Distributed Environments based on Resubmission Impact", Fifth IEEE International Conference on e-Science,Austria,2009

Anju Bala completed her B.Tech in Computer Science (1999) and M.Tech. in Information Technology from Pbi.University (2004) and pursuing Ph.D. in Cloud Computing from Thapar University, Patiala (2009) and has over eleven years of teaching experience. She is working as Assistant Professor in Computer Science and Engineering Department of Thapar University, Patiala. Her research interests lie in Fault Tolerance in Cloud Computing and Workflow management in Cloud Computing. She has over 14 publications in Conferences and International Journals of repute.

Dr. Inderveer Chana completed her B.Tech in Computer Science (1997) and M.E. in Software Engineering from TIET (2002) and Ph.D. in Resource Management in Grid Computing from Thapar University, Patiala (2009) and has over thirteen years of teaching and research experience. She is working as Assistant Professor in Computer Science and Engineering Department of Thapar University, Patiala. Her research interests include Grid computing, Cloud Computing and resource management challenges in Grids and Clouds. She has over 30 publications in International Journals and Conferences of repute. More than 20 Masters have been completed so far under her supervision and is currently supervising 9 Doctoral candidates in the area of Grid and Cloud Computing.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

A Study of Edge Detection Method for different Climate Condition Images using Digital Image Processing

P.Balakrishnan[1], S.Muruganand[2], K.Sriram[3]

[1]Department of Electronics and Instrumentation, Bharathiar University, Coimbatore (Dist.), Tamilnadu (State), India

[2]Department of Electronics and Instrumentation, Bharathiar University, Coimbatore (Dist.), Tamilnadu (State), India

[3]Department of Electronics and Instrumentation, Bharathiar University, Coimbatore (Dist.), Tamilnadu (State), India

Abstract

Human eye can sense different weather conditions as same as the sensor. The human eye gives good result. In this proposal work the six different edge detection methods are applied to analyze weather condition. In this work the special kind of changes in the picture for different weather condition are observed. The six different edge detection method like sobel, prewitt, Robert, canny, laplacian of Gaussian and zero crossing are used to analyze the weather condition and finally the result of three methods are compared for best result.

Key words: *edge detection, human eye, picture, sensor, weather condition.*

1. Introduction

Image processing has been applied in wide areas like computer vision, face detection, medical imaging, microscope image, raster operation and forensic etc. Human eye could analyze weather condition just by seeing the environment and location. By nature, Human eye have some sensing ability to find the weather condition by differentiating the brightness. Similar to that, different instruments are being used to find the weather condition as sunlight falls on the earth. Different changes occur in the intensity of sunlight radiations. These variations in the intensity of sunlight make changes in the pixel rating of the picture. When intensity of sunlight increases, the sharp edge of a picture also increases. In same way when the intensity falls, the sharp edge falls. The natures of human eyes analyze the two different conditions such as dark room and bright room. The human eye finds more sharp edge in the bright room than dark room.In the same way, this proposal work analyzes the images of different weather condition using six edge detection methods such as sobel, prewitt, Robert, canny, laplacian of Gaussian, and zero crossing.

2. Process

For processing, the pictures of the object for different climatic condition are taken on a rainy day at location *GPS* Longitude, 76 deg 51' 7.00" E in the same angle and distance. Time interval for each picture is one hour. The pictures of seven samples are taken from morning 6.00 am to 11.59 am. The different type edge detection methods are used for finding the area of each edge detected picture. These methods give the output by increasing intensity of light which in turn the area of the edge detected picture increases. The edge detected images are displayed below.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Figure 1. Original image

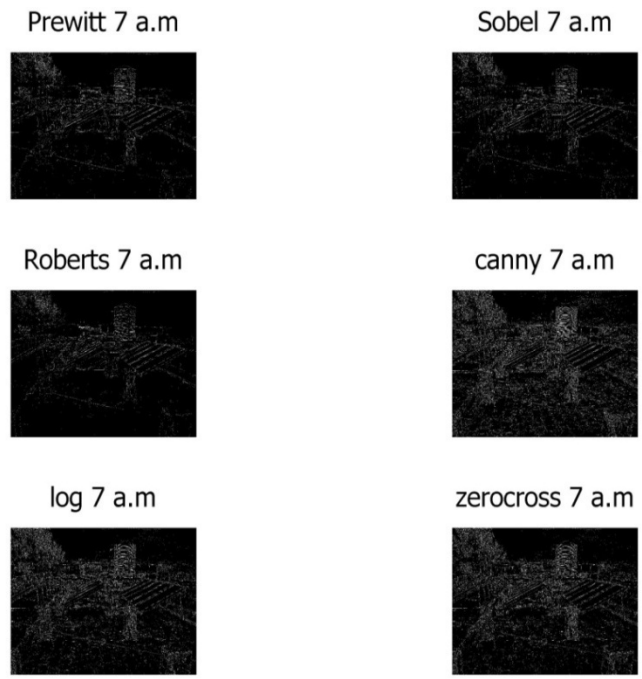


Figure 3.Result for 7.00 a.m sample image

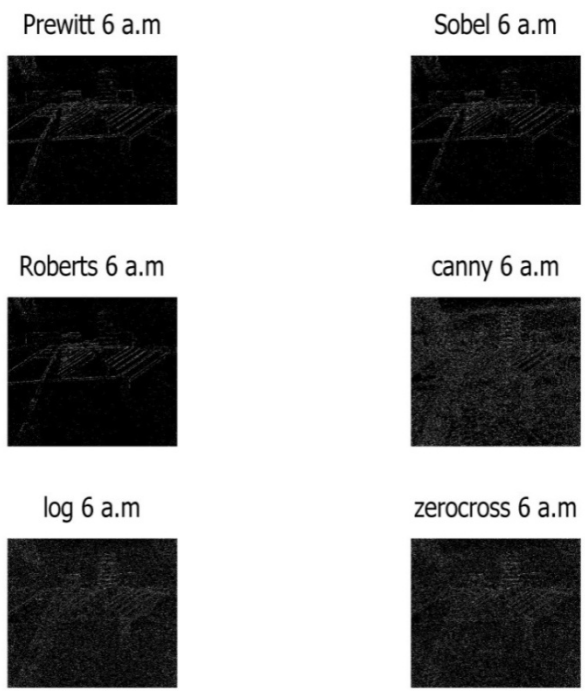


Figure 2.Result for 6.00 a.m sample image

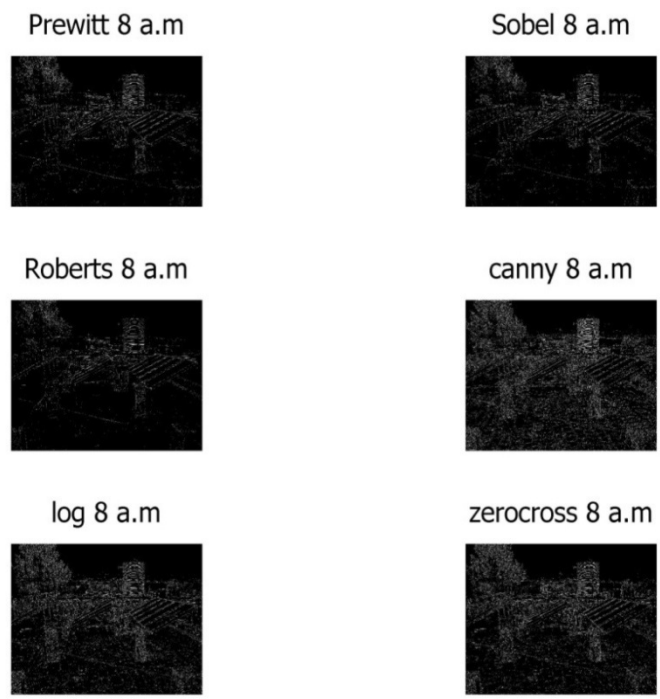


Figure 4.Result for 8.00 a.m sample image

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

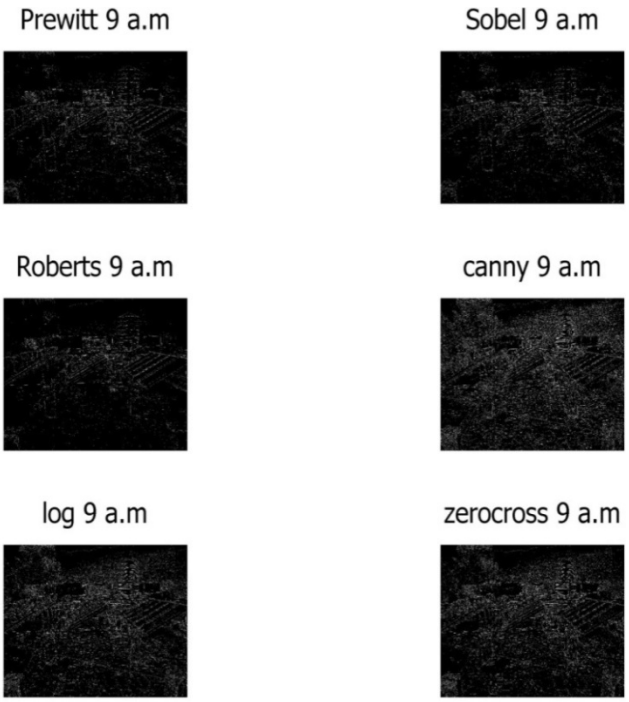


Figure 5.Result for 9.00 a.m sample image

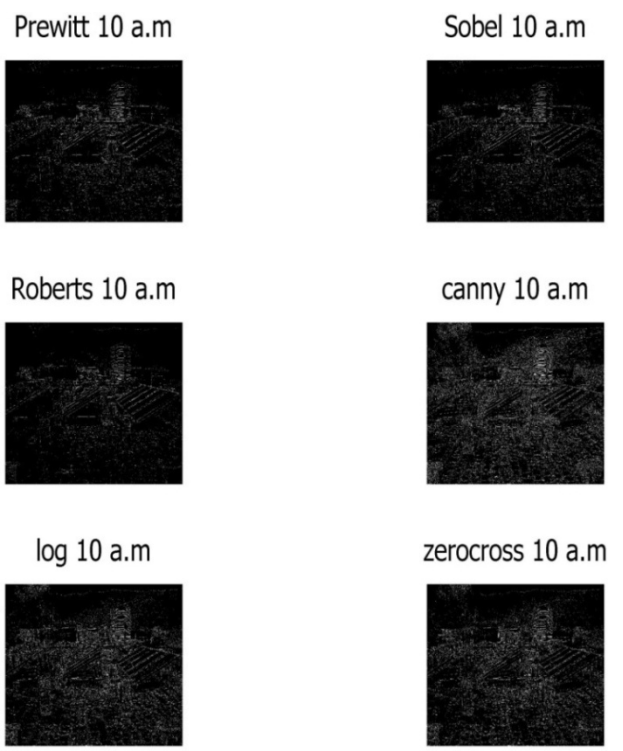


Figure 6.Result for 10.00 a.m sample image

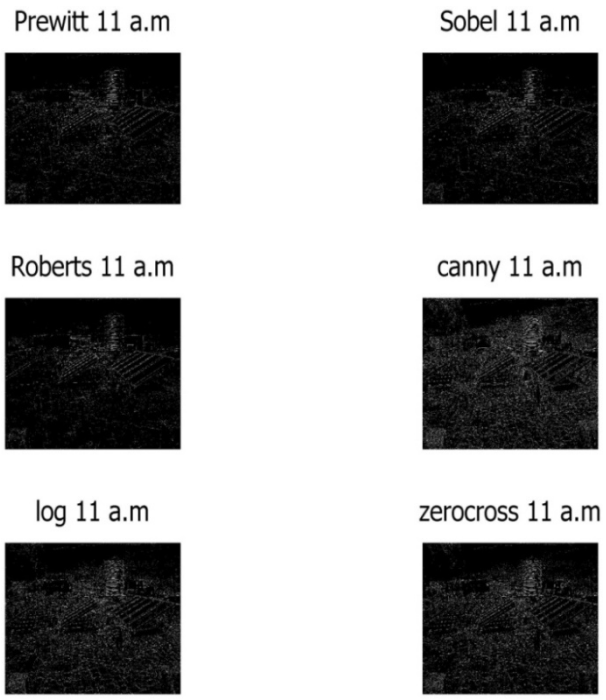


Figure 7.Result for 11.00 a.m sample image

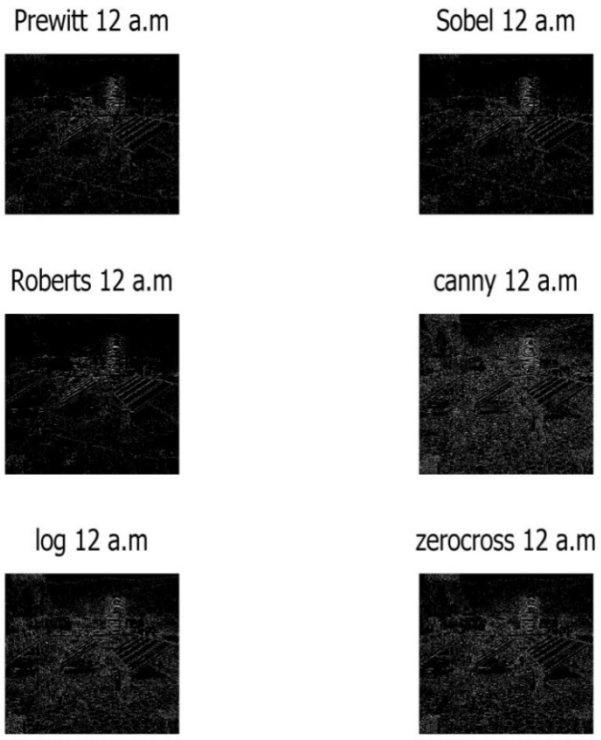


Figure 8.Result for 12.00 p.m sample image

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

3. Result and Discussions

Edge detection method is used for different images in regular interval. It gives different area of edge detected images and results. Figure (2) gives low area value because of the low intensity of light at that time. Basically in early morning, sunlight has low intensity. When the time get increase the light intensity, area of the edge detected picture also randomly increase. Suppose cloud around the place causes decrease in light intensity, also the area of image causes reduction. Here picture area is directly proportional to light intensity.

Every one hour time operator value for first three hours randomly increased and four and six hour slightly decreased in operator value and five and seven hour shows increase in operator value. Because of the five and seven hours image is taken when the cloud covered. Hence using this variation we can find what is climate condition and approximate climatic temperature at particular location by referring predefine measures of it Sobel and Prewitt Edge detecting method which gives good result whereas Laplacian of Gaussian and zero cross edge detection method gives same result but these methods approximately similar with sobel and prewitt. Roberts and canny edge detection methods do not match with other edge detection methods.

Table 1(different edge detection method results)

Time/ Operators	*Sobel*	*Prewitt*	*Roberts*	*Canny*	*Laplacian of Gaussian*	*Zero crossings*
6	1.8662	1.8615	1.3736	1.0341	7.8482	7.8482
7	2.8993	2.9004	2.2346	8.1532	6.0531	6.0531
8	3.0915	3.0878	2.3414	8.9793	6.5442	6.5442
9	2.6257	2.6253	2.1797	9.5311	6.3965	6.3965
10	2.9988	3.0028	2.3109	9.5647	6.7115	6.7115
11	2.7728	2.7743	2.2879	8.8621	6.1226	6.1226
12	2.8643	2.8670	2.2042	1.0175	6.7839	6.7839

Table 1 shows the result for edge detection image. In sobel and prewitt edge detection method are increasing the light intensity causes increasing operator value and decrease in light intensity lead to decrease operator value.

Robert edge detection method may gives output similar to the sobel and prewitt edge detection methods. But operator value changes are small, compared with sobel, prewitt edge detection methods. Canny edge detection method is totally different with all others edge detection methods. Laplacian of Gaussian and zero crossing methods are initially different from sobel prewitt and Robert. But increasing in the operator values gives same result as other methods. Finally three methods are suitable for finding weather conditions. They are Robert, sobel and prewitt methods.

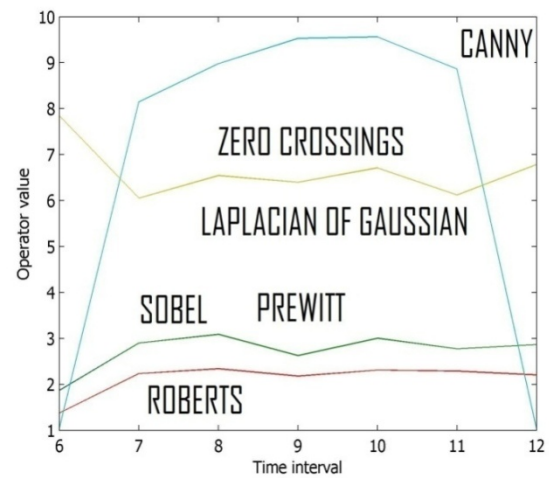


Figure 9 Graph model for different edge detection method

Graph model is drawn between operator value and time intervals. Each time interval has different weather condition. Sobel, Prewitt and Robert edge detection method have same track. Canny edge detection method have different track so it's not included for consideration. Laplacian of Gaussian and zero crossings methods give same value but they are different from other three methods such as Sobel, Prewitt and Robert.

4. Conclusion

After analyzing various edge detection methods prewitt and sobel methods produces best result.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Reference

[1] Krishna Kant Singh1 , Akansha Singh2, A STUDY OF IMAGE SEGMENTATION ALGORITHMS FOR DIFFERENT TYPES OF IMAGES, IJCSI International Journal of Computer Science Issues,Vol.7,Issue 5, September 2010

[2] Y.Ramadevi, T.Sridevi, B.Poornima, B.Kalyani, SEGMENTATION AND OBJECT RECOGNITION USING EDGE DETECTION TECHNIQUES, International Journal of Computer Science & Information Technology (IJCSIT), Vol 2, No 6, December 2010

[3] N.Senthilkumarn, R.Rajesh (2009),"EDGE DETECTION TECHNIQUES FOR IMAGE SEGMENTATION- A SURVEY OF SOFT COMPUTING APPROACHES", IJRTE, vol1, No2, 2009 250-254.

[4] Guangyu, L & Rensheng, C 2008, 'A SIMPLE AND EFFICIENT EDGE DETECTION ALGORITHM', paper presented at the Computer Science and Computational Technology, 2008. ISCSCT '08.

[5] He, Q & Zhang, Z 2007, 'A NEW EDGE DETECTION ALGORITHM FOR IMAGE CORRUPTED BY WHITE-GAUSSIAN NOISE', AEU - International Journal of Electronics and Communications, vol. 61, no. 8, pp. 546-550.

[6] Yu, Y-H & Chang, C-C 2006, 'A new edge detection approach based on Image context analyses, Image and Vision Computing, vol. 24, no. 10, pp. 1090-1102.

[7] www.mathworks.com

[8] www.code2learn.com

P.Balakrishnan pursuing PhD in electronics and instrumentation , Department of Electronics Instrumentation, Bharathiar University, Coimbatore, and he had finished M.Sc Applied Electronics in Ramakrishna mission vidyalaya college of arts and science in Coimbatore . He has one year industry experience and one and half year Research experience. His current research interest area is Image Processing

S.Muruganand is currently working as Assistant Professor, Department of electronics instrumentation, Bharathiar University, Coimbatore. He had finished PhD in Bharathiar University. He has twenty one year teaching experience and published research paper in journals. He has one UGC project work ongoing in industrial automation

K.Sriram pursuing PhD in electronics and instrumentation , Department of Electronics Instrumentation, Bharathiar University, Coimbatore, and he had finished M.Sc Applied Electronics in Ramakrishna mission vidyalaya college of arts and science in Coimbatore.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Survey of Tolerant Network in Mobile Communication

M. CHITHIK RAJA[#1] N. MOHAMED YASIN[#2] M.S.DURAI RAJAN[#3]

1. Department of Information Technology , Salalah College of Technology – Sultanate of Oman
2. Department of Information Technology , Salalah College of Technology – Sultanate of Oman
3. Department of Computer Science and Engineering –Affiliation of Anna University Madurai-India

***Abstract* - tolerant networks are a class of emerging networks that experience frequent and long-duration partitions. These net- works have a variety of applications in situations such as crisis environments and deep-space communication. In this paper, we study the problem of *multicasting* in TNs. Multicast supports the distribution of data to a group of users, a service needed for many potential TN applications. While multicasting in the Internet and mobile ad hoc networks has been studied extensively, due to the unique characteristic of frequent partitioning in TNs, multicasting in TNs is a considerably different and challenging problem. It not only requires new definitions of multicast semantics but also brings new issues to the design of routing algorithms. In this pa- per, we propose new semantic models for TN multicast and develop several multicast routing algorithms with different routing strategies. We present a framework to evaluate these algorithms in TNs. To the best of our knowledge, this is the first study of multicasting in TNs. Our objectives are to understand how routing performance is affected by the availability of knowledge about network topology and group membership and to guide the design of TN routing protocols. Using *ns* simulations, we find that efficient multicast routing for TNs can be constructed using only partial knowledge. In addition, accurate topology information is generally more important in routing than up-to-date membership information. We also find that routing algorithms that forward data along multiple paths achieve better delivery ratios, especially when available knowledge is limited.**

Keywords: **Multicast, delay tolerant networks, semantic model**

1. INTRODUCTION

Delay tolerant networks (DTNs) are a class of emerging net- works that experience frequent and long-duration partitions [9, 12]. There is no end-to-end path between some or all nodes in a DTN. These networks have a variety of applications in situations that include crisis environments like emergency response and military battlefields, deep-space communication, vehicular communication, and non-interactive Internet access in rural areas [1, 4, 10, 14, 19,20, 21, 22, 23].In this paper, we study the problem of multicasting in delay tolerant networks. Multicast service supports the distribution of data to a group of users. Many potential DTN applications operate in a group-based manner and require efficient network support for group communication. For example, in a disaster recovery scene, it is vital to disseminate information about victims and potential hazards among rescue workers. In a battlefield, soldiers in a squad need to inform each other about their surrounding environment. Although group communication can be implemented by sending a separate uncase packet to each user, this approach suffers from poor performance, which is confirmed in our simulations. The situation is especially acute in DTNs where resources such as connectivity among nodes, available bandwidth and storage are generally severely limited. Thus efficient multicast services are necessary for supporting these applications.

Multicasting in the Internet and mobile ad hoc networks (MANETs) has been studied extensively in the past . However, due to the unique characteristic of frequent partitioning in DTNs, multicasting in DTNs is a considerably different and challenging problem. It not only requires new definitions of multicast semantics but also brings new issues to the design of routing algorithms. The semantics of multicasting in traditional networks such as the Internet and MANETs are straightforward, specifying that packets sent to a multicast group be delivered to members of the group. Since data transfer delay in these networks is short (on the order of milliseconds), group membership changes during data transfer are rare and can be ignored. Thus the receivers of a multicast packet are well defined, i.e., all current group members. This, however, is no longer valid in DTNs. Due to frequent partitions and consequently large transfer delays in a DTN, membership changes during data transfer are the norm rather than the exception. Under these situations, it is not obvious how to define the receivers of a multicast packet, relative to the group membership over time. Consider a simple example where a source sends a message to a group at time t. Let t be the earliest time that other nodes could possibly receive this

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

message according to network topology limitations. Suppose that node A joins the group at time t_1 t and leaves at time t_2, t t_2 t . Node B joins at time t_3, t t_3 t and never leaves. From the perspective of traditional multicasting, it is not clear which nodes should receive this message, whether A, B, both or neither of them. For n ode A, it is a group member at the time of message generation but no longer a member at the earliest time of potential message delivery. The reverse is true for node *B*. To address this problem, new semantic models are needed for DTN multicasting.

In this paper, we develop new multicast semantic models for DTN environments that have explicit constraints on group membership and delivery action. These semantic models unambiguously define the receivers of a multicast packet and have various applications in DTN environments. With these semantic models, we study the problem of multicast routing in DTNs. DTNs introduce several challenges for routing. First, there may be no end-to-end path between nodes in DTNs. Traditional routing algorithms would fail to deliver data because no route is found to reach the destinations. Thus multicast routing in DTNs needs to operate in the presence of network partitions. Second, as proposed in [9], data transfer in DTNs is in application data units called *messages* (or bundles). This is different from the use of flows in traditional multicasting. Third, information about nodes joining or leaving a group may be available to nodes only after significant delays because of network partitions. Multicast routing algorithms need to handle these highly delayed join or leave requests. Finally, the multicast semantic models developed in this paper also introduce new requirements for message forwarding. We study four classes of multicast routing algorithms for DTN swith different routing strategies. To understand routing performance in DTN environments where available routing information may be significantly limited by network partitions, we present an evaluation framework that models different levels of available knowledge about network topology and group membership. This is an extension of the framework for unicast routing developed in [12]. Our objectives are to understand the impact of the availability of knowledge on routing performance and to guide the design of DTN routing protocols. With extensive *ns* [17] simulations, we evaluate various routing algorithms. We find that efficient routing for multi- cast can be constructed using only partial knowledge, as in the case of unicast [12]. In addition, accurate topology information is generally more important in routing than up-to-date membership information. Furthermore, routing algorithms that forward data along multiple paths achieve better delivery ratios, especially when avail- able knowledge is limited. Finally, our results confirm that unicast- based approaches that send a separate copy of messages to each receiver perform poorly in DTNs.

The rest of this paper is structured as follows. Section 2 describes the network model and briefly reviews unicast routing in DTNs. In Section 3, we present new semantic models for DTN multicasting. We describe an evaluation framework for multicast routing in Section 4. Section 5 presents the four classes of multicast routing algorithms. We present simulation results in Section 6 and review related work in Section 7. The paper is concluded in Section 8.

2. DTN NETWORK MODEL

In this section, we present the network model considered in this paper and briefly review unicast routing in DTNs.

2.1 Network Model

We assume that nodes in DTNs are identified by a unique ID. An endpoint is an entity at a node that acts as the source or destination of communication, e.g., an application at the node. An endpoint is identified by an endpoint ID which is a tuple node id entity id where ent it y *id* uniquely identifies an endpoint within a node[1] . We assume a message-oriented service, where endpoints communicate using application data units called messages. In addition, nodes are assumed to have loosely synchronized clocks.

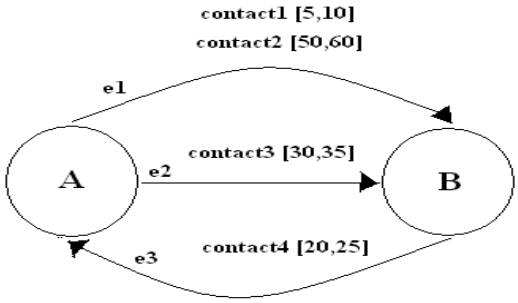


Figure:1 DTN Network Model

In our model, multicasting disseminates messages to a group of endpoints that are identified by a group ID. A group ID is a globally unique ID that has the same form as endpoint IDs. A multicast message encodes a group ID as the destination endpoint. In order to receive messages destined to a specific group, endpoints join the group by indicating a JOIN request with the group ID to the DTN routing agent at the node. Similarly, an endpoint leaves a group using a LEAVE request to stop receiving messages for a group. Routing agents in a DTN may

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

authenticate endpoints and authorize JOIN or LEAVE requests according to administrative policies. However, in this paper, we consider a general multicast model in which endpoints can join and leave groups autonomously.

In a DTN, network partitions may occur frequently. To overcome disconnections, data is forwarded in a store-carry-and-forward fashion, i.e., a node buffers messages in its storage until connections with other nodes become available. We assume that node storage is used for holding in-transit messages only. Delivered messages are stored in separate application buffers.

2.2 Unicast Routing in DTNs

We now briefly describe the unicast routing in DTNs developed by Jain et al. [12]. A DTN is represented as a directed multi-graph. Thus there may exist multiple edges between two nodes. Each edge represents a connection between nodes and has time-varying capacity and propagation delay that represent the properties of the connection over time. The capacity of an edge is zero when the corresponding connection is unavailable. A contact is defined as an opportunity to send data between nodes, i.e., an edge and the time interval during which the edge capacity is positive. Fig. 1 shows a DTN graph in which there are three edges between node A and node *B*. Contacts are shown along each edge with their time intervals. We can see that there are two contacts for edge e_1, from time 5 to 10 and from time 50 to 60 respectively.

Given the time-varying capacity and delay of edges in a DTN, routing decisions vary with time. Suppose that node A sends messages to node *B* and the routing objective is to minimize the message transfer delay. For simplicity of presentation, we ignore the transmission delay and propagation delay at the edges. If a message arrives at node *A* at time 0, the optimal route to node B is via contact 1 of edge e_1 which has the minimum delay of 5. If a message arrives at time 15, however, the optimal route would be via *contact* 3 of edge e_2 since contact 1 is no longer available. To compute the shortest (or minimum delay) paths in a DTN graph, Jain et al. [12] develop a modified Dijkstra's algorithm. The key difference in the modified algorithm is to take into account of the time of message arrivals at each node and only consider contact opportunities after message arrivals. In this paper, we assume that this algorithm is used to compute routes in a DTN graph and use the term "shortest path" and "minimum delay" of a message to refer to the path computed by this algorithm and the corresponding delay of forwarding the message along the computed path respectively. Readers should refer to [12] for more details about this algorithm.

3. MULTICAST SEMANTIC MODELS

In this section, we will present three new semantic models for multicasting in DTNs[2]. As discussed earlier, due to large transfer delays in DTNs, group membership may change during a message transfer, introducing ambiguity in multicast semantics. Under these situations, it is necessary to make a distinction between *group members* and the intended receivers of a message, i.e., end-points to which the message should be delivered. Group members may change with time as endpoints join and leave the group. The intended receivers, on the other hand, should be fixed for a message, even though they are defined based on group membership.

We develop three multicast semantic models that allow users to explicitly specify temporal constraints on group membership, unambiguously defining the intended receivers of a message. These models also specify constraints on the action of message delivery and have important applications in DTN environments.

3.1 Temporal Membership Model

To determine the receivers of a multicast message, we need to explicitly specify the time during which the intended receivers are defined. One straightforward approach is to define the receivers of a message as the group members at the time of message generation. In this paper, we consider a more general semantic model, called Temporal Membership (TM). In the TM model, a message includes a membership interval that specifies the period during which the group members are defined. For a message with group ID *G* and membership interval t_1 t_2, the intended receivers of the message consist of endpoints who are members of group *G* at any time during period t_1 t_2 [3]. Under the TM model, the receivers of a message are well defined. In the TM model, there is no delivery constraint so messages can be delivered at any time. Note that the intended receivers of a message may be different from the actual receivers which actually receive the message. The actual receivers are a subset of the intended receivers and dependent on the routing algorithm used and the traffic condition in the network.

The TM model allows users to flexibly specify the time-based characteristics of the receiving group of a message, which has some interesting applications in DTNs. One potential application of the TM model is in mobile sensor networks where mobile sensors record sensory data along their movement trajectories. Each region of interest is associated with a multicast group and sensor nodes join or leave multicast groups based on their locations. To query the

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

status of a given region during a specific period, a user can send a multicast message to the group that is associated with the region with a specific membership interval. The message will be delivered to sensors that are in the region during the specified period.

Consider an example in Fig. 2 which shows a DTN at time 0 when a message is generated at node S. If the membership interval of the message is $0\ 1$, the intended receivers are $R_1\ R_2\ R_3\ R_4$. If the membership interval is $15\ 20$, the intended receivers become $R_1\ R_3\ R_4$ since R_2 is no longer a group member during this period

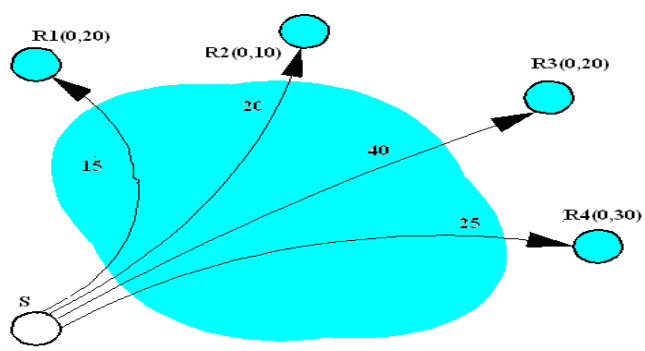


Figure 2: An example of multicast semantic models. The figure shows a DTN at time 0 when a message for group G is generated at node S. The dashed lines are the shortest paths from S to other endpoints with the minimum delay shown along each path. The time interval during which an endpoint is a member of group G is shown next to the endpoint

32 Temporal Delivery Model

Our second model is the Temporal Delivery (TD) model. In this model, messages specify additional constraints on the action of message delivery beyond the unconstrained TM model. A message specifies both a membership interval and a delivery interval. The delivery interval indicates the time period during which the message should be delivered to the intended receivers, as will be defined below. Note that the message can be delivered to nodes hosting the intended receivers *before* that period since nodes can delay forwarding the message to endpoints[4] .To be consistent with this delivery constraint, the intended receivers of a message should *exclude* endpoints that are not able to receive the message during the delivery interval. Let R be the set of all endpoints in the network and $member\ r\ t\ \ t$ be a predicate on whether endpoint r is a group member during period $t\ \ t$. Let t_0 be the message generation time and $d\ t\ r$ be the minimum delay from the source of the message to endpoint r starting at time t. As described in Section 2, $d\ t\ r$ can be computed using the modified Dijkstra's algorithm in [12]. For a message with group ID G, membership interval $t_1\ \ t_2$ and delivery interval $t_3\ \ t_4$, the set $I_{T\ D}$ ofintended receivers is defined as

$$I_{T\ D}\quad r\ member\ r\ t_1\ t_2\ \ true\ \textbf{and}\ d\ t_0\ r\quad t_0\qquad t_4\ r\quad R \quad (1)$$

Note that while the delivery interval specifies that the message be delivered no earlier than t_3, the definition of $I_{T\ D}$ does not require an earliest time for the message to reach a node. This is because nodes can delay forwarding the message to endpoints. The TD model is more general than the TM model, which is a special case with delivery interval $t_0\ \ \infty$.

The TD model enables users to have additional control on when messages are delivered. In addition, a delivery interval specifies an expiration time for a message. This enables routing algorithms to remove messages that are not able to meet the delivery intervals and reclaim storage space, which is crucial in DTNs since nodes may need to buffer messages for a significantly long period.
Consider an example using Fig. 2. For a message with membership interval $0\ 1$ and delivery interval $0\ 35$, the intended receivers are $R_1\ R_2\ R_4$. R_3 does not meet the delivery interval since it could receive the message no earlier than time 40, hence R_3 is not an intended receiver of this message.

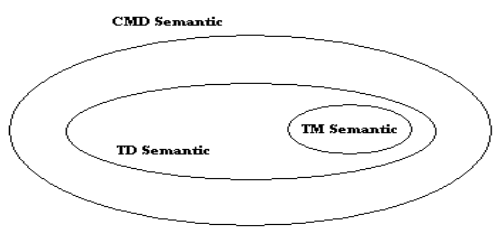


Figure 3: DTN multicast semantic models

3.3 Current-Member Delivery Model

In both the TM and TD models, receivers of a message

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

are not required to be group members at the time of message delivery. In our third model, the Current-Member Delivery (CMD) model, messages explicitly specify whether this requirement should be met. A message includes a CMD flag as well as a membership interval and a delivery interval. When the CMD flag is set, the receivers of the message should be group members at the time of message delivery. In addition, the message should be delivered during the delivery interval as in the TD model. When the CMD flag is not set, the CMD model reduces to the TD model, thus the CMD model is a more general model. Fig. 3 depicts the relationship among these semantic models.

We now define the intended receivers I_{CMD} of a message in the CMD model. When the CMD flag is set, I_{CMD} should exclude endpoints that are not able to be group members at the time of message delivery. Using the same notations as in the previous section, we define I_{CMD} as follows

$$I_{CMD}\ r\ r\ I_{TD}\ \text{and}\ member\ r\ t_m\ t_4 \quad true \qquad (2)$$

where t_m is max $d\ t_0\ r\ t_0\ t_3$, the earliest time that the message could be delivered to endpoint r because of the transfer delay from the source to r and the delivery interval constraint. In order to meet the CMD constraint, it is necessary that r be a group member during period $t_m\ t_4$ .Consider an example using Fig. 2. For a message with membership interval 0 1 , delivery interval 0 35 and the CMD flag set, the intended receivers are $R_1\ R_4$. R_2 is not an intended receiver because it could not be a group member at the time of message delivery which is at least time 20.

4. MULTICAST ROUTING FRAMEWORK

Given these semantic models, we now turn to the problem of multicast routing in DTNs. In this section, we present a framework for evaluating multicast routing in DTNs. This is an extension of the framework for unicast routing developed in [12]. Routing algorithms generally use various information about network conditions to achieve better performance. Due to network partitions in DTNs, however, there might not be complete or current knowledge available , degrading routing performance. In this paper, we study the fundamental trade-off between the amount of available knowledge and the achieved performance. To model the availability of knowledge, we use abstract *knowledge oracles* that encapsulate particular knowledge about network status to be used in routing algorithms [12]. While the use of knowledge oracles does not consider how such knowledge is actually disseminated in the network and how much overhead it causes, this approach isolates the effects of knowledge availability on routing performance, which would provide insight to guide the design of routing protocols in DTNs.

In the following, we first discuss the routing objectives in DTNmulticasting. We then describe various knowledge oracles and present an overview of four routing approaches.

4.1 Routing Objectives

For any routing algorithm, a basic objective is to maximize the probability of delivering messages. In this paper, we evaluate multicast routing algorithms by the message delivery ratio which is the ratio between the number of endpoints that receive a message and the number of intended receivers of the message according to the semantic model used. This metric measures how successful a routing algorithm is in delivering messages. In addition, we define the routing efficiency of an algorithm as the ratio between the total amount of delivered messages and the total amount of traffic generated in the network. This metric measures how efficient a routing algorithm is in utilizing resources. In this paper, we study several classes of routing algorithms that are expected to achieve different balance of delivery ratio and routing efficiency. For each class of algorithms, we focus on minimizing the delay for each intended receiver.

4.2 Knowledge Oracles

We consider knowledge oracles about both contact opportunities and group membership. Contact oracles provide information about network topology, while group membership oracles answer questions about group dynamics, e.g., the events of an endpoint joining or leaving a group. Due to space limits, we consider the following oracles in this paper. More oracles are considered in a longer version of this paper [24].

Contact Summary Oracle. This oracle can answer questions about the long-term statistics regarding network topology, i.e., average time between contact occurrences and average contact duration.

Complete Contact Oracle. This oracle can answer any question about network topology at any time, including the exact time when a contact occurs, the duration, capacity and delay of the contact.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Delayed Membership Oracle. For an endpoint r and a node S that queries the oracle, this oracle can answer questions about membership of endpoint r up to a specific time t. t is the latest time that satisfies $d\ t\ S\ \ t\ \ t_0$ where t_0 is the current time and $d\ t\ S$ is the minimum delay from endpoint r to node S starting at time t[5]. In other words, if endpoint r joins or leaves a group at or before time t and sends this information to other nodes by flooding, assuming no con- tending traffic in the network, node S should have received this information by the time of querying the oracle.

Complete Membership Oracle. This oracle can answer questions about group membership of all nodes at any time.

4.3 Routing Approaches

We now describe four approaches for multicast routing in DTNs which are adopted from multicasting in the Internet or MANETs. Fig. 4 depicts simple examples of these approaches.

Unicast-Based Routing *(UBR).* This approach implements multicast service by using unicast transfer, i.e., the source will send a copy of the message to every intended receiver.

Broadcast-Based Routing (BBR). In BBR or epidemic routing [20], messages will be flooded throughout the network in order to reach the intended receivers.

Tree-Based Routing (TBR). In TBR, messages are forwarded along a tree in the DTN graph that is rooted at the source and reaches all receivers. Messages are duplicated only at nodes that have more than one outgoing path.

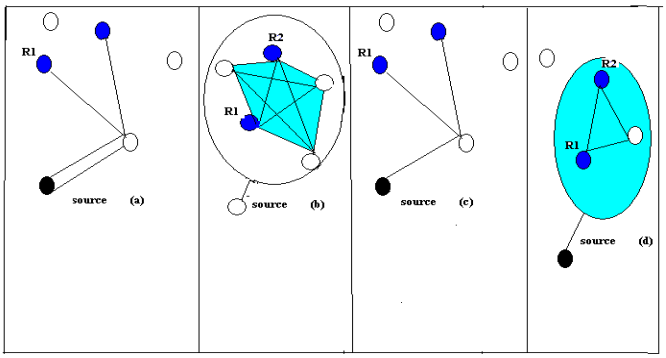


Figure 4: Routing approaches in DTNs. (a) unicast-based routing (b) broadcast-based routing (c) tree-based routing (d) group-based routing.

Group-Based Routing (GBR). GBR uses the concept of *for- warding group* [6] which is a set of nodes that are responsible for forwarding the message. Messages will be flooded within the forwarding group to increase the chance of delivery

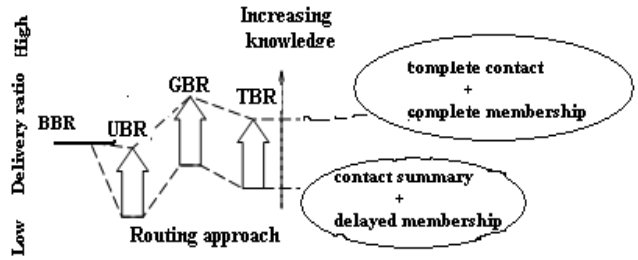


Figure 5: Conceptual performance of various routing approaches under different levels of knowledge

Fig. 5 summarizes the conceptual performance of various routing approaches under different levels of available knowledge. BBR is expected to achieve the same delivery ratio under different amount of available knowledge. The expected delivery ratio of other approaches (i.e., UBR, GBR and TBR) would improve with the in- creasing knowledge. GBR is expected to perform best while UBR is the worst. These approaches would also achieve different routing efficiency.

5. MULTICAST ROUTING ALGORITHMS

In this section, we describe multicast routing algorithms for DTNs based on the four routing approaches and various knowledge oracles described in the previous section. We first describe the common operations and then the specifics of each algorithm. More details can be found in [24].

General Operations

The general operations of these algorithms are sketched in Fig.6. When a message arrives, either generated by a local endpoint or received from another node, a node estimates the intended receivers of the message. If there are local intended receivers, the message is forwarded to these receivers according to the semantic model. The message is then buffered in node storage and forwarded to other nodes when contacts become available. Since data transfer in

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

DTNs is based on messages, nodes maintain forwarding state for each buffered message, which is updated as group membership changes. We now describe how nodes buffer messages, maintain forwarding state and forward messages

Message Buffering

In DTN multicast routing, messages will be buffered in node storage until being deleted due to buffer overflows or being expired according to the semantic models. This improves the availability of messages such that nodes other than the source can handle join requests and send buffered messages to new receivers. In this paper we adopt an age-based buffering policy which removes the oldest message when the buffer overflows, thus giving new messages opportunities to be delivered.

5.1.2 Forwarding State

Nodes maintain local forwarding state for each message buffered in node storage. Each message is associated with a NEXT-HOP list L_n that records nodes to which this message should be sent, and a SENT list L_s that consists of nodes that have already received thismessage. L_n is initialized upon the message arrival and updatedwhen group membership changes. To compute L_n, nodes need toestimate the intended receivers of a message based on the current available knowledge. We will describe how to calculate L_n for various algorithms in the next section. L_s is initially set to empty. In the rest of this paper, we use L_n m and L_s m to denote the NEXT-HOP and SENT lists for message *m* respectively.

5.1.3 Message Forwarding

Given message forwarding state L_n m and L_s m, nodes for- ward messages as follows. Suppose that a contact between node *A* and *B* becomes available. For each buffered message *m*, node *A* will try to forward message *m* to node *B* if *B* is in L_n m and not in L_s m. In other words, node *B* should be a next hop for this message and node *A* has not transmitted this message to node *B* before. After transmission, node *A* will add node *B* into L_s m. So node *A* will not send duplicate messages to node *B*. However, node *B* may still receive duplicate messages from different nodes, due to network dynamics or message flooding. To address this problem, nodes exchange control information first to determine which messages should be sent[6]. Specifically, node *A* will first send an ADV message including information about messages it wants to transmit.Upon reception of the ADV message, node *B* replies with a REQ message which lists only messages it currently does not have. Then *A* will send the messages listed in the REQ message. Since meta data is much smaller in size than the actual message, the overhead of ADV and REQ messages is generally not significant.

5.2 Specific Operations

In the following, we present the specific operations of each multicast routing algorithm.

5.2.1 Static Tree-Based Routing (STBR)

In STBR, nodes construct a shortest path tree in the DTN graph from the source to the estimated intended receivers of a message starting at the message generation time. L_n of a message would include nodes that are the next hops in the tree. This can be computed using the modified Dijkstra's algorithm in [12]. As group member-ship changes, nodes update the shortest path tree and L_n. Messages are then forwarded along the tree.

```
1. On arrival of message m
     Forward m to local receivers if any; Insert m in node storage;
     Initialize forwarding state Ln  m  and Ls  m ;
2. On contact with node B
     For each message m in storage
     IF B  Ln  m  and B  Ls m
        Send a copy of m to node B; Add B into Ls  m ;
3. On join/leave request for group G
     Update Ln  for messages destined for G;
```

Figure 6: Message forwarding.

In STBR, the route from the source to an intended receiver is static. Thus if a message misses a contact with a node in L_n, the message needs to wait for the next opportunity to connect to this node, which may significantly increase the message delay. In addition, the use of static routes disallows nodes to utilize local or more accurate information to forward messages along better paths.

5.2.2 Dynamic Tree-Based Routing (DTBR)

DTBR addresses the above problems with STBR using the explicit addressing approach [2], i.e., messages include the endpoint IDs of the receivers as well as the group ID in the message header. This way, nodes can determine the next-hops of a message dynamically based on current available information, such as local queuing information or newly

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

available contact information. Specifically, DTBR determines L_n by computing the shortest paths from the current node to endpoints embedded in the message. Nodes which are the next hops in these paths are added to L_n. When a message is forwarded, each copy of the message contains only the IDs of end- points to which it will be delivered.

5.2.3 Group-Based Routing (GBR)

In GBR, nodes construct a forwarding group for each message by computing a shortest path tree as in STBR and setting the forwarding group as the set of nodes in the tree including the receivers. Messages are then forwarded by flooding within the forwarding group. In other words, L_n consists of all nodes in the shortest path tree.

5.2.4 Broadcast-Based Routing (BBR)

In BBR, L_n always includes all nodes in the network. So messages are flooded throughout the network.

5.2.5 Unicast-Based Routing (UBR*)*

In UBR, when a multicast message is generated, the source node sends a unicast message, which encapsulates the original multicast message, to each of the estimated intended receivers. The source node also buffers the multicast message and sends out new unicast messages when being informed of new intended receivers. In this paper, we assume that unicast messages are forwarded using the shortest paths to the destinations. Unicast messages are removed from node storage after being transmitted to the next hop. Upon receiving a unicast message, the destination node will decapsulate the message and forward the original multicast message to the in- tended receiver according to the delivery constraints of the specified semantic model.

6. PERFORMANCE EVALUATION

In this section, we evaluate various multicast routing algorithms using *ns* simulations. We aim to compare these routing algorithms and understand how the availability of knowledge affects routing performance. We simulate a specific type of DTNs, *sparse mobile networks* that consist of mobile nodes communicating via wireless radios. In these networks, nodes are sparsely distributed such that the net- works experience frequent and long-duration partitions. We implement the four classes of routing algorithms and the contact and membership oracles in the *ns* simulator. Our simulations use the IEEE 802.11 MAC layer. The radio range and data rate are 250*m* and 2Mbps respectively. We use the following default settings unless specified otherwise. All simulations have 40 nodes on a 5000*m* 5000*m* area. Nodes move in the area according to the Random Way-Point (RWP) model [13] with a maximum speed 5m/s and a minimum speed 1m/s. The node storage capacity is 400 messages. In order to discover other nodes for communication, each node sends out beacon messages every 3 seconds. Each simulation lasts for 10000 seconds and each result is averaged over five runs with random seeds.

To understand how these routing algorithms perform, we consider only multicast traffic in the network. By default, there are4 multicast sessions and each session consists of a single source which transmits messages to a multicast group. Each multicast group has 10 potential members which join and leave the group dynamically. Both the source and the potential group members are chosen randomly. Messages are generated at each source ac- cording to a Poisson process with mean inter-arrival time 4 seconds. Each message has 1000 bytes, thus the traffic rate of each source is 2kbps. After an endpoint joins (leaves) a group, it will leave (join) the group after a duration that is exponentially distributed with mean 200 seconds. Messages use the TD semantic model with membership interval t_0 t_0 100 and delivery interval t_0 t_0 3000 where t_0 is the message generation time.

Simulation Results: With extensive *ns* simulations, we have evaluated these routing algorithms under different traffic rates, numbers of sessions, session sizes, mobility patterns, node buffers and se- mantic models. Due to space limits, we present only the results under different traffic rates in this paper, which illustrate the representative performance. Please refer to [24] for more details. In these simulations, the average message inter-arrival for all sources combined varies from 16, 4, 2, 1, to 0.5 seconds. Thus the total traffic load ranges from 0.5, 2, 4, 8, to 16 kbps. We first compare the performance between various algorithms. Fig. 7(a) and (b) show the delivery ratio when different oracles are used. We make the following observations. First, the delivery ratio decreases for all algorithms as the traffic load increases, which is as expected. Second, among the routing algorithms that utilize knowledge in computing routes, GBR achieves the best performance. This is be- cause in GBR, messages may be forwarded to receivers via multiple paths, which is better in exploiting available contact opportunities. UBR, on the other hand, has the worst delivery ratio because a separate unicast message is sent to each receiver which significantly in- creases contention for node storage and transmission opportunities, and results in message drops. This result

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

confirms the intuition that providing multicast service by sending multiple unicast messages is very inefficient in DTNs. The performance of both DTBR and STBR is between that of GBR and UBR. Since DTBR can adapt to network conditions, it performs slightly better than STBR. Third, GBR and BBR achieve the highest delivery ratio, depending on the level of available knowledge. GBR typically performs best when complete contact knowledge is available, while BBR achieves the best delivery ratio when the contact summary oracle is used. Both BBR and GBR utilize some form of flooding in message forwarding, which suggests that forwarding messages via multiple paths is a promising approach to achieve high delivery ratio in DTNs.We now study how each routing algorithm performs under different amount of available knowledge.the results for GBR, which are representative of other algorithms that utilize knowledge, i.e., UBR, STBR and DTBR. The labels "CC- xM"("SC-xM") in the figure represent scenarios where the complete contact (contact summary) oracle is used. We can see that the availability of up-to-date membership or exact contact knowledge has significant effect on routing performance. GBR performs poorly when such knowledge is not available. This suggests that a minimum amount of knowledge is required to achieve efficient routing for these approaches. In addition, the marginal improvement in performance for accurate contact information is more significant than that for up-to-date membership information.

We also evaluate the routing efficiency of various algorithms. Fig. 8(a) illustrates the results when the complete contact and complete membership oracles are used. We can see that BBR,which uses flooding to forward messages, has the lowest routing efficiency because it generates many redundant messages. Thus BBR is not suitable for mobile networks where nodes are equipped with limited power supplies. UBR is also inefficient in utilizing resources since it sends a separate copy of a multicast message to every receiver. STBR and DTBR achieve the best routing efficiency among all algorithms. The routing efficiency for GBR is slightly lower than that of the TBR algorithms. Fig. 8(b) depicts the routing efficiency when the contact summary and complete membership oracles are used. GBR achieves better efficiency than both TBR algorithms in this case. In addition, UBR achieves the highest routing efficiency, which, however, is obtained with a very low delivery ratio. Fig. 8(c) shows the average delay for delivered messages when the complete contact and complete membership oracles are used. We can see that for all algorithms, the message delay decreases as the traffic load increases. This is because as the network becomes more congested, messages of the same age are more likely to be removed from node storage. Thus messages tend to reach only receivers that are on a shorter forwarding path, resulting in lower message delay. BBR achieves slightly lower delay than other algorithms because messages are flooded to all nodes and it is more likely that messages follow a shorter path to the receivers. The de- lay of GBR is slightly larger than that of STBR and DTBR.

7. RELATED WORK

In this section, we review some related work on DTNs and multicasting in traditional networks. DTNs are a class of emerging net- works that experience frequent and long-duration partitions, such as military ad hoc networks [1], deep space communication [4] and vehicular communication [22]. To achieve interoperability between various types of DTNs, Fall [9] proposes an architecture that is based on an asynchronous message forwarding paradigm. This architecture operates as an overlay above the transport layers to connect different DTNs. Routing in frequent-disconnected networks has been studied relatively recently. In [12], Jain et al. study unicast routing in DTNs and develop several routing algorithms for scenarios where different levels of knowledge about network are available. The authors present a framework to evaluate these algorithms and find that efficient routing can be achieved using only limited amount of know edge. There is also other work that focuses on sparse mobile net- works and exploits node mobility to deliver data. For example, in Epidemic Routing [20], mobile nodes carry data and exchange data when they meet, essentially flooding data throughout the network. The Data Mules project [19] exploits mobile entities for data transportation to conserve energy in sensors. In the Message Ferrying project [23], special nodes called message ferries are used to provide communication services and controlled mobility is exploited to improve routing performance. Other work includes [10, 14, 21]. Multicasting has been studied extensively in the past, both in the Internet and in MANETs. Deering and Cheriton [7] first introduce the concept of IP multicasting. IP multicast assumes an open group model in which sources do not need to know the group membership or be group members to send data to a group. In addition, nodes can join or leave a multicast group at will. Various multicast protocols have been developed for the Internet, including DVMRP, MOSPF, PIM and CBT [3, 7, 8, 16]. These protocols construct a multicast tree to forward packets, using either a broad- and-prune (dense mode) or an explicit join (sparse mode) mechanism. In MANETs, node mobility introduces frequent topological changes which is different from the wired Internet. In addition, MANETs are resource-constrained in terms of bandwidth and en- ergy supplies. MANET multicast protocols (e.g., FGMP, MAODV and

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

ODMRP [6, 15, 18]) have been designed to address these is- sues, which normally use on-demand routing, localized repair of broken paths, and a mesh structure for packet forwarding.

In this paper, we study multicasting in DTNs. The semantic models we developed are based on the open group model used in IP multicasting but with additional temporal constraints to uniquely identify receivers of a message. Multicast routing in DTNs shares some commonalities with that in MANETs, e.g., dynamic topology and limited resources. On the other hand, DTNs differ from MANETs in the following aspects, namely frequent partitions and large message delivery delay. These factors affect not only data forwarding, but also the dissemination of control information. It would be interesting to study how MANET routing protocols or techniques can be adapted to DTNs, which is a topic of our future work.In [11], Huang et al. propose mob cast for sensor networks in which applications can specify spatiotemporal constraints on a mobile delivery zone for a packet. In contrast, the multicast models proposed in this paper define the intended receivers of messages, which are time-invariant, by specifying temporal constraints on group membership, instead of geographic regions that change over time.

8.CONCLUSION

In this paper, we studied the problem of multicasting in DTNs. We developed three multicast semantic models that allow users to explicitly specify temporal constraints on group membership and message delivery. These semantic models unambiguously define the intended receivers of messages and have various applications in DTN environments. We then developed four classes of routing algorithms for DTNs with different routing strategies. With extensive *ns* simulations, we compared these multicast algorithms and studied how routing performance is affected by the availability of knowledge. Based on our simulations, we obtained the following results. First, efficient routing for multicast can be constructed using only partial knowledge. In addition, the marginal improvement in performance for accurate contact information is generally more significant than that for up-to-date membership information. Second, GBR and BBR achieve the best delivery ratios depending on the amount of knowledge available. Both algorithms use some form of flooding, which suggests that forwarding messages along multiple paths is a promising approach for multicasting in DTNs. Third, UBR performs poorly in DTNs, confirming that multicast routing using multiple unicast messages is not efficient in DTNs.

In this paper, we studied the impact of available knowledge on routing performance using knowledge oracles which do not consider the overhead of disseminating such knowledge in the network. We are currently studying information dissemination in DTNs and how it might affect routing performance. We plan to develop multi- cast routing protocols for DTNs based on the semantic models and routing algorithms presented in this paper. In addition, we would like to extend our evaluation of the multicast algorithms in different networks, e.g., environments where node mobility is more predictable [12] or follows power-law distributions [5]. Furthermore, we are interested in extending our semantic models to incorporate spatial constraints as in geocast or mobicast.

9. References

1. L. Aguilar. Datagram routing for internet multicasting. In *Acmsigcomm84,* 1984.
2. T. Ballardie, P. Francis, and J. Crowcroft. Core Based Trees (CBT): An architecture for scalable inter-domain multicast routing. In *Acmsigcomm*, Sept. 1993.
3. S. Burleigh, A. Hooke, L. Torgerson, K. Fall, V. Cerf, B. Durst, K. Scott, and H. Weiss. Delay-tolerant networking – an approach to interplanetary internet. *IEEE Communications Magazine*, June 2003.
4. A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott.Pocket switched networks: Real-world mobility and its consequences for opportunistic forwarding. Technical Report *UCAM-CL-TR*-617, University of Cambridge, Computer Lab, February 2005.
5. C.-C. Chiang, M. Gerla, and L. Zhang. Forwarding group multicast protocol (FGMP) for multi hop, mobile wireless networks. *Cluster Computing*, 1(2):187–196, 1998.
6. S. Deering and D. Cheriton. Multicast routing in datagram internetworks and extended LANs. *ACM Transactions on Computer Systems (TOCS)*, 1990.
7. S. Deering, D. Estrin, D. Farinacci, V. Jacobson, C.-G. Liu, and L. Wei. An architecture for wide-area multicast routing. In *ACM SIGCOMM 1994*, 1994.
8. Fall. A delay-tolerant network architecture for challenged internets. In *ACM Sigcomm 2003*, 2003.
9. A. A. Hasson, R. Fletcher, and A. Pentland. DakNet: A road to universal broadband connectivity. *Wireless Internet UN ICT Conference Case Study*, 2003.
10. Q. Huang, C. Lu, and G.-C. Roman. Spatiotemporal multicast in sensor networks. In *ACM SenSys 2003*, pages 205– 217, 2003.
11. S. Jain, K. Fall, and R. Patra. Routing in a delay tolerant network. In *ACM Sigcomm 2004*, Portland, OR, 2004.
12. D. Johnson and D. Maltz. Dynamic source routing in ad-hoc wireless networks. In *ACM SIGCOMM*, August 1996.
13. P. Juang, H. Oki, Y. Wang, M. Martonosi, L. Peh, and D. Rubenstein.Energy-efficient computing for wildlife tracking: design

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

tradeoffs and early experiences with ZebraNet. In *ASPLOS'02*, October 2002.

14. S. Lee, M. Gerla, and C. Chiang. On-demand multicast routing protocol (ODMRP) for ad hoc networks. *Internet draft*, June 1999.

15. E. Royer and C. Perkins. Multicast operation of the ad-hocon-demand distance vector routing protocol. In *ACM Mobicom*, Seattle, WA, Aug. 1999.

16. R. Shah, S. Roy, S. Jain, and W. Brunette. Data MULEs: Modeling a three-tier architecture for sparse sensor networks. In *IEEE SNPA Workshop*, 2003.

17. A. Vahdat and D. Becker. Epidemic routing for partially-connected ad hoc networks. *Technical report, Duke University*, 2000.

18. R. Wang, S. Sobti, N. Garg, E. Ziskind, J. Lai, and A. Krishnamurthy. Turning the postal system into a generic digital communication mechanism. In *ACM SIGCOMM 2004*, August 2004.

19. H. Wu, R. Fujimoto, and G. Riley. Analytical models for data dissemination in vehicle-to-vehicle networks. In *IEEE VTC2004/Fall*.

20. W. Zhao, M. Ammar, and E. Zegura. A message ferrying approachfor data delivery in sparse mobile ad hoc networks. In *ACM MobiHoc2004*, Tokyo Japan, 2004.

21. W. Zhao, M. Ammar, and E. Zegura. Multicasting in delay tolerant networks: *Semantic models and routing algorithms.* *Technical report*, College of Computing, Georgia Institute of Technology, 2005.

22. DARPA Disruption Tolerant Networking Program:*http://www.darpa.mil/ato/solicit/dtn.*

23. J. Moy. Multicast extensions to OSPF. *RFC 1584*, Mar. 1994. [17] Network simulator *http://www.isi.edu/nsnam/ns/.*

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

An analytical model for Energy Consumption in Y-MAC Protocol

Ramchand V. and D.K.Lobiyal
School of Computer and Systems Sciences, Jawaharlal Nehru University
New Delhi, 110067, India.

Abstract

This paper presents an analytical model for estimating energy consumption in Y-MAC protocol for Wireless Sensor Networks (WSN). This protocol has been proposed to reduce the energy consumption of a node when involved in unicast or broadcast transmission of frames. In WSN a node consumes energy in transmitting and receiving of data, listening transmissions of other nodes, and in sleep mode. Therefore, energy consumption of a node has been estimated by adding up the energy consumed in each of the above activities. This has been achieved by estimating the time spent in each activity by a node. The protocol has been simulated using MATLAB. The simulation results show comparable energy savings as compared to Y-MAC protocol, therefore, the model validates the existing protocol.

Keywords: *Wireless sensor networks, Time Division Multiple Access, Poisson distribution, Binomial distribution, Energy efficiency.*

1. Introduction

Wireless Sensor Networks is a special kind of ad hoc networks. WSN is a technology that has grown very rapidly over the years with an increase in development of new protocols or methods to fulfill the needs. Though many protocols and methods were designed still WSN needs more contributions from researchers to meet the current requirements. Sensor Networks covers wider region, nodes have limited sensing region and processing power. Sensors consume energy in sensing the object, processing and transmitting. Energy consumed in processing is insignificant. However, energy consumed in other two tasks is significant that may result in partitioning of networks. WSN have been broadly applied in ecological, biological, event monitoring in remote areas, defense, health monitoring etc. WSN does not follow any infrastructure and sensor nodes in the network are equipped with battery, significant processing and wireless communication. Nodes are inexpensive with low data rate and are deployed in large numbers. Time synchronization in Wireless Sensor Networks has been studied in the literatures, by researchers yet there is no specific scheme to achieve accuracy. Applications of Sensor Networks require time for node to be synchronized within the network. Power control of a node is a vital issue in Wireless Sensor Networks. In this paper we propose Synchronization and power control model for Y-MAC protocol. Here we calculate the delay of a node involved in unicast and broadcast transmission and power consumption of nodes involved in these transmissions.

The rest of this paper is organized as follows. We discuss the related work in section 2. In section 3 design of proposed model is explained. Section 4 presents the simulation results. Finally, we conclude the paper in section 5.

2. Related Work

Numerous multichannel media access control protocols have been designed for ad hoc networks. However these protocols are not suitable for wireless sensor networks due to limited resources, power and cost. Y-MAC protocol was designed by **Y**oungminkin and implemented in **Y**onsei University and hence named as Y-MAC [6]. It is a multichannel media access control protocol designed with the help of Time Division Multiple Access method to support both unicast and broadcast transmission of data. It uses different time slot as transmit time slot and receive time slot to guarantee collision free access to the channels and reduces energy wastage to some extent. In this protocol every node in the network synchronizes their timing for upcoming events by exchanging the remaining time when transmitting in the current super-frame period among neighbors. It has been designed to work with heavy traffic conditions where several sink nodes are deployed in the network. Whenever a network partition is detected re-association is done by nodes where every node periodically turns on its radio to perform bootstrap. Timing Synchronization protocol for sensor networks is a multi-hop protocol attempts to reduce the uncertainty of sender node with low level time stamping packets in media access control layer. Greunen and Rabaey propose the Light weight Tree-based Synchronization protocol that decreases the complexity of synchronization, and the accuracy is achieved with minimal complexity [3].

Jeremy Elson and Deborah Estrin's propose Time Synchronization for wireless sensor networks where Post-facto synchronization method is designed for extremely low power consumption [4]. Gradient Time Synchronization Protocol (GTSP) is a distributed time synchronization protocol. In gradient time synchronization protocol the logical clock of nodes converges to a common logical clock and this protocol relies on local information making it robust to node

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

failures and changes in the network topology. The remaining synchronization error between neighbors is reduced while still maintaining acceptable global skew. Nodes in gradient time synchronization protocol can convert the current hardware clock reading into a logical clock value and vice versa [5].

MAC layer has a vast impact on the energy consumption of sensor nodes, because of radio communication. Communication is a major source of power consumption and the MAC layer design manages the transmission and reception of data over the wireless medium using the radio The MAC layer is responsible for access to the shared medium. MAC protocols assist nodes in deciding when to access the channel. Pattern-MAC (P-MAC) for sensor networks adaptively determines the sleep-wake up schedules for a node based on its own traffic, and the traffic patterns of its neighbors. This protocol achieves a better throughput at high loads, and conserves more energy at light loads. In P-MAC, the sleep-wake up times of the sensor nodes is adaptively determined. The schedules are decided based on a node's own traffic and that of its neighbors. The improved performance of P-MAC suggests that 'pattern exchange' is a promising framework for improving the energy efficiency of the MAC protocols used in sensor networks [6].

3. Proposed Work

The total power consumption of a node is the combination of overheads introduced due to hardware design and software complexity. Solution to hardware overhead is up to the engineers who design sensor devices. Software overheads occur because of delay in context switching and execution time. The time synchronization between the sender clock and the receiver clock may have a symmetrical nominal delay. Following is algorithm for calculating nominal delay.

3.1 Symmetrical Nominal Delay Algorithm

Step 1: Receiver calculates nominal delay
Step 2: Sender calculates nominal delay
Step 3: Nominal delays are calculated by all nodes, and delays are stored and forwarded to one hop neighbors. Nominal delay packet has information such as Sender ID, Receiver ID, time delay and Next sleep time of a node.
Step 4: Nominal delays are calculated after fixed time interval and disseminated to neighbors to alert them for future transmission.

Nominal delay calculation by receiver.

1. Node A transmits a request packet to node B at r_1 µsec
2. Node B receives request packet at r_2 µsec
3. Node B calculates and stores the nominal delay (ND) on its delay table

Nominal delay calculation by sender.

1. Node B transmits a response packet to node A at t_1 µsec
2. Node A receives response packet at t_2 µsec
3. Nodes A calculate and store the nominal delay (ND) on its table

This protocol is designed to supports unicast & broadcast communications with multichannel transmission. The communication may happen between sender and the nearest devices in the range of receivers in the network. Events occurring are independent of time. Traffic may be control packets or data packet. Traffic may be exponentially increasing or decreasing. Every node in a network senses the events in a given area and forwards its data to the group of neighbors. The sensed data which is to be transmitted may exceed the maximum frame size. In this case a node sends frames by dividing it into different frames of equal size.

3.2 Power level and exchange of control packets and frames

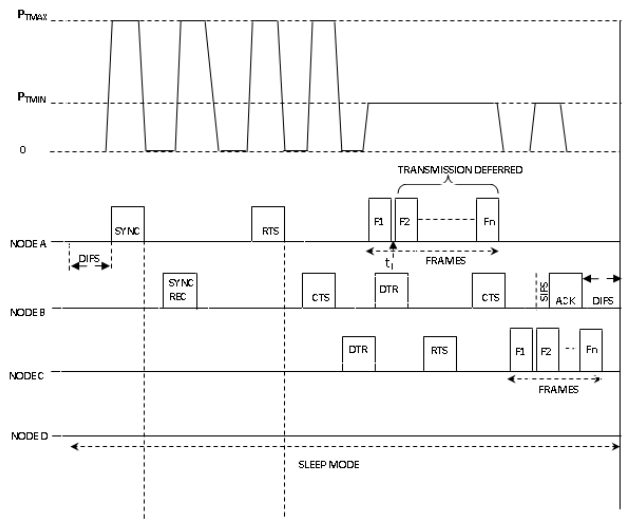


Fig. 1 Power level of nodes

In the above figure node A sense the status of the wireless medium for DIFS duration. If it found that the medium is continuously idle for DIFS duration then it transmit SYNC packet to node B. Node B records the SYNC information. Once the synchronization is done Node A sends RTS packet to Node B with maximum transmission power level P_{TMAX}. Node B calculates the minimum required power, and replies the CTS packet with minimum power level. But the minimum power level should lie within the nominal limits of transceivers. The ratio of the received power and the transmission power is given by [1]

$$\frac{P_r}{P_t} = \left[\sqrt{Gl\lambda} \Big/ 4\pi d \right]^2 \qquad (1)$$

Where $\sqrt{Gl\lambda}$ is product of transmit and receive antenna field radiation patterns in the Line of Site direction, λ is the

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

wavelength, and d is the distance. Minimum transmission power of a node is given by

$$P_{TMIN} \geq \left[\Pr_{threshold} * 16\pi^2 d^2 \Big/ Gl\lambda \right] \quad (2)$$

There are 'm' sensors are distributed uniformly over a circular area A.

The node density of network is m/A.

Therefore, average distance between one-hop neighbor nodes is $\sqrt{A/m}$ [11].

Let 'e' is the latent error in the node position from the center. The distance between one-hop neighbor nodes is calculated as $d = \sqrt{A/m} \pm e$

Therefore, the minimum power level of a node is

$$P_{TMIN} \geq \left[\Pr_{threshold} * 16\pi^2 \left(\sqrt{A/m} \pm e \right)^2 \Big/ \mathrm{Gl}\lambda \right] \quad (3)$$

The network is designed for high priority jobs and any node in the network is prone to receive High Priority Request (HPR). In this case the ongoing transmission is blocked and the remaining data is saved in buffer and priority is set in priority scheduler table (PST). Priority scheduler table has information such as Sender ID, Receiver ID, Priority of transmission and sleeping time. Here t_l is the time lapse between the reception of ACK of n-1th frame and beginning of transmission of nth frame where t_l is less than SIFS time. Node C has High Priority Request sent to node B. so node C is sending a deferred transmission request (DTR) to node B. Node B forwards the data to node A during t_l time and interrupts the transmission. Now B and node C involves in transmission of frames, once the transmission is over. Node B initiates a clear to send (CTS) frame to node A to receive the remaining frames. Node A checks the priority scheduler table (PST) to send the pending frames based on priority.

When a sender node broadcast a frame to group of neighbors, the frame traversing through shorter path reaches the receiver earlier than those on longer paths. The direct effect of these non uniform arrivals of frames causes delay in the whole network. The delay puts a limitation on the maximum transmission capacity on the wireless channel. Specifically, if the transmission time is larger than that of delay, inter-symbol interference will occur at the receiver. It is a type of distortion due to multipath resulting in spreading out of the received frame. It occurs when identical frame arrive via different paths and have different time delays. The sender chooses the path through which a frame reaches to the destination with minimum delay.

3.3 Delay Calculation

When a node has a packet to transmit, it competes with its neighbor nodes that are ready with packets. Therefore, it also has to consider certain delays like carrier sense and back-off delay s. These delays are the same as IEEE 802.11 protocol. A node can also suffer from packet transfer delay.

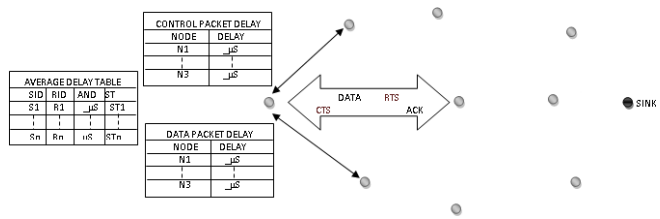


Fig. 2 Network Scenario

In wireless sensor networks events can happen at any point of time. Any node in the network can forward the sensed data to its neighbor or destination node. When a node has a frame to transmit it initiates the transmission with control packet exchange.

To transmit an RTS packet a node takes μ_R time, therefore the average arrival rate of receiving an RTS packet $\lambda = \frac{1}{\mu_R}$.

We assume that the packet arrival follows Poisson distribution. Therefore, probability of transmitting an RTS packet is calculated as

$$P(X=1) = \frac{e^{-(1/\mu_R)} * (1/\mu_R)^1}{1!} \quad (4)$$

Similarly, the probability of transmitting a CTS packet can be represented

$$P(X=1) = \frac{e^{-(1/\mu_C)} * (1/\mu_C)^1}{1!} \quad (5)$$

3.4 Round-trip delay of control packet and frame exchange

c_0 is the time of RTS transmission by a sender,
c_1 is the time of RTS reception by a receiver,
c_2 is the time of CTS transmission by a receiver and
c_3 is the time of CTS reception by a sender.
$c_e = c_3 - c_0$ is the time elapsed between the release of the RTS and the reception of the CTS packet,
$c_w = c_2 - c_1$ is the time the receiver waited before sending the CTS.
The round-trip delay of RTS and CTS exchange is

$$c_{rd} = (c_e - c_w)$$

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

f_0 is the transmission time of FRAME
f_1 is the reception time of FRAME by the receiver,
f_2 is the transmission time of an ACK and
f_3 is the reception time of an ACK.
$f_e = f_3 - f_0$ is the time elapsed between the release of the FRAME and the reception of the ACK frame,
$f_w = f_2 - f_1$ is the time a receiver wait before sending the ACK.
The round-trip delay of FRAME and ACK exchange is

$$f_{rd} = (f_e - f_w)$$

After exchange of control packets and FRAMEs, the round trip delay, total power consumption and number of frames to be transmitted is shared by both the sender and receiver. The receiver forward a REQUEST packet which has information about as round trip delay, total power consumed, and number of frames to be transmitted to one hop neighbors for unicast communication. On receiving the request packet the receiving node checks P_{AVAIL} (Power Available) and reply with ACCEPT packet. If P_{AVAIL} is less than the required one, a node goes to sleep mode. In case of broadcast communication the REQUEST packet is broadcast to the nodes in the circular region. On receiving the request, nodes which has sufficient power reply by sending the ACCEPT packet with P_{AVAIL}. Once the sender receives ACCEPT packet, it transmits the frame to those nodes which have maximum power and forms the route to the sink node. Nodes which are not involved in broadcast transmission back-off until transmission is over. In the active state a node may sense the events, then transmits the sensed events as frames or receives frame from neighbor nodes. If no events are sensed the node goes to sleep state.

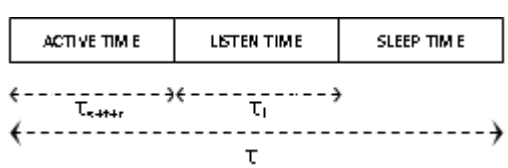


Fig. 3 Time Cycle of a node

Sleep time $= \tau - (\tau_{(s+t+r)} + \tau_l) => (1 - ((s+t+r)+l))\tau$

Probability of **transmitting** 'n' number of frames by a node to its neighbor is

$$P(T_{f=1,2...n}) = \left(\frac{e^{-(1/\mu_f)} * (1/\mu_f)^n}{n!} \right) \quad (6)$$

where $1/\mu_f$ is the average arrival rate of transmitting a frame. During transmission of frames from a source node to a destination node or a sink node, the frames may experience store and forward delays. However a first frame may not experience delay. Therefore the probability of successful transmission of first frame is

$$P(T_{f=1}) = \left(\frac{e^{-(1/\mu_f)} * (1/\mu_f)^1}{1!} \right) \quad (7)$$

As the network is designed for high priority jobs, a neighbor of the transmitting node or receiving node may block the ongoing transmission by initiating high priority request. Therefore the probability of remaining frames being queued in the buffer is

$$P(B_{f=n,...N}) = \left[1 - \sum_{i=1}^{N} \frac{e^{-(1/\mu_f)} * (1/\mu_f)^{N-n}}{N-n!} \right] \quad (8)$$

where 'n' is a frame from 2nd frame till n-1th frame and 'N' is the last frame. The probability of receiving 'k' number of frames from total of N frames can be represented by using binomial distribution

$$p_p(k/N) = \frac{N!}{k!(N-k)!} p^k (1-p)^{N-k} \quad (9)$$

$(1/\mu_f) = N * p$ Where $(1/\mu_f)$ is the average arrival rate and 'p' is the probability of successful transmission,

$$p = \frac{1/\mu_f}{N}$$

$$P_{1/\mu_f/N}(k/N) = \frac{N!}{k!(N-k)!} * \left(\frac{1/\mu_f}{N} \right)^k * \left(1 - \frac{1/\mu_f}{N} \right)^{N-K} \quad (10)$$

For large N binomial distribution becomes Poisson bistribution

$$\begin{aligned} P_{(1/\mu_f)}(k) &= \lim_{N\to\infty} p_p(k/N) \\ &= \lim_{N\to\infty} \frac{N(N-1)...(N-k+1)}{k!} * \frac{(1/\mu_f)^k}{N^k} * \left(1 - \frac{1/\mu_f}{N}\right)^N * \left(1 - \frac{1/\mu_f}{N}\right)^{-k} \\ &= \lim_{N\to\infty} \frac{N(N-1)...(N-k+1)}{N^k} * \frac{(1/\mu_f)^k}{k!} * \left(1 - \frac{1/\mu_f}{N}\right)^N * \left(1 - \frac{1/\mu_f}{N}\right)^{-k} \\ &= \left(\frac{(1/\mu_f)^k}{k!} . e^{(-1/\mu_f)} \right) \end{aligned} \quad (11)$$

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Power consumption of a node varies when involved in unicast or broadcast transmission. Power consumption of a node when involved in **unicast** transmission is expressed as

$$p_{ut} = \tau_{sn} * p_{sn} + \left(\frac{e^{-(1/\mu_R)} * (1/\mu_R)^1}{1!} \right) * P_{TMAX} + \left(\frac{e^{-(1/\mu_C)} * (1/\mu_C)^1}{1!} \right) * P_{TMAX} + \left(\frac{e^{-(1/\mu_f)} * (1/\mu_f)^n}{n!} \right) * P_{TMIN} + t_{ACK} * P_{TMIN} + \tau_i * p_i + (1 - ((s+t+r)+l))\tau * p_{sp} \quad (12)$$

τ_{sn} and p_{sn} is the time spent and power consumed in sensing the events. τ_i and p_i is the time spent and power consumed in listen mode. $(1-((s+t+r)+l))\tau$ is the time spent in sleep mode. Power consumption of a node when involved in **broadcast** transmission is given as

$$p_{bt} = \tau_{sn} * p_{sn} + \left[1 - \sum_{i=1}^{N} \frac{e^{-(1/\mu_f)} * (1/\mu_f)^{N-n}}{N-n!} \right] * p_b + \left(\frac{e^{-(1/\mu_{Rq})} * (1/\mu_{Rq})^n}{n!} \right) * P_{TMAX} + \left[\sum_{i=1}^{n} \frac{e^{-(1/\mu_{At})} * (1/\mu_{At})^1}{1!} \right] * P_{TMAX} + \left(\frac{e^{-(1/\mu_f)} * (1/\mu_f)^n}{n!} \right) * P_{TMIN} + t_{ACK} * P_{TMIN} + \tau_i * p_i + (1 - ((s+t+r)+l))\tau * p_{sp} \quad (13)$$

where $1/\mu_{Rq}$ is the time taken by a node for broadcasting the request message and $1/\mu_{At}$ is the time taken by the receiver nodes for replying accept message. Power consumption of frames being received by a receiver is

$$p_r = \tau_{sn} * p_{sn} + \left(\frac{e^{-(1/\mu_R)} * (1/\mu_R)^1}{1!} \right) * P_{TMAX} + \left(\frac{e^{-(1/\mu_C)} * (1/\mu_C)^1}{1!} \right) * P_{TMAX} + \left(\frac{(1/\mu_f)^k}{k!} . e^{(-1/\mu_f)} \right) * P_{TMIN} + t_{ACK} * P_{TMIN} + \tau_i * p_i + (1 - ((s+t+r)+l))\tau * p_{sp} \quad (14)$$

4. Simulation and Results

The proposed analytical model for estimating energy consumption in Y-MAC protocol is implemented in MATLAB. This protocol has been designed to support high priority transmission request. The proposed protocol has been tested for both single hop and multi-hop transmission of frames. The protocol also supports irregular time interval whereas Y-MAC protocol supports constant time interval between events. The Simulation parameters are listed in the table below.

Table 1: Simulation parameters

Parameters	Values
Number of nodes	50
Simulation time	150 seconds
Transmission Range	22 meters
Bandwidth	11 Mb
Transmitting and Receiving antenna gain	Gt=1, Gr=1
Transmission power	0.031622777W
Carrier Sense Power	5.011872e- 12W
Received Power Threshold	5.82587e-09W
Traffic type	CBR
Initial Energy	1000 Joule
Time synchronization message interval	10 seconds
Contention window	12ms
Packet header length	16 bytes
Payload length	32 bytes
Control packet interval	10 seconds
Number of retransmissions	1
Number of channels	6

The performance of the proposed work is analyzed and compared with Y-MAC protocol. The performance is evaluated in terms of two metrics - energy efficiency and reception rate. Every node broadcast its remaining time in the current frame period after every 10 seconds. Each node records the time difference when it receives time synchronization messages from the neighbors. It reports average time difference to the sink node at fixed time intervals. The number of transmitting nodes is varied to analyze the relationship between performance and node density. The sink node in Y-MAC protocol continuously hops

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

to the next channel so that it does not interfere with data reception by other nodes. Energy saving by sink node in Y-MAC protocol is not considered since the sink node is normally powered by rechargeable battery or by other supply source. The transmission rate is restricted by varying the traffic load in the network.

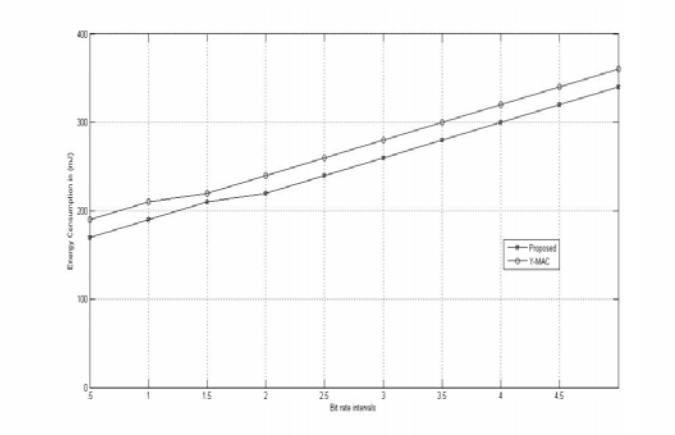


Fig. 4 Comparison of Energy Consumption

The above figure shows the comparison of energy consumption between Y-MAC and proposed protocol. This figure shows-the impact of the bit rate interval between the nodes on the energy consumption of nodes in milli-joules. It is quite clear from the figure that the proposed protocol out performs the Y-MAC in energy consumption. It consumes approximate of 60 milli-joules lesser than energy consumed by Y-MAC protocol.

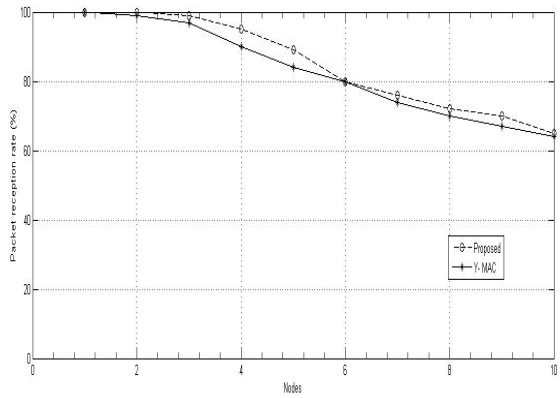


Fig. 5 Single hop average Data reception

Fig 2 shows the average data reception percentage for single hop transmission. In Y-MAC protocol every node transmits message once in every second. But, the proposed protocol transmits 'n' number of messages at different time interval. The data reception rate of proposed protocol is better than Y-MAC protocol. In both cases it decreases gradually with the increase in number of nodes.

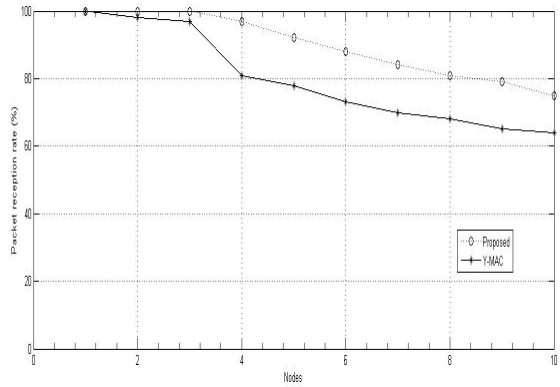


Fig. 6 Multi hop average Data reception

Fig 3 shows the average data reception rate at the sink node for multi-hop transmission. Nodes in Y-MAC protocol generate messages after a fixed time interval, whereas in this work messages can be generated at any time. From the figure, it is clear that data reception rate declines faster than the proposed protocol. Therefore, we can infer that under heavy traffic conditions the data reception rate of proposed protocol is better than Y-MAC protocol.

5. Conclusions

In this work, we have proposed analytical model for estimating energy consumption in Y-MAC protocol for Wireless Sensor Networks. The power consumption for an individual node is calculated for both unicast and broadcast transmission with multichannel transmission of frames. A node in the network saves its energy by changing its mode periodically. The proposed protocol calculates active, listen and sleep time of a node and shows results comparable with Y-MAC protocol in terms of energy consumption. However, analytical model as expected performance marginally better since it does include the complexities of the system.

.

References

[1] Andrea Goldsmith, Wirless Communication. Cambridge University Press, 2007.
[2] C. Siva Ram Murthy and B.S.Manoj, Ad hoc Wireless Sensor Networks, Department of Computer Science & Engineering, IIT Madras, Pearson Education 2005.
[3] Jeremy Elson and Deborah Estrin "Time Synchronization for Wireless Sensor Networks", Workshop on parallel and Distributed Computing Issues in Wireless Networks and Mobile Computing San Francisco, CA USA 2001.
[4] Prakesh Ranganathan and Kendall Nygard , "Time Synchronization in Wireless Sensor Networks: A Survey", International Journal of UbiComp (IJU), Vol.1 No.2, April 2010.
[5] Ramchand V and D.K.Lobiyal, "An Analytic model for Power Control T-MAC protocol", International Journal of Computer Applications (0975-8887) Volume 12-No.1, December 2010.
[6] Youngmin Kim, Hyojeong Shin and Hojung Cha, "Y-MAC An Energy Efficient Multi-channel MAC Protocol for Dense

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Wireless Sensor Networks", International Conference of Information Processing in Sensor Networks 2008.

[7] Philipp Sommer and Roger Wattenhofer, "Gradient Clock Synchronization in wireless sensor networks" ACM IPSN'09, San Francisco, California, USA, April 13-16, 2009.

[8] J.V.Greunen and J.Rabaey, "Light Weight Time Synchronization for Wireless Sensor Networks", Proceedings of the 2nd ACM International Conference on Wireless Sensor Networks and Applications (WSNA), SanDieg, CA, September 2003.

[9] Tao Zheng, Sridhar Radhakrishnan and Venkatesh Sarangan, "PMAC: An Adaptive Energy-Efficient MAC Protocol for Wireless Sensor Networks", - Workshop 12 - Volume 13, Proceedings of the 19th IEEE International Parallel and Distributed Processing Symposium (IPDPS 2005)

[10] Yaser Mahmood A. Hamid, and D. K. Lobiyal "IPCM/COMPOW: An Efficient Power Saving Scheme for Multi-Hop Wireless Ad Hoc Networks", pp.452-458, ICWN 2008.

[11] Stephen F. Bush "Low-Energy Sensor Network Time Synchronization as an Emergent property", Proceedings. 14th International Conference on Digital Object Identifier: 10.1109/ICCCN.2005.1523817, pp.93-98, ICCCN 2005.

First Author Mr. Ramchand V. received his M.Sc. (Computer Science) and M.Tech (Computer Science and Engineering) from Bishop Heber College, Tamilnadu, India. and S.R.M University, Tamilnadu, India in 2005 and 2007 respectively. Currently Mr. Ramchand is pursuing Ph.D (Computer Science) in the area of Transmission power control in Wireless Sensor Networks in the School of Computer & Systems Sciences, Jawaharlal Nehru University, New Delhi, India.

Second Author Dr. D.K. Lobiyal received his Ph.D and M. Tech (Computer Science) from School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi, India in 1996 and 1991, respectively, and B. Tech. (Computer Science and Engineering) from Lucknow University, India in 1988. He is currently working as an Associate Professor in the School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi, India. His research interest includes Mobile Ad hoc Networks, VoD, Bioinformatics and Natural Language Processing. He has published papers in International Journals and Conferences including IEEE, Wiley & Sons, Springer, Inder Science, WSEAS, IGI Global, and ACTA Press.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Performance Based Novel Techniques for Semantic Web Mining

Mahendra Thakur[1] , Geetika S. Pandey[2]

[1] MTECH Scholar, Computer Science Department
Samrat Ashok Technological Institute, Vidisha, M.P, INDIA

[2]Asst. Professor, Computer Science Department
Samrat Ashok Technological Institute, Vidisha, M.P, INDIA

Abstract

The explosive growth in the size and use of the World Wide Web continuously creates new great challenges and needs. The need for predicting the users' preferences in order to expedite and improve the browsing though a site can be achieved through personalizing of the websites. Most of the research efforts in web personalization correspond to the evolution of extensive research in web usage mining, i.e. the exploitation of the navigational patterns of the web site's visitors. When a personalization system relies solely on usage-based results, however, valuable information conceptually related to what is finally recommended may be missed. Moreover, the structural properties of the web site are often disregarded. In this paper, we propose novel techniques that use the content semantics and the structural properties of a web site in order to improve the effectiveness of web personalization. In the first part of our work we present standing for Semantic Web Personalization, a personalization system that integrates usage data with content semantics, expressed in ontology terms, in order to compute semantically enhanced navigational patterns and effectively generate useful recommendations. To the best of our knowledge, our proposed technique is the only semantic web personalization system that may be used by non-semantic web sites. In the second part of our work, we present a novel approach for enhancing the quality of recommendations based on the underlying structure of a web site. We introduce UPR (Usage-based PageRank), a PageRank-style algorithm that relies on the recorded usage data and link analysis techniques. Overall, we demonstrate that our proposed hybrid personalization framework results in more objective and representative predictions than existing techniques.

Keywords- *Web personalization, Semantic web and Recommender systems.*

1.Introduction

During the past few years the World Wide Web has become the biggest and most popular way of communication and information dissemination. It serves as a platform for exchanging various kinds of information, ranging from research papers, and educational content, to multimedia content, software and personal logs. Every day, the web grows by roughly a million electronic pages, adding to the hundreds of millions pages already on-line. Because of its rapid and chaotic growth, the resulting network of information lacks of organization and structure. Users often feel disoriented and get lost in that information overload that continues to expand. On the other hand, the e-business sector is rapidly evolving and the need for web market places that anticipate the needs of their customers is more than ever evident. Therefore, the ultimate need nowadays is that of predicting the user needs in order to improve the usability and user retention of a web site. This paper presents novel methods and techniques that address this requirement. In brief, web personalization can be defined as any action that customizes the information or services provided by a web site to an individual user, or a set of users, based on knowledge acquired by their navigational behavior, recorded in the web site's logs, in other words, its usage. This information is often combined with the content and the structure of the web site, as well as the interests/preferences of the user, if they are available. The web personalization process is illustrated in Figure 1. Using the four aforementioned sources of information as input to pattern discovery techniques, the system tailors the provided content to the needs of each visitor of the web site. The personalization process can result in the dynamic generation of recommendations, the creation of index pages, the highlighting of existing hyperlinks, the publishing of targeted advertisements or emails, etc. In this paper we focus on personalization systems that aim at providing personalized recommendations to the web site's visitors. Furthermore, since the personalization algorithms we propose in this work are generic and applicable to any web site, we assume that no explicit knowledge involving the

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

users' profiles, such as ratings or demographic information is available [1] and [2] and [5].

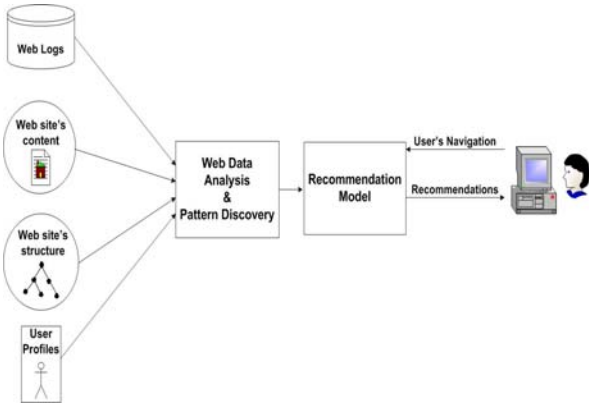


Figure: 1 the web personalization process

The problem of providing recommendations to the visitors of a web site has received a significant amount of attention in the related literature. Most of the research efforts in web personalization correspond to the evolution of extensive research in web usage mining, taking into consideration only the navigational behavior of the (anonymous or registered) visitors of the web site. Pure usage-based personalization, however, presents certain shortcomings. This may happen when, for instance, there is not enough usage data available in order to extract patterns related to certain navigational actions, or when the web site's content changes and new pages are added but are not yet included in the web logs. Moreover, taking into consideration the temporal characteristics of the web in terms of its usage, such systems are very vulnerable to the training data used to construct the predictive model. As a result, a number of research approaches integrate other sources of information, such as the web content or the web structure in order to enhance the web personalization process. As already implied, the users' navigation is largely driven by semantics. In other words, in each visit, the user usually aims at finding information concerning a particular subject. Therefore, the underlying content semantics should be a dominant factor in the process of web personalization. The web site's content characterization process involves the feature extraction from the web pages. Usually these features are keywords subsequently used to retrieve similarly characterized content. Several methods for extracting keywords that characterize web content have been proposed. The similarity between documents is usually based on exact matching between these terms. This way, however, only a binary matching between documents is achieved, whereas no actual semantic similarity is taken into consideration. The need for a more abstract representation that will enable a uniform and more flexible document matching process imposes the use of semantic web structures, such as ontology's. By mapping the keywords to the concepts of an ontology, or topic hierarchy, the problem of binary matching can be surpassed through the use of the hierarchical relationships and/or the semantic similarities among the ontology terms, and therefore, the documents. Finally, we should take into consideration that the web is not just a collection of documents browsed by its users. The web is a directed labeled graph, including a plethora of hyperlinks that interconnect its web pages. Both the structural characteristics of the web graph, as well as the web pages' and hyperlinks' underlying semantics are important and determinative factors in the users' navigational process. The main contribution of this paper is a set of novel techniques and algorithms aimed at improving the overall effectiveness of the web personalization process through the integration of the content and the structure of the web site with the users' navigational patterns. In the first part of our work we present the semantic web personalization system standing for Semantic Web Personalization that integrates usage data with content semantics in order to compute semantically enhanced navigational patterns and effectively generate useful recommendations. Similar to previously proposed approaches, the proposed personalization framework uses ontology terms to annotate the web content and the users' navigational patterns. The key departure from earlier approaches, however, is that standing for Semantic Web Personalization is the only web personalization framework that employs automated keyword-to-ontology mapping techniques, while exploiting the underlying semantic similarities between ontology terms. Apart from the novel recommendation algorithms we propose, we also emphasize on a hybrid structure-enhanced method for annotating web content. To the best of our knowledge, standing for Semantic Web Personalization is the only semantic web personalization system that can be used by any web site, given only its web usage logs and a domain-specific ontology [1], [2], [4] and [6].

2. Background

The main data source in the web usage mining and personalization process is the information residing on

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

the web site's logs. Web logs record every visit to a page of the web server hosting it. The entries of a web log file consist of several fields which represent the date and the time of the request, the IP number of the visitor's computer (client), the URI requested, the HTTP status code returned to the client, and so on. The web logs' file format is based on the so called "extended" log format. Prior to processing the usage data using web mining or personalization algorithms, the information residing in the web logs should be preprocessed. The web log data preprocessing is an essential phase in the web usage mining and personalization process. An extensive description of this process can be found. In the sequel, we provide a brief overview of the most important pre-processing techniques, providing in parallel the related terminology. The first issue in the pre-processing phase is data preparation. Depending on the application, the web log data may need to be cleaned from entries involving page accesses that returned, for example, an error or graphics file accesses. Furthermore, crawler activity usually should be filtered out, because such entries do not provide useful information about the site's usability. A very common problem to be dealt with has to do with web pages' caching. When a web client accesses an already cached page, this access is not recorded in the web site's log. Therefore, important information concerning web path visits is missed. Caching is heavily dependent on the client-side technologies used and therefore cannot be dealt with easily. In such cases, cached pages can usually be inferred using the referring information from the logs and certain heuristics, in order to re-construct the user paths, filling out the missing pages. After all page accesses are identified, the pageview identification should be performed. A pageview is defined as "the visual rendering of a web page in a specific environment at a specific point in time". In other words, a pageview consists of several items, such as frames, text, graphics and scripts that construct a single web page. Therefore, the pageview identification process involves the determination of the distinct log file accesses that contribute to a single pageview. Again such a decision is application-oriented. In order to personalize a web site, the system should be able to distinguish between different users or groups of users. This process is called user profiling. In case no other information than what is recorded in the web logs is available,

this process results in the creation of aggregate, anonymous user profiles since it is not feasible to distinguish among individual visitors. However, if the user's registration is required by the web site, the information residing on the web log data can be combined with the users' demographic data, as well as with their individual ratings or purchases. The final stage of log data pre-processing is the partition of the web log into distinct user and server sessions. A user session is defined as "a delimited set of user clicks across one or more web servers", whereas a server session, also called a visit, is defined as "a collection of user clicks to a single web server during a user session". If no other means of session identification, such as cookies or session ids is used, session identification is performed using time heuristics, such as setting a minimum timeout and assumes that consecutive accesses within it belong to the same session, or a maximum timeout, assuming that two consecutive accesses that exceed it belong to different sessions [3] and [4].

2.1 Web Usage Mining and Personalization

Web usage mining is the process of identifying representative trends and browsing patterns describing the activity in the web site, by analyzing the users' behaviour. Web site administrators can then use this information to redesign or customize the web site according to the interests and behavior of its visitors, or improve the performance of their systems. Moreover, the managers of e-commerce sites can acquire valuable business intelligence, creating consumer profiles and achieving market segmentation. There exist various methods for analyzing the web log data. Some research studies use well known data mining techniques such as association rules discovery, sequential pattern analysis, clustering, probabilistic models, or a combination of them. Since web usage mining analysis was initially strongly correlated to data warehousing, there also exist some research studies based on OLAP cube models. Finally some proposed web usage mining approaches that require registered user profiles, or combine the usage data with semantic meta-tags incorporated in the web site's content. Furthermore, this knowledge can be used to automatically or semi-automatically adjust the content of the site to the needs of specific groups of users, i.e. to personalize the site. As already mentioned, web personalization may include the provision of recommendations to the users, the creation of new index pages, or the generation of targeted advertisements or product promotions. The usage-based personalization systems use association rules and sequential pattern discovery, clustering, Markov models, machine learning algorithms, or are based on collaborative filtering in order to generate recommendations. Some research studies also

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

combine two or more of the aforementioned techniques [3] and [7].

2.2 Integrating Structure in Web Personalization

Although the connectivity features of the web graph have been extensively used for personalizing web search results, only a few approaches exist that take them into consideration in the web site personalization process. To use citation and coupling network analysis techniques in order to conceptually cluster the pages of a web site. The proposed recommendation system is based on Markov models. In previous, use the degree of connectivity between the pages of a web site as the determinant factor for switching among recommendation models based on either frequent itemset mining or sequential pattern discovery. Nevertheless, none of the aforementioned approaches fully integrates link analysis techniques in the web personalization process by exploiting the notion of the authority or importance of a web page in the web graph [2] and [10].
In a very recent work, address the data sparsity problem of collaborative filtering systems by creating a bipartite graph and calculating linkage measures between unconnected pairs for selecting candidates and make recommendations. In this study the graph nodes represent both users and rated/purchased items. Finally, subsequent work, proposed independently two link analysis ranking methods, SiteRank and PopularityRank which are in essence very much like the proposed variations of our UPR algorithm (PR and SUPR respectively). This work focuses on the comparison of the distributions and the rankings of the two methods rather than proposing a web personalization algorithm [2], [3] and [11].

3. Proposed Personalization Techniques

In this paper we present standing for Semantic Enhancement for Web Personalization, a web personalization framework that integrates content semantics with the users' navigational patterns, using ontologies to represent both the content and the usage of the web site. In our proposed framework we employ web content mining techniques to derive semantics from the web site's pages. These semantics, expressed in ontology terms, are used to create semantically enhanced web logs, called C-logs (concept logs). Additionally, the site is organized into thematic document clusters. The C-logs and the document clusters are in turn used as input to the web mining process, resulting in the creation of a broader, semantically enhanced set of recommendations. The whole process bridges the gap between Semantic Web and Web Personalization areas, to create a Semantic Web Personalization system.

3.1 Standing for Semantic Enhancement for Web Personalization System Architecture

Standing for Semantic Enhancement for Web Personalization uses a combination of web mining techniques to personalize a web site. In short, the web site's content is processed and characterized by a set of ontology terms (categories). The visitors' navigational behavior is also updated with this semantic knowledge to create an enhanced version of web logs, C-logs, as well as semantic document clusters. C-Logs are in turn mined to generate both a set of URI and category-based association rules. Finally, the recommendation engine uses these rules, along with the semantic document clusters in order to provide the final, semantically enhanced set of recommendations to the end user.

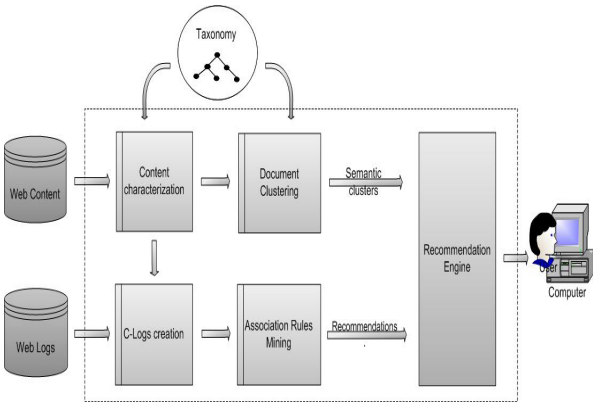


Figure 2 Standing for Semantic Enhancement for Web Personalization architecture

As illustrated in Figure 2, Standing for Semantic Enhancement for Web Personalization consists of the following components:

• **Content Characterization*:*** This module takes as input the content of the web site as well as a domain-specific ontology and outputs the semantically annotated content to the modules that are responsible for creating the C-Logs and the semantic document clusters. The content characterization process consists of the keyword extraction, keyword translation and semantic characterization sub-processes.
• **Semantic Document Clustering***:* The semantically annotated pages created by the previous component are grouped into thematic clusters. This categorization is achieved by clustering the web

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

documents based on the semantic similarity between the ontology terms that characterize them.

• **C-Logs Creation & Mining:** This module takes as input the web site's logs as well as the semantically annotated web site content. It outputs the semantically enhanced C-logs (concept logs) which are in turn used to generate both URI and category-based frequent itemsets and association rules. These rules are subsequently matched to the current user's visit by the recommendation engine.

• **Recommendation Engine:** This module takes as input the current user's path and matches it with the semantically annotated navigational patterns generated in the previous phases. The recommendation engine generates three different recommendation sets, namely, original, semantic and category-based ones, depending on the input patterns used.

The creation of the ontology as well as the semantic similarity measures used as input in the aforementioned web personalization process are orthogonal to the proposed framework. We assume that the ontology is descriptive of the web site's domain and is provided / created by a domain expert. In what follows we describe the key components of our architecture, starting by introducing the similarity measures we used in our work.

3.2 Content Characterization

A fundamental component of the Standing for Semantic Enhancement for Web Personalization architecture is the automated content characterization process. Our Personalization is the only web personalization framework enabling the automated annotation of web content with ontology terms without needing any human labeling or prior training of the system. The keywords' extraction is based both on the content of the web pages, as well as their connectivity features. What is more, our technique enables the annotation of multilingual content, since it incorporates a context-sensitive translation component which can be applied prior to the ontology mapping process. In the subsections that follow we describe in detail the aforementioned processes, namely, the keyword extraction, keyword translation and semantic characterization modules.

3.3 Keyword Extraction

There exists a wealth of methods for representing web documents, most of which have emerged from the area of searching and querying the web. The most straightforward approach is to perform text mining in the document itself following standard Information Retrieval (IR) techniques. This approach, however, has been shown insufficient for the web content, since it relies solely on the information included in the document ignoring semantics arising from the connectivity features of the web. It is difficult to extract keywords from web documents that contain images, programs etc. Additionally, many web pages do not include words that are the most descriptive ones for their content (for example rarely a portal web site includes the word "portal" in its home page). Therefore, in many approaches information contained in the links that point to the document and the text near them - defined as "anchor-window"- is used for characterizing a web document. This approach is based on the hypopaper that the text around the link to a page is descriptive of the page's contents and overcomes the problems of the content-based approach, since it takes into consideration the way others characterize a specific web page. In our work, we adopt and extend this approach, by also taking into consideration the content of the pages that are pointed by the page that is processed, based on the assumption that in most web pages the authors include links to topics that are of importance in the page's context.

More specifically, the keywords that characterize a web page p are extracted using:

1. raw term frequency of p
2. raw term frequency of a selected fraction (anchor-window) of the web pages that point to p (in links)
3. raw term frequency of the web pages that are pointed by p (out links) The three keyword extraction methods can be applied interchangeably or in combination. We should explain at this point the decision concerning term weighting phase, when the extracted keywords are given weights in order to use the most important ones. Term weighting, extensively used in the vector space model for document clustering, is carried out using several methods, such as raw term frequency. Raw term frequency is based on the term statistics within a document and is the simpler way of assigning weights to terms. The method used for collections of documents, i.e. documents that have similar content. In the case of a Web site however, this assumption is not always true since a Web site may contain documents that refer to different thematic categories (especially in the case of Web portals) and this was the reason for choosing raw term frequency as the term weighting method of our approach.

3.4 Keyword Translation

As already mentioned, the recommendation process is based on the characterization of all web documents using a common representation. Since many web sites contain content written in more than one language, this raises the issue of mapping keywords from different languages to the terms of a common

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

domain-ontology. For instance, our technique makes an implicit assumption of "one sense per discourse", i.e., that multiple appearances of the same word will have the same meaning within a document. This assumption might not hold in several cases, thus leading to erroneous translations. Our technique constitutes a first step toward the automated mapping of keywords to the terms of a common concept hierarchy; clearly, a more extensive study is required in order to provide a complete and more precise solution.

```
Procedure translateW(Gr,En)
1. K ← Ø ;
2. for all g ∈Gr(D) do
3. for all s ∈Sn(g) do
4. score[s] = 0;
5. for all w ∈En(D)U Gr(D)-{g} do
6. sim = max(WPsim(s, Sn(w)));
7. score[s] += sim;
8. done
9. done
10. s_max = s';
(score[s'] = max(score[s]), s ∈Sn(g))
11. K ← e, e ∈ En(g), e contains s_max;
12. done
```

Figure 3 The keyword translation procedure

3.4 Semantic Characterization

In order to assist the remainder of the personalization process (C-logs cretion, semantic document clustering, semantic recommendations) the n most frequent (translated) keywords that where extracted in the previous phase, are mapped to the terms O = {c1, …, ck.} of a domain ontology (in our approach we need the concept hierarchy part of the ontology). This mapping is performed using a thesaurus. If the keyword belongs to the ontology, then it is included as it is. Otherwise, the system finds the "closest" (i.e. most similar) term (category) to the keyword through the mechanisms provided by the thesaurus. Since the keywords carry weights according to their frequency, the categories' weights are also updated.

We should stress here that the selection of the ontology influences the outcome of the mapping process. For this purpose, it should be semantically relevant to the content to be processed. In order to find the closest term in the ontology O for a keyword k that describes document, we compute the similarity between all senses of k, Sn(k) and all senses of all the categories c in O, Sn(ci). At the end of this process, each keyword is mapped to every category with a similarity s respectively. We select the (k,c) pair that gives the maximum similarity s. This process is shown in Figure 4.

```
Procedure CategoryMapping(k, O)
1. for all sns ∈Sn(k) do
2. for all ci ∈O do
3. scsim_max ← maxarg_sc ∈Sn(ci)(WPsim(sns, sc));
4. done
5. ssim_max= max({scsim_max});
6. cmax = c ∈O, for which (scsim_max == ssim_max);
7. done
8. sim = max({ssim_max});
9. cat = c'∈ {c_max}, for which (ssim_max == sim);
10. return(cat, sim);
11. done
```

Figure 4 the semantic characterization process

3.5 Semantic Recommendations

Navigational patterns: We use the Apriori algorithm to discover frequent itemsets and/or association rules from the C-Logs, *CLg*. We consider that each distinct user session represents a different transaction. We will use $S = \{I_m\}$, to denote the final set of frequent itemsets/association rules, where $Im = \{(uri_i)\}$, $uri_i \in CLg$.

Recommendations: In brief, the recommendation method takes as input the user's current visit, expressed a set of URIs: $CV = \{(uri_j)\}$, $uri_j \in WS$, (*WS* is the set of the web site's URIs. Note that some of these may not be included in *CLg*). The method finds the itemset in *S* that is most similar to *CV*, and recommends the documents (labeled by related categories) belonging to the most similar document cluster $Cl_m \in Cl$ (*Cl* is the set of document clusters). In order to find the similarity between URIs, we perform binary matching (denoted as *SIM*). This procedure is shown in Figure 5.

```
Procedure SemanticRec(CV)
1. CM ← Ø ;
2. Im = maxargI∈S SIM(I,CV);
3. for all d ∈ I_m do
4. for all cj ∈d do
5. if cj ∈CM then
6. rj' += rj;
7. CM ← (cj,rj');
8. else
9. CM ← (cj,rj);
10. done
11. done
12. return D = {d}, {d} ∈ Cl_m,
maxargCln∈Cl WPsim(CL_m,CM);
```

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Figure 5 the *semantic* recommendation method

The method finds the itemset in *C* that is most similar to *CV,* creates a generalization of it and recommends the documents (labeled by related categories) belonging to the most similar document cluster $Cl_n \in Cl$ (*Cl* is the set of document clusters). To find the similarity between categories we use the metric, whereas in order to find similarity between sets of categories, we The same procedure can be run by omitting the weights in one or all the phases of the algorithm. On the other hand, in case weights are used, an extension of the Apriori algorithm, which incorporates weights in the association rules mining process, can be used.

Procedure **CategoryRec(CV)**
1. I**k** = maxarg**I**∈**S** THEsim(I,CV);
2. for all c**j** ∈ CV do
3. c**i** = maxarg**c**∈**Ik** WPsim(c,c**j**);
4. c**n** = least_common_ancestor(c**i**,c**j**), r**n** = max(r**i**,r**j**);
5. CI ← (c**n**, r**n**);
6. done
7. return D = {d}, {d} ∈ Cl**n**, maxarg**Cln**∈**Cl** WPsim(CL**n**,CI);

Figure 6 the category-based recommendation method

Let us also stress that even though this description of the method focuses on sets' representation (derive frequent itemsets and use them in the recommendation method), it can also be applied (with no further modification) to the association rules that can be derived by those sets. If association rules are derived, then the user's activity is matched to the LHS of the rule (step 2), and recommendations are generated using the RHS of the rule (step 7).

So far, we have described the framework for enhancing the recommendation process through content semantics. Our claim, that the process of semantically annotating web content using terms derived from a domain-specific taxonomy prior to the recommendation process enhances the results of web personalization, is intuitive. Since the objective of the system is to provide useful recommendations to the end users, we performed an experimental study, based on blind testing with 15 real users, in order to validate the effectiveness of our approach. The results indicate that the effectiveness of each recommendation set (namely, *Original, Semantic, Category*), depends on the model, incorporating all three types of recommendations, generates the most effective results.

4. Methodology

Data Set: The two key advantages of using this data set are that the web site contains web pages in several formats (such as pdf, html, ppt, doc, etc.), written both in Greek and English and a domain-specific concept hierarchy is available (the web administrator created a concept-hierarchy of 150 categories that describe the site's content). On the other hand, its context is rather narrow, as opposed to web portals, and its visitors are divided into two main groups: students and researchers. Therefore, the subsequent analysis (e.g. association rules) uncovers these trends: visits to course material, or visits to publications and researcher details. It is essential to point out that the need for processing online (up-to-date) content, made it impossible for us to use other publicly available web log sets, since all of them were collected many years ago and the relevant sites' content is no longer available. Moreover, the web logs of popular web sites or portals, which would be ideal for our experiments, are considered to be personal data and are not disclosed by their owners. To overcome these problems, we collected web logs over a 1-year period (01/01/10 – 31/12/10). After preprocessing, the total web logs' size was approximately 105 hits including a set of over 67.700 distinct anonymous user sessions on a total of 360 web pages. The sessionizing was performed using distinct IP & time limit considerations (setting 20 minutes as the maximum time between consecutive hits from the same user).

Keyword Extraction – Category Mapping: We extracted up to 7 keywords from each web page using a combination of all three methods (raw term frequency, inlinks, outlinks). We then mapped these keywords to ontology categories and kept at most 5 for each page.

Document Clustering*:* We used the clustering scheme described in recent, i.e. the DBSCAN clustering algorithm and the similarity measure for sets of keywords. However, other web document clustering schemes (algorithm & similarity measure) may be employed as well.

Association Rules Mining*:* We created both URI-based and category-based frequent itemsets and association rules. We subsequently used the ones over a 40% confidence threshold.

4.1 Link Analysis for Web Personalization

The connectivity features of the web graph play important role in the process of web searching and navigating. Several link analysis techniques, based on

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

the popular PageRank algorithm [BP98], have been largely used in the context of web search engines. The underlying intuition of these techniques is that the importance of each page in a web graph is defined by the number and the importance of the pages linking to it. In this paper, we introduce link analysis in a new context, that of web personalization. Motivated by the fact that in the context of navigating a web site, a page/path is important if many users have visited it before, we propose a new algorithm UPR (Usage-based PageRank). UPR is based on a personalized version of PageRank, "favoring" pages and paths previously visited by many web site users. We apply UPR to a representation of the web site's user sessions, termed Navigational Graph in order to rank the web site's pages. This ranking may then be used in several contexts:

➔Use it as a "global ranking" of the web site's pages. The computed rank probabilities can serve as the prior probabilities of the pages when recommendations are generated using probabilistic predictive models such as Markov Chains, higher-order Markov models, tree synopses etc.

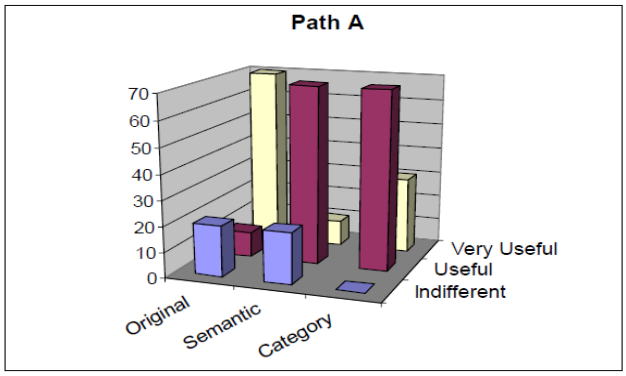
➔Apply UPR to small subsets of the web site's navigational graph (or its approximations), which are generated based on each current user's visit. This localized version of UPR (named l-UPR) provides localized personalized rankings of the pages most likely to be visited by each individual user. In what follows we illustrate our approach through a motivating example. We then provide the required theoretical background on link analysis before presenting the proposed algorithm. We prove that this hybrid algorithm can be applied to any web site's navigational graph as long as the graph satisfies certain properties. We then proceed with describing the two proposed frameworks in which UPR can be applied, namely, the localized personalized recommendations with l-UPR and the hybrid probabilistic predictive models (h-PPM). We conclude with an extensive experimental evaluation we performed on both frameworks (l-UPR and h-PPM), proving our claim that the underlying link structure of the web sites should be taken into consideration in the web personalization process, and details on the system prototype we used.

5. Results

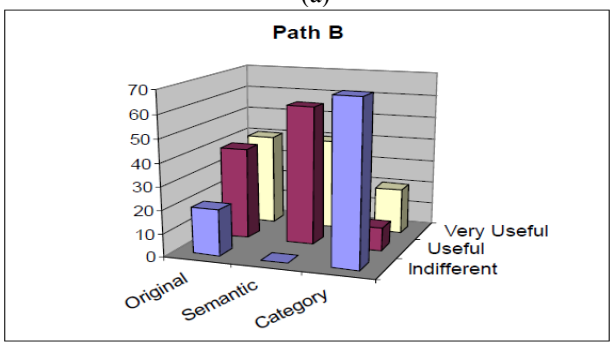
We created three different sets of recommendations named Original, Semantic, and Category (the sets are named after the respective recommendation methods). We presented the users with the paths and the three sets (unlabeled) in random order and asked them to rate them as "indifferent", "useful" or "very useful". The outcome is shown in Figure a, b, and c.

The results of the first experiment revealed the fact that depending on the context and purpose of the visit the users profit from different source of recommendations. More specifically, in visit A, both Semantic and Category sets are mostly evaluated as useful/very useful. The Category recommendation set performs better, and this can be explained by the fact that it's the one that recommends "hub" pages, which seems to be the best after a "random walk" on the site.

On the other hand, in visits B and C, Semantic performs better. In visit B, the path was focused to specific pages and the same held for the recommendations' preferences. In visit C the recommendations that were more relevant to the topics previously visited were preferred.

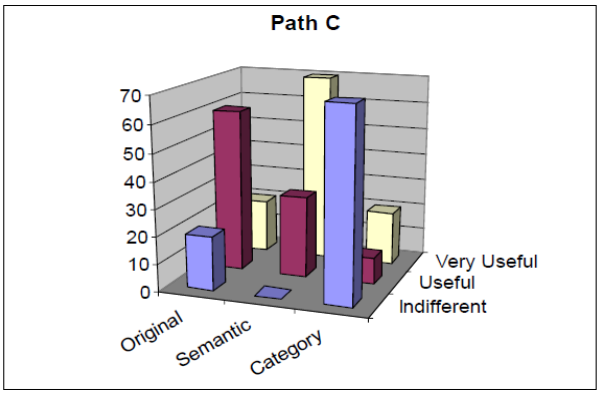


(a)



(b)

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.



(c)

Figure 7 a, b, c: Recommendation sets' evaluation

Users had to select between the Category-based recommendation set and the Hybrid one. The outcome is shown in Figure d.

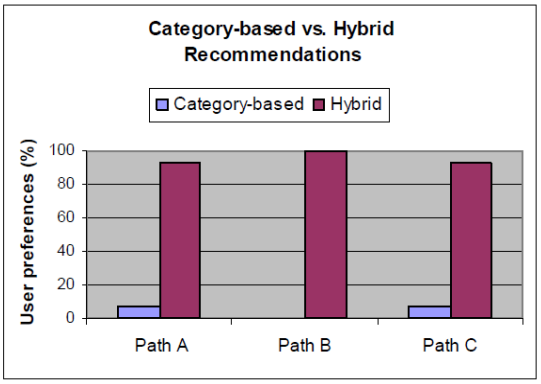


Figure7 d: Category-based vs. Hybrid Recommendations

The results of this experiment demonstrate the dominance of the Hybrid recommendation set over the Category-based one. One explanation for this would be that in the second case, important information may be lost during the generalization (convert user's current path to categories) back to specialization (convert categories to URIs) process.

Based on these experimental results, we observe that what is characterized as useful by the users depends on the objective of each visit. Out of the three possible recommendation sets, the Semantic recommendation set, generated after the semantic expansion of the most popular association rule performs better. Comparing all three recommendation sets with the Hybrid one, we observe that it dominates the other three, since the hybrid recommendations are preferred by the users in most cases. Therefore, we conclude that Standing for Semantic Enhancement for Web Personalization semantic enhancement of the personalization process improves the quality of the recommendations in terms of complying with the users' needs.

6. Conclusion

Most of the research efforts in web personalization correspond to the evolution of extensive research in web usage mining, i.e. the exploitation of the navigational patterns of the web site's visitors. When a personalization system relies only on usage-based results, however, valuable information conceptually related to what is finally recommended may be missed. Moreover, the structural properties of the web site are often disregarded. In this paper, we present novel techniques that incorporate the content semantics and the structural properties of a web site in the web personalization process. In the first part of our work we present a semantic web personalization system. Motivated by the fact that if a personalization system is only based on the recorded navigational patterns, important information that is semantically similar to what is recommended might be missed, we propose a web personalization framework that integrates usage data with content semantics, expressed in ontology terms, in order to compute semantically enhanced navigational patterns and effectively generate useful recommendations. The diversity of our specializations verifies the potential of our approach in providing an integrated framework for applications of link analysis to web personalization.

ACKNOWLEDGMENT

This research is supported by the Computer Science and Engineering department, SATI, Vidish

References

[1] Dario Vuljani, Lidia Rovan, and Mirta Baranovi, "Semantically Enhanced Web Personalization Approaches and Techniques", IEEE Proceedings of the *ITI 2010 32nd Int. Conf. on Information Technology Interfaces,* June 21-24, 2010, Cavtat, Croatia.

[2] Raymond Y.K. Lau, "Inferential Language Modeling for Selective Web Search Personalization and Contextualization", IEEE 2010 3rd International Conference on Advanced Computer Theory and Engineering(1CACTE).

[3] Esteban Robles Luna, Irene Garrigos, and Gustavo Rossi, "Capturing and Validating Personalization Requirements in Web Applications", IEEE 2010.

[4] Annappa B, K Chandrasekaran, K C Shet, "Meta-Level Constructs in Content Personalization of a Web Application", IEEE *Int'l Conf. on Computer & Communication Technology*-ICCCT'10.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

[5] Pedro J. Muñoz-Merino, Carlos Delgado Kloos and Martin Wolpers, and Martin Friedrich, "An Approach for the Personalization of Exercises based on Contextualized Attention Metadata and Semantic Web technologies", 2010 10th IEEE International Conference on Advanced Learning Technologies.

[6] Xiaogang Wang and Yan Bai and Yue Li, "An Information Retrieval Method Based On Sequential Access Patterns", IEEE 2010 Asia-Pacific Conference on Wearable Computing Systems.

[7] Xiangwei Mu, Yan Chen and Taoying Li, "User-Based Collaborative Filtering Based on Improved Similarity Algorithm", IEEE 2010.

[8] Dimitris Antoniou, Mersini Paschou, Efrosini Sourla, Athanasios Tsakalidis, "A Semantic Web Personalizing Technique The case of bursts in web visits", 2010 IEEE Fourth International Conference on Semantic Computing.

[9] Rong Shan and Zhibin Ren, "Research on Personalized Recommendation System in E-Iearning", IEEE 2010 2nd International Conforence on Education Technology and Computer (ICETC).

[10] Yan Gao, Bin Zhang, Shao-wei Shi, Hong-ning Zhu, Jun Na, Fu-cai Zhou, "A User Requirement-driven Service Dynamic Personalized QoS Model", IEEE 2010 Third International Conference on Dependability.

[11] Muhammad Shoaib, Amna Basharat, "Ontology based Knowledge Represeenation and Sematnic Profiling In Personalized Semantic Social Networking Framework", IEEE 2010.

AUTHORS PROFILE

Geetika S. Pandey presently working as Asst. Professor in Computer Science & Engineering at Samrat Ashok Technological Institute Vidisha M.P India. The degree of B.E. (Hons) secured in Computer Science & engineering. She secured M.Tech in Computer science and Engineering from Banasthali University. She is currently pursuing PHD in Computer science and engineering.

Mr. Mahendra Thakur is a research scholar pursuing M.Tech in Computer Science & Engineering from Samrat Ashok Technological Institute Vidisha M.P India. He secured degree of B.E. in IT from Rajiv Gandhi Technical University, Bhopal (M.P.) India in 2007.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Optimized Energy-Aware Algorithm for Collaboration in Visual Sensor Networks

[1]Harish.H.Kenchannavar, [2]Umakant.P.Kulkarni

[1]Department of Computer Science & Engineering,
Gogte Institute of Technology, Belgaum

[2]Department of Computer Science & Engineering,
SDM College of Engineering and
Technology, Dharwad

Abstract

Network lifetime is one of the critical performance parameters in Visual Sensor Networks. In order to avoid transmission of redundant information to the base station, it becomes necessary to select among the nodes in the neighboring region having the same area of coverage for data transfer. This work presents an efficient approach aimed at reducing the extra amount of information to the controlling station. The proposed algorithm is based on the concept of image overlapping and energy conservation technique which increases the network lifetime.

Key words: *visual sensor network, overlapping based energy aware, horizontal collaboration, network lifetime, redundant*

1. Introduction

In wireless sensor networks, a sensor node comprises of sensing computing, communication, actuation and power components. A typical WSN incorporates around tens to thousands of such nodes, which share information and perform tasks in a co-operative manner using the wireless medium. Wireless sensor networks can be utilized at a global level for monitoring the environment and habitat study, in the battle field for military surveillance and scrutinizing the battle field for useful military information, for search and rescue operations during natural calamities and mishaps, for maintenance in industries, for infrastructure health monitoring, to realize smart homes and in human bodies for monitoring the patient's health conditions. In WSNs, the sensor nodes have the self – organizing ability that allows them to get arranged in an appropriate network infrastructure which results in ad-hoc nature of WSNs. Also the WSNs behave as a distributed database networks where users can retrieve information from a WSN by sending queries to and retrieving information from the base stations.Recent progress in imaging technologies, prove that the design of small low power and low cost image/video capturing devices is achievable in near future.

For most of the above listed applications, the QoS plays a major role while designing the sensor networks responsible for delivering the multimedia content for critical applications such as health monitoring, military surveillance etc. Research carried out in the field of sensor networks concentrates on curtailing the energy consumed by the sensor nodes for processing, problems associated with preserving the coverage area etc. The primary goal in the coverage optimization algorithm is to retain the coverage area when a sensor node fails and hence conserve the energy by making a sensor node active and the other nodes in the neighboring area having same field

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

of view to go into the sleep mode. This feature helps to prolong the lifetime of the network [4]. Remaining part of this paper contains the following sections. Related work is presented in Section 2. Section 3 describes the network model and problem definition. Section 4 and Section 5 discuss the flowchart for the novel approach and the simulation setup. Section 6 presents the result analysis.

2. Related Work

Research carried out in the field of cluster–based WSN is massive focusing on the issues of energy efficiency and scalability related to clustering protocols. In statistical techniques, for selecting the cluster head [5, 6], all the sensor nodes in the network share equally the role of cluster head which adds to the life span of the sensor node. HEED Clustering algorithm [7] implements a hybrid model to select the cluster head based on the residual energy of the node and the node's proximity to its neighbors- an additional parameter involved the selection process. The issues related to the power of balanced energy consumption among the cluster head nodes has been discussed in [8, 9]. The ACE clustering algorithm [10] partitions a network into uniformly dispersed clusters. Another clustering technique known as autonomous clustering makes use of coverage estimation parameter which is presented in [11].

One of the basic problems faced in sensor network is the network coverage, as it critically affects the performance of the network sensing task. Therefore in recent years, strong research is being done on the coverage aspect in the sensor management area [12, 13]. A scheduling scheme for activating the sensor nodes depending on the data collected from the nodes in the neighboring region without compromising on the coverage area has been proposed by the authors in [14]. Authors Wu and Chen in their work [15] have suggested a collaborative processing technique which exploits the spatial correlation among the different images captured by the adjacent cameras in the neighborhood. Applications of fundamentals of projective geometry for detection and tracking have been proposed by the authors in [17, 18, 19] and include the discussion on image mapping technique using the sample Euclidean co-ordinate transformation.

3. Network Model and Problem Description

Let a multi-hop wireless VSN be represented by a graph G, in which set N consists of visual sensor nodes given by N= $\{n_1, n_2, n_3 \ldots n_k\}$ situated at random positions within the specified network boundary. The network area can be bifurcated into various regions, represented by the set R= $\{R_1, R_2, R_3 \ldots R_m\}$ where regions $R_1, R_2, R_3 \ldots R_m$ comprise of a group of visual sensor nodes covering the region.

Each sensor node in the VSN is defined by its sensing area and the energy component specified by N_{si} and N_{ei} respectively. Let $N_{s1}, N_{s2}, N_{s3} \ldots N_{si}$ specify the node sensing areas, $N_{e1}, N_{e2}, N_{e3} \ldots N_{ei}$ present the energy levels and $N_{I1}, N_{I2}, \ldots\ldots N_{Ii}$ denote the images captured at all the nodes respectively. Thus a node N_i in the VSN network be characterized by (N_{si}, N_{ei}, N_{Ii}).

This work focuses on node collaboration where the coverage area is overlapped among a group of sensor nodes in the surrounding area and only one node from the cluster is activated for transmitting the captured image to the controlling station. The other nodes go into standby/sleep mode- which is a low power energy state thus preserving its energy which increases the network lifespan. Decision as to which nodes will be activated for transmission is based on the overlapping data provided by the sensor nodes in the specified region. This paper presents an analysis carried out using the 2 approaches which affect network lifetime.

3.1 Without using overlapping information from the sensor nodes

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

This case considers that all the nodes in the network are active at any given instance. As stated, every sensor node n_i in the network has the energy component associated with it denoted by E_i, where i= 1, 2, 3…n

Total energy at all nodes is given by

$$E_{total} = E_1+E_2+E_3\ldots E_n=\sum_{i=0}^{n} E_i \qquad (1)$$

where, every node has up to E_{max} threshold energy.

Considering the scenario where all the sensor nodes in the network are in active mode, the total energy dissipated in the network by all the nodes after time 't' secs is given by equation (1).

Energy dissipated at each node = E_{total}/n Joules/sec – (2)

where n = number of sensor nodes

Since all the nodes are active simultaneously and each node dissipates a quantum of energy after processing all the energy in the nodes gets exhausted after a time interval of 't' secs. Therefore the network life time can be calculated to be 't'secs as shown in Figure 2(a).

3.2 Issue of Data Redundancy and Energy Conservation

In a sensor network, sensor nodes are closely placed to each other for a given application. This leads to data duplication (i.e. sensors capturing and sending the same information since they have same field of view) at the computing center which results in more energy dissipation at the nodes. The approach presented here mainly addresses the data redundancy and energy preservation aspects in the visual sensor networks which affect the lifetime of the network.

3.3 Activating the nodes based on overlapping scheme

This case considers the situation where the sensor nodes in the network have the same field of view for capturing the image information. Then based on the amount of overlapping, the sensor nodes forming the cluster in the region are selected and made active for image transmission to the base station. This approach helps in conservation of energy dissipated by the nodes and increases the network lifetime.

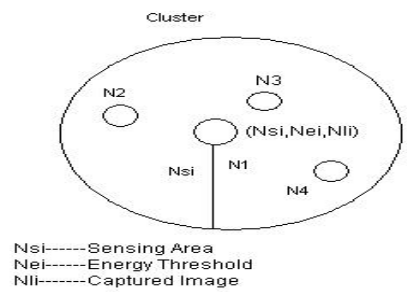


Figure 1. Cluster Formation

Cluster formation in the network occurs based on the fact that sensor nodes n_i and n_j have the same sensing area such that the distance D_{ij} between the nodes n_i and n_j should be less than the sensing area of node n_i given by N_{si} , as shown in Figure 1.

3.3.1 Calculation of Overlapping Factor

Consider a digital image f be represented by a 2 dimensional array of M x N pixels represented by {0,1,2,…..M-1} x {0,1,2,……N-1} having gray levels mentioned in the set {0,1,2…….L-1}. Let X = {X_i / i=1, 2, 3 …N} denote the reference image and Y= {Y_i/ i=1, 2, 3 …N} denote the captured image by the sensor nodes.

In VSN networks, the sensor nodes capturing the images are located at very short distances from each other. Hence the image magnifications and view points with respect to images taken from the neighboring sensor nodes are assumed to be similar. From the work presented in [15], it is considered that the overlapping content among the captured pictures can be carried out geometrically and also the energy information present in the overlapping portions is less in comparison to the other tasks in the sensor network.

3.3.2 Node Selection and Activation

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Consider the nodes $\{N_1, N_2, N_3 \ldots\ldots\ldots N_n\}$ forming the cluster C_i in the VSN. A sensor node N_i from the cluster characterized by the following properties N_{si}, N_{ei} and N_{Ii} overrides another sensor node N_j with characteristics N_{sj}, N_{ej}, N_{Ij}, iff the following conditions are fulfilled. **Condition 1**: The Overlapping factor of the Image calculated should be greater for node N_i than for node N_j.

Condition 2: Energy at node N_i is greater than node N_j.

Employing the image overlapping technique to the nodes in the cluster, the network lifetime can be calculated as follows:

Best Case:

For a VSN consisting of n nodes: n/2 nodes are activated for transmission. n/2 nodes are in low power/standby mode.

At clock interval 1 unit, energy at the nodes reduces by amount E_i energy

At clock interval 2 units, energy at the nodes reduces by amount E_{i+1} energy

At clock interval t1 units, energy at the nodes reduces by amount | E|/n X n/2 active nodes = | E|/ 2 energy units.

At clock interval t2 units, energy at the nodes reduces by |E|. Total life time of visual sensor network can be given by energy at intervals 1…t1 and t1…t2.

$$= |E|/2 + |E|/2 \qquad (3)$$

= | E|Same energy is utilized for two intervals 1…t1 and t1…t2 as shown in Figure 2(b).

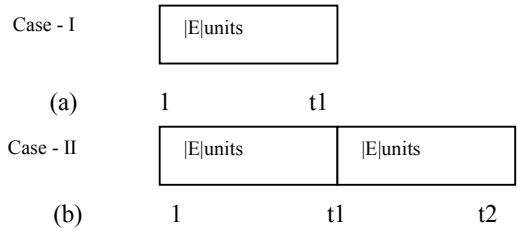


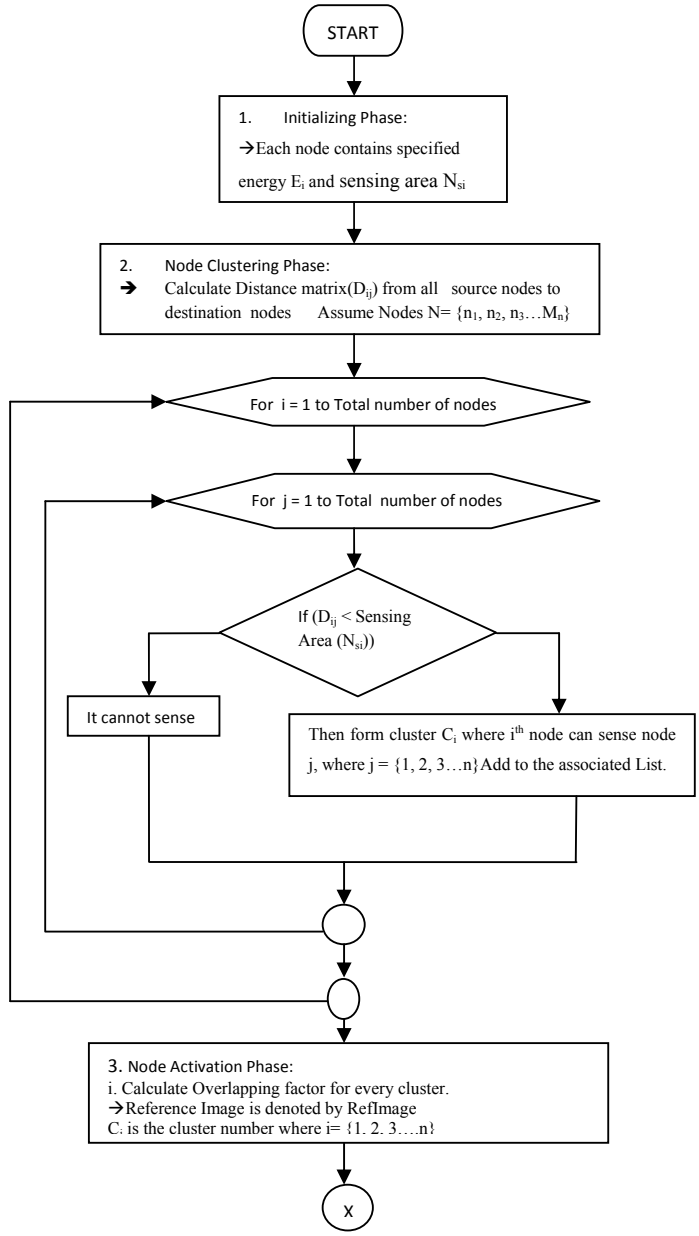
Figure 2 . Time Slot Diagram

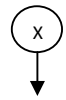
4. **Flowchart and System Model**

The sensor network lifetime can be increased by forming the sensor node clusters, based on quality metrics (Energy, distance and Overlapping factor) that is applied to spatially close visual sensor nodes in order to find correlation between the images to get rid of information redundancy that exists between the nodes.

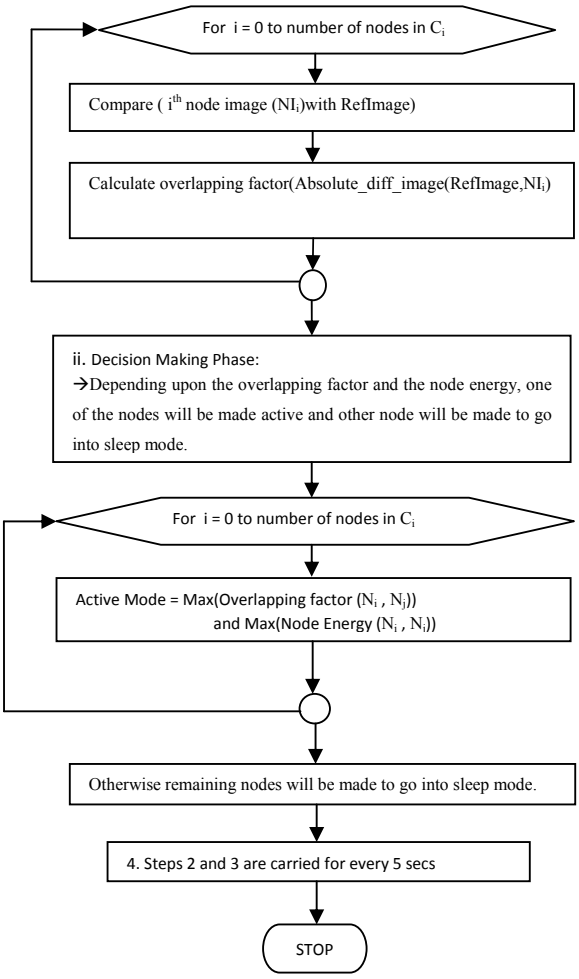
4.1 Flowchart Using Overlapping Based Energy-Aware Technique

Consider in a visual sensor network,





Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.



4.2. ADVANTAGES:

1. Redundant information capturing and transmission of image can be avoided which helps in saving energy.
2. It increases the life time of the network.

4.3. LIMITATIONS:

1. Some amount of energy is dissipated while calculating the overlapping factor between the images from the nodes.
2. There is a small overhead of image transmission along with NodeID.

Following are the information formats used in the visual sensor network:

Packet format used for image transmission to the co-ordinator

NodeID	Image	Energythreshold	Timestamp

Packet format used to instruct the node to be active or to sleep

NodeID	Control Signal Active/Sleep

5. Simulation Set-Up

In this section, we discuss the set-up for VB .Net simulations performed with Image overlapping based energy aware approach and without overlapping of image.

5.1 Network Scenarios

We carried out simulations with 100 visual sensor nodes deployed over an area of size 640 X 480 pixels**,** which is programmed using VB .net. The nodes are deployed either randomly or non-uniformly in the network using a graphical interface as shown in figure 3. In case of the random deployment, the (x, y) locations of sensor nodes in the network are randomly chosen. The non-uniform deployment corresponds to the case where user manually places nodes within a certain area of network.

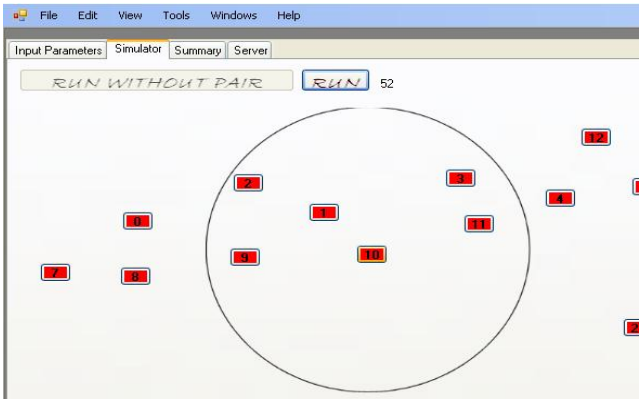


Figure 3. GUI For Visual Sensor Arrangement In The Network

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Parameters	Value
Tx/Rx electronics Constant	50 nJ/bit
Amplifier Constant	10pJ/bit/m^2
Path Loss Exponent	2
Energy Threshold	10^{-4} J
Packet Size	30 bytes
Packet Rate	1 Packet/Sec
Sensing Range	15mts
Cluster Range	30mts

Table 1. Simulation Parameters

The input parameters considered for simulation is as given in Table 1.

5.2 Energy Model

We have used the free-space energy model defined in [13], where the energy required to transmit a *p*-bit packet is equal to

$$E_{tx} = p.(E_{amp} + \varepsilon_{fs} .d^n) \quad (4)$$

The energy required to receive a *p*-bit packet is

$$E_{rx} = p.E_{amp} \quad (5)$$

The parameters E_{amp} and ε_{fs} are the parameters of the transmission / reception circuitry and n is the path-loss exponent as listed in Table 1.

6. Result Analysis

Here, we have compared performance parameters such as network life time, number of dead nodes, active nodes and sleep nodes in the analysis.

6.1. Comparison of Network Life Time

Overlapping based energy aware algorithm is compared with normal simulation without overlapping and it is observed that network lifetime approximately doubles in case of overlapping based energy aware approach as seen from figure 4. Figure 5 and figure 6 show the comparative analysis of as to at what instance of time the sensor nodes become dead in both cases - with overlapping based energy aware algorithm and without overlapping.

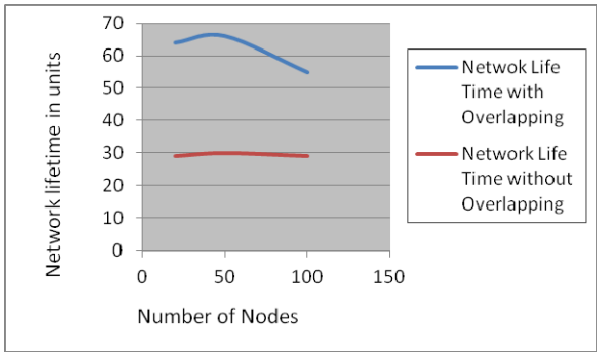


Figure 4. Comparison of Network Life Time (Overlapping Based Energy Aware Algorithm Vs Without Overlapping)

6.2 Analysis of Dead Nodes

Consider different sets of total number of nodes such as 20, 30, 60 and 100 present within the visual sensor network. For these sets the observation is made as to when the sensor node loses its energy and the time interval at which the sensor node becomes a dead node. From figure 8, in case of without overlapping, it is seen for total of 20 nodes in the network, at approximately 27 units of time, 4 nodes become dead and gradually all the sensor nodes go dead at 29th time unit, i.e. the network life ends at 29 time units. From figure 7, in case of overlapping based energy aware algorithm, for total of 20 nodes in the network, approximately at 27 units of time none of the nodes dissipate their energy. The situation starts changing approximately from the 30th time units where dead nodes begin to occur and gradually increase in number. And finally at 50th time units all the nodes in the network are dead, i.e. life time of sensor network is 50 units of time, which is nearly double the value as compared to without overlapping.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

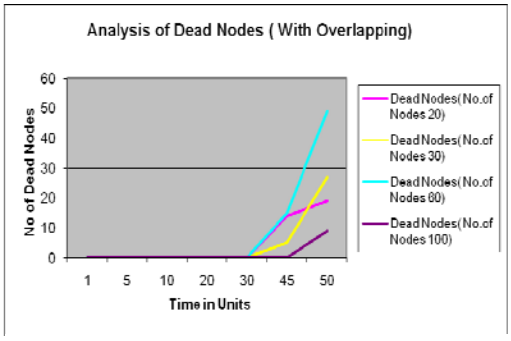


Figure 5. Shows The Analysis Of When The Sensor Nodes Will Become Dead Using Overlapping Based Energy Aware Approach For Different Sets Of VSNNodes.

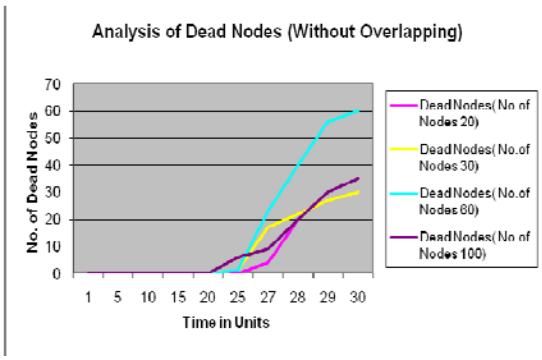


Figure6. Analysis Of When VSN Nodes Will Become Dead.

6.3 Analysis of Active Nodes

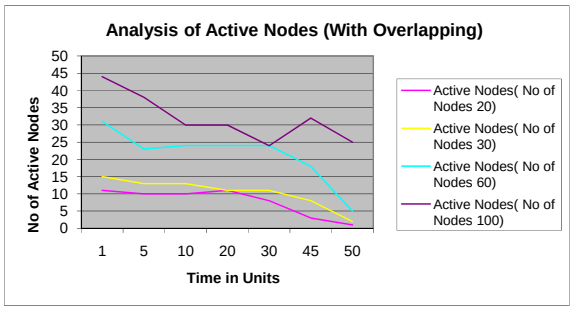


Figure 7. Analysis of Active Nodes with Overlapping

Figure 7 gives the analysis of the number of active nodes in case with overlapping. In case of overlapping based energy aware algorithm the number of active nodes is almost half the total nodes while the other half are in sleep mode at the initial stage and then gradually decrease in number as the time progresses due to the overlapping concept thus increasing the lifetime of the network to almost double the value achieved for without overlapping case. For example consider the total sets of nodes as 100 within the network. At time unit 1, the number of active nodes is nearly 40, which drops down to 35 at 5 time units. At 7th time unit the active node number drops to 30 and remains constant at this value from this time unit till 17 time units and then at 20th time unit the number of active nodes decreases to 27 and then finally at 50th time unit the number of active nodes is 25. Similar observations can be made for other sets of total nodes as shown in figure 9. In this case it is seen that the lifetime of the network is 50 time units for all the sets of total nodes which shows an increase in network life time.

6.4. Analysis of Sleep Nodes

Figure 8 gives the analysis of sleep nodes in case of overlapping based energy aware algorithm. Consider the scenario when there are 100 total nodes in the network. Observe from the graph at 1 time unit the number of sleep nodes is 56 i.e. the rest 44 nodes are active and are engaged in image capturing. At time unit 10 the number of sleep nodes is 70 and the remaining 30 nodes are active. And finally at 50th time unit the number of sleep nodes is 46.

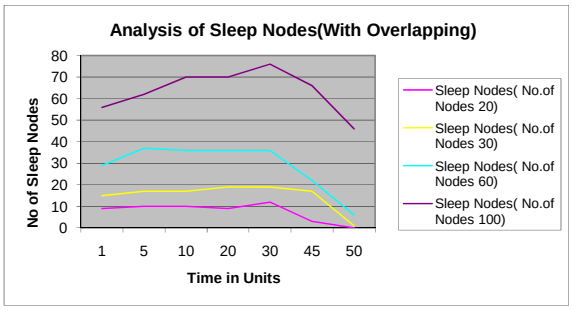


Figure 8. Analysis of Sleep Nodes

7. CONCLUSION

We propose algorithm with overlapping based energy aware concept for selection of visual sensor node from cluster whose aim is to increase network life time. This

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

can be done based on calculating overlapping factor of image between visual sensor node within their sensing area and node energy. Thus the sensor nodes with maximum overlapping factor and energy in the network are made active and other nodes are made to go into sleep mode. Using this approach, simulations and analysis carried out show increase in the network life time and reduction in data redundancy when compared with non-overlapping based energy aware technique.

References

[1] Sanislava Soro and Wendi Heinzelman, "A Survey of Visual Sensor Networks" Advances in Multimedia Volume 2009, Article ID 640386, 21 pages.

[2] Youssef Chari, Bell Canada, Naoki Wakamiya, "Challenging Issues in Visual Sensor Networks", IEEE Wireless Communication, April 2009, page no 44-49.

[3] F.Akyildiz "A Survey on Sensor Networks", IEEE Communications Magzine., Vol.40, Aug.2002,pp. 102-114.

[4] F.Akyildiz, T.Melodia and K.R.Chowdhary, "A Survey on wireless Multimedia Sensor Networks", Elsevier, Computer Network, Vol.51, March 2007, pp.921-960.

[5] W.Heinzelman, A.Chandrakasan, Balakrishnan, "An application specific protocol architecture for Wireless micro sensor networks", IEEE Transactions on Wireless Communications 1 (4) 2002.

[6] M.J. Handy, M.Haase, D.Timmermann, "Low energy clustering hierarchy with deterministic cluster head selection", in: proceedings of IEEE MWCN, 2002.

[7] O. Younis, S.Fahmy, Distributed Clustreing in ad-hoc sensor networks: a hybrid, energy efficient approach, IEEE Transaction on Mobile computing 3 (4) (2004).

[8] T.Shu, M.Krunz, S.Vrudhula, "Power balanced coverage-time optimization for clustered wireless sensor networks", in: proceedings of MobiHoc'05, 2005.

[9] S.Soro, W.Heinzelman, Prolonging the lifetime of wireless sensor networks via unequal clustering, in: Proceeding of the Fifth International workshop on Algorithm for wireless, Mobile, Ad Hoc and Sensor Networks (IEEE WMAN'05), 2005.

[10] H.Chan , A.Perrig, ACE: "An emergent algorithm for highly uniform cluster formation", in Proceedings of EWSN, 2004.

[11] K.Bae, H.Yoon, "Autonomous Clustering Scheme for wireless sensor networks using coverage estimation self-pruning", IEICE Transcation on Communication E88-B (3) (2005).

[12] M.T. Thai, F.Wang, D.Du, "Coverage problems in wireless sensor networks- designs and analysis", International Journal of Sensor networks, Special Issue on Coverage Problems in Sensor networks, 2005.

[13]Stanislava Soro and Wendi Heinzelman, "Cluster head election techniques for coverage preservation in wireless sensor networks", Ad Hoc Networks, Elsevier, Sept 2008.

[14] D. Tian, N.D. Georgans, "A Coverage-Preserving node scheduling scheme for large wireless sensor networks", in proceeding of ACM wireless Sensor Network and Application Workshop, 2002.

[15] Min Wu and Chang Wen Chen, "Collaborative Image Coding and Transmission over Wireless Sensor Networks", EURASIP Journal on Advances in Signal Processing, Volume 2007, Article ID 70481, 9 pages.

[16] K-Y Chow, K-S Lui and E.Y. Lam, "Balancing Image Quality and Energy Consumption in Visual Sensor Networks", IEEE, 2006.

[17] Aswin C. Sankaranarayanan, Ashok Veeraraghavan and Rama Chellappa, "Object Detection, Tracking and Recognition for Mutiple Smart Cameras",

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Proceedings of the IEEE, Vol.96, No.10, October 2008.

[18] Mohammad Hammoudeh, Robert Newman, "An approach to Data Extraction and Visualization for Wireless Sensor Networks", 2009 Eigth International Conference on Networks,pp 156-160.

[19] Huadong Ma and Yonghe Liu, "Correlation based video processing in video sensor networks", 2005 International Conference on wireless Networks, communications and Mobile Computing.

Asst. Prof. Harish H. Kenchannavar received his Masters degree (M.E) from Shivaji University, Walchand College of Engineering & Tech, Sangli, Maharashtra. He has registered for Phd degree under Visvesvaraya Technological University, Belgaum. Research Center:- SDMCET, Dharwad, Karnataka under the guidance of Prof. Umakant P. Kulkarni. He is currently working as Assistant Professor in Computer Science Department, Gogte Institute of Technology, Belgaum, Karnataka. He has presented and published papers at national and IEEE International conferences. His research interests include computer networking, simulation and modeling, visual sensor networks and queuing algorithms.

Prof. Umakant P. Kulkarni received his PhD degree from Shivaji University, Kolhapur. Research Center Walchand College of Engineering & Tech, Sangli, Maharashtra in Nov 2007 . He has completed his Masters degree (M.E) from PSG, Coimbatore. He is currently working as Professor in Computer Science Department, SDMCET, Dharwad, Karnataka. He has published and presented a number of papers in many reputed journals and at International IEEE conferences as well. This apart he also has presented technical talks and tutorials on various aspects of Mobile Agents. His research interests include mobile agents, computer networking, and distributed networking.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Comment on "Quantum Multiplexer Designing and Optimization applying Genetic Algorithm"

Subhasis Pal, Arijit Roy and Dibyendu Chatterjee

Department of Electronics, West Bengal State University Department
Berunanpukuria, Barasat, Kolkata 700 126, India

Abstract

We have noticed that many fundamental laws and rules of quantum-circuit are violated in a work published in International Journal of Computer Science Issues [1]. The authors of this published work presented a quantum multiplexing circuit which is composed of half-adders, i.e. quantum multiplexing functionality is claimed to achieve by using few units of quantum half-adder. The authors performed 'qubit copying', 'fanout' etc. in the quantum circuit which are prohibited. Apart from these, two quantum circuits are claimed to be equivalent to each other though their operational matrices are not same. In this article, we have explained the mentioned violations in details.

Keywords: *Quantum Circuit, Quantum Multiplexer, Quantum Bit Copying.*

1. Comment – I

1.1 Section A

The 4:1 QMUX proposed in the ref. [1] is constructed using four quantum half-adder circuit. According to the authors of Ref. [1], the dotted blocks in the QMUX circuits shown in Fig. 3 and Fig. 2 (fourth QHA section and its equivalent circuit) of Ref. [1] are equivalent circuit of QHA [2]. But if we simplify the QHA using CNOT, Controlled-square-root of NOT gate (i.e. CV gate) and Controlled-square-root of NOT adjoint (i.e. $CV^{\dagger}$ gate) then the resulted circuit does not matches with the dotted portion of the circuits in Ref. [1]. Additionally, the tensor like symbol '⊗' used the quantum circuits in Ref. [1] provides no meaningful operation.

The fourth QHA section of the multiplexer circuit and its equivalent circuit are shown in Fig. 2 of Ref. [1] and these circuits are repeated for ease of explanation and shown in Fig. 1 of this article.

The authors of Ref. [1] claim that the bottom circuit of the circuit shown Fig. 1 (a) is equivalent of the top circuit of the same figure. However, we found that this is not true. Next, we have provided an explanation to proof that the said circuits are not equivalent.

Proof: Since, the authors used non-conventional circuit symbol, first we assume tensor like symbol represents CNOT gate and Controlled-Controlled tensor like symbol represents a Toffoli gate. Based on these assumptions we have re-drawn the circuit which is shown in Fig. 1 (b). Note that the assumed circuit is not too an equivalent of QHA circuit. The authors used some two-qubit gates to represent a desired circuit. This combination of the primitive gates can be optimized and such an optimized circuit is shown in Fig. 2.

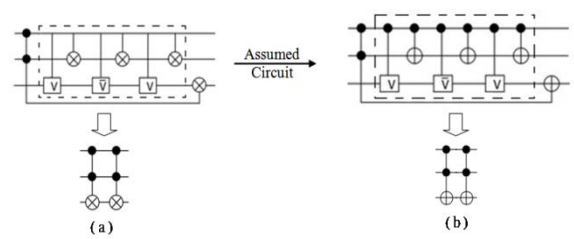


Fig. 1 (a) Fig. 2 of Ref. [1]. (b) Our assumed circuit of Fig. 1 (a).

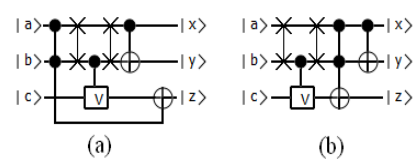


Fig. 2 (a) Optimized circuit of circuit shown in Fig. 1. (b) Equivalent of circuit of (a).

The equivalent circuit shown in Fig. 2(b) is obtained by applying moving rule and deletion rule [3] on the circuit shown in Fig. 2(a). The operational matrix of the circuit shown in Fig. 2 can be expressed as (with self explanatory notation):

$M_1 = (U_{SWAP} \otimes ID) \cdot (ID \otimes U_V) \cdot (U_{SWAP} \otimes ID) \cdot U_{Toffoli} \cdot (U_{CN} \otimes ID)$

$$M_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1+i}{2} & \frac{1-i}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1-i}{2} & \frac{1+i}{2} \\ 0 & 0 & 0 & 0 & \frac{1-i}{2} & \frac{1+i}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1+i}{2} & \frac{1-i}{2} & 0 & 0 \end{pmatrix} \quad \text{.....Eq. (1)}$$

On the other hand, the operational matrix (M_2) of the

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

equivalent circuit shown in Fig. 1(b) (bottom circuit) can be expressed as:

$$M_2 = \begin{pmatrix} 1&0&0&0&0&0&0&0\\ 0&1&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0\\ 0&0&0&1&0&0&0&0\\ 0&0&0&0&1&0&0&0\\ 0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&0&1\\ 0&0&0&0&0&0&1&0 \end{pmatrix} \cdot \begin{pmatrix} 1&0&0&0&0&0&0&0\\ 0&1&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0\\ 0&0&0&1&0&0&0&0\\ 0&0&0&0&1&0&0&0\\ 0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&0&1\\ 0&0&0&0&0&0&1&0 \end{pmatrix}$$

$$= \begin{pmatrix} 1&0&0&0&0&0&0&0\\ 0&1&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0\\ 0&0&0&1&0&0&0&0\\ 0&0&0&0&1&0&0&0\\ 0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&1&0\\ 0&0&0&0&0&0&0&1 \end{pmatrix} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{Eq(2)}$$

Since, the symbol used in the equivalent circuit (see Fig. 1(a)) is non conventional, we have assumed an equivalent circuit shown at the bottom of the Fig. 1(b) to derive the Eq. (2).

From Eq. (1) and (2), one can see that $M_1 \neq M_2$ and hence, the circuits are not equivalent to each other. Moreover, the matrix M_2 is an identity matrix, i.e. it represents quantum wire (since successive C^2NOT operation gives the output same as input) in quantum circuit and it does nothing to the inputs. In short, the operational matrices of the circuits shown in Fig. 1 (top and bottom) are not identical and hence, they cannot be equivalent to each other.

1.2 Section B

Moreover, the authors claimed the dotted portion in the Fig 1(a) is an equivalent circuit of QHA, which is not true. The QHA circuit and its equivalent circuits are shown in Fig 3.

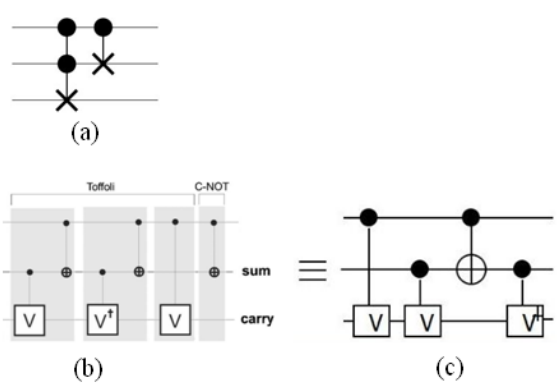
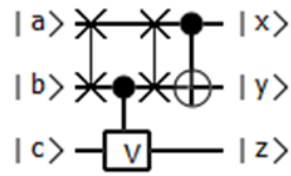


Fig. 3 (a) The QHA circuit. (b) and (c) are equivalent circuits of QHA.

The equivalent circuit of the dotted portion of the Fig. 1 (a) is shown in Fig. 4 (a). The operational matrix (M_3) of the dotted portion of the Fig 1(a) is computed and is shown in Fig. 4 (b). The operational matrix (M_4) of the QHA (circuit in Fig. 3) is shown in Fig. 4 (c). It is therefore evident from the Fig. 4 that, $M_3 \neq M_4$. Hence, the dotted portion of the circuit does not represent a QHA which has been used in Ref. [1] to synthesize the multiplexer circuit.



(a)

$$M_3 = \begin{pmatrix} 1&0&0&0&0&0&0&0\\ 0&1&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0\\ 0&0&0&1&0&0&0&0\\ 0&0&0&0&0&0&\frac{1+i}{2}&\frac{1-i}{2}\\ 0&0&0&0&0&0&\frac{1-i}{2}&\frac{1+i}{2}\\ 0&0&0&0&\frac{1+i}{2}&\frac{1-i}{2}&0&0\\ 0&0&0&0&\frac{1-i}{2}&\frac{1+i}{2}&0&0 \end{pmatrix}$$

(b)

$$M_4 = \begin{pmatrix} 1&0&0&0&0&0&0&0\\ 0&1&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0\\ 0&0&0&1&0&0&0&0\\ 0&0&0&0&0&0&1&0\\ 0&0&0&0&0&0&0&1\\ 0&0&0&0&0&1&0&0\\ 0&0&0&0&1&0&0&0 \end{pmatrix}$$

(c)

Fig. 4 (a) Equivalent circuit of dotted portion of Fig 1(a). (b) Its matrix. (c) Matrix of QHA.

2. Comment – II

In any quantum circuit; the number of input lines must equal to number the number of output lines along any cross-section of the circuit [4,5]. This rule is violated by the authors. For example, one can see the circuit shown in Fig. 3 of Ref. [1] (see page no. 365). In that circuit, the number of input line is seven while the number of output lines is nine. From this point of view, we can say that some copy operations exists in the circuit. Therefore, the proposed circuit is irreversible.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

3. Comment – III

Copying (or Fanout) of qubit in quantum circuit is not allowed. This is because; copying of qubit from one quantum wire to another does not represent same state transformation to the copied quantum wire [5]. The transferred states could be different from the intended states. The authors violated this rule and some copy operations are considered into their circuits [1]. For example, one can see Fig. 3 on page 365 of the Ref. [1], in which many qubits are copied. In this circuit, the sates like $|S_0\rangle$, $|S_1\rangle$ etc. are copied and shared among the circuit. The copying of qubit is shown by dotted arrow in Fig. 5.

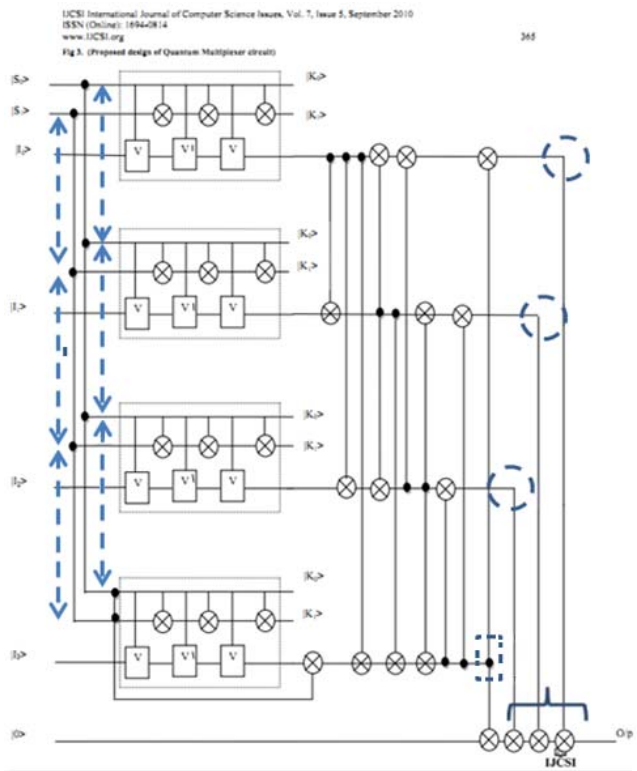


Fig. 5 A figure from Ref. 1 (page 365). The copying of qubits are shown by dotted arrows.

4. Comment – IV

The end of the line shown in Fig. 5 is prohibited in quantum circuit. In Ref. 1, the authors have represented a gate by '⊗' symbol. According to the description mentioned in Ref. 1, the gates shown by a brace in Fig. 5, are controlled gates for which '●' symbol should be used. However, no such symbol is used at position indicated by dotted circle in Fig. 5. Apart from these, the end of a line (shown by dotted circles and dotted square in Fig. 5) is presented in such a way which is possible in classical circuit but not in quantum circuit. Hence, four more output lines are hidden in this circuit and the actual number of output lines raise to thirteen and number of input lines is seven, which in turn represents an irreversible quantum circuit and hence violating the law of quantum mechanics.

References

[1] D. Mukhopadhyay and A. Si, "Qunatum multiplexer design and optimization applying genetic algorithm", *Int. J. Comp. Sci. Issues*, Vol. 7 (5), Start Page 360 (2010).

[2] Geraldo A. Barbosa, "Quantum half-adder", Phy. Rev. A 73, 052321 (2006).

[3] D. Maslov, C. Young, D.M. Miller, G.W. Dueck, "Quantum circuit simplification using templates", *Proc. of the conference on Design, Automation and Test in Europe, Munich, Germany*, Vol. 2, Start Page 1208 (2005).

[4] V.V. Shende, A.K. Prasad, I.L. Markov, John P. Hayes, "Synthesis of Reversible Logic Circuits", *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, Vol. 22 (6), Start Page 710 (2003).

[5] M.A. Nielsen, I.L. Chuang I., Quantum computation and quantum information, Cambridge University Press (2002).

First Author Mr. Subhasis Pal is graduated as a student of Electronics (Hons.) from Vidyasagar University (West Bengal, India) in 2006 and M.Sc in Electronic Science from Jadavpur University (West Bengal, India) in 2008. He got first class first position (gold medalist) in graduation. Presently, he is pursuing Ph.D. at the Department of Electronics, West Bengal State University (Barasat, India). His area of research is Quantum Computation.

Second Author Dr. Arijit Roy has attended many prestigious institutes like IITs and Infineon Technologies in his career. He earned Ph.D. degree from School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. His publication includes regular paper, review article, conference article etc. in international journals and conferences. His paper in Symposium On Electronics – 2004 (Singapore) won the second best paper award. Among his various publications, a 75-journal-page-size review article on "Electromigration" is published in a journal of impact factor 17.731. So far his research works have been cited one hundred times (excluding self citations) and his present research h-index is five. In 2011, Dr. Roy authored a research monograph, titled "Electromigration in Cu Interconnects, The Driving Force Formalism: Modeling and Experiments", published by Lambert Academic Publishing, Germany (ISBN: 978-3-8454-1292-4). Dr. Roy is also involved as reviewer, advisory board member etc. for many international publishers. His research interests include Micro- and Nano-electronics, Microelectronic Reliability, Quantum Circuits and Single Electron Transistor. Presently he is Assistant Professor and HOD of the Electronics Department at West Bengal State University (Barasat, India).

Third Author Mr. Dibyendu Chatterjee completed his B.Sc. degree in Electronics (Hons.) and stood first class first (gold medalist) from Bankura Christian College (under Burdwan University, West Bengal, India) in 2006 and M.Sc. degree in Electronic Science from Jadavpur University (West Bengal, Kolkata, India) in 2008. He is also qualified prestigious examinations such as NET (National Eligibility Test, conducted by UGC govt. of India) and GATE (Graduate Aptitude Test in Engineering, conducted by IITs) in Electronics. Presently Mr. Chatterjee is pursuing Ph.D. at the Department of Electronics, West Bengal State University (Barasat, India). His main area of research is Quantum Computation.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Software Size Estimation Using Fuzzy Backpropagation Network Method

B V A N S S PRABHAKAR RAO[1] & P SEETHA RAMAIAH[2]
[1] Associate Professor, Dept. of CSE, Miracle Educational Society Group of Institution
Vizianagaram, Andhra Pradesh 535 216, India

[2] Professor, Dept. of CS & SE, College of Engineering (A), Andhra University
Visakhapatnam, Andhra Pradesh 530 003, India

Abstract

This paper presents the problems of effort estimation for software development projects. Most of the research has focused on the construction of formal software effort estimation models. All most all the models were typically based on the regression analysis or mathematically derived from theories from other domains. This paper focuses on Soft Computing Hybrid Systems in general and Fuzzy Back-propagation Network method in particular. Though Hybrid Systems have a tremendous potential to solve problems, an inappropriate use of the technology can backfire. Hybrid systems are those which employ integrated technologies to effectively solve the problem. Software development efforts estimation is the process of predicting the most realistic use of effort required to develop or maintain software product in an optimized way for the benefit of all the stakeholders. Effort estimates may be used as input to the plans, budgets, investment, pricing processes and bidding rounds. In this context we are suggesting the Fuzzy based Backpropagation Network to solve the problem and to produce the better results.

Keywords**:** *Artificial Neural Network, Backpropagation, Neuro-Fuzzy, Software Estimation, Fuzzy Logic, LOC, Sizing.*

1. Introduction

Software Project planning encompasses five major activities – estimation, scheduling, risk analysis, quality management planning, and change management planning. Estimation includes your attempt to determine how much money, effort, resources, and time it will take to build a specific software-based product. Software project managers using information solicited from project stakeholders and software metrics data collected from past projects [1, 6]. Most of the software researchers and practitioners have been addressing the problems of effort estimation for software development projects since at least the 1960s; Published surveys on estimation practice suggest that expert estimation is the dominant strategy when estimating software development effort. Typically, effort estimates are over-optimistic and there is a strong over-confidence in their accuracy. The mean effort overrun seems to be about 30% and not decreasing over time. Currently the term "effort estimate" is used to denote as different concepts as most likely use of effort (modal value), the effort that corresponds to a probability of 50% of not exceeding (median), the planned effort, the budgeted effort or the effort used to propose a bid or price to the client. This is believed to be unfortunate, because communication problems may occur and because the concepts serve different goals [2] [3].

Estimation begins with a description of the scope of the problem. The problem is then decomposed into a set of smaller problems, and each of these is estimated using historical data and experience as guides. Problem complexity and risk are considered before a final estimate is made.

2. Challenges

The advantage of buying software is automation, better analyzing of business, thus making life easier. Software Development is intangible, Customer can see the benefits of buying only when he uses the software for quiet a period of time and then says ' its worth of it'. We do not quote too less, programmers work for overnight that leads to lose the project or end doing social service, or loss. Do not quote too high that lose the project. So, be fair to ourselves and our customers. Hence, there is need to use of a repeatable, clearly defined and well understood software development process that has to be the most effective method.

- It's easy to estimate what you know.
- It's hard to estimate what you know you don't know.
- It's very hard to estimate things that you don't know you

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

don't know.

3. Estimation methods

The uncertainty of an effort estimate can be described through a prediction interval (PI). An effort PI is based on a stated certainty level and contains a minimum and a maximum effort value.

3.1 Selection of Estimation approach

The use of artificial neural networks is a cross-disciplinary field that integrates neuroscience, information science and computer sciences. Among many neural network models, the back-propagation (BP) neural network displays a strong learning ability using nonlinear models with a high fault tolerance. The evidence on differences in estimation accuracy of different estimation approaches and models suggest that there is no "best approach". There are many ways of categorizing estimation approaches, see for example [4][5]. The top level categories are the following:

3.2 Expert estimation: The quantification step based on judgmental processes.

3.3 Formal estimation model: The quantification step is based on mechanical processes. Formal estimation models not tailored to a particular organization's own context, may be very inaccurate. Use of own historical data is consequently crucial if one cannot be sure that the estimation model's core relationships (e.g., formula parameters) are based on similar project contexts [6, 7].

3.4 Combination-based estimation: The quantification step is based on a judgmental or mechanical combination of estimates from different sources.

4. Artificial Neural Network

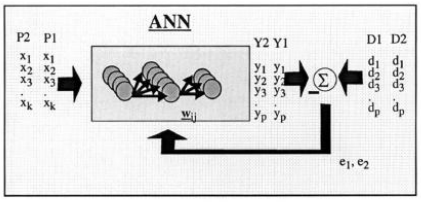
An artificial neural network is a system based on the operation of biological neural networks, in other words, is an emulation of biological neural system. Why would be necessary the implementation of artificial neural networks? Although computing these days is truly advanced, there are certain tasks that a program made for a common microprocessor is unable to perform; even so a software implementation of a neural network can be made with their advantages and disadvantages.

Advantages:

- A neural network can perform tasks that a linear program cannot.
- When an element of the neural network fails, it can continue without any problem by their parallel nature.
- A neural network learns and does not need to be reprogrammed.
- It can be implemented in any application.
- It can be implemented without any problem.

Of course the neural network needs training to operate.

In the world of engineering, neural networks have two main functions: Pattern classifiers and as non linear adaptive filters. As its biological predecessor, an artificial neural network is an adaptive system. By adaptive, it means that each parameter is changed during its operation and it is deployed for solving the problem in matter. This is called the training phase. An artificial neural network is developed with a systematic step-by-step procedure which optimizes a criterion commonly known as the learning rule. The input/output training data is fundamental for these networks as it conveys the information which is necessary to discover the optimal operating point.



The style of neural computation.

Basically, an artificial neural network is a system. A system is a structure that receives an input, process the data, and provides an output. Commonly, the input consists in a data array which can be anything such as data from an image file, a WAVE sound or any kind of data that can be represented in an array. Once an input is presented to the neural network, and a corresponding desired or target response is set at the output, an error is composed from the difference of the desired response and the real system output. The error information is fed back to the system which makes all adjustments to their parameters in a systematic fashion (commonly known as the learning rule). This process is repeated until the desired output is acceptable.

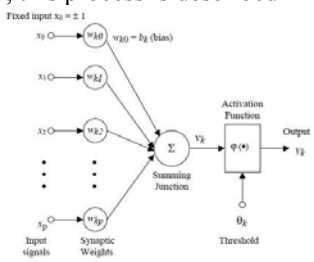
4.1 The Mathematical Model

Once modeling an artificial functional model from the biological neuron, we must take into account three basic components. First off, the synapses of the biological neuron are modeled as weights. Let's remember that the synapse of the biological neuron is the one which

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

interconnects the neural network and gives the strength of the connection. For an artificial neuron, the weight is a number, and represents the synapse. A negative weight reflects an inhibitory connection, while positive values designate excitatory connections. The following components of the model represent the actual activity of the neuron cell. All inputs are summed altogether and modified by the weights. This activity is referred as a linear combination. Finally, an activation function controls the amplitude of the output. For example, an acceptable range of output is usually between 0 and 1, or it could be -1 and 1.

Mathematically, this process is described in the figure



From this model the interval activity of the neuron can be shown to be:

$$v_k = \sum_{j=1}^{p} w_{kj} x_j$$

The output of the neuron, yk, would therefore be the outcome of some activation function on the value of vk.

4.2 Activation functions

As mentioned previously, the activation function acts as a squashing function, such that the output of a neuron in a neural network is between certain values (usually 0 and 1, or -1 and 1). In general, there are three types of activation functions, denoted by Φ(.) . First, there is the Threshold Function which takes on a value of 0 if the summed input is less than a certain threshold value (v), and the value 1 if the summed input is greater than or equal to the threshold value.

$$\varphi(v) = \begin{cases} 1 & \text{if } v \geq 0 \\ 0 & \text{if } v < 0 \end{cases}$$

Secondly, there is the Piecewise-Linear function. This function again can take on the values of 0 or 1, but can also take on values between that depending on the amplification factor in a certain region of linear operation.

$$\varphi(v) = \begin{cases} 1 & v \geq \frac{1}{2} \\ v & -\frac{1}{2} > v > \frac{1}{2} \\ 0 & v \leq -\frac{1}{2} \end{cases}$$

Thirdly, there is the sigmoid function. This function can range between 0 and 1, but it is also sometimes useful to use the -1 to 1 range. An example of the sigmoid function is the hyperbolic tangent function.

$$\varphi(v) = \tanh\left(\frac{v}{2}\right) = \frac{1-\exp(-v)}{1+\exp(-v)}$$

4.3 A framework for distributed representation

An artificial neural network consists of a pool of simple processing units which communicate by sending signals to each other over a large number of weighted connections. A set of major aspects of a parallel distributed model can be distinguished:

- a set of processing units ('neurons,' 'cells');
- a state of activation yk for every unit, which equivalent to the output of the unit;
- connections between the units. Generally each connection is defined by a weight wjk which determines the effect which the signal of unit j has on unit k;
- a propagation rule, which determines the effective input sk of a unit from its external inputs;
- an activation function Fk, which determines the new level of activation based on the efective input sk(t) and the current activation yk(t) (i.e., the update);
- an external input (aka bias, offset) øk for each unit;
- a method for information gathering (the learning rule);
- an environment within which the system must operate, providing input signals and if necessary error signals.

4.4 Neural Network topologies

In the previous section we discussed the properties of the basic processing unit in an artificial neural network. This section focuses on the pattern of connections between the units and the propagation of data. As for this pattern of connections, the main distinction we can make is between:

- Feed-forward neural networks
- Recurrent neural networks

Classical examples of feed-forward neural networks are the Perceptron and Adaline.

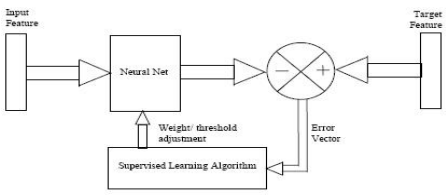
4.5 Training of artificial neural networks

A **neural network** has to be configured such that the application of a set of inputs produces (either 'direct' or via a relaxation process) the desired set of outputs. Various methods to set the strengths of the connections exist. One way is to set the weights explicitly, using a priori knowledge. Another way is to **'train' the neural network**

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

by feeding it teaching patterns and letting it change its weights according to some learning rule.
We can categorise the learning situations in two distinct sorts. These are:

- **Supervised learning** or Associative learning in which the network is trained by providing it with input and matching output patterns. These input-output pairs can be provided by an external teacher, or by the system which contains the neural network (self-supervised).



- **Unsupervised learning** or Self-organisation in which an (output) unit is trained to respond to clusters of pattern within the input. In this paradigm the system is supposed to discover statistically salient features of the input population. Unlike the supervised learning paradigm, there is no a priori set of categories into which the patterns are to be classified; rather the system must develop its own representation of the input stimuli.

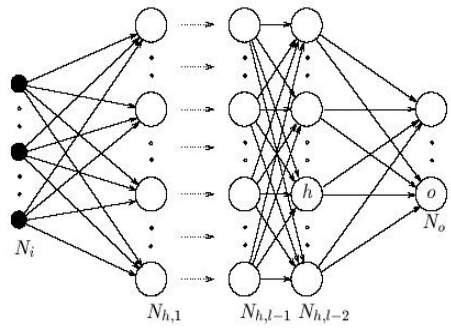
- **Reinforcement Learning** This type of learning may be considered as an intermediate form of the above two types of learning. Here the learning machine does some action on the environment and gets a feedback response from the environment. The learning system grades its action good (rewarding) or bad (punishable) based on the environmental response and accordingly adjusts its parameters. Generally, parameter adjustment is continued until an equilibrium state occurs, following which there will be no more changes in its parameters. The self organizing neural learning may be categorized under this type of learning.

4.6. Multi-layer feed-forward networks

A feed-forward network has a layered structure. Each layer consists of units which receive their input from units from a layer directly below and send their output to units in a layer directly above the unit. There are no connections within a layer. The Ni inputs are fed into the first layer of Nh;1 hidden units. The input units are merely 'fan-out' units; no processing takes place in these units. The activation of a hidden unit is a function Fi of the weighted inputs plus a bias, as given in in eq

$$y_k(t+1) = \mathcal{F}_k(s_k(t)) = \mathcal{F}_k\left(\sum_j w_{jk}(t)\, y_j(t) + \theta_k(t)\right),$$

The output of the hidden units is distributed over the next layer of Nh;2 hidden units, until the last layer of hidden units, of which the outputs are fed into a layer of No output units .



Although backpropagation can be applied to networks with any number of layers, just as for networks with binary units it has been shown (Hornik, Stinchcombe, & White, 1989; Funahashi, 1989; Cybenko, 1989; Hartman, Keeler, & Kowalski, 1990) that only one layer of hidden units success to approximate any function with finitely many discontinuities to arbitrary precision, provided the activation functions of the hidden units are non-linear (the universal approximation theorem). In most applications a feed-forward network with a single layer of hidden units is used with a sigmoid activation function for the units.

5. Understanding Backpropagation

The equations derived in the previous section may be mathematically correct, but what do they actually mean? Is there a way of understanding back-propagation other than reciting the necessary equations? The answer is, of course, yes. In fact, the whole back-propagation process is intuitively very clear. What happens in the above equations is the following. When a learning pattern is clamped, the activation values are propagated to the output units, and the actual network output is compared with the desired output values, we usually end up with an error in each of the output units. Let's call this error eo for a particular output unit o. We have to bring eo to zero The simplest method to do this is the greedy method: we strive to change the connections in the neural network in such a way that, next time around, the error eo will be zero for this particular pattern. We know from the delta rule that, in order to reduce an error, we have to adapt its incoming weights according to. $\Delta w_{ho} = (d_o - y_o) y_h$.

That's step one. But it alone is not enough: when we only apply this rule, the weights from input to hidden units are never changed, and we do not have the full representational power of the feed-forward network as promised by the universal approximation theorem. In order to adapt the weights from input to hidden units, we again want to apply the delta rule. In this case, however, we do not have a value for δ for the hidden units. This is

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

solved by the chain rule which does the following: distribute the error of an output unit o to all the hidden units that is it connected to, weighted by this connection. Differently put, a hidden unit h receives a delta from each output unit o equal to the delta of that output unit weighted with (= multiplied by) the weight of the connection between those units.

5.1 Working with back-propagation

The application of the generalized delta rule thus involves two phases: During the first phase the input x is presented and propagated forward through the network to compute the output values yp o for each output unit. This output is compared with its desired value do, resulting in an error signal δp o for each output unit. The second phase involves a backward pass through the network during which the error signal is passed to each unit in the network and appropriate weight changes are calculated.

5.2 Weight adjustments with sigmoid activation function.

- The weight of a connection is adjusted by an amount proportional to the product of an error signal δ, on the unit k receiving the input and the output of the unit j sending this signal along the connection: $\Delta_p w_{jk} = \gamma \delta_k^p y_j^p.$
- If the unit is an output unit, the error signal is given by $\delta_o^p = (d_o^p - y_o^p)\,\mathcal{F}'(s_o^p).$ Take as the activation function F the 'sigmoid' function as defined $y^p = \mathcal{F}(s^p) = \frac{1}{1+e^{-s^p}}.$ In this case the derivative is equal to

$$\begin{aligned}\mathcal{F}'(s^p) &= \frac{\partial}{\partial s^p}\frac{1}{1+e^{-s^p}}\\ &= \frac{1}{(1+e^{-s^p})^2}(-e^{-s^p})\\ &= \frac{1}{(1+e^{-s^p})}\frac{e^{-s^p}}{(1+e^{-s^p})}\\ &= y^p(1-y^p).\end{aligned}$$

such that the error signal for an output unit can be written as: $\delta_o^p = (d_o^p - y_o^p)\, y_o^p(1-y_o^p).$

- The error signal for a hidden unit is determined recursively in terms of error signals of the units to which it directly connects and the weights of those connections. For the sigmoid activation function:

$$\delta_h^p = \mathcal{F}'(s_h^p)\sum_{o=1}^{N_o}\delta_o^p w_{ho} = y_h^p(1-y_h^p)\sum_{o=1}^{N_o}\delta_o^p w_{ho}.$$

5.3 Learning rate and momentum

The learning procedure requires that the change in weight is proportional to $\partial E^p/\partial w$. True gradient descent requires that in nitesimal steps are taken. The constant of proportionality is the learning rate . For practical purposes we choose a learning rate that is as large as possible without leading to oscillation. One way to avoid oscillation at large , is to make the change in weight dependent of the past weight change by adding a momentum term:

$\Delta w_{jk}(t+1) = \gamma \delta_k^p y_j^p + \alpha \Delta w_{jk}(t),$ where t indexes the presentation number and F is a constant which determines the effect of the previous weight change.

6. Design of the Fuzzy Backpropagation Network

This study aimed at building and evaluating an Artificial Intelligence System in general and neuro-fuzzy model in particular to estimate software size. Neural network techniques are based on the principle of learning from historical data, whereas fuzzy logic is a method used to make rational decisions in an environment of uncertainty and vagueness. However, fuzzy logic alone does not enable learning from the historical database of software projects. Once the concept of fuzzy logic is incorporated into neural network, the result is a neuro-fuzzy system that combines the advantages of both techniques.

Let $x_j \in [0,1] \subset \Re$, $j = 1, \ldots, n_0$, and all network weights $w_{ij}^{(l)} \in [0,1] \subset \Re$, where $l = 1, 2$. Let $P(J)$ be the power set of the set J of the network inputs (indices), where J_{ij} is the jth input of the ith neuron. Further on for simplicity, network inputs with their indices are identified. Let $g : P(J) \to [0,1]$ be a function defined as follows:

$$g(\emptyset) = 0,$$

$$\vdots$$

$$g(\{J_{ij}\}) = w_{ij}^{(1)}, \quad 1 \leqslant j \leqslant n_0,$$

$$\vdots$$

$$g(\{J_{ij_1}, J_{ij_2}, \ldots, J_{ij_r}\}) = w_{ij_1}^{(1)} \vee w_{ij_2}^{(1)} \vee \cdots \vee w_{ij_r}^{(1)},$$

$$\text{where } \{j_1, j_2, \ldots, j_r\} \subset \{1, 2, \ldots, n_0\},$$

$$\vdots$$

$$g(\{J_{i1}, J_{i2}, \ldots J_{ij}, \ldots, J_{in_0}\}) = 1,$$

6.1 Neuro-Fuzzy ystem

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

The hybridization of neural networks and fuzzy logic is the basic idea behind the neuro-fuzzy system. Neuro-fuzzy hybridization is done in two ways: fuzzy neural networks - is a neural network equipped with the capability of handling fuzzy information and neuro-fuzzy systems - is a fuzzy system augmented by neural networks to enhance some characteristics like flexibility and adaptability.

The fuzzy neural network was integrated with an artificial neural network in this study. Fuzzy logic can express the logical meanings commonly used by humans in a more natural and direct way. Logical decision-making was performed according to the language rules proposed by experts, which can solve nonlinear questions that cannot be addressed using rigorous modeling methods. The integration of fuzzy logic with a neural network can facilitate self-adaptation through a learning function and automatically acquire an algorithm for the information expressed as fuzzy or precise data. This method can overcome the difficulties in expressing time-varying knowledge and processes, a feature of fuzzy logic. The integration of these two components can compensate for the insufficiency of a neural network in fuzzy data processing and the deficiencies of pure fuzzy logic in learning functions. To construct a fuzzy neural network structure, this system adds a fuzzification layer before the input layer of the neural network and a defuzzification layer after the output layer, which results in the fuzzification of the input information and defuzzification of the output information, respectively. The premises and the confidence level of the conclusions in fuzzy logic, provided by experts in the field, were used as the inputs and expected outputs of the learning samples for the neural network. With the aid of the strong learning and associative memory capabilities of the artificial neural network, the neural network is trained for its learning function to automatically acquire fuzzy rules that are stored in the network in the forms of weights and thresholds. Thus, our fuzzy neural network obtains the capabilities of analyzing fuzzy questions and making diagnoses and achieves an effective combination of fuzzy logic with a neural network.

Software estimation accuracy is among the greatest challenges for software developers. Software cost estimation is the process of predicting the effort required to develop a software system. Many estimation models have been proposed. This paper provides a general overview of software cost estimation methods. As a number of these models rely on a software size estimate as input, we first provide an overview of common size metrics then we are using Fuzzy Backpropagation Network for estimating the size.

6.2 Software sizing

The software size is the most important factor that affects the software cost. The lines of code and function point are the most popular software size metrics used in practice.

Lines of Code (LOC): This is the number of lines of the *delivered source code* of the software, excluding comments and blank lines and is commonly known as *LOC*. Although LOC is programming language dependent, it is the most widely used software size metric. Most models relate this measurement to the software cost. However, exact LOC can only be obtained after the project has completed. Estimating the code size of a program before it is actually built is almost as hard as estimating the cost of the program.
A typical method for estimating the code size is to use experts' judgment together with a technique called Project Evaluation and Review Technique (*PERT)*.

It involves experts' judgment of three possible code-sizes:
l, the lowest possible size;
h the highest possible size;
and m, the most likely size.
Estimation of the code size, S = (l+h+4m)/6.

PERT can also be used for individual components to obtain an estimate of the software system by summing up the estimates of all the components. It is generally popular among the users. Here, we are suggesting the Fuzzy Backpropagation Network (instead of PERT) for producing better results.

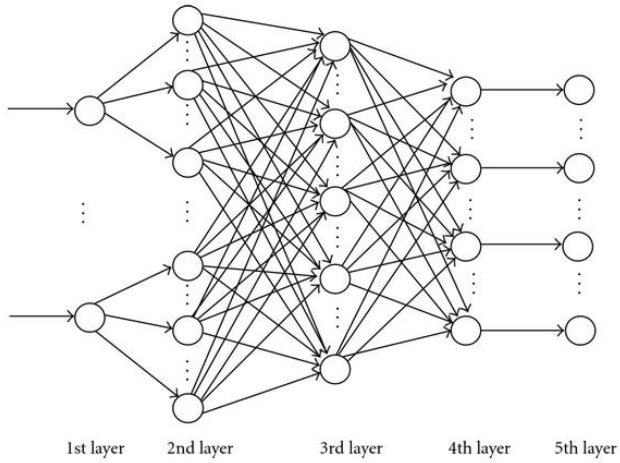


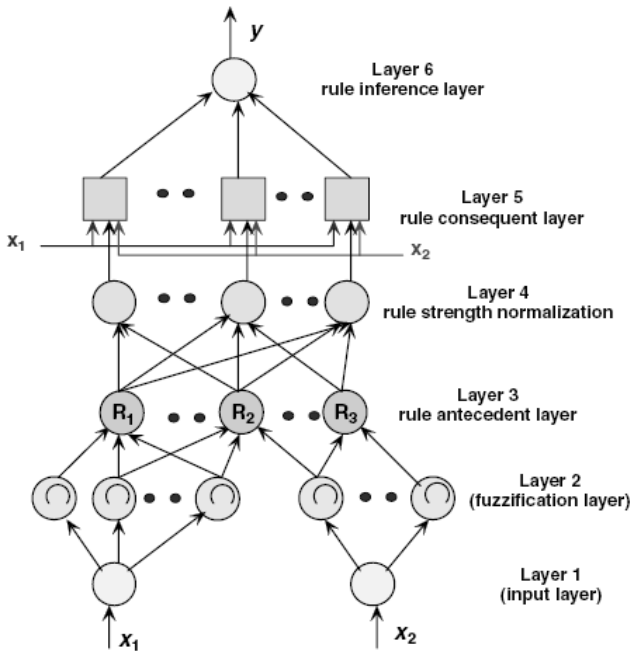
Figure: The structure of the series of fuzzy BP neural network consisting of five layers.

The entire network is composed of two parts; fuzzy processing and a conventional BP network, with the network input the same as a conventional BP network

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

input. Fuzzy processing was performed on the network input through the membership function in the fuzzy processing part, and the processed data were submitted directly to the BP network for further processing. The output data were compared with the expected output and reversely adjusted based on the mean square error to specify the network connection weights.

The structure of a series of fuzzy BP neural networks is shown below.



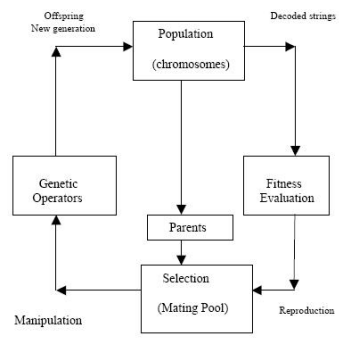
This network consists of five layers. The processing procedure of each layer is as follows.

(1) The first layer is the input layer. The nodes in this layer receive input from the outside and send it to the next layer. The connection weight constant between the first and second layer is 1.
(2) The second layer is the fuzzification layer. It performs fuzzy processing on the input and calculates the membership function value for each input component.
(3) The third layer is connected to the output of fuzzy processing through the weights. This layer is equivalent to the hidden layer of a three-layered BP network.
(4) The fourth layer is the defuzzification layer. It performs defuzzification processing on the output of the BP network.
(5) The fifth layer is the output layer.

7. Optimization of the Network Structure Using the Genetic Algorithm Approach

Professor John Holland in 1975 proposed an attractive class of computational models, called Genetic Algorithms (GA), that mimic the biological evolution process for solving problems in a wide domain. A Genetic Algorithms operates through a simple cycle of stages:
i) Creation of a “;population” of strings,
ii) Evaluation of each string,
iii) Selection of best strings and
iv) Genetic manipulation to create new population of strings.

The cycle of a Genetic Algorithms is presented below



Each cycle in Genetic Algorithms produces a new generation of possible solutions for a given problem. In the first phase, an initial population, describing representatives of the potential solution, is created to initiate the search process. The elements of the population are encoded into bit-strings, called chromosomes.
The crossover points of any two chromosomes are selected randomly. The second step in the genetic manipulation process is termed mutation, where the bits at one or more randomly selected positions of the chromosomes are altered. The mutation process helps to overcome trapping at local maxima. The offsprings produced by the genetic manipulation process are the next population to be evaluated.

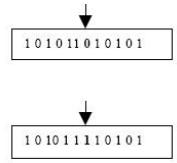


Fig.: Mutation of a chromosome at the 5th bit position.

The Genetic Algorithms cycle is illustrated in this example for maximizing a function f(x) = x2 in the interval 0 = x = 31. In this example the fitness function is f (x) itself. The larger is the functional value, the better is the fitness of the string. In this example, we start with 4 initial strings. The fitness value of the strings and the percentage fitness of the total are estimated in Table A. Since fitness of the second string is large, we select 2 copies of the second string and one each for the first and fourth string in the mating pool. The selection of the partners in the mating pool is also done randomly. Here in table B, we selected

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

partner of string 1 to be the 2-nd string and partner of 4-th string to be the 2nd string. The crossover points for the first-second and second-fourth strings have been selected after o-th and 2-nd bit positions respectively in table B. The second generation of the population without mutation in the first generation is presented in table C.

Table A:

Initial population and their fitness values

string no.	initial population	x	f(x)	strength fitness (% of total)
1	01101	13	169	14.4
2	11000	24	576	49.2
3	01000	08	64	5.5
4	10011	19	361	30.9
Sum-fitness			1170	100.00

Table B:

Mating pool strings and crossover

String No.	Mating Pool	Mates string	Swapping	New population
1	01101	2	0110[1]	01100
2	11000	1	1100[0]	11001
2	11000	4	11[000]	11011
4	10011	2	10[011]	10000

Table C:

Fitness value in second generation

Initial population	x	f(x) (fitness)	strength (% of total)
01100	12	144	8.2
11001	25	625	35.6
11011	27	729	41.5
10000	16	256	14.7
Sum-fitness =		1754	100.00

A Schema (or schemata in plural form) / hyperplane or similarity template is a genetic pattern with fixed values of 1 or 0 at some designated bit positions.

7.1 . The Markov Model for Convergence Analysis

For the sake of understanding, let us now consider the population size = 3 and the chromosomes are 2-bit patterns, as presumed earlier. The set S now takes the following form.

S = {(00, 00, 00), (00, 00, 01), (00, 00, 10), (00, 00, 11), (00, 01, 00), (00, 01, 01), (00, 01, 10), (00, 01, 11), ... (11, 11, 00), (11, 11, 01), (11, 11, 10), (11, 11, 11) }

It may be noted that the number of elements of the last set S is 64. In general, if the chromosomes have the word length of m bits and the number of chromosomes selected in each Genetic Algorithm cycle is n, then the cardinality of the set S is 2 ^ mn. The Markov transition probability matrix P for 2-bit strings of population size 2, thus, will have a dimension of (16 x 16), where the element pij of the matrix denotes the probability of transition from i-th to j-th state. A clear idea about the states and their transitions can be formed from fig.

It needs mention that since from a given i-th state, there could be a transition to any 16 j-th states, therefore the row sum of P matrix must be 1. Formally, $\sum_{\forall j} P_{ij} = 1$,

for a given i.

Now, let us assume a row vector πt, whose k-th element denotes the probability of occurrence of the k-th state at a given genetic iteration (cycle) t; then πt + 1 can be evaluated by

$$\pi_{t+1} = \pi_t \,.\, P$$

Thus starting with a given initial row vector π0 , one can evaluate the state probability vector after n-th iteration πn by

$$\pi_n = \pi_0 \,.\, P^n$$

The first part of the present study focused on improving the optimization of the momentum terms and structure of the BP network, to eliminate the disadvantages of BP network algorithms such as their liability to fall into a local minimum, difficulties in determining the number of hidden layer nodes, slow convergence rate in algorithm learning, poor generalization of the network, and excessive sensitivity to initial values.

The optimization was focused on the number of hidden layers and the number of nodes in each layer of the network.

(1) Number of hidden layers: an increase in the number of hidden layers can form more complex decision-making domains, which can enhance the ability of the network to solve nonlinear problems. An appropriate number of hidden layers can also minimize the system error of the network. Based on the results of a large number of experiments, a three-layered network structure can solve most complex problems.

(2) Number of nodes in the hidden layers: the choice of neuron numbers in the hidden layers is a very complex issue, which is dependent on the experience of the designer and the results from multiple experiments. No optimal analytical expression exists. The number of hidden layer units is directly related to the requirements of the problem and the number of input/output units. If the

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

number is too small, the information obtained through the network will be insufficient to solve the problem. If the number is too large, it will lead to increased training time, longer learning time, a nonoptimal error rate, poor fault tolerance, failure to recognize samples that were not involved in the previous training set, and the possibility of the so-called "transitional agreement" issue. Therefore, the selection of an optimal number of hidden layer units is crucial.

The optimal boundary numbers of the hidden layer units, min and max, were first determined through the incorporated use of formulas. The network training then started from the minimum unit number min, followed by a gradual increase in unit number until the maximum unit number max was validated. For each number of hidden layer units, the network convergence speeds were compared after network convergence was achieved using the same training samples. Finally, the optimal number of hidden layer units was determined based on the training and testing errors of the training results. This approach can effectively reduce the verification time and provide the fastest way to identify the optimal number of hidden layer units.

Conclusion

Today, almost no model can estimate the cost of software with a high degree of accuracy. For the time being we suggest this approach for better understanding. Software size estimation is the key of entire software program project, and the accurate estimation immediately affects the success of project. An improved function point analysis (FPA) method was proposed for analyzing the software size. The method combined the advantage of the fuzzy rules and back propagation (BP) network. Firstly, fuzzy inference system based on the complexity weight matrix of function component was established. Then the adjusted complexity weight was used for modifying the software function point. The adjusted data as samples were transferred to BP network. By the advantage of BP network function approaching, the relationship between software components and software size was established. Finally, BP network was used for estimating software size. The experiment results show that the method could eliminate discontinuity among the different complexity grades, and could make the best of history data, which enhances the accuracy of function point estimation. Further research work with sufficient amount of data should be conducted to increase the accuracy of this method.

References

[1] Roger S Pressman, Seventh Edition, Software Engineering, A Practitioner.s Approach; McGraw Hill International Edition.

[2] Fu Limin (1994), Neural Networks in Computer Intelligence, McGraw Hill Inc.

[3] SRajasekharan, G A Vijayalakshmi Pai (2009), Neural Networks, Fuzzy Logic and Genetic Algorithms – Synthesis and Applications, PHI Learning Private Limited.

[4] B Yegnanarayana (2010), Artificial Neural Netwoks, PHI Learning Private Limited.

[5] C H Chen, Fuzzy Logic and Neural Netwokk Handbook, New York: McGraw-Hill Inc.,1996.

[6] F L Chung and T Lee, "Fuzzy competitive learning", Neural Networks, Vol 7, No 3, pp.539-552, 1994.

[7] M.T. Su, T.C.Ling, K.K.Phang, C.S.Liew, P.Y.Man, "Enhanced Software Development Effort and Cost Estimation Using Fuzzy Logic Model", Malaysian Journal of Computer Science, Vol. 20, No. 2, 2007, pp. 199-207.

[8] V Cherkassky and N Vassilas, "Performance of Backporpagation networks for associative database retrieval", in Proceedings of International Joint Conference on Neural Networks, Washington D C, Vol 1, pp. 77-84, 1989.

[9] Edwards, J.S. Moores, T.T. (1994), "A conflict between the use of estimating and planning tools in the management of information systems.". European Journal of Information Systems 3(2): 139-147.

[10] Briand, L. C. and I. Wieczorek (2002). Resource estimation in software engineering. Encyclopedia of software engineering. J. J. Marcinak. New York, John Wiley & Sons: 1160-1196.

[11] P D Wasserman, Neural Computing, Theory and Practice, New York; Van Nostrand Reinhold, 1989.

[12] S.N. Sivanandam, S. Sumathi, S.N. Deepa, "Introduction to fuzzy logic using MATLAB", Springer, 2007.

[13] P D Wasserman, Advanced Methods in Neural Computing, New York; Van Nostrand Reinhold, 1993.

[14] R A Jacobs, "Increaed rates of convergence through learning rate adaptation", Neural Networks, vol 1, No 4, pp. 295-307, 1988.

[15] Adeli H. and Wu M. (1998) "Regularization neural network for construction cost estimation", Journal of Construction Engineering and Management, ASCE 124(1), pp 18-24.

[16] J Von Neumann, The Computer and the Brain, New Haven, CT:Yale University Press, 1958.

[17] Hill Peter (ISBSG) - Estimation Workbook 2 - published by International Software Benchmarking Standards Group ISBSG - Estimation and Benchmarking Resource Centre

[18] A. J. Albrecht, "Measuring Application Development Productivity," Proceedings of the Joint SHARE, GUIDE, and IBM Application Development Symposium, Monterey, California, October 14–17, IBM Corporation (1979), pp. 83–92.

[19] Flyvbjerg B. H., Mette S., and Buhl S. (2002), "Underestimating costs in public works projects, error or lie", Journal of the American Planning Association, Vol. 68, No.3, pp 279-292.

[20] Jorgensen, M.. "A Review of Studies on Expert Estimation of Software DevelopmentEffort". .

[21] Morris Pam - Overview of Function Point Analysis Total Metrics - Function Point Resource Centre.

[22] A.A. Moataz, O.S.Moshood, A.Jarallah, "Adaptive fuzzy-logic-based framework for software development effort prediction", Information and Software Technology, Vol. 47, Issue 1, 2005, pp. 31-48.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

[23] Jørgensen, M. Shepperd, M.. "A Systematic Review of Software Development Cost Estimation Studies".
[24] Goodwin, P. (1998). Enhancing judgmental sales forecasting: The role of laboratory research. Forecasting with judgment. G. Wright and P. Goodwin. New York, John Wiley & Sons: 91-112.
[25] N K Bose and P Liang, Neural Netwrok Fundamentals with Graphs, Algorithms and Applications, McGraw-Hill, Inc., Editions, 1996.
[26] S. Mitra, Y.Hayashi, "Neuro-Fuzzy Rule Generation: Survey in Soft Computing Framework", IEEE Transactions on Neural Networks, Vol.11, No.3, 2000, pp. 748-768.
[27] D. Nauck, F. Klawonn, R. Kruse, "Foundations of Neuro-Fuzzy Systems", Wiley, Chichester, 1997.

Biography of Author(s)

B V A N S S Prabhakar Rao is obtained his Bachelor's Degree in Electronics, Master of Computer Applications & M.Tech (CST) from Andhra University and also pursuing his Ph.D through JNTUK, Kakinada. Previously he worked with Govt. Degree College, Tekkali & Dr. V S Krishna Govt. Degree College, Visakhapatnam as Lecturer in Computer Applications and Science. And with GITAM University as Assistnat Professor. Presently he working as Associate Professor with Department of Computer Science Engineering, Miracle Educational Society Group of Institutions, Vizianagaram. He is Life Member in ISTE, ISCA, IACSIT, CSTA and IAENG.

Dr. P. Seetha Ramaiah is presently working as Professor in the Department of Computer Science and Systems Engineering, College of Engineering (A), Andhra University. He received his Ph.D in Computer Science from Andhra University in 1990. He is the Principal Investigator for several Defense R&D projects and Department of Science and Technology (DST) projects of the Government of India in the areas of Embedded Systems and Robotics. He has published several Journal papers, and presented around Fifteen International Conference papers in addition to twenty one papers at National Conferences in India. His areas of research include Safety-Critical Computing- Software Safety, Computer Networks, VLSI and Embedded Systems, Real-Time Systems, Microprocessor-based System Design, Robot Hand-Eye Coordination, Signal Processing Algorithms on fixed-point DSP processors, and Bio-Electronics Systems.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Improving Web Performance by a Differencing and Merging System

Aleksandar Jevremovic[1], Ranko Popovic[1], Dejan Zivkovic[1], Mladen Veinovic[1] and Goran Shimic[2]

[1] Department of Informatics and Computing, Singidunum University, Belgrade, 11000, Serbia

[2] Military Academy, Belgrade, 11000, Serbia

Abstract

In this paper we consider the problem of improving Web performance and propose an efficient differencing and merging system (DMS) based on an HTTP protocol extension. To provide for faster information exchange over the Web, the system tries to transfer only computed differences between requested documents and previously retrieved documents from the same site. Analysis and experimental results prove the effectiveness of DMS, but also show bigger processor and memory load on servers and clients. DMS is compatible with most of the existing solutions for improving Web performance. Moreover, SSL security system may be used to provide Web privacy and authenticity. The DMS model is simple to use and can be relatively easily integrated in Web servers and browsers.

Keywords: **Web performance, HTTP protocol, HTML differencing.**

1. Introduction

Web performance is the key ingredient to an effective use of the Internet. In case of static Web pages and information portals, better Web performance exerts positive influence on the user's impression about the content. Also, dynamic Web pages and Web applications directly depend on Web performance, because speed of information exchange ultimately determines the kind of functionality that is implementable in a Web application. More generally, Web performance has a great impact on smooth migration from desktop-based to Web-based applications, as well as on success of the cloud-computing model.

Improving Web performance has been the goal of many different attempts. Popular approaches include both direct and indirect methods. The direct methods are based on contents cashing, contents compression, and AJAX calls, while the indirect methods are based on new network protocols such as SCTP, MUX/SMUX, and SPDY. The next section presents some advantages and drawbacks of these popular approaches to improve Web performance.

Main contribution of this paper is a model for a novel approach regarding the Web performance improvement. The proposed model is based on a differencing and merging system, hence it is called DMS. In a nutshell, DMS makes use of structural similarities between HTML pages. Consequently, for a client to get a page from a server, it makes sense that the client requests the smallest differences between the desired page and previously retrieved pages from the same server. If the server's difference computation procedure is not time intensive, the desired page in this way may be delivered . much faster to the client because the server transfers smaller amounts of data. The client on its part, however, needs to merge received differences to reproduce the correct version of the page.

2. Related Work

In the last few decades there has been a growing interest in improving Web performance. Many existing solutions to the problem basically try to reduce the amount of data transfered between servers and clients. Basic technique to reduce the network traffic is contents cashing, i.e., reuse of previously retrieved Web pages that are kept locally for a fixed time interval. The main drawback of contents cashing is possibility of using stale contents or retrieving unchanged pages. Among many attempts to improve efficiency of contents cashing [22], [18], [20], [21], [11], the most prominent approach is a farm of proxy servers that create a network of cached pages. This relieves the burden on central servers and speeds up contents distribution to clients. However, maintaining consistency of such a network requires specific architecture based on agents [11].

Another standard technique to reduce the network traffic is data compression. Support for data compression is built into the HTTP protocol from version 1.1 and it is supported by all modern Web browsers. Compression of textual Web contents may save up to 75% of the contents

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

size, while the overall savings figure for general contents amounts to about 37% [12]. Reducing the network trafic by means of data compression is compatible with the DMS model.

A more complex method for improving Web performance is contents fragmentation and asynchronous transfer of fragments based on JavaScript calls (AJAX). Advantages of this method lie mainly in its compatibility with the service oriented architecture, while drawbacks include necessity to redesign existing applications.

Two leading attempts to improve Web performance by protocol replacement are SMUX protocol [5] and HTTP over SCTP protocol [16]. SMUX is an attempt of the W3C consortium to design a lightweight application protocol and to make migration to future Web protocols simpler. However, SMUX protocol is still in an experimental phase. HTTP over SCTP protocol works by replacing traditional TCP transport protocol with more efficient SCTP protocol [4]. Using HTTP over SCTP is particularly relevant to the DMS model, since HTTP over SCTP gives good results in developing regions [15], i.e., regions with poor communication infrastructure. In such an environment DMS could be very useful and greatly speed up the information exchange.

In an attempt to improve Web performance, Google is making an important step forward with its SPDY protocol [6]. Goal of this protocol is to enable multiple data streams within one HTTP connection. In addition, this protocol compresses data during data transfer and excludes the “host” field from the request header. Another interesting method for decreasing contents lag, which is used in Google’s Chrome browser, is so called DNS prefetching. This function actively performs domain name resolutions for the domain name references found in hyper links of each document.

Opera’s attempt to improve Web performance is based on its technology called Opera Turbo [2]. This is a commercial technology that uses specialized intermediary servers and data compression to reduce the network traffic.

The most common system for tracking and viewing changes on the Web is probably WebVigiL [9]. This system automates the change detection of HTML/XML documents and timely notifies users about changes based on user-specified changes of interest. In WebVigiL, two separate change detection algorithms are used to detect changes occurring to HTML and XML documents.

The problem of comparing two HTML documents to find the differences between them can be simply solved by comparing textual content of the documents, e.g. using traditional UNIX diff algorithm. However, comparing HTML documents at the lexical level, as if they were ordinary text documents, is neither informative nor intuitive. VDiff algorithm [13] gives better accuracy by taking documents’ internal tree structure into account. Still better results can be obtained in case of well-structured XML documents by using their tree representation and identifying common nodes and subtrees [7]. The hierarchical structure of XML documents has been exploited recently to detect their differences by using semantic analysis of such documents [8], [19], [14].

3. DMS Model

This section describes the proposed system and its role in the client-server communication. The central idea around DMS is that transferring differences between documents instead of full documents would reduce the network traffic. Upon receiving request for a document from a client, a server computes the differences between the document it needs to send and documents that the client already retrieved from the server. The server then selects the smallest size difference and sends it to the client. The client finally merges received difference with the corresponding local document to reconstruct the correct version of the document.

A DMS enabled server must keep track of served documents (and differences) and assign a unique ID to each document it serves. Similarly, a client must keep track of retrieved documents (and differences). Figure 1 illustrates DMS role in the client-server networking model in the simplest case of static Web setting.

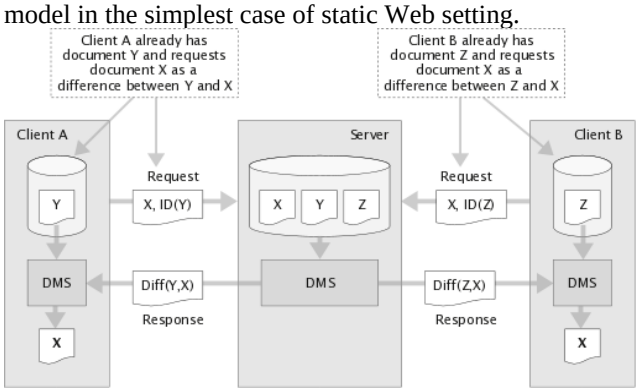


Fig. 1. DMS model in the static Web setting

Figure 1 depicts the state when clients A and B already requested some documents from the same server in the past and retrieved and stored them locally: client A retrieved document Y and client B retrieved document Z. Now both clients in Figure 1 make request for document X to the server. Along with the request for X, client A also sends to server the document ID of Y it has locally as a reference document for computing differences. Server as its response computes differences between the documents X and Y and sends them to client A. Client A

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

then merges received differences with its local copy of Y to reconstruct the document X. Similarly, to request the same document X, client B sends along the document ID of Z it has locally, and the server response are differences between the documents X and Z.

More interesting environment for DMS is the dynamic Web setting and Web applications. In this case, URL address is not sufficient resource identifier, since various requests with the same URL address may produce different results. Moreover, Web applications often dynamically customize user interface as a function of the user account type and user activities. Thus, different users can get different results for the same request.

This specific features of dynamically created pages and Web applications imply that a DMS enabled server should limit computation of differences between pages transferred only within user sessions. Such a server also needs to assign a special unique ID to each page obtained as a result of the dynamic document creation procedure. This means that a DMS enabled server should reserve separate space for each client to maintain already transfered pages per session.

Figure 2 illustrates DMS role in the dynamic Web setting. For each client, a DMS enabled server keeps history of pages delivered within current session. A client in subsequent requests sends IDs of some of the pages as a reference for difference computation. In Figure 2, client A retrieved three pages P1, P2, and P3 within current session, while client B retrieved two pages. Subsequently, client A requests fourth page P4 and sends along IDs of P1, P2, and P3. The server then computes differences between P4 and each of P1, P2, and P3, and returns the smallest difference as its response. The pages Q1 and Q2 delivered to client B within current session are kept separately, and client B's request for page Q3 is serviced with the smallest difference between Q3 and Q1.

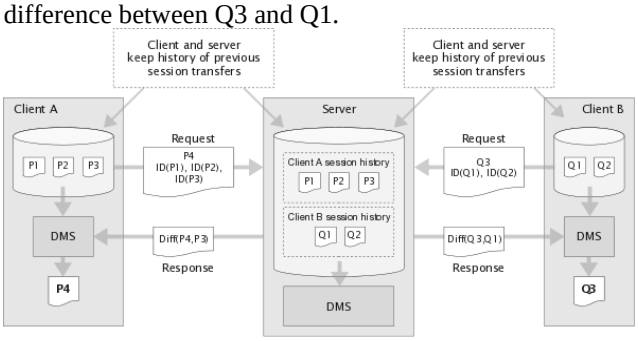


Fig. 2. DMS model in the dynamic

Web setting Note that DMS functionality is not limited to HTML pages. DMS may be also used with other textual resources (CSS, JavaScript...), as well as with binary resources (pictures, sounds...) using appropriate difference computation algorithms.

2.1 DMS Implementation

To implement DMS, it is necessary to modify the HTTP protocol itself [3] and both sides of the client-server networking model. Changes to the HTTP protocol include support for the document IDs inside HTTP messages. A new field "Accept-Diff", for example, may be added to the header of a HTTP message. This field in a HTTP request informs server that client supports DMS and contains values of document IDs as parameters for computing differences. This field in a HTTP response contains ID of the document that server has used to compute (the smallest) returned differences.

In case that server does not contain any of the documents specified in an HTTP request, it sends the requested document in its fullness. This also applies if server does not support DMS at all.

The server side implementation of DMS may take two basic forms: direct integration in the server software, or indirect support through a special service. The first approach is compatible with data compression and SSL security, but requires redesign of the Web server software. On the other hand, an independent service (which listens to a port, forwards received requests to the Web server, and sends computed differences to a client) does not necessitate changes to the original software, but it is not compatible with data compression and SSL security.

Web browser on the client side must support sending HTTP requests with the "Accept-Diff" field whose values are documents IDs that the client has stored locally. Thus, the client needs to select one or many (perhaps all) of the local documents and include their IDs in a HTTP request. Also, the client from a HTTP response needs to extract received differences and merge them with the document whose ID is specified in the "Accept-Diff" field.

A server or a client may choose to store full pages or only page differences for later use. The choice is independent and depends on the particular server and client processing power, storage capacity, and other factors that determine response time and storage requirements. For example, storing only page differences favors storage to response time, because it might be necessary to apply the merging procedure repeatedly to get the correct page.

4. DMS on the Web

This section investigates two applications of DMS on the Web, besides the direct client-server communication. First, if the client-server communication goes indirectly through a proxy server, then the proxy server must support both the client and server side of

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

DMS functionality. Figure 3 shows the general information flow between a client and a server in the presence of one or more proxy servers.

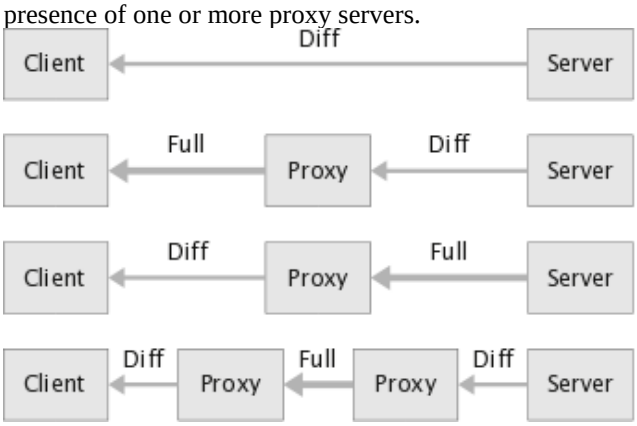


Fig. 3. DMS in the proxy server setting

Using a DMS enabled proxy server may prove especially efficient when retrieving documents from a Web server on behalf of clients. A proxy server in this case can use many already retrieved documents to request the smallest difference from the server. Moreover, to fulfill a request for one client, the proxy server may use retrieved documents received for other clients.

Second important application where DMS may find its use is the distributed contents delivery to users in geographical proximity through close distribution servers. Figure 4 illustrates DMS usage in the distributed contents delivery setting.

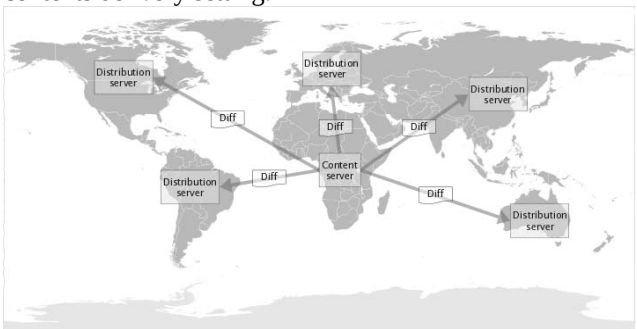


Fig. 4. DMS in the distributed contents delivery setting

Aim of using DMS in the distributed contents delivery setting is not taking the load off of the central content server, since distribution servers still forward client requests to the central server. Instead, DMS usage can result in lower datatraffic volume because of the transfer of smaller document differences from the central server to distribution servers. Similarly to a proxy server, a distribution server may request differences based on already received documents from all sessions.

Main benefit of this solution is central processing of HTTP requests with efficient distributed contents delivery. Moreover, this efficiency is achieved simply without complex schemes involving load balancing among distribution servers and their contents synchronization.

5. DMS Performance

This section analyzes DMS efficiency by evaluating its impact on the data-traffic volume. This is done by means of two experiments—one for a Web site and the other for a Web application.

It is clear that DMS usage is justified if the total time for computing, transferring, and merging document differences is less than the time for transferring a full document. DMS efficiency thus depends mostly on the similarity of requested and local documents, as well as on the efficiency of an algorithm for computing and merging differences. However, an algorithm for merging differences is determined by the corresponding algorithm for computing differences, so we only consider the latter algorithm efficiency.

If an HTTP request contains only single local document ID, then DMS efficiency on the server side is dominated by the efficiency of an algorithm for computing differences. In general, this algorithm efficiency can be measured using two opposing criteria: the time for computing differences and the size of computed differences. However, an algorithm for computing differences may choose a trade-off between the two criteria, since larger difference computation time can be made up for smaller transfer time, and vise versa. Moreover, an adaptive algorithm may be further tuned up to base its optimal decision on the processor load and the communication bandwidth.

On the other hand, if an HTTP request contains many local document IDs, then DMS performance is a function of the total time needed for computing differences with respect to all given local documents. In this case, better performance is still possible using an advanced method to select the best document that gives the smallest difference.

5.1 Experimental Results

Since the primary goal of this paper is to show soundness of the DMS concept, we have conducted two basic experiments to confirm DMS effectiveness. To get a first glimpse at the DMS performance, in the experiments we have not tested any of the real DMS functionality needed in a client and a server. Instead, we have restricted ourselves to analyze only the DMS impact on the amount of transfered network data.

To this end, we have used two of the most common contents publishing models on the Web—a Web site and a Web application. For the Web site example we have

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

chosen the Times Online site www.timesonline.co.uk. Since we could not use DMS with commercial servers and clients, testing is performed by first retrieving some pages using ordinary browser, and then computing the page differences using the GNU diff implementation of the basic difference computation algorithm [1]. On the other hand, for the Web application example we have developed custom chat application in PHP language.

5.2 The Web Site Example

To estimate DMS impact on the speed of data transfer from a Web site, we have taken ten randomly selected documents from the science section of the Times Online site at the address http://www.timesonline.co.uk/ tol/news/science/:

1. physics/article7121202.ece
2. physics/article7131244.ece
3. medicine/article7121219.ece
4. medicine/article7125803.ece
5. eureka/article7110849.ece
6. genetics/article7120516.ece
7. biology_evolution/article7132299.ece
8. biology_evolution/article7127753.ece
9. genetics/article7128357.ece
10. eureka/article7115855.ece

These documents were taken from the site in this order on May 23, 2010 in time frame of two minutes. As each document's HTML page was stored in a local file, differences between the page and all previous pages were computed. Table I summarizes the page sizes and computed differences in bytes, where the smallest difference is highlighted in bold.

Table 1: Sizes of ten pages and mutual differences

-	Size	1	2	3	4	5	6	7	8	9	10
1	180238	-	-	-	-	-	-	-	-	-	-
2	179681	**83186**	-	-	-	-	-	-	-	-	-
3	156594	**83140**	84726	-	-	-	-	-	-	-	-
4	162835	82595	83613	**68327**	-	-	-	-	-	-	-
5	148804	91125	91527	**67453**	74039	-	-	-	-	-	-
6	187365	92658	96554	**89766**	94969	97095	-	-	-	-	-
7	185722	**89562**	91740	92577	90015	96531	99034	-	-	-	-
8	179537	86922	87551	88563	**86143**	89534	96001	86520	-	-	-
9	150343	87284	87229	61810	70055	**56766**	88474	92523	85916	-	-
10	146289	83352	88486	63694	72074	**54365**	94110	90900	87517	56801	-

Figure 5 gives the smallest difference obtained for each page in terms of the percentage of the original size.

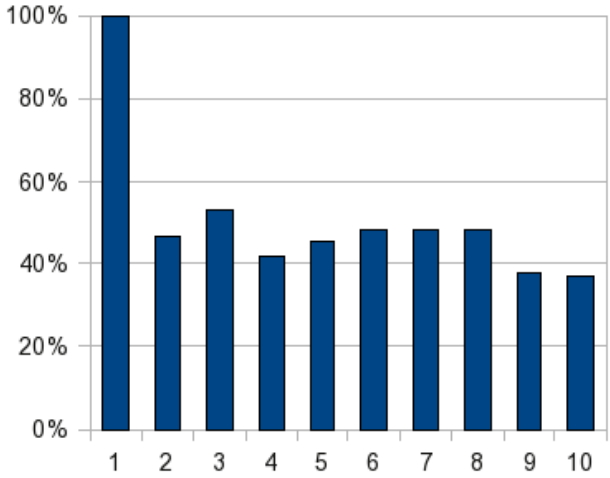


Fig. 5. The smallest differences with respect to the page sizes

The test results show that using DMS even for relatively unrelated documents would reduce the network data traffic more than 50% in most cases (except the first). Thus, taking into account that

- the Times Online site has over three million visitors each day (source: www.statbrain.com)
- the visitors of this site open 2.5 pages on average per visit (source: www.alexa.com)

and extrapolating the results obtained in the testing scenario that

- the average page size is about 160 KB
- the expected savings in the data transfer is about 50%

we may conclude that DMS would overall save about 360 GB in the data-traffic volume each day in case when the first page is fully transfered, or 600 GB in case when pages from previous visits are used as references to difference computations.

5.3 The Web Application Example

To estimate DMS efficiency in the context of a Web application, we have chosen an application for short message exchange between users. Motivated by the Lift Web framework [17], this minimalistic chat application has been easily developed in PHP.

The main window of the chat application contains at most 10 messages. When a server adds a new message to the full window, it removes the oldest message from the window. A client sends messages using an HTML form,

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

and the server embeds them into a page using PHP. For the sake of simplicity, a client retrieves a complete page each time it checks whether a new message appeared in the meantime.

Testing scenario for the application has been devised to comprise three steps—in each step a new message is added to the application and then the relevant HTML pages are compared. Initially, there have been nine messages exchanged as shown in Figure 6, and size of the the corresponding HTML page that contains the nine messages is 1063 bytes.

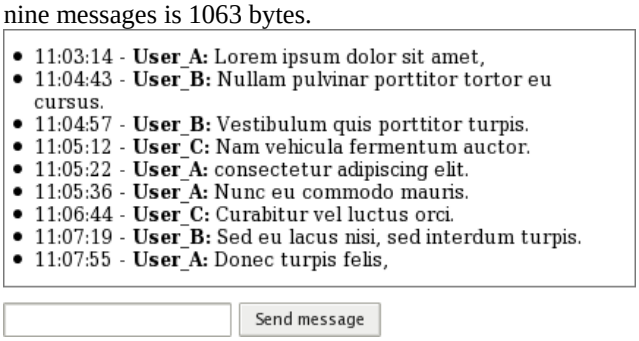


Fig. 6. Initial nine messages in the chat application

In the second step, a new message is added to the application for the total of ten messages (Figure 7). The size of the corresponding HTML page has increased to 1148 bytes. However, by using DMS only differences between the first and the second page would be transferred. The size of the differences in the diff format is mere 93 bytes.

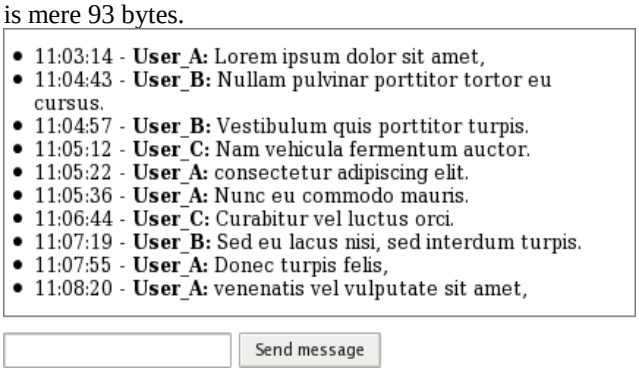


Fig. 7. The chat application after the second step

In the third step, once a new message is added, the oldest message is removed (Figure 8). The size of the corresponding HTML page has been decreased to 1136 bytes. The sizes of the differences between the third and the second page and the third and the first page are 161 bytes and 251 bytes, respectively.

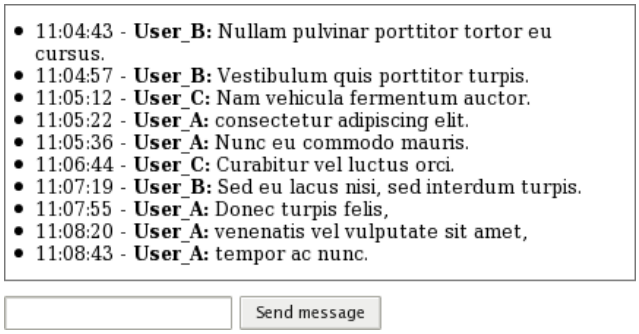


Fig. 8. The chat application after the third step

The test results show that using DMS in the chat application would reduce the network traffic to almost 90%. Thus, it is fair to expect that DMS can be successfuly applied in other Web applications that do not use service calls and do not get page changes as service messages (SOAP, JSON, . . .).

6. Web Security and Privacy

The DMS model is compatible with standard systems that provide HTTP security and privacy. For example, the SSL security system can be used with DMS without modifications, because SSL is part of the message transport layer, while DMS acts on the application level above HTTP.

There is one situation, however, where basic DMS can be abused by a malicious user to get hold of a private document. Consider the scenario when client requests a private document protected by the SSL system and assume that server sends full version of the document as response. Client then leaves protected communication and makes a new non-SSL request for some other public document. Assume further that, for some reason, client in the request specifies ID of the protected document received during secure communication. As a response then client receives clear differences between the private document and the pubic document.

An eavesdropper can thus learn two items: URL address of the public document and clear differences between the private document and the public document. In this case, an attacker can independently request full version of the public document and use the differences in a reverse-merging procedure to obtain the protected document.

This scenario reveals a potential problem of DMS abuse that should be kept in mind when implementing DMS for a secure environment. More specifically, a DMS enabled client implementation should

1. distinguish IDs of unsecured and secured documents, i.e., those received using HTTP and HTTPS protocols, and should not mix them;

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

and
2. limit document IDs in a HTTPS request to those received only in the current session.

7. Conclusion and Future Work

This paper proposes DMS as a solution for an effective use of the Web. Since Web pages are well-structured and contain many common HTML language constructs, an intuition is that differences between Web pages are small despite their content. Thus, to reduce the network traffic, it pays off to transfer differences between pages instead of full pages. This is the main idea behind DMS whose soundness is justified by experimental results.

DMS is compatible with most of the existing solutions for improving Web performance. For example, clients may use ordinary document caching methods before turning to DMS, or servers may compress differences before sending them. Moreover, SSL security system may be used to provide Web privacy and authenticity.

The DMS model is simple to use and can be relatively easily integrated in Web servers and browsers. It is also general in the sense that its main idea with small modifications can be used in other problem domains such as routing, load balancing, and computer forensics [10].

Obvious subject of future work is a software implementation of DMS and evaluation of its efficiency in production environments. The software design will certainly include heuristics and algorithms for selection of documents being compared and for computation of document differences. An efficiency analysis of different design decisions that are made in this regard is also necessary for broad DMS adoption on the Web.

Acknowledgments

This work has been supported by the Serbian Ministry of Education and Science (projects III44006 and TR32054)

References

[1] GNU Diffutils. Available from: http://www.gnu.org/software/diffutils
[2] Opera Turbo. Available from: http://www.opera.com/business/solutions/ turbo
[3] RFC2616: Hypertext transfer protocol – HTTP/1.1. Available from: http://www.w3.org/Protocols/rfc2616/rfc2616.html
[4] RFC4960: Stream control transmission protocol. Available from: http: //tools.ietf.org/html/rfc4960
[5] SMUX protocol specification. Available from: http://www.w3.org/TR/ WD-mux
[6] SPDY protocol. Available from: http://www.chromium.org/spdy/ spdy-protocol
[7] Al-Ekram, R., Adma, A., Baysal, O.: diffx: An algorithm to detect changes in multi-version XML documents. In: Conference of the Centre for Advanced Studies on Collaborative Research (2005)
[8] Al-Namiy, A.Q., Majeedl, F.S.: Towards automatic extracted semantic annotation (ESA) for Web documents. In: APCIP 2009, vol. 2 (2009)
[9] Chakravarthy, S., Hara, S.: Automating change detection and notification of Web pages. In: 17th International Workshop on Database and Expert Systems Applications (DEXA 2006) (2006)
[10] Chen, L., Wang, G.: An efficient piecewise hashing method for computer forensics. In: Proceedings of the First International Workshop on Knowledge Discovery and Data Mining (2008)
[11] Kannammal, A., Padmanabhan, R., Iyengar, N.: Web cache consistency maintenance through agents. In: Proceedings of the Second International Conference on Communication Software and Networks, pp. 329–333 (2010)
[12] King, A.B.: Speed Up Your Site: Web Site Optimization. New Riders (2003). Available from: http://www.websiteoptimization.com/speed/18/ 18-2t.html
[13] Mikhaiel, R., Stroulia, E.: Accurate and efficient HTML differencing. In: 13th IEEE International Workshop on Software Technology and Engineering Practice (2005)
[14] Moon, H.J., Yoo, J.W.: Free-traversing syntactic and semantic comparison on semi-structured languages. In: Convergence and Hybrid Information Technology (ICHIT 2008) (2008)
[15] Natarajan, P., Amer, P.D., Stewart, R.: Multistreamed Web transport for developing regions. In: The second ACM SIGCOMM workshop on Networked systems for developing regions (2008)
[16] Natarajan, P., Iyengar, J.R., Amer, P.D., Stewart, R.: SCTP: An innovative transport layer protocol for the Web. In: Proceedings of the International World Wide Web Conference (2006)
[17] Pollak, D., Vinoski, S.: A chat application in Lift. IEEE Internet Computing 14(3), 88–91 (2010)
[18] Ramaswamy, L., Liu, L., Iyengar, A.: Cache clouds: Cooperative caching of dynamic documents in edge networks. In: Proceedings of the 25th IEEE International Conference on Distributed Computing Systems (ICDCS 2005) (2005)
[19] Santos, R.C., Hara, C.S.: A semantical change detection algorithm for XML. In: SEKE 2007 (2007)
[20] Sharman, R., Ramanna, S.S., Ramesh, R., Gopal, R.: Cache architecture for on-demand streaming on the Web. ACM Transactions on the Web (TWEB) 1 (2007)
[21] Sheu, T.L., Yang, C.H.: A novel hierarchical cache architecture for WWW servers. In: Proceedings of the 15th International Conference on Information Networking (ICOIN 2001), p. 863 (2001)
[22] Yin, J., Alvisi, L., Dahlin, M., Iyengar, A.: Engineering Web cache consistency. ACM Transactions on Internet Technology 2(3) (2002)

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Verdict Accuracy of Quick Reduct Algorithm using Clustering and Classification Techniques for Gene Expression Data

T.Chandrasekhar[1], K.Thangavel[2] and E.N.Sathishkumar[3]

[1] Department of Computer Science, Periyar University, Salem, Tamilnadu-636 011, India

[2] Department of Computer Science, Periyar University, Salem, Tamilnadu-636 011, India

[3] Department of Computer Science, Periyar University, Salem, Tamilnadu-636 011, India

Abstract

In most gene expression data, the number of training samples is very small compared to the large number of genes involved in the experiments. However, among the large amount of genes, only a small fraction is effective for performing a certain task. Furthermore, a small subset of genes is desirable in developing gene expression based diagnostic tools for delivering reliable and understandable results. With the gene selection results, the cost of biological experiment and decision can be greatly reduced by analyzing only the marker genes. An important application of gene expression data in functional genomics is to classify samples according to their gene expression profiles. Feature selection (FS) is a process which attempts to select more informative features. It is one of the important steps in knowledge discovery. Conventional supervised FS methods evaluate various feature subsets using an evaluation function or metric to select only those features which are related to the decision classes of the data under consideration. This paper studies a feature selection method based on rough set theory. Further K-Means, Fuzzy C-Means (FCM) algorithm have implemented for the reduced feature set without considering class labels. Then the obtained results are compared with the original class labels. Back Propagation Network (BPN) has also been used for classification. Then the performance of K-Means, FCM, and BPN are analyzed through the confusion matrix. It is found that the BPN is performing well comparatively.

Keywords: *Rough set theory, Feature Selection, Gene Expression, Quick Reduct, K-means, Fuzzy C means, BPN.*

1. Introduction

Feature selection is the process of choosing the most appropriate features when creating the model of the process. Most of the feature selection methods are applied across the entire data set. Once such genes are chosen, the creation of classifiers on the basis of the genes is another undertaking. If we survey the established investigations in this field, we will find that almost all the accurate classification results are obtained based on more than two genes. Rough sets have been used as a feature selection methods by many researchers among them Jensen and Schen, Zhong et al, Wang and Hu et al. The Rough set approach to feature selection consists in selecting a subset of features which can predict the classes as well as the original set of features. The optimal criterion for Rough set feature selection is to find shortest or minimal reducts while obtaining high quality classifiers based on the selected features. Here we propose a feature selection method based on rough set theory for reducing genes from large gene expression database [1, 4].

Discriminant analysis is now widely used in bioinformatics, such as distinguishing cancer tissues from normal tissues. A problem with gene expression analysis or with any large dimensional data set is often the selection of significant variables (feature selection) within the data set that would enable accurate classification of the data to some output classes. These variables may be potential diagnostic markers too. There are good reasons for reducing the large number of variables:
1) An opportunity to scrutinize individual genes for further medical treatment and drug development. 2) Dimension reduction to reduce the computational cost. 3) Reducing the number of redundant and unnecessary variables can improve inference and classification. 4) More interpretable features or characteristics that can help identify and monitor the target diseases or function types [5].

The rest of the paper is organized as follows: Section 2, briefs about the Rough set theory. Section 3 describes the clustering techniques. Section 4 briefs about classification techniques. Section 5 describes the confusion matrix for

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

two class problem. Section 6 explains briefly about experimental analysis and results. Section 7 presents a conclusion for this paper.

2. Rough Set Theory

Rough set theory (Pawlak, 1991) is a formal mathematical tool that can be applied to reducing the dimensionality of datasets. The rough set attribute reduction method removes redundant input attributes from datasets of discrete values, all the while making sure that no information is lost. The approach is fast and efficient, making use of standard operations from conventional set theory [3].

Definition: Let U be a universe of discourse, $X \subseteq U$, and R is an equivalence relation on U. U/R represents the set of the equivalence class of U induced by R. The *positive region* of X on R in U, is defined as $pos(R,X) = U\ \{Y \in U/R \mid Y \subseteq X\}$.

The partition of U, generated by $IND\ (P)$ is denoted U/P. If $(x, y) \in IND\ (P)$, then x and y are indiscernible by attributes from P. The equivalence classes of the P-indiscernibility relation are denoted $[x]p$. The indiscernibility relation is the mathematical basis of rough set theory. Let $X \subseteq U$, the P-lower approximation $\underline{PX}$ and P-upper approximation PX of set X can be defined as:

$$\underline{P}X = \{ x \in U \mid [x]p \subseteq X \} \quad (1)$$

$$\overline{PX} = \{ x \in U \mid [x]p \cap X \neq \varphi \} \quad (2)$$

Let P, $Q \subseteq A$ be equivalence relations over U, then the positive, negative and boundary regions can be defined as:

$$\mathrm{POS}_P(Q) = \bigcup_{X \in U/Q} \underline{P}X \quad (3)$$

$$\mathrm{NEG}_P(Q) = U - \bigcup_{X \in U/Q} \overline{P}X \quad (4)$$

$$\mathrm{BND}_P(Q) = \bigcup_{X \in U/Q} \overline{P}X - \bigcup_{X \in U/Q} \underline{P}X \quad (5)$$

An important issue in data analysis is discovering depende ncies between attributes dependency can be defined in the following way. For P, $Q \subseteq A$, P depends totally on Q, if and only if $IND\ (P) \subseteq IND\ (Q)$. That means that the partition generated by P is finer than the partition generated by Q. We say that Q depends on P in a degree $0 \leq k \leq 1$ denoted $P \Rightarrow k\ Q$, if

$$k = \gamma_P(Q) = \frac{|\mathrm{POS}_P(Q)|}{|U|} \quad (6)$$

If $k = 1$, Q depends totally on P, if $0 \leq k \leq 1$, Q depends partially on P, and if $k=0$ then Q does not depend on P. In other words, Q depends totally (partially) on P, if all (some) objects of the universe U can be certainly classified to blocks of the partition U/Q, employing P. In a decision system the attribute set contains the condition attribute set C and decision attribute set D, i.e. $A = C \cup D$. The degree of dependency between condition and decision attributes, $\gamma c(D)$, is called the quality of approximation of classification, induced by the set of decision attributes[6,10].

2.1 Quick Reduct Algorithm

The reduction of attributes is achieved by comparing equivalence relations generated by sets of attributes. Attributes are removed so that the reduced set provides the same quality of classification as the original. A reduct is defined as a subset R of the conditional attribute set C such that γR(D)=γC(D). A given dataset may have many attribute reduct sets, so the setR of all reducts is defined as:

$$R_{all} = \{X \mid X \subseteq C, \gamma X(D) = \gamma C(D);\ \gamma X - \{a\}(D) \neq \gamma X(D),\ \forall a \in X\}. \quad (7)$$

The intersection of all the sets in Rall is called the core, the elements of which are those attributes that cannot be eliminated without introducing more contradictions to the representation of the dataset. For many tasks (for example, feature selection), a reduct of minimal cardinality is ideally searched for. That is, an attempt is to be made to locate a single element of the reduct set Rmin⊆Rall:

$$R_{min} = \{X \mid X \in R_{all},\ \forall Y \in R_{all}, |X| \leq |Y|\}. \quad (8)$$

The Quick Reduct algorithm shown below [8, 9], it searches for a minimal subset without exhaustively generating all possible subsets. The search begins with an empty subset; attributes which result in the greatest increase in the rough set dependency value are added iteratively. This process continues until the search produces its maximum possible dependency value for that dataset (γc(D)). Note that this type of search does not guarantee a minimal subset and may only discover a local minimum.

QUICKREDUCT(C, D)
C, the set of all conditional features;
D, the set of decision features.
(a) $R \leftarrow \{\}$
(b) ***Do***
(c) $T \leftarrow R$
(d) $\forall x \in (C-R)$
(e) ***if*** $\gamma_{R \cup \{x\}(D)} > \gamma_{T(D)}$

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Where $\gamma_{R(D)} = card(POSR(D)) / card(U)$

(f) $T \leftarrow R \cup \{x\}$

(g) $R \leftarrow T$

(h) ***until*** $\gamma_{R(D)} = = \gamma_{C(D)}$

(i) ***return*** R

It starts off with an empty set and adds in turn, one at a time, those attributes that result in the greatest increase in the rough set dependency metric, until this produces its maximum possible value for the dataset. Other such techniques may be found in [8, 9]

3. Clustering Techniques

Clustering is the process of grouping data into clusters, where objects within each cluster have high similarity, but are dissimilar to the objects in other clusters. Similarities are assessed based on the attributes values that best describes the objects. Often distance measures are used for the purpose. Clustering has its roots in many areas, including data mining, statistics, biology, and machine learning. In this work K-Means, FCM and BPN algorithms which are used to classify the data.

3.1 K-Means Algorithm

K-Means algorithm (MacQueen, 1967) is one of a group of algorithms called partitioning methods[11]. The k-mean algorithm is very simple and can be easily implemented in solving many practical problems. The k-means algorithm is the best-known squared error-based clustering algorithm.

Input:

Set of sample patterns $\{x1, x2, \ldots, x_m\}$, $x_i \in R^n$

Output:

Set of code vectors of quantization $z_1, z_2, \ldots, z_K$, which are centers of the clusters $\{C_1, C_2, \ldots, C_K\}$.

Step 1: Choose K initial cluster centers $z_1, z_2, \ldots, z_K$ randomly from the m patterns $\{x_1, x_2, \ldots, x_m\}$ where $K < m$.

Step 2: Assign pattern x_i to cluster C_j, where $i = 1, 2, \ldots, m$ and $j \in \{1, 2, \ldots, K\}$,
if and only if $\|x_j - z_j\| < \|x_j - z_p\|$, $p = 1, 2, \ldots, K$ and $j \neq p$.
Ties are resolved arbitrarily. And compute cluster centers for each point x_i as follows,
$z_i = (1/n_i) \sum x_j$,
$i = 1, 2, \ldots, K$.
$xj \in Ci$
Where n_i is the number of elements belongs to cluster C_i.

Step 3: Assign each pattern x_i to cluster C_j,
where $i = 1, 2, \ldots, m$ and $j \in \{1, 2, \ldots, K\}$
if and only if $\|x_j - z_j\| < \|x_j - z_p\|$, $p = 1, 2, \ldots, K$ and $j \neq p$,
where $\| \bullet \|$ is an Euclidean metric norm.
Ties are resolved arbitrarily, without changing the cluster centers z_j, $j = 1, 2, \ldots, K$

Step 4: Stop.

The k-means algorithm is the most extensively studied clustering algorithm and is generally effective in producing good results. The major drawback of this algorithm is that it produces different clusters for different sets of values of the initial centroids. Quality of the final clusters heavily depends on the selection of the initial centroids [12].

3.2 Fuzzy C Means

Fuzzy clustering allows each feature vector to belong to more than one cluster with different membership degrees (between 0 and 1) and vague or fuzzy boundaries between clusters. Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. This method (developed by Dunn in 1973 and improved by Bezdek in 1981) is frequently used in pattern recognition [15].

Algorithm Steps:

Step-1: Randomly initialize the membership matrix using this equation,

$$\sum_{j=1}^{C} \mu_j(x_i) = 1 \qquad i = 1,2,\ldots..k \qquad (9)$$

Step-2: Calculate the Centroid using equation,

$$Cj = \frac{\sum_i [\mu_j(x_i)]^m x_i}{\sum_i [\mu_j(x_i)]^m} \qquad (10)$$

Step-3: Calculate dissimilarly between the data points and Centroid using the Euclidean distance.

Step-4: Update the New membership matrix using the equation,

$$\mu_j(x_i) = \frac{[\frac{1}{d_{ji}}]^{1/m-1}}{\sum_{k=1}^{c} [\frac{1}{d_{ki}}]^{1/m-1}} \qquad (11)$$

Here **m** is a fuzzification parameter, The range **m** is always {1.25, 2}

Step-5: Go back to Step 2, unless the centroids are not changing.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

4. Classification Techniques

4.1 Back Propagation Networks (BPN)

Back Propagation learns by iteratively processing a data set of training tuples, comparing the network's prediction for each tuple with the actual known target value. The target value may be the known class label of the training tuple (for classification problems) or a continuous value (for prediction). For each training tuple, the weights are modified so as to minimize the mean squared error between the network's prediction and the actual target value. These modifications are made in the "backwards" direction, that is, from the output layer, through each hidden layer down to the first hidden layer (hence the name backpropagation) [13,14].

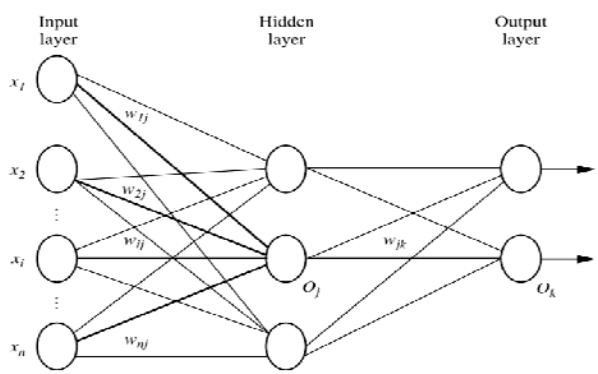


Fig. 1: A multilayer feed-forward neural network.

5. Confusion Matrix for a Two-Class Problem

A confusion matrix (Kohavi and Provost, 1998) 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 [7].

Table 1: Confusion Matrix for a Two-Class Problem

	Positive prediction	**Negative prediction**
Positive class	True positive (TP)	False negative (FN)
Negative class	False positive (FP)	True negative (TN)

Table 1 shows a confusion matrix for a two-class problem with positive and negative class values. From such a matrix it is possible to extract a number of widely used metrics to measure the performance of a classifier, such as error rate, defined as in the following equation,

$$Err = \frac{FP + FN}{TP + FN + TN + FP} \quad (12)$$

and overall accuracy, defined as in the following equation,

$$Acc = \frac{TP + TN}{TP + FN + TN + FP} \quad (13)$$

It is possible to derive four performance metrics from Table 1 to measure the classification performance on the positive and negative classes independently:

True positive rate: TPrate is the percentage of positive samples correctly classified as belonging to the positive class.

$$TPrate = \frac{TP}{TP + FN}$$

True negative rate: TNrate is the percentage of negative samples correctly classified as belonging to the negative class.

$$TNrate = \frac{TN}{TN + FP}$$

False positive rate: FPrate is the percentage of negative samples misclassified as belonging to the positive class.

$$FPrate = \frac{FP}{FP + TN}$$

False negative rate: FNrate is the percentage of positive samples misclassified as belonging to the negative class.

$$FNrate = \frac{FN}{FN + TP}$$

The advantages of the four performance measures are of being independent of class costs and a priori probabilities. The aim of a classifier is to minimize false positive and negative rates, or similarly to maximize true negative and positive rates.

6. Experimental Results

6.1 Data Sets

We use four datasets: leukemia, breast cancer, lung cancer and prostate cancer which are available in the website: http://datam.i2r.a-star.edu.sg/datasets/krbd/, [2]. the gene number and class contained in four datasets are listed in Table 2.

Table2: Summary of the four gene expression datasets.

Dataset	**#Gene**	**Class**	**# Samples**
Leukemia	7129	ALL/AML	34 (20/14)
Prostate	12600	Tumor/Normal	21 (8/13)
Breast	24481	Relapse/Non Relapse	19 (12/7)
Lung	7129	Tumor/Normal	96 (86/10)

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

The data studied by rough sets are mainly organized in the form of decision tables. One decision table can be represented as $S = (U, A=C \cup D)$, where U is the set of samples, C the condition attribute set and D the decision attribute set. We can represent every gene expression data with the decision table like Table 3.

Table 3. Microarray data decision table.

Samples	Condition attributes(genes)				Decision attributes
	Gene 1	Gene 2	...	Gene q	Class label
1	g(1,1)	g(1,2)	…	g(1,q)	Class(1)
2	g(2,1)	g(2,2)	…	g(2,q)	Class(2)
…	…	…	…	…	…
p	g(p,1)	g(p,2)	…	g(p,q)	Class(p)

In the decision table, there are p samples and q genes. Every sample is assigned to one class label. Each gene is a condition attribute and each class is a decision attribute. $g(p, q)$ signifies the expression level of gene q in sample p. [2].

6.2 Data Pre-processing, Gene Selection

Before applying feature selection algorithm all the conditional attributes (samples) are normalized using Z-Score normalization and then discretized using K-Means discretization [16]. Let us considered U is the set of samples, C the condition attribute set and D the decision attribute set. By applying Quick Reduct Algorithm, In leukemia dataset gene 4 and gene 3252 are identified, where as in prostate gene dataset, gene 20 and gene 11154 are identified, in breast cancer dataset gene 3 and gene 22019 are identified, finally in lung cancer dataset gene 4817 as best attribute for finding appropriate decision.

Table 4: Features selected by Quick Reduct Algorithm

Gene Data	Identified Attributes (Genes)
Leukemia Cancer	Gene 4, Gene 3252
Prostate Cancer	Gene 20, Gene 11154
Breast Cancer	Gene 3, Gene 22019
Lung Cancer	Gene 4817

6.3 Classification Performance

In this section the selected data is clustered by the K-Means and FCM algorithm. The data presented in Table 5 and 6 shows the classification performance of True Positive (TP) rate, True Negative (TN) rate, False Positive (FP) rate, and False Negative (FN) rate as previously described. Table 7 shows classification performance of Back Propagation Network. Results are presented both in terms of classification accuracy and classification error [7].

Table 5: K-Means Classification Performance Rate

Gene Data	K-Means			
	TP	FP	TN	FN
Leukemia Cancer	0.9820	0.0180	0.2428	0.7571
Prostate Cancer	0.6750	0.3250	0.0769	0.9230
Breast Cancer	0.6667	0.3333	0.4286	0.5714
Lung Cancer	0.8883	0.1116	0.0180	0.9820

Table 6: FCMs Classification Performance Rate

Gene Data	FCM			
	TP	FP	TN	FN
Leukemia Cancer	0.9820	0.0180	0.2428	0.7571
Prostate Cancer	0.3250	0.7750	0.0769	0.9230
Breast Cancer	0.6667	0.3333	0.4286	0.5714
Lung Cancer	0.8883	0.1116	0.0180	0.9820

When comparing classification results, where the BPN method shows a high in classification accuracy, which is demonstrated in Fig. 2.

Table 7: K-Means, FCM and BPN Classification Accuracy

Gene Data	K-Means	FCM	BPN
Leukemia Cancer	0.9721	0.9721	1.0000
Prostate Cancer	0.7421	0.6370	0.8210
Breast Cancer	0.6521	0.6521	0.9160
Lung Cancer	0.9695	0.9695	1.0000

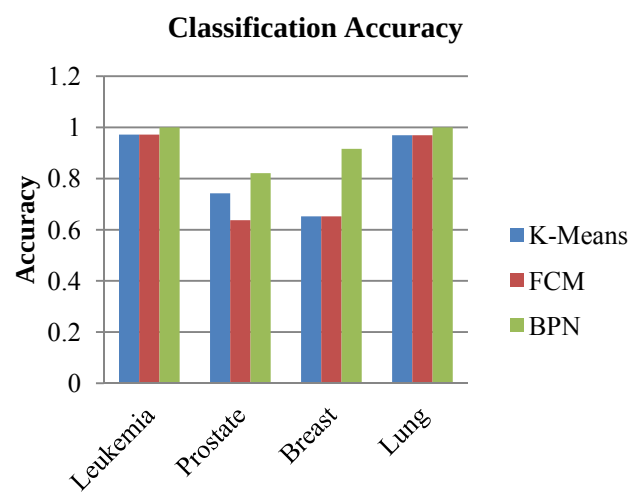


Fig 2: K-Means, FCM and BPN Classification Accuracy

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Table 8: K-Means, FCM and BPN Classification Error

Gene Data	K-Means	FCM	BPN
Leukemia Cancer	0.0279	0.0279	0.0000
Prostate Cancer	0.2579	0.3630	0.1790
Breast Cancer	0.3479	0.3479	0.0840
Lung Cancer	0.0305	0.0305	0.0000

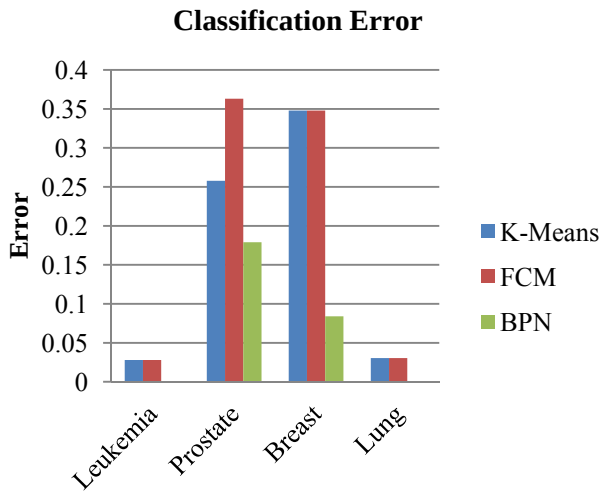


Fig 3: K-Means, FCM and BPN Classification Error

Fig. 2 and Fig.3 demonstrated the classification accuracy and error rate of Quick Reduct algorithm for gene expression data sets.

7. Conclusion

In this paper, Quick reduct algorithm based on rough set theory has been studied for gene expression datasets. The Quick reduct algorithm has been implemented and the obtained reduct features has been applied to find the clusters using K-Means and FCM clustering algorithm without class labels. The performance was evaluated using confusion matrix with positive and negative class values. Further, the selected features with class labels were classified using Back Propagation Network to find the accuracy of the algorithm. The BPN shows best performance for selected features over the K-Means and FCM algorithms.

References

[1] Jensen, R. and Shen, Q. (2003) 'Finding rough set reducts with ant colony optimization', Proceedings UK Workshop on Computational Intelligence, pp. 15–22.

[2] Xiaosheng Wang, Osamu Gotoh, "Cancer Classification Using Single Genes", pp 179-188.

[3] Pawlak, Z. (2002) 'Rough Sets and Intelligent Data Analysis', Information Sciences, Vol. 147, pp. 1–12.

[4] Changjing Shang and QiangShen, "Aiding Classification of Gene Expression Data with Feature Selection: A Comparative Study", International Journal of Computatonal Intelligence Research. ISSN 0973-1873 Vol.1, No.1 (2005), pp. 68–76

[5] Liang Goh, Qun Song, and Nikola Kasabov,"A Novel Feature Selection Method to Improve Classification of Gene Expression Data", Conferences in Research and Practice in Information Technology, Vol. 29.

[6] PradiptaMaji and Sankar K. Pal, "Fuzzy–rough sets for information measures and Selection of relevant genes from microarray data", IEEE transactions on systems, man, and cybernetics—part b: cybernetics, vol. 40, no. 3, June 2010

[7] C.Velayutham, K.Thangavel, "Unsupervised Feature Selection Using Rough Set". Proceeding on International Conference, Emerging Trends in Computing(ICETC-2011), 17-18 Mar 2011.

[8] K.Thangavel, P. Jaganathan, A. Pethalakshmi, M.Karnan,"Effective Classification with Improved Quick Reduct For Medical Database Using Rough System", BIME Journal, Volume (05), Issue (1), Dec., 2005.

[9] K.Thangavel, A. Pethalakshmi, "Feature Selection for Medical Database Using Rough System", AIML Journal, Volume (6), Issue (1), January, 2006

[10] QiangShen, Alexios Chouchoulas, "A Rough Fuzzy Approach For Generating Classification Rules",ww.elsevier.com/locate/patcog, Pattern Recognition 35 (2002) 2425 – 2438

[11] Parvesh Kumar, SiriKrishanWasan,"Comparative Analysis of k-mean Based Algorithms". IJCSNS International Journal of Computer Science and Network Security, VOL.10 No.4, April 2010.

[12] K. A. Abdul Nazeer, M. P. Sebastian, "Improving the Accuracy and Efficiency of the k-means Clustering Algorithm". Proceedings of the World Congress on Engineering 2009 Vol I WCE 2009, July 1 - 3, 2009, London, U.K.

[13] AshaGowdaKaregowda, A.S. Manjunath, M.A. Jayaram," Application of Genetic Algorithm Optimized Neural Network Connection Weights for Medical Diagnosis of Pima Indians Diabetes". International Journal on Soft Computing (IJSC), Vol.2, No.2, May 2011.

[14] Ping Chang and Jeng-Shong Shih," The Application of Back Propagation Neural Network of Multi-channel

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Piezoelectric Quartz Crystal Sensor for Mixed Organic Vapours".Tamkang Journal of Science and Engineering, Vol. 5, No. 4, pp. 209-217 (2002).

[15] Binu Thomas, Raju G., and SonamWangmo, "A Modified Fuzzy C-Means Algorithm for Natural Data Exploration". World Academy of Science,Engineering and Technology 49 2009.

[16] SellappanPalaniappan, Tan Kim Hong, "Discretization of Continuous Valued Dimensions in OLAP Data Cubes".IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.11, November 2008.

T. Chandrasekhar was born in 1980 at Karur, Tamilnadu, India. He is received the Master of Science in information technology and management in 2003 and his M.Phil (Computer Science) Degree in 2004, from Bharathidasan University, Trichy, India.

Currently he is working as Guest lecturer, Department of Computer Science, Periyar University, Salem, Tamilnadu, India. He is pursuing his Ph.D., from Bharathiar University in Computer Science under the guidance of Dr. K.Thangavel. His area of interests includes Medical Data Mining, Rough Set and Bioinformatics.

K.Thangavel was born in 1964 at Namakkal, Tamilnadu, India. He received his Master of Science from the 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 Institute-Deemed University, Gandhigram, India in 1999. Currently he is working as Professor and Head, Department of Computer Science, Periyar University, Salem. He is a recipient of Tamilnadu Scientist award for the year 2009. His area of interests includes Medical Image Processing, Artificial Intelligence, Neural Network, Fuzzy logic, Data Mining and Rough Set.

E.N. Sathishkumar was born in 1986 at Namakkal, Tamilnadu, India. He received his Master of Science in Information Technology from Anna University, Coimbatore in 2009. He obtained his M.Phil (Computer Science) Degree from Periyar University, Salem, India in 2011.

His area of interests includes Data Mining, Rough Set and Neural Network.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

A Quantitative Study and Comparison of AODV, OLSR and TORA Routing Protocols in MANET

Tamilarasan-Santhamurthy

Department of Information Technology, LITAM,
Dullipala (village), Sattenpalli (Mandal), Guntur, Andhra Pradesh, 522412, India

Abstract

Mobile Ad-Hoc Network (MANET) is a collection of wireless mobile hosts forming a temporary network without the aid of any stand-alone infrastructure or centralized administration. Most of the proposed MANET protocols do not address security issues. In MANETs routing algorithm is necessary to find specific routes between source and destination. The primary goal of any ad-hoc network routing protocol is to meet the challenges of the dynamically changing topology and establish an efficient route between any two nodes with minimum routing overhead and bandwidth consumption. The existing routing security is not enough for routing protocols. An ad-hoc network environment introduces new challenges that are not present in fixed networks. A several protocols are introduced for improving the routing mechanism to find route between any source and destination host across the network. In this paper present a logical survey on routing protocols and compare the performance of AODV, OLSR and TORA

Keywords: *AODV, OLSR, TORA, MANET, Routing*

1. Introduction

A MANET is a collection of mobile nodes that can communicate with each other without the use of predefined infrastructure or centralized administration. Due to self-organize and rapidly deploy capability, MANET can be applied to different applications including battlefield communications, emergency relief scenarios, law enforcement, public meeting, virtual class room and other security-sensitive computing environments. There are 15 major issues and sub-issues involving in MANET such as routing, multicasting/broadcasting, location service, clustering, mobility management, TCP/UDP, IP addressing, multiple access, radio interface, bandwidth management, power management, security, fault tolerance, QoS/multimedia, and standards/products. Currently, the routing, power management, bandwidth management, radio interface, and security are hot topics in MANET research. The routing protocol is required whenever the source needs to transmit and delivers the packets to the destination. Many routing protocols have been proposed for mobile ad hoc network. In this paper we present a number of ways of classification or categorization of these routing protocols and the performance comparison of an AODV, OLSR and TORA routing protocols.

2. Routing Protocol

MANET protocols are used to create routes between multiple nodes in mobile ad-hoc networks. IETF (Internet Engineering Task Force) MANET working group is responsible to analyze the problems in the ad-hoc networks and to observe their performance. There are different reasons for designing and classifying routing protocols for wireless ad-hoc networks. The MANET protocols are classified into three huge groups, namely Proactive (Table-Driven), Reactive (On-Demand) routing protocol and hybrid routing protocols. The following figure shows the classification of protocols [3].

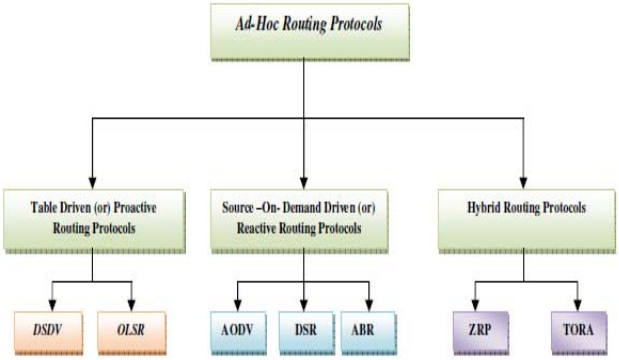


Fig.1: Different type of routing protocols in wireless Ad-hoc network

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Proactive (Table-Driven) routing protocol: - In proactive routing protocol perform reliable and up-to-date routing information to all the nodes is maintained at each node.
Reactive (On-Demand) routing protocol: - This type of protocols find route on demand by flooding the network with Route Request packets.

2.1. Proactive vs. Reactive Routing

Proactive methods determine the routes to various nodes in the network in advance, so that the route is already present whenever needed. Route Discovery overheads are larger in such schemes as one has to discover all routes. Examples of such schemes are the conventional routing schemes, Destination Sequenced Distance Vector (DSDV).Reactive methods determine the route when needed. Therefore they have smaller Route Discovery overheads. Examples for such schemes are Ad Hoc On-Demand Distance Vector (AODV) routing protocol.

2.2. Single-Path vs. Multi-Path

There are several decisive factors for comparing single-path routing and multi-path routing in ad-hoc networks. First, the overhead of route discovery in multi-path routing is much more than that of single-path routing. On the other hand, the frequency of route discovery is much less in a network which uses multi-path routing, since the system can still operate even if one or a few of the multiple paths between a source and a destination fail. Second, it is commonly believed that using multi-path routing results in a higher throughput. Third, multi-path networks are fault tolerant when dynamic routing is used, and some routing protocols, such as OSPF (Open Shortest Path First), can balance the load of network traffic across multiple paths with the same metric value.

2.3. Proactive vs. Source Initiated

A proactive (Table-Driven) routing protocols are maintaining up-to-date information of both source and destination nodes. It is not only maintained a single node's information, it can maintain information of each and every nodes across the network. The changes in network topology are then propagated in the entire network by means of updates. Some protocols are used to discover routes when they have demands for data transmission between any source nodes to any destination nodes in network, such protocol as DSDV(.Destination Sequenced Distance Vector) routing protocol. These processes are called initiated on-demand routing. Examples include DSR (Dynamic Source Routing) and AODV (Ad-hoc On Demand Distance Vector) routing protocols [3].

3. Ad-hoc on demand Vector Protocol (AODV)

AODV combines some properties of both DSR and DSDV. It uses route discovery process to cope with routes on-demand basis. It uses routing tables for maintaining route information. It is reactive protocol; it doesn't need to maintain routes to nodes that are not communicating. AODV handles route discovery process with Route Request (RREQ) messages. RREQ message is broadcasted to neighbor nodes. The message floods through the network until the desired destination or a node knowing fresh route is reached. Sequence numbers are used to guarantee loop freedom. RREQ message cause bypassed node to allocate route table entries for reverse route. The destination node uncast a Route Reply (RREP) back to the source node. Node transmitting a RREP message creates routing table entries for forward route [2] [5] and [7]. Figure (Fig.2) shows, AODV routing protocol with RREQ and RREP message.

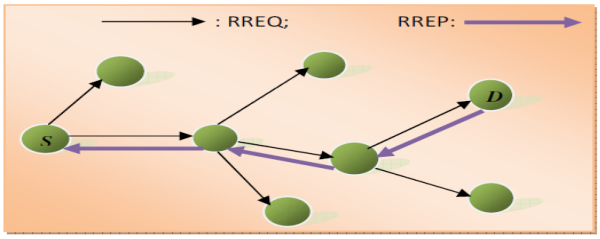


Fig. 2: AODV routing protocol with RREQ and RREP message

For route maintenance nodes periodically send HELLO messages to neighbour nodes. If a node fails to receive three consecutive HELLO messages from a neighbour, it concludes that link to that specific node is down. A node that detects a broken link sends a Route Error (RERR) message to any upstream node. When a node receives a RERR message it will indicate a new source discovery process. Figure (Fig.3) shows AODV routing protocol with RERR message.

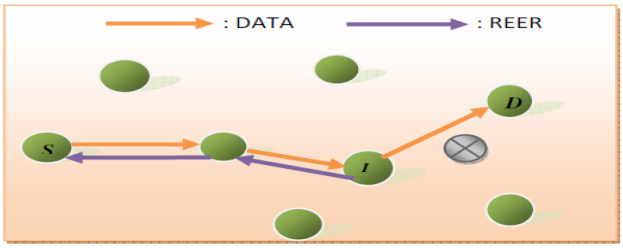


Fig.3: AODV routing protocol with RERR message

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Advantages:

- Routes are established on demand and destination sequence numbers are used to find the latest route to the destination.
- Lower delay for connection setup.

Disadvantage:

- AODV doesn't allow handling unidirectional links.
- Multiple Route Reply packets in response to a single Route Request packet can lead to heavy control overhead.
- Periodic beaconing leads to unnecessary bandwidth consumption.

4. Optimized Link State Routing Protocol (OLSR)

OLSR is a proactive routing protocol for mobile ad hoc networks. The protocol inherits the stability of the link state algorithm and has the advantage of having routes immediately available when needed due to its proactive nature. OLSR minimizes the overhead caused by flooding of control traffic by using only selected nodes, called Multi-Point Relays (MPR), to retransmit control messages. This technique significantly reduces the number of retransmissions required to flood a message to all nodes in the network. Upon receiving an update message, the node determines the routes (sequence of hops) toward its known nodes. Each node selects its MPRs from the set of its neighbors saved in the Neighbor list. The set covers nodes with a distance of two hops. The idea is that whenever the node broadcasts the message, only the nodes included in its MPR set are responsible for broadcasting the message [5] [6].

OLSR uses HELLO and TC messages. The Topology Control (TC) messages for continuous maintain of the routes to all destinations in the network, the protocol is very efficient for traffic patterns where a large subset of nodes is communicating with another large subset of nodes, and where the [source, destination] pairs change over time. The HELLO messages are exchanged periodically among neighbor nodes, to detect the identity of neighbors and to signal MPR selection. The protocol is particularly suited for large and dense networks, as the optimization is done by using MPRs which work well in this context. The larger and more dense a network, the more optimization can be achieved as compared to the classic link state algorithm. OLSR uses hop-by-hop routing, i.e., each node uses its local information to route packets [5].

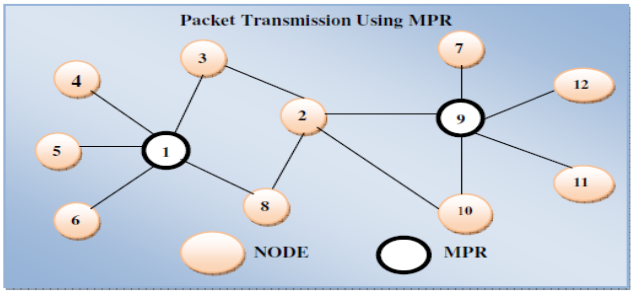


Fig.4: Packet Transmission Using MPR

Advantages:

- OLSR does not need central administrative system to handle its routing process.
- The link is reliable for the control messages, since the messages are sent periodically and the delivery does not have to be sequential.
- OLSR is suitable for high density networks.
- It does not allow long delays in the transmission of packets.

Disadvantages:

- OLSR protocol periodically sends the updated topology information throughout the entire network.
- It allows high protocol bandwidth usage.

5. Temporary Ordered Routing Algorithm (TORA)

The Temporally Ordered Routing Algorithm (TORA) is a highly adaptive, efficient and scalable distributed routing algorithm based on the concept of link reversal. TORA is proposed for highly dynamic mobile, multi-hop wireless networks. It is a source-initiated on-demand routing protocol. It finds multiple routes from a source node to a destination node. The main feature of TORA is that the control messages are localized to a very small set of nodes near the occurrence of a topological change. To achieve this, the nodes maintain routing information about adjacent nodes. The protocol has three basic functions: Route creation, Route maintenance and Route erasure. TORA can suffer from unbounded worst-case convergence time for very stressful scenarios. TORA has a unique feature of maintaining multiple routes to the destination so that topological changes do not require any reaction at all. The protocol reacts only when all routes to the destination are lost. In the event of network partitions

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

the protocol is able to detect the partition and erase all invalid routes.

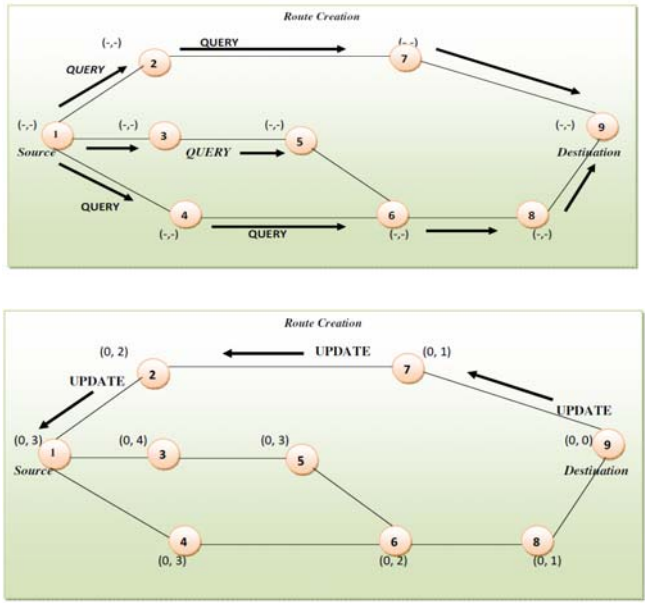


Fig.4: Route Creation

The figure shows, source node (1) broadcasts QUERY to its neighbor's node. Node (6) does not propagate QUERY from node (5) as it has already seen and propagated QUERY message from node (4). A source node (1) may have received a UPDATE each from node (2), it retains that height. When a node detects a network partition, it will generate a CLEAR packet that results in reset of routing over the ad-hoc network. The establishment of the route mechanism based on the Direct Acyclic Group (DAG). Using DAG mechanism, we can ensure that all the routes are loop free. Packets move from the source node having the highest height to the destination node with the lowest height like top-down approach.

Advantages:

- TORA supports multiple routes between source and destination. Hence, failure or removal of any of the nodes quickly resolved without source intervention by switching to an alternate route to improve congestion.
- TORA does not require a periodic update, consequently communication overhead and bandwidth utilization is minimized.
- TORA provides the supports of link status sensing and neighbor delivery, reliable, in-order control packet delivery and security authentication.

Disadvantages:

- It depends on synchronized clocks among nodes in the ad hoc network.
- The dependence of this protocol on intermediate lower layers for certain functionality presumes that the link status sensing, neighbor discovery, in order packet delivery and address resolution are all readily available. This solution is to run the Internet MANET Encapsulation Protocol at the layer immediately below TORA.
- This will make the overhead for this protocol difficult to separate from that imposed by the lower layer.

6. Comparative Study of Ad Hoc Routing Protocols

6.1. Metrics for Performance Comparison

MANET has number of qualitative and quantitative metrics that can be used to compare ad hoc routing protocols. The table-I illustrates the comparison of OLSR, AODV and TORA routing protocols. This paper has been considered the following metrics to evaluate the performance of ad hoc network routing protocols.

- Packet delivery ratio: The ratio of the data packets delivered to the destinations to those generated by the CBR sources.
- Optimal path length: It is the ratio of total forwarding times to the total number of received packets.
- Optimal path length: It is the ratio of total forwarding times to the total number of received packets.
- Average end to end delay: This is the difference between sending time of a packet and receiving time of a packet. This includes all possible delays caused by buffering during route discovery latency, queuing at the interface queue, retransmission delays at the MAC, and propagation and transfer tirnes.
- Media Access Delay: The time a node takes to access media for starting the packet transmission is called as media access delay. The delay is recorded for each packet when it is sent to the physical layer for the first time.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Table 1. Comparison of ad-hoc routing protocols

SL. NO	***Performance Constraints***	***OLSR***	***AODV***	***TORA***
1.	Multi-cost Routes	No	No	Yes
2.	Distributed Environment	Yes	Yes	Yes
3.	Unidirectional Link	Yes	No	Yes
4.	Multicast	Yes	Yes	No
5.	Periodic Broadcast	Yes	Yes	Yes
6.	QoS Support	Yes	No	Yes
7.	Routes Information Maintained in	Route Table	Route Table	Route Table(Adjacent nodes on-hop knowledge
8.	Reactive	No	Yes	Yes
9.	Proactive	Yes	No	Yes
10.	Hybrid	No	No	Yes
11.	Provide Loop-Free Routers	Yes	Yes	Yes
12.	Scalability	Yes	Yes	Yes
13.	Route Reconfiguration	Control Messages sent in advance to increase the reactivenes	Erase Route notify source	Link reversed route repair
14	Routing Philosophy	Flat	Flat	Flat
15.	Route Optimization	Yes	Yes	Yes
16.	Protocol Type	Link State scheme	Distance Vector	Link Reversed
17.	Message Overhead	Minimum	Larger	Moderate
18.	Protocol Suite	Large and Dense networks	Dynamic Self-Starting networks	Large and Dense networks
19.	Summary	Control messages for link sensing, Neighbor (MPR) Detection, Multiple Interface Detection, Route Calculation	Route Discover, Expanding Ring, Search, Setting Forward Path.	Link Reversal, Route Discovery, Route Update Packets.

Table. 2. Routing performance in low mobility

Low Mobility and Low traffic					
Sl. No	***Protocol***	***End-to-End Delay***	***Packet Delivery Ratio***	***Route Over head***	***Path Optimality***
1.	OLSR	Low	High	Low	Good
2.	AODV	Average	High	Low	Average
3.	TORA	Low	High	Average	Good

Table.3. Routing performance in high mobility

High Mobility and High traffic					
Sl. No	***Protocol***	***End-to-End Delay***	***Packet Delivery Ratio***	***Route Over head***	***Path Optimality***
1.	OLSR	Low	Average	Average	Good
2.	AODV	Average	Average	Average	Average
3.	TORA	High	Low	Average	Good

7. Conclusion

In this article, we presents the comparative study and performance analysis of three mobile ad hoc routing protocols (OLSR, AODV and TORA) on the basis of end-to-end delay, packet delivery ratio, media access delay, path optimality, routing overhead performance metrics. The quantitative study of these routing protocols shows that OLSR is more competent in high density networks with highly sporadic traffic. OLSR requires that it continuously have some bandwidth in order to receive the topology updates messages. AODV keeps on improving in packet delivery ratio with dense networks. The performance of all protocols was almost stable in sparse medium with low traffic. TORA performs much better in packet delivery owing to selection of better routes using acyclic graph. It has been concluded that performance of TORA is better for dense networks. The AODV is better for moderately dense networks where as the OLSR performs well in sparse networks. The future work suggested that the effort will be made to enhance ad hoc network routing protocol by tackling core issues.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

References

[1] Sachin Kumar Gupta and R.K.Saket; "PERFORMANCE METRIC COMPARISON OF AODV AND DSDV ROUTING PROTOCOLS IN MANETs USING NS-2", IJRRAS 7 (3). JUNE 2011, PP: 339 – 350.

[2] C. E. Perkins and E. M. Royer; "Ad-Hoc On Demand Distance Vector Routing", Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications (WMCSA), PP: 90-100, 1999.

[3] S.Tamilarasan; "A Performance Analysis of Multi-hop Wireless Ad-Hoc Network Routing Protocols in MANET', International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 2 (5), 2011, PP: 2141 – 2146.

[4] Preeti Nagrath, Bhawana Gupta; "Wormhole Attacks in Wireless Adhoc Networks and their Counter Measurements: A Survey" 2011, IEEE, PP: 245 – 250.

[5] Zhan Haawei, Zhou Yun; "Comparison and analysis AODV and OLSR Routing Protocols in Ad Hoc Network", 2008, IEEE.

[6] JIRI HOSEK, KAROL MOLNAR; "Investigation on OLSR Routing Protocol Efficiency", Recent Advances in Computers, Communications, Applied Social Science and Mathematics, ISBN: 978-1-61804-030-5, PP: 147 – 153.

[7] Anuj K. Gupta, Member, IACSIT, Dr. Harsh Sadawarti, and Dr. Anil K. Verma; "Performance analysis of AODV, DSR & TORA Routing Protocols", IACSIT International Journal of Engineering and Technology, Vol.2, No.2, April 2010, ISSN: 1793-8236, PP: 226 – 331.

[8] Alexander Klein, "Performance Comparison and Evaluation of AODV, OLSR and SBR in Mobile Ad-Hoc Networks", Innovation Works, IEEE 3rd International Symposium on Wireless Pervasive Computing, 2008.

[9] J. Broch, D.A. Maltz, D. B. Johnson, Y-C. Hu, J. Jetcheva, "A performance comparison of Multi-hop wireless ad-hoc networking routing protocols", in the proceedings of the 4th International Conference on Mobile Computing and Networking (ACM MOBICOM '98), pp. 85-97, October 1998.

[10] Md. Golam Kaosar, Hafiz M. Asif, Tarek R. Sheltami, Ashraf S. Hasan Mahmoud, "Simulation-Based Comparative Study of On Demand Routing Protocols for MANET", available at http://www.lancs.ac.uk, Internaional Conference on Wireless Networking and Mobile Computing, Vol. 1, pp. 201 – 206, December 2005.

[11] S. Gowrishankar, T.G. Basavaraju, Subir Kumar Sarkar "Simulation Based Overhead Analysis of AOMDV, TORA and OLSR in MANETUsing Various Energy Models",Proceedings of the World Congress on Engineering and Computer Science 2010 Vol.I , October 2010.

[12] V. Park and S. Corson, "Temporally Ordered Routing Algorithm (TORA) Version 1, Functional specification", IETF Internet draft, http://www.ietf.org/internet-drafts/draft-ietf-manet-tora-spec-01.txt, 1998.

[13] V. D. Park and M. S. Corson, "A Highly Adaptive Distributed Routing Algorithm for Mobile Wireless Networks", Proceedings of the IEEE International Conference on Computer Communications (INFOCOM), Kobe, Japan, pp. 1405-1413, 1997.

[14] Z. J. Hass and M. R. Pearlman, "Zone Routing Protocol (ZRP)", Internet draft available at www.ietf.org, November 1997.

[15] H. Ehsan and Z. A. Uzmi (2004), "Performance Comparison of Ad HocWireless Network Routing Protocols", IEEE 8th International Multitopic Conference, Proceedingsof INMIC, pp.457 – 465, December 2004.

[16] Charles E.Perkins, Elizabeth M.Royer, Samir R.Das, "Performance comparison of two on-demand Routing Protocols for Ad-hoc Networks", IEEE Personal Communications, pp. 16-28, February 2001.

[17] C. E. Perkins and E. M. Royer, "Ad-Hoc On Demand Distance Vector Routing", Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications (WMCSA), pp. 90-100, 1999.

Authors

S. Tamilarasan, M.E.
Associate professor cum Head of Department, Loyola institute of Technology and management,
Dullipala (village), Sattenapalli (Mandal), Guntur, Andhra Pradesh, India.
Specialization:
Mobile computing, Advanced Data Structure, Design and Analysis of algorithm, Computer networks

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Proposal of a New Block Cipher reasonably Non-Vulnerable against Cryptanalytic Attacks

Abhijit Chowdhury[1], Angshu Kumar Sinha[1], Saurabh Dutta[2]

[1] **Department of Computer Applications, NSHM College of Management and Technology, Durgapur, West Bengal, INDIA**

[2]**Dr. B. C. Roy Engineering College, Durgapur-713206, West Bengal, INDIA**

Abstract

This paper proposes a new block cipher termed as "Modular Arithmetic based Block Cipher with Varying Key-Spaces (MABCVK)" that uses private key-spaces of varying lengths to encrypt data files. There is a simple but intelligent use of theory of modular arithmetic in the scheme of the cipher. Based on observed implementation of the proposed cipher on a set of real data files of several types, all results are tabulated and analyzed. The schematic strength of the cipher and the freedom of using a long key-space expectedly can make it reasonably non-vulnerable against possible cryptanalytic attacks. As a part of the future scope of the work, it is also intended to formulate and implement an enhanced scheme that will use a carrier image to have a secure transmission of the private key

Keywords: *Cryptography, Modular Arithmetic, Carrier Image, Data Security*

1. Introduction

Module arithmetic [1], a strong backbone in numerous existing ciphering protocols [2], has been applied in an intelligent manner to develop "Modular Arithmetic based Block Cipher with Varying Key-Spaces" with the acronym as MABCVK. Although the encryption of data using MABCVK results in encrypted data of extended dimension, its schematic strength and the liberty of using a long key-space while encrypting using it expectedly can make MABCVK reasonably non-vulnerable against possible cryptanalytic attacks.

Section 2 presents the scheme followed in MABCVK. A report of implementation of MABCVK on a sample plaintext is demonstrated in section 3. Results of a set of real implementations are presented in section 4. Section 5 is an analytical observation to MABCVK. A conclusion is drawn in section 6.

2. The Scheme

The data to be encrypted is divided in blocks each of L-bits. All these blocks are converted into equivalent set of integers of base-10.Next, two sufficiently large prime integers K1 and K2 are taken as keys satisfying the conditions:

1. K1, K2 being sufficiently bigger than the number or numbers to be encrypted

2. K1 < K2

A variable α with its value being in the range of 0.1 and 0.9 is taken to be used in due course of time to shift the sequence of encrypted set of integers. Sections 2.1 and section 2.2 respectively describe the encryption and the decryption processes with conventionally *Alice* being considered as the sender and *Bob* being considered as the receiver

2.1 Encryption

The set of computations to be done in the encryption end for P being a positive integer to be encrypted is as follows:

1. Set K3 = φ (K1) – 1.
2. Set A = Chosen random number from the set of numbers, which when used as quotient to divide K2, leaves P as remainder.
3. Set B = A^{K3} mod K1
4. Set C = Product of B and K2

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

5. Set $D = C^{K3} \bmod K1$
6. Set E = Result of shifting D by α %, which is the encrypted form for P.

The function φ (K1) is a totient function [3]. The totient φ (K1) of a positive integer K1 defines the number of integers less than or equal to K1 that are relatively prime to K1. φ (K1) = K1 – 1, when K1 is a prime number.

2.2 Decryption

Based on the set of values for K1, K2, α extracted from the private key, the following set of operations are carried out in the decryption end.

1. Set K3 = φ (K1) – 1.
2. Set D = Result of shifting E by (1- α)%
3. Set $B = (D \bullet K2)^{K3} \bmod K1$
4. Set $A = B^{K3} \bmod K1$
5. Set P = K2 mod A

The function φ (K1) is a totient function as was described in section 2.1.

3. Implementation

This section describes the process of encryption and decryption using a sample tiny plaintext WORLD considering K1 = 263, K2 = 317 and α = 0.1.
Table 1 enlists different intermediate and the final result of implementation following the encryption algorithm stated in section 2.1. Consequently, the cipher text "U≤a¼E" is generated for the taken plaintext

Table 1 : Implementation of Encryption of the Plaintext "WORLD"

CHARACTER	ASCII VALUE	Value obtained by following step 2 of the Encryption Algorithm	Value obtained by following step 3 of the Encryption Algorithm	Value obtained by following step 4 of the Encryption Algorithm	Value obtained by following step 5 of the Encryption Algorithm	Character corresponding to value obtained through step 5 of the Encryption Algorithm	Character obtained by following step 6 of the Encryption Algorithm for an assumed value of 0.1 for α
W	87	230	255	80835	14	E	U
O	79	119	42	13314	85	U	≤
R	82	235	216	68472	243	≤	a
L	76	241	251	79567	97	a	¼
D	68	83	244	77348	172	¼	E

A reverse process is followed to employ the decryption algorithm stated in section 2.2.

4. Results

The implementation of the encryption / decryption algorithm described in section 2.1 and section 2.2 respectively was accomplished using C language and MIRACL (**M**ulti precision **I**nteger and **R**ational **A**rithmetic **C**/C++ **L**ibrary) library [4] for big integer calculations. The plaintext to be encrypted was read as blocks of 8-bits, a pair of private keys KEY1 and KEY2, each 10-bits long, and a small value **α** was taken for operation on the plaintext data.

Ciphertext data produced after encryption process was written as blocks of 32-bits i.e. size of integer type variable, to avoid overflow, as in some cases 8-bit data after encryption is to yield a value greater than 255. We have encrypted and also successfully decrypted file types .txt, .doc, .docx, .xls, .pdf, .mp4, .ocx, .jpg, .bmp, .dll, .gif, .sys etc.

A comparison of file type, source file size, decrypted file size, encryption time and decryption time is depicted in the table 2.

It is seen from data in Table 2 that the relation between encryption time and file size is linear. The average time, based on tabulated data in table 2, to encrypt 1 KB of data is remarkably found to be 0.028852538 seconds. Similarly, it was noted that the average time required for decryption of 1 KB of data is 0.007198586 seconds.

Each 8-bit source data is written as 32-bit encrypted data, so the encrypted data file size (in bytes) is approximately four times the source data file size (in bytes). Thus it was observed that the proposed cipher carries a schematic overhead of causing data expansion. Having computational complexities, an operational overhead of being marginally slow in execution is observed during its real implementation.

TABLE 2:
file type, source file size, and encryption time

sl no	Source filename	size (in kb)	encryption time (sec)	Encrypted file size (in KB)	Encrypted filename	Decryption Time (sec)	Decrypted File size (in KB)
1	abc.docx	47	1.516	188	abc.docx.enc	1.5	47
2	hawaii.jpg	18	0.64	71	hawaii.jpg.enc	0.656	18
3	3rd sem2011.docx	47	1.532	188	3rd sem2011.docx.enc	1.5	47
4	cv.doc	60	1.219	238	cv.doc.enc	1.235	60
5	EffectCtrl.ocx	405	11.203	1617	EffectCtrl.ocx.enc	11.172	405
6	Flower.bmp	17	0.625	68	Flower.bmp.enc	0.625	17
7	hmmapi.dll	59	1.532	236	hmmapi.dll.enc.enc	1.516	59
8	internal marks.xls	85	1.781	338	internal marks.xls.enc	1.719	85
9	Crypto.zip	66	2.172	261	Crypto.zip.enc	2.188	66
10	test.txt	2	0.125	7	test.txt.enc	0.11	2
11	Univ_tabulator.pdf	434	13.625	1736	Univ_tabulator.pdf.enc	13.688	434
12	Vivekananda.gif	25	0.875	100	Vivekananda.gif.enc	0.875	25
13	BuzzingBee.wav	144	4.829	573	BuzzingBee.wav.enc	4.765	144
14	clock.avi	81	2.578	324	clock.avi.enc	2.484	81
15	EffectCtrl.ocx	405	11.219	1617	EffectCtrl.ocx.enc	11.266	405
16	Gone Fishing.bmp	17	0.625	68	Gone Fishing.bmp.enc	0.641	17
17	Greenstone.bmp	26	0.875	104	Greenstone.bmp.enc	0.844	26
18	handsafe.reg	1	0.031	3	handsafe.reg.enc	0.031	1
19	himem.sys	5	0.234	19	himem.sys.enc	0.234	5
20	IE7Eula.rtf	73	2.469	292	IE7Eula.rtf.enc	2.485	73
21	jview.exe	151	3.156	602	jview.exe.enc	3.14	151
22	Keyboard.drv	2	0.094	8	Keyboard.drv.enc	0.109	2
23	Mscomm32.ocx	102	2.828	406	Mscomm32.ocx.enc	2.765	102
24	normnfd.nls	39	1.063	154	normnfd.nls.enc	1.031	39
25	notepad.exe	68	1.797	270	notepad.exe.enc	1.734	68
26	PreConvertLite.dll	103	2.969	412	PreConvertLite.dll.enc	2.953	103

Figure 1 shows plotting of source file size (in KB) and encryption time (in Sec).

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

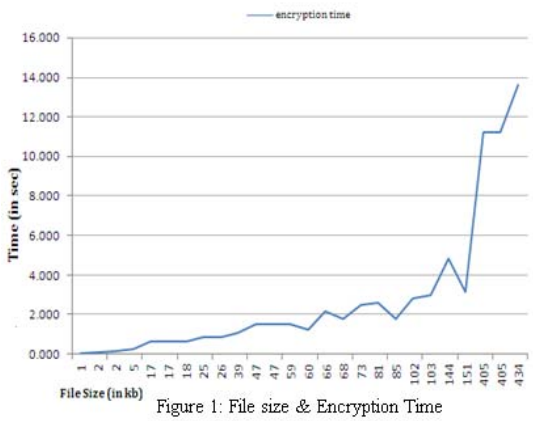


Figure 1: File size & Encryption Time

Figure 2 shows plotting of file size and decryption time.

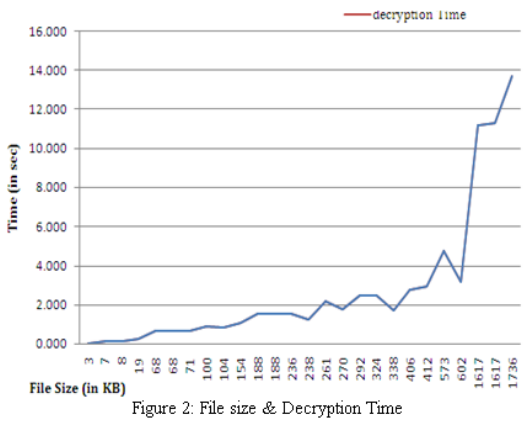


Figure 2: File size & Decryption Time

The encryption and decryption process being almost symmetrical, the encryption time for any specific file is nearly equal to the time required to decrypt it. This is validated from the graph plotting shown in figure 1 and figure 2.

5. Analysis

Although small keys were used to tabulate results of encryption using MABCVK, it can be used with much larger key-spaces with the support of appropriate computing platform for large exponential computations.

The substitution of plaintext data with random integers generates different ciphers for the same set of keys. If the number of integers to be encrypted is N and if there can be at least one substitution for each then there can be at least N different cipher texts generated using the same key sets. This property makes frequency analysis attacks difficult.

The effective key length is equal to sum of length of KEY1 and KEY2. Brute-force search will require searching entire search space of $2^{\text{(sum of length of KEY1 \& KEY2)}}$.

The average time required for such extensive search is given in table 3 [5].

TABLE 3: Average Required Time for Exhaustive Key Search

Key Size(Bits)	Number of Alternative Keys	Time Required at 1 Encryption /µs	Time Required at 10^6 encryptions /µs
56	$2^{56} = 7.2 \times 10^{16}$	2^{55} µs =1142 years	10.01 hours
128	$2^{128} = 3.4 \times 10^{38}$	2^{127} µs = 5.4×10^{24} years	5.4×10^{18} years
168	$2^{168} = 3.7 \times 10^{50}$	2^{167} µs = 5.9×10^{36} years	5.9×10^{30} years
26 characters	$26! = 4 \times 10^{26}$	2×10^{26} µs = 6.4×10^{12} years	6.4×10^{6} years

Data expansion noted during encryption is expected as the proposed cipher uses exponentiation. We can adopt to write encrypted data as bit stream and device an appropriate padding scheme to identify separate blocks to reduce the encrypted file size considerably. The computational complexity of calculating large integer powers of integers can be simplified by adopting faster methods like exponentiation by squaring, square and multiply or binary exponentiation, which leads to reduction of both encryption and decryption time. Encryption process can be augmented to be faster if read block size (in bits) is increased but it should be less than the size of key (in bits). The proposed scheme restricts use of the read block size greater than the key size.

6. Conclusion

The proposed MABCVK is flexible enough for implementation with larger keys. The substitution of source data with random integers produces different sets of encrypted data for the same private key pairs, which, in turn, makes the task of cryptanalysis a challenging one. Encryption and decryption using MABCVK require almost same execution time. If a pair of 128-bit keys is used for encryption and if brute-force attack is applied for decryption then at least all the primes in the range of integers from 1 to $\mathbf{2^{128}}$ are to be tested as a probable key. There are $\mathbf{2^{128}}$ / Log (2^{128}) = 383534127545935 x $\mathbf{10^{22}}$ prime numbers in the range of integers from 1 **to** $\mathbf{2^{128}}$. It is observed that if one trillion of these numbers per second can be checked, more than 121,617,874,031,562 x $\mathbf{10^3}$ years is required to check all such primes. The requirement of time as shown in table 3, to calculate the keys makes it reasonably non-vulnerable against cryptanalytic attacks. As a part of the future scope of the work, it is also intended to formulate and implement an enhanced scheme that will use a carrier image to have a secure transmission of the private keys.

References

[1] Menezes, P. van Oorschot, and S. Vanstone – "Handbook of Applied Cryptography ", CRC Press, 1996.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

[2] Bruce Schneier - "Applied Cryptography, Second Edition: Protocols, Algorithms, and Source Code in C", Publisher: John Wiley & Sons, Inc.

[3] Clay S. Turner, Euler's Totient Function and Public Key Cryptography - *http://web.cs.du.edu/~ramki/courses/security/2011Winter/notes/RSAmath.pdf*.

[4] MIRACL Library web link: - *http://www.shamus.ie*

[5] Saurabh Dutta - An Approach to Design New Efficient Encryption Techniques, ISBN: 978-3-8443-0082-6, Lap Lambert Academic Publishing, Germany.

Abhijit Chowdhury, MCA, Assistant Professor, Department of Computer Applications – NSHM College of Management & Technology, Durgapur, West Bengal, INDIA. Member of IAENG (International Association of Engineers, http://www.iaeng.org). Research Area of interest is cryptography, network security.

Angshu Kumar Sinha, M.Sc.-Mathematics, Department of Computer Applications – NSHM College of Management & Technology, Durgapur, West Bengal, INDIA.

Dr. Saurabh Dutta, Ph.D. (Computer Science), Dr. B. C. Roy Engineering College, Durgapur-713206, West Bengal, INDIA.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Hybridizing PSM and RSM Operator for Solving NP-Complete Problems: Application to Traveling Salesman Problem

Otman ABDOUN[1], Chakir TAJANI[2] and Jaafar ABOUCHABKA[3]
LaRIT, Department of Computer Science
IBN Tofail University, Kenitra, Morocco

Abstract

In this paper, we present a new mutation operator, Hybrid Mutation (HPRM), for a genetic algorithm that generates high quality solutions to the Traveling Salesman Problem (TSP). The Hybrid Mutation operator constructs an offspring from a pair of parents by hybridizing two mutation operators, PSM and RSM. The efficiency of the HPRM is compared as against some existing mutation operators; namely, Reverse Sequence Mutation (RSM) and Partial Shuffle Mutation (PSM) for BERLIN52 as instance of TSPLIB. Experimental results show that the new mutation operator is better than the RSM and PSM.

Keywords: *NP-complete problem, Traveling Salesman Problem, Mutation operator*.

1. Introduction

NP-Complete is a class of problems that are so difficult that even the best solutions cannot consistently determine their solutions in an efficient way. Specifically, NP Complete problems can only possibly be solved in polynomial time using a nondeterministic Turing machine. Problems such as the Traveling Salesman problem fall into the realm of NP Complete problems. Because of the problem with finding the "best" solution, programs are often developed to find a usually reasonable solution.
The traveling salesman problem (TSP) is a well known and important combinatorial optimization problem. In the traveling salesman problem (TSP) we are given ***n*** vertices ***1, .., n*** and all ***n(n-1)/2*** distances between them, as well as a budget ***b***. We are asked to find a tour, a cycle that passes through every vertex exactly once, of total cost b or less or to report that no such tour exists. The number of solutions becomes extremely large for even moderately large n so that an exhaustive search is impracticable.

The methods that provide the exact optimal solution to the problem are called exact methods. An implicit way of solving the TSP is simply to list all the feasible solutions, evaluate their objective function values and pick out the best. However it is obvious that this "exhaustive search" is grossly inefficient and impracticable because of vast number of possible solutions to the TSP even for problem of moderate size. Since practical applications require solving larger problems, hence emphasis has shifted from the aim of finding exactly optimal solutions to TSP, to the aim of getting, heuristically, 'good solutions' in reasonable time and 'establishing the degree of goodness'. Genetic algorithm (GA) is one of the best heuristic algorithms that have been used widely to solve the TSP instances [1].

In Genetic Algorithm, a population of potential solutions termed as chromosomes and individuals are evolved over successive generations using a set of genetic operators called selection, crossover and mutation. First of all, based on some criteria, every chromosome is assigned a fitness value and then the selection operator is applied to choose relatively fit chromosomes to be part of the reproduction process. In reproduction process new individuals are created through application of operators. Large number of operators has been developed for improving the performance of GA, because the performance of algorithm depends on the ability of these operators [15]. One of the operators, mutation operator is used to maintain adequate diversity in the population of chromosomes and avoid premature convergence. Whereas crossover operator, blends the genetic information between chromosomes to explore the search space[16].

Researchers have shown that the mutation operator plays an important role in genetic algorithm, so many mutation operators have been proposed for the TSP. Partial Shuffle Mutation (PSM) [4], Inversion Mutation [8], Exchange mutation [9], Displacement Mutation [10], Insertion Mutation [11], Heuristic mutation [12], Greedy Swap Mutation (GSM) [13], Reverse Sequence Mutation (RSM), etc.

In this paper, a new mutation operator named Hybrid Mutation (HPRM) is developed for solving the TSP.
This paper is organized as follows: Section 2 presents the NP-complete problem. Section 3 proofs TSP is an NP-complete problem. Section 4 presents the proposed operator. Section 5 describes computational experiments for four mutation operators. Finally, Section 6 presents comments and concluding remarks.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

2. NP-Complete Problem

NP-complete problems are a class of hard problems, which so far cannot be solved in polynomial time. Many problems from our everyday life are NP-complete, and although we might not be able to find an optimal solution within reasonable time, different methods exist to find a satisfactory solution. These methods include approximation, where a near optimal solution can be guaranteed, randomization, where an optimal solution can be found with a certain probability, or heuristics where a good solution can be found, but where there is no guarantee, it will be found fast. For many problems it is difficult to devise a complete algorithm that guaranties an optimal solution. A higher level of heuristics called meta-heuristics can then be used to combine the solution given by heuristics and solution strategies to obtain a better solution.

Problems that can be solved by algorithms in polynomial time are considered to be so called easy problems. For a problem of size ***n*** the time needed to find a solution is a polynomial function of ***n***. Harder problems requires on the other hand an exponential function of ***n***, which of course means that the execution time grows much faster than for an easy problem, when the size of the problem increases. NP-complete problems are hard problems to solve. They belong to a class of computational problems, for which no deterministic polynomial algorithm has been found.

NP-complete problems are a subset of the class ***NP*** (Non-Deterministic Polynomial). A Non-deterministic algorithm is able to find a correct solution, but it is not always guaranteed. The solution is found by making a series of guesses, and the algorithm will only arrive at a correct solution, if the right guesses are made along the way. A problem is called ***NP***, if its solution can be found and verified by a non-deterministic algorithm in polynomial time. The class has the following definition according to [5]:

Definition: A yes-no-problem is in ***NP*** if there is a polynomial ***p*** and a randomized.
p-bounded algorithm ***A*** such that for every input ***X*** the following holds:
True answer for ***X*** is **YES** then PR[A(X;R) = YES] > 0
True answer for ***X*** is **NO** then PR[A(X;R) = YES] = 0
where P*R*[*Z*] denotes the probability of event *Z* over uniform distribution of *R*.

The list of *NP*-complete is long, there exits several thousand problems, such as The Travelling Salesman problem (TSP), Constraint Satisfaction Problems (CSP), A Satisfiability (SAT), The Job Shop Scheduling problem, The k-Graph Partitioning problem, The vertex cover problem, etc.

The question now is, the Travelling Salesman Problem is an NP-Complete problem?

3. TSP is an NP-complete problem?

The TSP is a classical NP-Complete combinatorial optimization problem which comes up in different situations in our world .The TSP was first introduced by Karl Menger, in Vienna, and Harvard Universities and its significance was raised at Princeton University in 1930's. [2]. It is also the most studied problem for finding optimal solution. It can be stated as – ***N*** points (cities) as well as the cost of traveling between each pair of them are given. Assume that a sales person starting from a given city has to visit each city exactly once and should come back to the starting city to complete the tour. The aim is to find out the optimum tour in which the total cost is minimized.

More formally TSP can be formulated as problem of graph theory. Given a graph ***G*** on a set of ***N*** vertices and the distances (costs of travelling) between each pair of cities as cost matrix matrix ***C = [cij]*** of order ***N*** associated with ordered node pairs ***(i, j)***. The objective is to find a close tour with a minimum-cost that visits each city once returning to the starting city i.e. finding the shortest Hamiltonian cycle through ***G***. On the basis of the structure of the cost matrix, the TSPs are classified into two groups symmetric and asymmetric. The TSP is symmetric if ***Cij = Cji***, for all ***i, j*** and asymmetric otherwise. For an n-city asymmetric TSP, there are ***(N−1)!*** Possible solutions, one or more of which gives the minimum cost. For an n-city symmetric TSP, there are ***(N −1)!/2*** possible solutions along with their reverse cyclic permutations having the same total cost. In either case the number of solutions becomes extremely large for even moderately large ***N*** so that an exhaustive search is impracticable. Since TSP is ***NP*** complete the corresponding optimization problems are therefore NP hard.

The travelling salesman problem (TSP) is an NP-hard problem in combinatorial optimization studied in operations research and theoretical computer science [5].

Theorem: The subset-sum problem is NP-complete [6].

Proof : We first show that TSP belongs to NP. Given an instance of the problem, we use as a certificate the sequence of n vertices in the tour. The verification algorithm checks that this sequence contains each vertex exactly once, sums up the edge costs, and checks whether the sum is at most k. This process can certainly be done in polynomial time.

To prove that TSP is NP-hard, we show that ***HAM-CYCLE ≤ P*** TSP. Let ***G = (V, E)*** be an instance of HAM-CYCLE. We construct an instance of TSP as follows. We form the complete graph ***G' = (V, E'),*** , where ***E'={ (i,j) : i, j ∈ V*** and ***i ≠j }***, and we define the cost function ***c*** by

$$c(i,j) = \begin{cases} 0 & if\ (i,j) \in E \\ 1 & if\ (i,j) \notin E \end{cases} \quad (5)$$

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

(Note that because ***G*** is undirected, it has no self-loops, and so ***c(v, v)=1*** for all vertices ***v*∈*V***.) The instance of TSP is then ***(G'***, ***c, 0),*** which we can easily create in polynomial time.

We now show that graph ***G*** has a Hamiltonian cycle if and only if graph ***G'*** has a tour of cost at most ***0***. Suppose that graph ***G*** has a Hamiltonian cycle ***h***. Each edge in ***h*** belongs to ***E*** and thus has cost ***0*** in ***G'***. Thus, ***h*** is a tour in ***G'*** with cost ***0***.

Conversely, suppose that graph ***G'*** has a tour ***h'*** of cost at most ***0***. Since the costs of the edges in ***E'*** are ***0*** and ***1***, the cost of tour ***h'*** is exactly ***0*** and each edge on the tour must have cost ***0***. Therefore, ***h'*** contains only edges in ***E***. We conclude that ***h'*** is a Hamiltonian cycle in graph ***G***.
The best known algorithms have exponential (deterministic) rum time complexity. Such combinatorial optimization problems are in the domain of Genetic algorithms that's why TSP has been solved through GA though, as far as the heuristic approach is concerned, it has been provided many algorithms for finding near optimal solutions for symmetric as well as asymmetric TSP.

Genetic algorithms (GAs) [17] are population based search techniques which mimics the principles of natural selection and natural genetics laid by Charles Darwin. Since Holland introduced the genetic algorithm (GA) in the early 1970's, many researchers have become interested in it as a new method of solving real life problems. As it is a promising heuristic approach to locate near optimal solution in large spaces, it is not surprising that it is a target of many researchers [16].
I
n Genetic Algorithm, a population of potential solutions termed as chromosomes and individuals are evolved over successive generations using a set of genetic operators called selection, crossover and mutation. Large number of operators has been developed for improving the performance of GA, because the performance of algorithm depends on the ability of these operators [15].

One of the operators, mutation plays the role of recovering the lost genetic materials as well as for randomly disturbing genetic information and to maintain adequate diversity in the population of chromosomes [16].

4. Hybrid Mutation Operator

Mutation involves the modification of the value of each 'gene' of a solution with some probability pm, (the mutation probability). The role of mutation in GAS has been that of restoring lost or unexplored genetic material into the population to prevent the premature convergence of the GA to suboptimal solutions.

Mutation plays the role of recovering the lost genetic materials as well as for randomly disturbing genetic information. Mutation prevents the algorithm to be trapped in a local minimum. It is an insurance policy against the irreversible loss of genetic material. Mutation is viewed as a background operator to maintain genetic diversity in the population. It introduces new genetic structures in the population by randomly modifying some of its building blocks. Mutation helps escape from local minima's trap and maintains diversity in the population.

There are many different forms of mutation for the different kinds of representation. For binary representation, a simple mutation can consist in inverting the value of each gene with a small probability. The probability is usually taken about 1/L, where L is the length of the chromosome. It is also possible to implement kind of hill-climbing mutation operators that do mutation only if it improves the quality of the solution. Such an operator can accelerate the search. But care should be taken, because it might also reduce the diversity in the population and makes the algorithm converge toward some local optima.

Partial Shuffle Mutation (PSM) [4] changes part of the order of the genes in the genotype. Inversion Mutation [8] selects two positions within a chromosome/tour and then inverts the substring between these two positions. Exchange mutation by [9] selects two positions at random and swaps the cities on these positions.

Displacement Mutation [10] takes a sub tour at random and inserts it in a random. Insertion Mutation [11] selects a city at random and inserts it in a random position. Insertion can be viewed as a special case of displacement in which the substring contains only one city. Heuristic mutation by [12] was designed with neighborhood technique in order to produce an improved offspring. Greedy Swap Mutation (GSM) [13] also selects better result and therefore comes up closer to the solution quickly and Reverse Sequence Mutation (RSM).

Here two most known existing mutation operators are explained below. In addition to these the proposed mutation operators will be presented. The two most known existing mutation operators are Partial Shuffle Mutation (PSM) and Reverse Sequence Mutation (RSM) [4]. The proposed mutation operator is Hybridizing PSM and RSM Mutation operator (HPRM).

4.1 The most known existing mutation operators

a- Reverse Sequence Mutation (RSM)

In the reverse sequence mutation operator, we take a sequence S limited by two positions ***i*** and ***j*** randomly chosen, such that ***i<j***. The gene order in this sequence

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

will be reversed by the same way as what has been covered in the previous operation. The algorithm (Fig. 5) shows the implementation of this mutation operator.

Table 1: Mutation operator RSM

		*			*								
Parent	1	2	3	4	5	6	→ Child	1	5	4	3	2	6

Input: **Parents** $x_1=[x_{1,1},x_{1,2},\ldots\ldots,x_{1,n}]$
Output: **Children** $x_1=[x_{1,1},x_{1,2},\ldots\ldots,x_{1,n}]$

Choose two mutation points a and b such that $1 \le a \le b \le n$;
Repeat
  Permute (x_a, x_b);
  a = a + 1;
  b = b − 1;
until a<b

Fig. 1. Algorithm of RSM operator

b- Partial Shuffle Mutation (PSM)

The Partial Transfer Shuffle (PSM) as its name suggests, change part of the order of the genes in the genotype. The algorithm (Fig. 6) describes in detail the stages of change.

Input: **Parents** $x=[x_1,x_2,\ldots\ldots,x_n]$and ***Pm*** is Mutation probability
Output: **Children** $x=[x_1,x_2,\ldots\ldots,x_n]$
---
i = 1;
Repeat
  Choose **p** a random number between ***0*** and ***1***
  if p < ***Pm*** **then**
    Choose ***j*** a random number between ***1*** and ***n;***
      Permute (x_i, x_j);
  End if
Until $i \le n$

Fig. 2. Algorithm of Mutation operator PSM

4.2 Hybridizing PSM and RSM Operator (HPRM)

The Hybridizing PSM and RSM Operator (HPRM) constructs an offspring from a pair of parents by hybridizing two mutation operators, PSM and RSM. The algorithm (Fig. 3) presents the stages of change in our proposed operator.

Input: **Parents** $x=[x_1,x_2,\ldots\ldots,x_n]$and ***Pm*** is Mutation probability
Output: **Children** $x=[x_1,x_2,\ldots\ldots,x_n]$

Choose two mutation points a and b such that $1 \le a \le b \le n$;
Repeat
  Permute (x_a, x_b);
  Choose **p** a random number between ***0*** and ***1***
  if p < ***Pm*** **then**
    Choose ***j*** a random number between ***1*** and ***n;***
      Permute (x_a, x_j);
  End if
  a = a + 1;
  b = b − 1;
Until a < b

Fig. 3. Algorithm of the proposed mutation operator, HPRM

5. Result and discussion

In this section the results of the proposed operator HPRM and the best mutation operators, RSM and PSM, are compared and discussed. To resolve a real Travelling Salesman Problem, we use the test problem BERLIN52 to 52 locations in the city of Berlin (Fig. 4). The only optimization criterion is the distance to complete the journey. The optimal solution to this problem is known, it's 7542 m (Fig. 5).

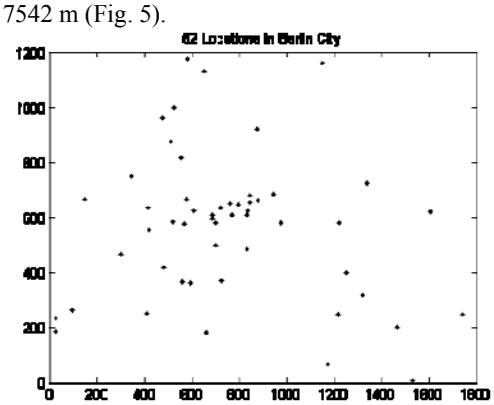


Fig. 4. The 52 locations in the Berlin city

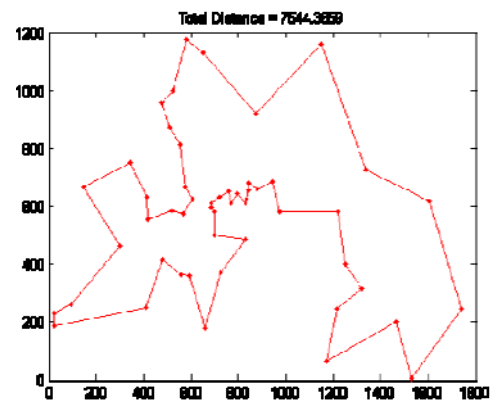


Fig. 5. The optimal solution of Berlin52

We change at a time one parameter and we set the others and we execute the genetic algorithm fifty times. The programming was done in *C++* on a PC machine with *Core2Quad 2.4GHz* in ***CPU*** and *2GB* in ***RAM*** with a *CentOS 5.5 Linux* as an ***operating system***.

To compare statistically the three operators, HPRM, RSM and PSM, these are tested one by one on 50 different initial populations after that those populations are reused for each operator. In the case of the comparison of crossover operators, the evolutionary algorithm is presented in Figure 6 which the operator of variation is given by the crossover algorithm OX [18] and the selection is made by Roulette for choosing the shortest route.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Generate the initial population P_0
i = 0
Repeat
P'_i = Variation (P_i);
Evaluate (P'_i);
P_{i+1} = Selection ([P'_i, P_i]);
Until i<Itr

Fig.6. Evolutionary algorithm

Figure 7 shows the statistics of the experiments relating to the operators of mutation. The obtained results indicate the performance of the HPRM operator to found the better minimal values than the remaining mutations used in this work.

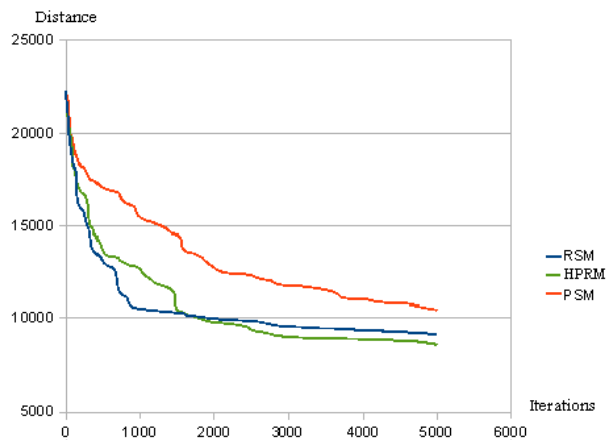


Fig. 7. Convergence comparison of the mutations in Berlin52

6. Conclusion

In this paper a new mutation operators known as HPRM (Hybridizing PSM and RSM) is presented. The HPRM operator is a combination of two mutation operators, Partial Shuffle Mutation and Reverse Sequence Mutation The performance of these mutation operators is compared with four existing mutations. For this comparison berlin52, the instance of TSPLIB, is used. The obtained results indicate the performance of the HPRM operator to found the better minimal values than the remaining mutations used in this work.
Based on this study, it is expected that in future the HPRM mutation operator shows a great potential for future research using in other NP-complete problem.

References

[1] H. Zakir Ahmed, "Genetic Algorithm for the Traveling Salesman Problem Using Sequential Constructive Crossover Operator", Journal of Biometrics & Bioinformatics (IJBB) Vol. 3 - Issue (6) ,2010, pp.96-105.
[2] D .Goldberg, "Web Courses", http://www.engr.uiuc.edu/OCEE 2000.
[3] S. Bourazza "Variantes d'algorithmes génétiques appliquées aux problèmes d'ordonnancement". PhD Thesis, Havre University, France. p. 33-34 (2008).
[4] O. Abdoun, J. Abouchabaka, and C. Tajani, "Analyzing the Performance of Mutation Operators to Solve the Travelling Salesman Problem", in Wotic, 2011!;
[5] P. Fischer. Computationally Hard Problems. IMM, DTU 2006.
[6] T. H. Cormen, C. E. Leiserson, R. L. Rivest and C. Stein, Introduction to Algorithms, 3rd edition, MIT press , 2010.
[7] D. Kusum and M. Hadush "Combined Mutation Operators of Genetic Algorithm for the Travelling Salesman problem", IJCOPI Vol. 2, No. 3, 2011, pp. 1-23 .
[8] D. Fogel, "A parallel processing approach to a multiple travelling salesman problem using evolutionary programming", Proceedings of the Fourth annual Symposium on Parallel Processing, (Fullerton, California), 1990, pp. 318–326.
[9] W. Banzhaf, "The molecular traveling salesman, Biological Cybernetics", 1990, Vol. 64, 7-14.
[10] Z. Michalewicz, "Genetic algorithms + data structures = evolution programs". Berlin: Springer. 1992.
[11]D. B. Fogel, "An evolutionary approach to the travelling salesman problem", Biological Cybernetics, 1988, Vol. 60, No. 2 139-144 .
[12]M. Gen, and R. Cheng, "Genetic Algorithm and Engineering design", John wiley and sons, New York, 1997, 118-127.
[13]S. J. Louis, and R. Tang, "Interactive genetic algorithms for the traveling salesman problem". In Proceedings of the 1999 genetic and evolutionary computing conference (GECCO), 1999, pp. 385-392.
[14]A;Christian, and O. H. Mikael. "A Multi-Agent Approach to Solving NP-Complete Problems". Masters Thesis, Technical University of Denmark, 2008, pp. 1-5.
[15]S.M. Soak, and B. H. Ahn. "New Genetic Crossover Operator for the TSP". ICAISC, 2004, pp.480-485.
[16]D. Kusum, and M. Hadush , "A. New Variations of Order Crossover for Travelling Salesman Problem". IJCOBI, Vol. 2, No. 1, 2011, pp:2-13.
[17]D.E.Goldberg, "Genetic Algorithms in Search. Optimization and Machine Learning". Addison-Wesley 1989.
[18]O. Abdoun O. and J. Abouchabaka, "A Comparative Study of Adaptive Crossover Operators for Genetic Algorithms to Resolve the Traveling Salesman Problem". IJCA, Vol. 31, No. 11, 2011.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Toward an example-based machine translation from written text to ASL using virtual agent animation

Mehrez Boulares[1] and Mohamed Jemni[2]

Research Lab. UTIC, University of Tunis
5, Avenue Taha Hussein, B. P. : 56, Bab Menara, 1008 Tunis, Tunisia

Abstract

Modern computational linguistic software cannot produce important aspects of sign language translation. Using some researches we deduce that the majority of automatic sign language translation systems ignore many aspects when they generate animation; therefore the interpretation lost the truth information meaning. This problem is due to sign language consideration as a derivative language [9, 15], but it is a complete language with its own unique grammar [1, 2]. This grammar is related to semantic-cognitive [9, 18] models of spatially, time, action and facial expression to represent complex information to make sign interpretation more efficiently, smooth, expressive and natural-looking human gestures. All this aspects give us useful insights into the design principles that have evolved in natural communication between people.

In this work we are interested in American Sign Language, because it is the simplest and most standardized sign language [2]. Our goals are: to translate written text from any language to ASL animation; to model maximum raw information using machine learning and computational techniques; and to produce a more adapted and expressive form to natural looking and understandable ASL animations.

Our methods include linguistic annotation of initial text and semantic orientation to generate the facial expression. We use genetic algorithms [6, 14, 19] coupled to learning/recognized systems to produce the most natural form. To detect emotion we based on fuzzy logic [20] to produce the degree of interpolation between facial expressions. Roughly, we present a new expressive language Text Adapted Sign Modeling Language TASML that describes all maximum aspects related to a good sign language interpretation.

This paper is organized as follow: the next section is devoted to present the comprehension effect of using Space/Time/SVO form in ASL animation based on experimentation. In section 3, we describe our technical considerations. We present the general approach we adopted to develop our tool in section 4. Finally, we give some perspectives and future works.

Keywords: *American Sign Language, Animation, Natural Language Generation, Accessibility Technology for the Deaf, biological algorithms, facial expression, emotion, machine learning, Fuzzy logic.*

1. Introduction

According to many studies, reading levels of hearing impaired is lower than the reading level of hearing student. Moreover, in 1996 Marschark and Harris confirmed that their learning progress is extremely slow. Furthermore, the gain of experience collected by deaf children in four years is equivalent to the gain of one year for hearing children.

Many deaf people in the world have difficulties with reading and writing; in fact some of them do not read. A text-based system would make it impossible for the people to follow and understand a conversation in real-time. If spoken language is rendered as text, all the information related to the semantic-cognitive meaning, time, space, emotion, intensity of the signs and facial expression is lost [2, 11].

Using ICT we can improve the classic pedagogical methods by new e-learning methods [7, 8, 12] based on multi-media contents [10, 16] and sign language specificities for a better discourse interpretation. Research in computational linguistics, computer graphics and machine learning systems has led to the development of increasingly sophisticated system based on virtual character over the past few years, bringing new perspective to machine translation research [3].

As the primary communication means used by members of deaf community, sign language is not a derivative language. It is a complete language with its own unique grammar [1, 13]. But the majority of automatic interpretation system ignores this important aspect therefore the interpretation lost the truth information meaning. In our study, we considered the American Sign Language as a target and any textual language as an input. We use the Google translator to translate all input language text to be entered in English before starting.

In this paper, we present a new approach developed in our Research Laboratory which can provide benefits for deaf people with low English literacy to improve their social integration and communication. As a first step we focus on Sign Language specificities to generate automatically the text description to more adaptive form represented in new descriptive language Text Adaptive Sign Modeling Language TASML. This new approach uses mainly a web-based interpreter of sign language developed in our research laboratory and called WebSign [7]. It is a tool that permits to interpret automatically written texts in visual-gestured-spatial language using avatar technology.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

2. Comprehension effect of the use of Space/Time/Subject/Action form in sign language:

In order to determine the more adequate textual form to be interpreted in sign language. We conducted experiments with Sign Language animations in which some animations are conformed to the place time subject action form and other versions of the animations have these linguistic phenomena removed. For experimentation, we have a specialist sign language interpreter which interprets the news program and some deaf participants.
Before discussing this latest study, we will first briefly survey some earlier experiments about general specificities of sign language. The methods and results of this earlier study served as a foundation for the design of our most recent study.

2.1 Earlier work on sign language specificities:

In order for us to conduct studies in which we focus on specific linguistic phenomena, we first needed to ensure that our animations have a baseline level of understandability. To this end, we needed to prove the existence of phenomena models representing sign language specificities. Furthermore, Patrice dale gives us an idea about signing space model for the interpretation of sign language interactions. Also the experimental study of Matt Huenerfauth [5] to improving spatially references, given us the following results:
Participants in the study viewed 16 ASL stories of two versions: “space” and “no-pace” animations. The average comprehension question scores for the space animation are greater than no space animations. This explains the importance of spatial reference in discourse interpretation. As other aspects we find speed/timing [4] configuration and iconicity in sign language. Model of the speed, timing and iconicity [2, 18] of ASL animations is very important to produce more understandable animation for better text interpretation [17].

2.2 Current experiments evaluating use of Place/Time/Subject/Action form with Emotion facial expression:

In our laboratory, we conducted a new study to determine whether there is a comprehension benefit when Sign Language animations include signs that organize the item that represent place reference at first position, time reference at second, subject at third and finally the action at fourth position.

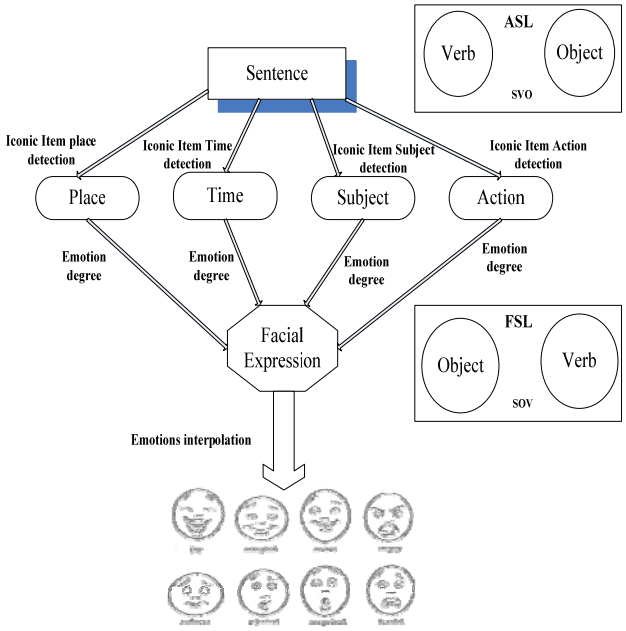


Fig. 1: Sign Language Linguistic form

We designed for explication, 7 sentences in American Sign Language and French Sign Language, and our interpreter translates these sentences to 7 deaf participants. The first version was produced in the original text order without facial animation to describe Emotion. The second version of each sentence was produced in our abstract form using facial expression to describe emotion in every sentence. If sign language target is the FSL we respect the grammar related to SOV (Subject Object Verb) [2, 15] form and if animation result is in ASL we respect the SVO (Subject Verb Object) [9, 15] form. The action items are followed or preceded by Object according to target sign language.

Original Text Order

1- The Viking period in history stretched for about three to four hundred years from 790 after Jesus-Christ to 1100 after Jesus-Christ.
2- During this period, Viking warriors raided nearby lands, explored uncharted seas, and searched for and found trade routes throughout Britain, Ireland, Southern Europe, North Africa, and Central Asia.
3- Vikings or Norsemen originated from what are now Norway, Sweden, and Denmark.
4- The geographical location of these three Norse countries determined the routes their inhabitants took to raid, explore, and trade.
5- The first recorded raids occurred in 793 after Jesus-Christ. When warriors raided a small island on the northeast coast of Britain and attacked the monastery on Lindisfarne.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

6- The last recorded battle took place in 1066 when the Norwegian king, Harald Hardrada, invaded England to claim the throne of Edward the Confessor.
7- The new Anglo-Saxon king, Harold Godwinson defeated Harald in Yorkshire on September 25.

1- The Viking period in history stretched for about three to four hundred years from 790 after Jesus-Christ to 1100 after Jesus-Christ.

ASL-SVO Order

Place: [null]
Time: [Past: for about three to four hundred years from 790 to 1100 after Jesus-Christ [**pause: 1Second]** à 1100]
Subject: [The Viking period in history] **[pause: 0.5Second]**
Verb: [to stretch]
Object: [null]
Emotion: [Surprise] [Fear] [Surprise-Fear]
Finger Spelling: Name
Speed: 1 Sign/Second

2- Au cours de cette période, les guerriers viking envahissent les terres avoisinantes, ils explorent des mers inconnues, cherchent et trouvent des routes commerciales à travers la Grande-Bretagne, Irlande, Europe du Sud, l'Afrique du Nord et en Asie centrale.

FSL-SOV Order

Lieu : [les terres avoisinantes]
Temps : [Passé : au cours de cette période] **[pause : 1Seconde]**
Sujet : [les guerriers viking]
Objet : [les terres avoisinantes] **[pause : 1Seconde]**

Verbe : [envahir]

Objet : [des mers inconnues] **[pause : 1Seconde]**
Verbe : [explorer]
Objet : [des routes commerciales **[pause : 1Seconde]** à travers la Grande- Bretagne, Irlande, Europe du Sud, l'Afrique du Nord et en Asie centrale.]
Verbe : [chercher et trouver]
Emotion: [Colère] [surprise] [Colère-surprise]
Finger Spelling : Noms propres
Vitesse : 1 Signe/Seconde

Fig. 2: Transcription of three different versions of the same sentences

Fig. 2 contains transcripts of three different versions of same sentences here we put only some sentences to just explain the output form to be interpreted. In original text we use only the input text order to be interpreted in this way. Respectively in SVO and SOV form the interpreter translates the sentences according to the pivot language related to each sign language with emotion, sign speed [4], and finger spelling for nouns. According to Matt Huenerfauth study of the calculating linguistically motivated durations for the signs in an ASL animation, we put some pauses between signs and sign speed is parameterized at 1 Sign/Second. Our sign language expert interprets first the original text and then each one of 7 participants describes in his own way what he has understood. We repeat this operation for ASL and FSL [2] form next we compare the results. We used the two sign language ASL and FSL, to prove that our form is not related to some specific sign language but is a more general approach

As results of the study, Figure 3, 4 and 5 displays the average comprehension scores for participants in the study for the Pivot/Emotion vs. no- Pivot/Emotion. As shown in Fig. there was a statistically significant difference in Understandability scores between the Abstract/Emotion order and no-Abstract/Emotion version of the animations with considering some other parameters like pauses delay and sign speed [4]. Figures 3, 4 and 5 shows the participants average interpretation scores for the 1-to-7 participant scale.

Each participant interprets the sentence in his own way and our expert gives us an approximate percentage value in understandability/Grammaticality/Naturalness [2]. The differences between the Pivot/Emotion and no-Pivot/Emotion version is very clear. Participants give higher scores for the Pivot grammatical order, understandability, and naturalness of this structure according to the animations. The comprehension benefit observed in this study is related to the use of the more appropriate text structure and emotion when translating into sign language which facilitates the participants to understanding information in the story;

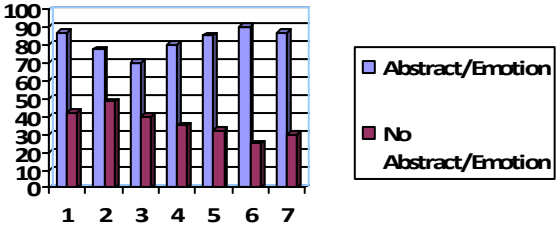


Fig. 3: Understandability

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

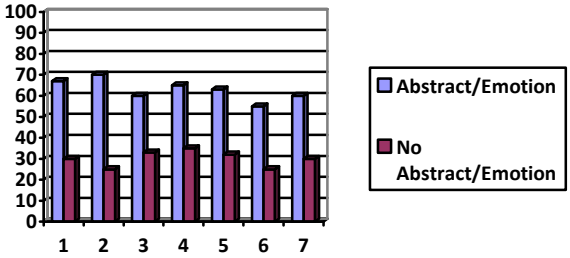


Fig. 4: Grammaticality

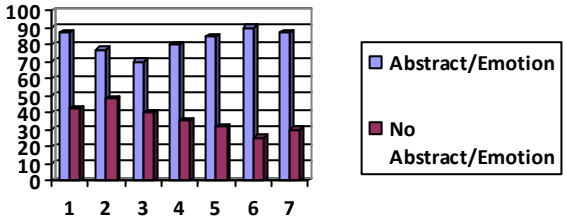


Fig. 5: Naturalness

3. Technical considerations

The objective of our project is to develop a Web based interpreter of Sign Language (SL). This tool would enable people who do not know SL to communicate with deaf individuals. Therefore, contribute in reducing the language barrier between deaf and hearing people.

Our secondary objective is to distribute this tool on a non-profit basis to educators, students, users, and researchers and to encourage its wide use by different communities. The developing team has adopted Web 3D technologies [7] which have surpassed traditional approaches such as video and still images. We use the Sign Modeling Language SML as transcript language to animate our 3D humanoid interpreter.

SML is based on XML structure to describe animation such as X3D format [7]. Actually our system is based on automatic translation from the written text without considering the sign language specificities such as the cognitive form according to the sign construction, facial expression, timing and speed parameters. In this work we focus on the transformation of initial text to more adapted form related to target sign language. Furthermore, to aim more comprehensive animation produced by our system and to represent a general text form whatever the initial text language and the target sign language which can be used by other transcription systems.

3.1 Translation system architecture

Our transcription system architecture is based on sign modeling language data base. The system sends the input text as a request to Google translator service. This service give' us English translation to be an input to the Text Adapted Sign Modeling Language module. After this operation we obtain a text form adapted to the sign language specificities and which represent an input to WebSign kernel [7]. Finally this engine generates the sign modeling language SML [7] according to this text form to be interpreted by our virtual character (Figure 6).
At first, the system support all input text language according to Google translator but the result will be in American Sign Language ASL of course the system can be parameterized to support other sign language

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

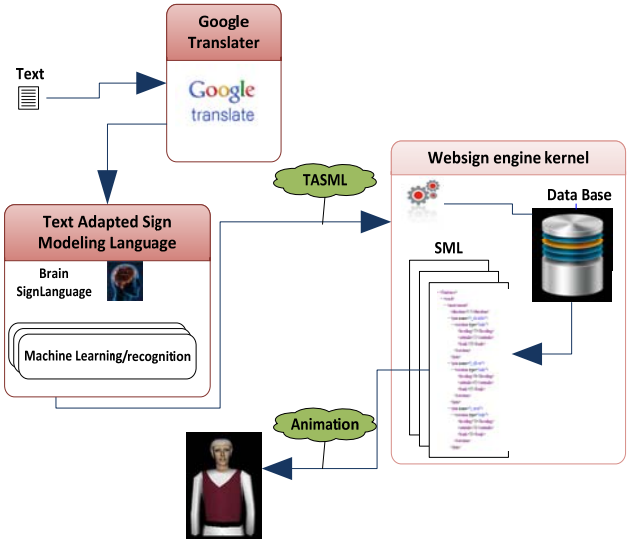


Fig. 6: Translation system architecture

4. Our approach

Based on machine learning systems, biological algorithms and fuzzy logic concept. We have built algorithms for generating the different elements that constitute our adapted text form and algorithms for emotion detection

4.1 Text abstract form generator module Acknowledgments

Our solution is based on Example-Based machine learning system associated to genetic algorithms. The aim is to use the power of genetic algorithms for detecting different sequences existing in the initial sentence. Furthermore, we proceed to fetch if the input sentence was learned by our system or no. If our system doesn't recognize this sentence, the proximity search is launched to get the better sentence proximity.

According to the threshold, the system makes decision about the degree of correctness recognized form. If this part of decision is not enough the template (Example) detection module takes over to detect different part that are used to learn our system.

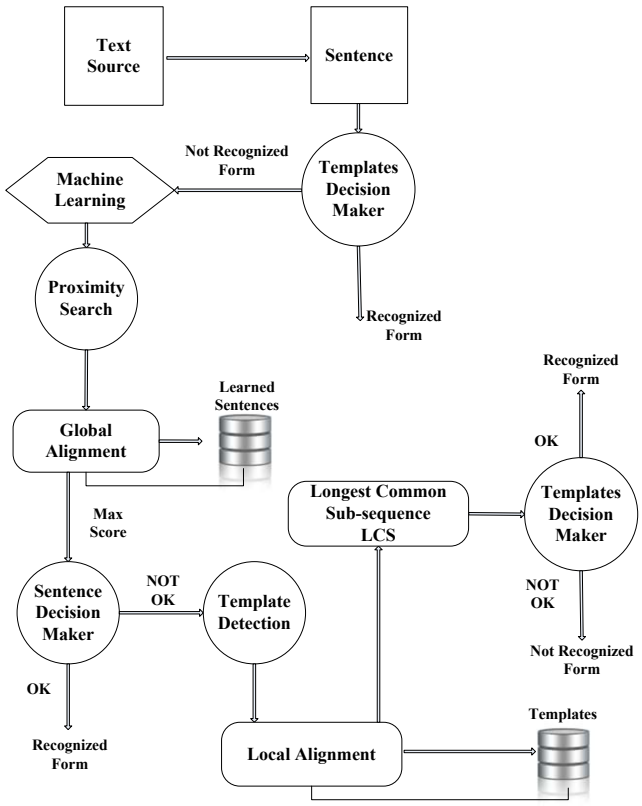


Fig. 7: Global system architecture

4.1.1 Global proximity search

We have implemented an algorithm for detecting the global proximity between the input sentence and the learned sentences. This part of system allows us to detect the more appropriate form according to input sentence. However, Bioinformatics is the solution for calculating the proximity degree between different sequences.

The genetic algorithms are created for comparing DNA and protein sequences and exploit it to calculate the global proximity degree between character sequences. The idea is to use the power of these algorithms through the search heuristic that is routinely used to generate useful solutions to optimization and search problems.

They are many algorithms for comparing different sequence, the better one is The Needleman–Wunsch algorithm [14] performs a global alignment on two sequences. It is commonly used in bioinformatics to align protein or nucleotide sequences. This algorithm belongs to the dynamic programming for solving complex problems by breaking them down into simpler sub problems, and

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

was the first application of dynamic programming to biological sequence comparison.

We start first by splitting the paragraph into sentences, next we proceed to a global alignment between each learned sentence and the first sentence. After this step we calculate the max proximity score between all obtained score. The algorithm is described in Figure 8.

Algorithm: Proximity-Search

F: Matrix [length (SEQ1) +1, length (SEQ2) +1]

Basis:

F0j = d * j

Fi0 = d * i

S (Ai, Bi) is the similarity of characters Ai and Bi

Recursion, based on the principle of optimality:

Fij = max (Fi-1, j-1 + S (Ai, Bi), Fi, j-1 + d, Fi-1, j + d)

Score = F [length (SEQ1), length (SEQ2)]

Compare (Max (scorei), threshold)

If SentenceDecisionMaker = OK then

Return TargetForm

Return LearningStep

Fig. 8: Proximity Search Pseudo Code

In order to perform a Needleman-Wunsch alignment, a matrix is created which allows us to compare the two sequences. The score as determined through use of the above calculation is placed in the corresponding cell. This algorithm performs alignments with a time complexity of O (mn) and a space complexity of O (mn) [14].

4.1.2 Local template (Example) proximity search

Our algorithm tries to search first if the input sentence was learned or no. If the system doesn't recognize automatically the input sentence, the algorithm builds a global alignment to determine the closest sequence, if this sequence is very similar according to the threshold, the system gives similar output related to the founded sentence and this is how the system learns the new sentences.

Now, if the sentence is not recognized, also our algorithm try automatically to fetch all different templates (Examples) related to the abstract (Pivot) form to be clustered. We recall that our abstract form is composed by all the elements related to the subject place indication at first position, at second position we find the time parameter, at third position the subject takes his place in abstract form and finally the action can begin and all elements according to the action are clustered in this position.

In this step, the objective is to learn system how gives us the desired target form. However, we use the local proximity search and consequently the local alignment algorithm. This approach is related to the more appropriate algorithm between local and global alignment according to the context.

The global alignment considers all elements of a sequence. If the lengths of sequences are different, insertions / deletions are introduced in the shortest sequence for the two sequences are aligned along their entire length.

This algorithm is most useful when the sequences in the query set are similar and of roughly equal size. We use it at first step because we just need a few characters differences between the initial sentence and the learned sentence.

However, the local alignment looks between the two sequences to align the sequence fragments that align best. This kind of alignments is more useful for dissimilar sequences that are suspected to contain regions of similarity or similar sequence motifs within their larger sequence context.

The Smith-Waterman algorithm [19] is a general local alignment method also based on dynamic programming. However, is a well-known algorithm for performing local sequence alignment; this algorithm gives very good results when we need to determining similar regions between two nucleotide or protein sequences.

In our context we use this algorithm to determining the different elements that can be clustered according to proximity between the learned cluster (Example) and the input sentence. Furthermore, Instead of looking at the total sequence, the Smith–Waterman algorithm compares segments of all possible lengths and optimizes the similarity measure.

Using this approach, our recognizer requires a set of templates for matching the input sentence with all learned models. At the end of this operation we obtain a new sentence conformed to the place, time, subject, action form and this sentence will be a template for the next input text.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Algorithm: Pivot form recognition

S: Input sentence

WS: {$word_1$, $word_2$,…,$word_n$}, word ∈ S

TS : {$Template_1$, $Template_2$,…,$Template_n$}

T : Template, T ∈ TS

WT: {$word_1$, $word_2$,…,$word_n$}, word ∈ T

LS: Learned Sentence

For Template in TS

For wordT in WT (Template)

For wordS in WS

Sc = **Smith-Wat(**wordT, wordS) **Max(Sc)**

end.

$\sum$(Sc) **Max$\sum$(Sc)**

end.

end.

$$\mathbf{LS = Max\ (\sum Smith\text{-}Waterman\ (}WT, WS\mathbf{))}$$

If length (LCS (SL, S)) ≥ length(S) – threshold Then

Return True

Return False

Fig. 9: Pseudo code abstract (Pivot) form recognition

Figure 9 shows the pseudo code abstract form recognition from the input sentence using all learned templates. For each template word a local alignment applied with every word of the initial sentence. Then the system tries to find the maximum score alignment according to all proximity search operation. For each template word a local alignment applied with every word of the initial sentence. Then the system tries to find the maximum score alignment according to all proximity search operation. After this step we will have a set of score which represents the proximity between all words in the two sentences.

In order to output the best form related to the initial sentence, our algorithm repeats this operation and compares the added score of all template and choose the highest score. Therefore, the system gives us the most appropriate form to our initial sentence but this may be a wrong result. In order to ensure the learning process automatically, we implements an algorithm that finds the longest common subsequence LCS [6] between the source sentence and the result.

This step gives us very good results by comparing the LCS length and input sentence length to make decision to conserve this result or no. Certainly this phase requires a long execution time for all new sentences but after the recognition process of many sentences the system response time improves considerably due to the learning already done by machine learning system.

The LCS computation complexity, is evidently in O (m*n) time and space related to the matrix size for two sequences. The global alignment Needleman-Wunsch complexity is in O (m*n) time and pace also the local alignment Smith-Waterman is in O (m*n) [6].

4.1.3 Automatic detection of simple/compound emotion

Our approach in generating abstract (pivot) form related to the place, time, subject, action order detects also the different elements that reference to the emotion in sentence. But sometimes we find many emotions like surprised, feared and other in the same sentence. When the system interprets the text automatically, we lost the exact meaning related to the emotion and this is a big problem.

In order to find solution, we are based on the observation of the natural deaf interpretation for this phenomenon. With collaboration of our sign language specialist, we deduced that the sign language interpreter read first the text, he understand it and he translate the global text meaning. Therefore, when we focus on his interpretation we presume that some emotions are mixed automatically to produce the truth meaning.

According to studies of psychiatrist Jacques Cosnier expert in human and animal emotion a basic emotion classification is built figure.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

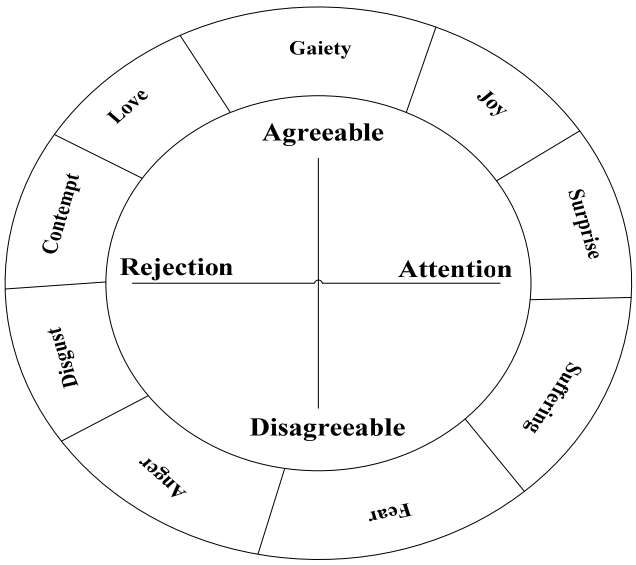


Fig. 10: Organizing axes emotions

This figure shows that there is some overlap between the basic emotions, for example it makes sense to find the emotion of joy and surprise, surprise and suffering, suffering and fear, fear and anger, anger and disgust, disgust and contempt, contempt and love, love and gaiety, gaiety and joy.

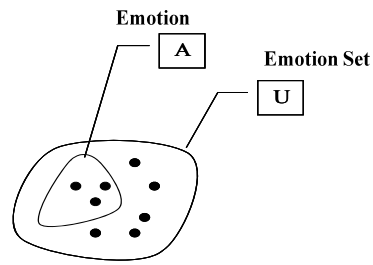
Therefore, this classification gives us the most appropriate emotion related to the context because in most of cases emotions are confused and mixed. We can find a mix between two or even three emotion in a single facial expression which gives us more precision.

For generating automatically the mixed emotions, we need a mathematical model to represent the desired results. For this reason we focus on the fuzzy logic which represents as best as possible our problem. Fuzzy logic emerged as a consequence of the 1965 proposal of fuzzy set theory by Lotfi Zadeh [20] though fuzzy logic has been applied to many fields, from control theory to artificial intelligence.

Fuzzy logic is a form of many-valued logic derived from fuzzy set theory to deal with reasoning that is fluid or approximate rather than fixed and exact. In our context we need to represent fuzzy information's related to the emotion classification. Furthermore, to make possible to detect one or more emotions in the same sentence. Consequently to make decision and to produce automatically the appropriate emotion.

In contrast with "classical logic", where binary sets have two-valued logic, if we use this binary logic the system produce wrong results according to the comprehension meaning. Fuzzy logic variables may have a truth value that ranges in degree between 0 and 1. In other meaning, fuzzy logic is a superset of conventional (Boolean) logic that has been extended to handle the concept of partial truth, where the truth value may range between completely true and completely false.

Consequently, we can represent the result with the emotion overlap criterion. In this theory we use the linguistic variables which are ideally suited to our problem. Furthermore, the problem may be managed by specific functions.



Classical set theory:

if μA is the membership function of the Emotion A set

$\forall x \in U$, $\mu A(x) = \text{Value1}$ if $x \notin$ Emotion A

$\mu A(x) = \text{Value2}$ if $x \in$ Emotion A

$x \in$ Value1 **or** Value2

If $\mu A(x) = 0.4$

x : is the interpolation value between two emotions

Value $_n$: Emotion value

$x \in$ Value1

$x \notin$ Value2

Fuzzy set concept:

if μA is the membership function of the Emotion A set

$\forall x \in U$, $\mu A(x) \in [\text{Value1},\text{Value2}]$

If $\mu A(x) = 0.4$

x : is the interpolation value between two emotions

Value $_n$: Emotion value

$x \in$ Value1 **and** Value2

$x \in$ Value1 at 40% and $\in$ Value2 at 60%

Fig. 11: Emotion fuzzy concept

This approach has yielded good results which will be useful in systems that use interpolation between the different facial expression models to produce emotion. For example: sentence which contains suffering and fear. Suffering intensity value detected from sentence is bigger than fear intensity value:

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

X ∈ suffering at 70% and ∈ fear at 30%, the result is an interpolation value between suffering and fear emotion which can be parameterized according to the target system.

5 Conclusion and future works

This research has identified a solution to improve the comprehension of Sign Language animations when using the automatic text interpretation into sign language. By modifying some parameters that have received little systematic research attention before: reorganizing the input text to an appropriate more adapted form based on surrounding sign language factors and automatic detection of simple and compound emotion.

This project is classified among the first studies to prove that deaf perceives a big difference in quality of Sign Language animations based on this specific form and automatic compound emotion detection from text. To confirm the theoretical approach of this contribution, we based on experimental support to shows that deaf perceive a difference in understandability, grammaticality and naturalness according to these specific form and emotion interpolation factors.

While this study focused on French Sign Language and American Sign Language, these techniques should be applicable to animations of other sign languages around the world. However, we just need to train system with a desired corpus language which contains the output templates (Examples) and sentences. Now that this study has shown that Sign Language animations can be made more understandable, our approach can be used in a real-system setting as part of a WebSign tool [7], we are going to integrate this system in WebSign to improve the animation quality.

In future work, we will focus on other sign language specificities to be learned by our system. We will study the space factor related to the objects location in discussion. The automatic management of the sign area is very important factor to give us a more significant animation closest to the natural animation. The promising initial results of this study open the question of whether there may be standardization of sign language when using a collaborative approach to generates all sign language corpus to be learned by this system. Therefore, an investigation approach based on data meaning can be applied to gives us the more appropriate standardized sign language.

References

[1] Bouchard, D. 1996. Sign Languages & Language Universals: The Status of Order & Position in Grammar. Sign Language Studies. Sign Language Studies, v91 p101-60.
[2] Diane, B. 2010. Sign languages. Cambridge University Press. 20:100.
[3] Elliott, R., Glauert, J., Kennaway, J., Marshall, I., Safar, E. 2008. Linguistic modeling and language-processing technologies for avatar-based sign language presentation. Univ Access Inf Soc, 6:4.
[4] Huenerfauth, M. 2008. Evaluation of a Psycholinguistically Motivated Timing Model for Animations of American Sign Language. In Proc. of ACM SIGACCESS Conf. on Comp. and Accessibility (Assets '08).
[5] Huenerfauth, M. 2009 Improving Spatial Reference in American Sign Language Animation through Data Collection from Native ASL Signers. In Universal Access in HCI, Part III, HCII 2009, LNCS 5616, pp. 530–539, 2009. Springer-Verlag Berlin Heidelberg.
[6] Hirschberg, D. S. 1977. Algorithms for the Longest Common Subsequence Problem. In Journal of the Association for Computing Machinery, Vol 24, No 4, October 1977.
[7] Jemni, M., Elghoul, O. 2008. A System to Make Signs Using Collaborative Approach. In ICCHP 2008 Proceedings, Linz, Austria. Lecture Notes in Computer Science 5105 Springer 2008 pp. 670-677.
[8] Karpouzis, K., Caridakis, G., Fotinea, S.-E. and Efthimiou, E. 2007. Educational Resources and Implementation of a Greek Sign Language Synthesis Architecture. Computers and Education, v49 n1 p54-74.
[9] Karen, E. 2002.Language, Cognition, and the Brain Insights from Sign Language Research 2002. By Lawrence Erlbaum Associates, Inc.
[10] Kervajan, L., Breton, G. 2007. Du texte en français à l'avatar signant. TALN 2007, Toulouse, 12–15 juin 2007.
[11] Loïc, K. 2006. Problèmes de représentation de la Langue des Signes Française en vue du Traitement Automatique, TALN 2006, Leuven, 10–13 avril 2006.
[12] Liwei, Z., Karin, K., William, S.,Christian V., Norman, B. and Martha, P.. 2000. A Machine Translation System from English to American Sign Language. Lecture Notes in Computer Science, 2000, Volume 1934/2000, 191-193, DOI= http://doi.acm.org/10.1007/3-540-39965-8_6.
[13] Michael, F. 2010. Stabilisation d'un modèle à contraintes pour la représentation lexicale des langues des signes. TALS 2010, Montréal, 23 juillet 2010.
[14] Needleman, S. B., Wunsch , C. D. 1969. A general method applicable to the search for similarities in the amino acid sequence of two proteins. Reprinted from J. Mol. Biol. (1981) 48, 443-453.
[15] Olivier, DE L., Pierre, G., Henri, P., Christian, R. A propos des structures OSV en langue des signes française.
[16] Pengfei, L, Huenerfauth, M. 2010. Collecting a Motion-Capture Corpus of American Sign Language for Data-Driven Generation Research. In First Workshop on Speech and Language Processing for Assistive Technologies SLPAT.
[17] Sylvie, G, Alexis, H. 2007. Formalisme de description des gestes de la langue des signes française pour la génération du

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

mouvement de signeurs virtuels. revue Traitement Automatique des Langues 48, 3 (2007) 115-149.
[18] Sarah, F.2004. Language from the body Iconicity and Metaphor in American Sign Language. Cambridge University Press 2004. 19:152
[19] SMITE, T. F., WATERMAN, M. S. 1981. Identification of Common Molecular Subsequences. Reprinted from J. Mol. Biol. (1981) 147, 195-197.
[20] Zadeh, L.A., 1965. Fuzzy sets. From information and control 8, 338-353 (1965).

Mehrez Boulares is currently a PhD student under the supervision of Prof. Mohamed Jemni. He received in September 2009 the Master degree on Computer Science from Tunis College of Sciences and Techniques (ESSTT), University of Tunis in Tunisia. His research interests are in the areas of Sign Language Processing. His current topics of interests include Computer graphics and Accessibility of ICT to Persons with Disabilities.

Mohamed Jemni is a Professor of ICT and Educational Technologies at the University of Tunis, Tunisia. He is the Head of the Laboratory Research of Technologies of Information and Communication (UTIC). Since August 2008, he is the General chair of the Computing Center El Khawarizmi, the internet services provider for the sector of the higher education and scientific research. His Research Projects Involvement are tools and environments of e-learning, Accessibility of ICT to Persons with Disabilities and Parallel & Grid Computing.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Mercator as a web crawler

Priyanka-Saxena[1]
[1] Department of Computer Science Engineering, Shobhit University,
Meerut, Uttar Pradesh-250001, India

Abstract

The Mercator describes, as a scalable, extensible web crawler written entirely in Java. In term of Scalable, web crawlers must be scalable and it is important component of many web services, but their design is not well-documented in the literature. In this paper, we enumerate the major components of any scalable web crawler, comment on alternatives and tradeoffs in their design, and describe the particular components used in Mercator. We also describe Mercator's support for extensibility and customizability. Finally, we comment on Mercator's performance, which we have found to be more efficient and comparable to that of other crawlers.

Keywords: *Introduction, Related Work, Architecture, Components, Extensibility, Conclusion.*

1. Introduction

Designing a scalable web crawler comparable to general crawler used by the major search engines is a complex endeavor. Due to the competitive nature of the search engine business, there are few papers in the literature form for describing the challenges and tradeoffs of inherent web crawler design. This paper describes, Mercator as a scalable, extensible web crawler written entirely in Java for filling the gap between the challenges comparable to that of other crawlers.

By *scalable is define as*, Mercator is designed to scale up to the entire web, and used for fetching millions (tens of) of web documents. We achieve scalability by implementing our data structures with bounded amount of memory and regardless for the size of the crawl. Vast majority of our data structures are stored on disk, and small parts of them are stored in memory for efficiency.

By *extensible is define as*, Mercator is designed in a modular way, with the expectation that new functionality will be added by third parties. In practice, it has been used to collect a variety of statistics about the web, and to perform a series of random walks of the web.

The initial motivations is to collect statistics about the web such as the size and the evolution of the URL space, the distribution of web servers over top-level domains, the lifetime and change rate of documents, and so on. However, it is hard to know *a* exact priority for which statistics are interesting, and the topics of interest may change overtime. Mercator makes it easy to collect new statistics—to configure for different crawling tasks—by allowing users to provide their own modules for processing downloaded documents.

The remainder of the paper is structured as follows. The next section surveys related work. Section 3 describes the main components of a scalable web crawler, the alternatives and tradeoffs in their design, and the particular choices we made in Mercator. Section 4 describes Mercator's support for extensibility.

2. Related work

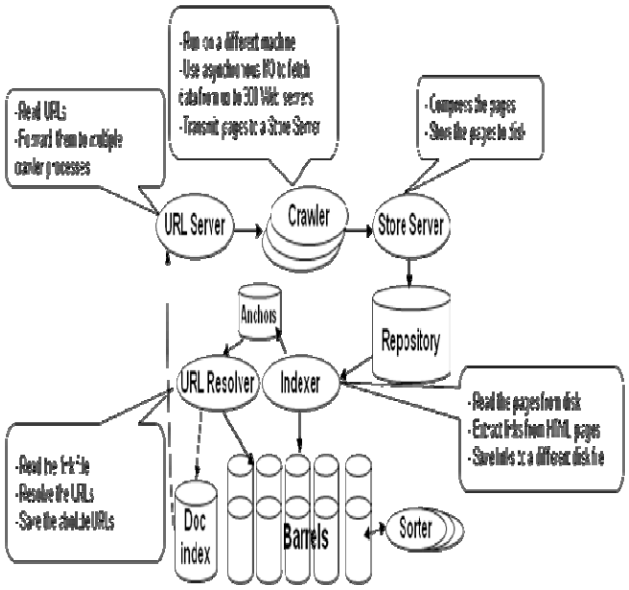
Web crawlers— are almost as old as the web itself. Several papers about web crawling were
presented at the first two World Wide Web conferences. However, time to time, the web was considered challenges like two to three orders of magnitude smaller. Today, those systems did not address the scaling problems inherent in a web crawler. Obviously, all of the popular search engines use crawlers that must scale up to substantial portions of the web such as two notable exceptions: the Google crawler and the Internet Archive crawler.

The Google search engine is a distributed system that uses multiple machines for crawling. The crawler consists of five functional components running in different processes.

- A *URL server process* reads URLs out of a file and forwards them to multiple crawler processes.
- Each *crawler process* runs on a different machine, is single-threaded, and uses asynchronous I/O to fetch data from up to 300 web servers in parallel.
- The crawlers transmit downloaded pages to a single *StoreServer process*, which compresses the pages and stores them to disk.
- The pages are then read back from disk by an *indexer process*, which extracts links from HTML pages and saves them to a different disk file.
- A *URL resolver process* reads the link file, derelativizes the URLs contained therein, and saves the

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

absolute URLs to the disk file that is read by the URL server.



The Internet Archive also uses multiple machines to crawl the web and each crawler process is assigned up to 64 sites to crawl, but no site is assigned to more than one crawler. Each single-threaded crawler process reads a list of seed URLs for its assigned sites from disk into per-site queues, and then uses asynchronous I/O to fetch pages from these queues in parallel. Once a page is downloaded, the crawler extracts the links contained in it. If a link refers to the site of the page it was contained in, it is added to the appropriate site queue; otherwise it is logged to disk. Periodically, a batch process merges these logged "cross-site" URLs into the site-specific seed sets, filtering out duplicates in the process.

In the area of extensible web crawlers, SPHINX system provides some of the same customizability features as Mercator. It provides a mechanism for limiting which pages are crawled, and it allows customized document processing code to be written. However, SPHINX is targeted towards site-specific crawling, and therefore is not designed to be scalable.

3. Architecture of a scalable web crawler

The basic algorithm executed by any web crawler takes a list of *seed URLs* as a input and repeatedly execute the following steps. Remove a URL from the URL list; and determine the IP address of its Host Name, download the corresponding document, and extract any links contained in document. For each of the extracted links, ensure that it is an absolute URL, and add URL to the list of URLs to download, provided it has not been encountered before.

This basic algorithm requires a number of functional components:

_ a component (called the *URL frontier*) for storing the list of URLs to download;

_ a component for resolving host names into IP addresses;

_ a component for downloading documents using the HTTP protocol;

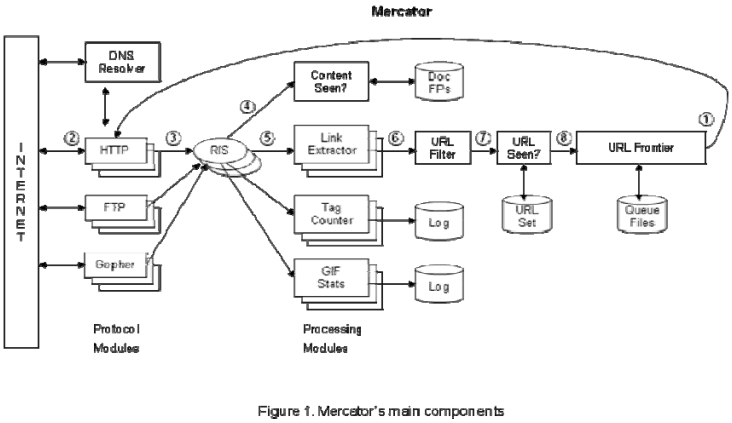


Fig. 1. Mercator's Main components

4. Mercator's components

Mercator's main components are described in figure 1. The Web Crawling is performed by multiple worker threads and each worker repeatedly performs the steps needed to download and process a document. In the above figure 1, Mercator's components working shown in the form of steps are as follows:-

(1) The first step is to remove an absolute URL from the shared URL frontier for downloading.

An absolute URL begins with a *scheme* (e.g., "http"), which identifies the network protocol that are implemented by *protocol modules*. The protocol modules is to be used in a crawl are specified in a user-supplied configuration file which is dynamically laded at the start of the crawl, and there is a separate instance of each protocol module per thread, which allows each thread to access local data without any synchronization.

(2) The second step is based on the URL's scheme, the worker selects the appropriate protocol module for downloading the document and the protocol module's *fetch* method, which downloads the document from the Internet.

(3) The third step is that the downloaded document from the internet after fetching by protocol module is write into a per-thread *RewindInputStream* (RIS) . A RIS is an I/O abstraction that is initialized from an arbitrary input stream, and that subsequently al-

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

lows that stream's contents to be re-read multiple times.(ADVANTAGE OF MERCATOR)

(4) After successful completion of step third then the forth step is the worker thread invokes the *content-seen test* to determine whether this document (associated with a different URL) has been seen before . If document is not processed any further, and the worker thread removes the next URL from the frontier.

Every downloaded document has an associated MIME type and a Mercator configuration file also associates MIME types with one or more *processing modules* is an abstraction for processing downloaded documents, for instance extracting links from HTML pages, counting the tags found in HTML pages, or collecting statistics about GIF images. Like protocol modules, there is a separate instance of each processing module per thread.

(5) Based on the downloaded document's MIME type, the fifth step of the worker invokes the *process* method of each processing module associated with that MIME type.

Note:-The Link Extractor and Tag Counter processing modules in Figure 1 are used for text/html documents, and the GIF Stats module is used for image/gif documents.

(6) The *process* method of the processing module extracts all links from an HTML page and each link is converted into an absolute URL, and the sixth step is tested against a user-supplied *URL filter* to determine if it should be downloaded.(ADVANTAGE OF MERCATOR)

(7) If the URL passes from the URL filter, the seventh step is the worker performs the *URL-seen test* , which checks if the URL has been seen before.

(8) if it is in the URL frontier or has already been downloaded. If the URL is new, it is added to the frontier.

4.1 The url frontier

The URL frontier is the data structure that contains all the URLs that remain to be downloaded. Most crawlers work by performing a breath-first traversal of the web, starting from the pages in the seed set and this type of traversals are easily implemented by using a FIFO queue.

In a standard FIFO queue, elements are dequeued in the order they were enqueued. In the context of web crawling matters are complicated by the multiple HTTP requests pending to the same server. If multiple requests are in parallel, the queue's *remove* operation should not simply return the head of the queue, but rather than the URL close to the head whose host has no outstanding request.

To implement this politeness constraint, the default version of Mercator's URL frontier is implemented by a collection of distinct FIFO sub queues. First, there is one FIFO sub queue per worker thread. Second, when a new URL is added, the FIFO sub queue in which it is placed is determined by the URL's canonical host name. This design prevents Mercator from overloading a web server, also handle a bottleneck of the crawl.(ADVANTAGE OF MERCATOR).

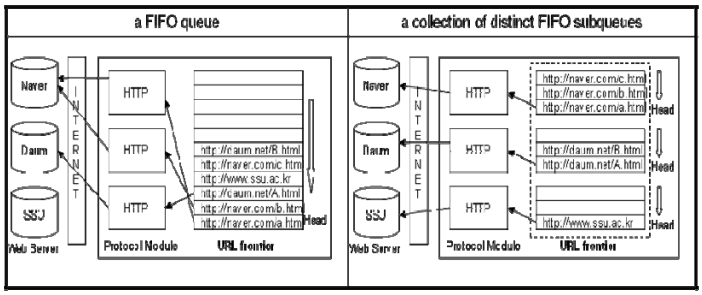


Fig. 2. Data Structure of URL Frontier

4.2 The http protocol module

The purpose of a protocol module is to fetch the document corresponding to a given URL using the appropriate network protocol which is supported by Mercator include HTTP, FTP, and Gopher.

The Mercator implement the Robots Exclusion Protocol, which allows web masters to declare the wed crawler to fetch a special document containing these declarations from a web site before downloading any real content from it. To avoid downloading the Robot Exclusion File(Robot.txt) file on every request, Mercator's HTTP protocol module maintains a fixed-sized cache mapping host names to their robots exclusion rules and due to this by default, the cache is limited to 218 entries, and uses an LRU replacement strategy.

Mercator uses its own "lean and mean" HTTP protocol module; its requests time out after 1 minute, and it has minimal synchronization and allocation overhead.

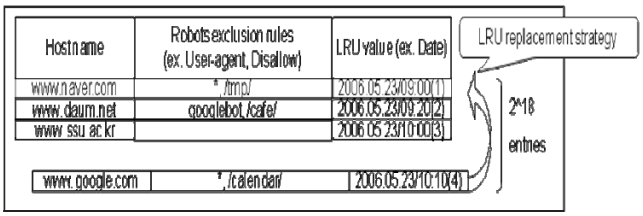


Fig. 3. The HTTP Protocol Module

4.3 Rewind input stream

Mercator's design allows the same document to be processed by multiple processing modules. To avoid reading a document over the network multiple times, Mercator cache the document locally using an abstraction called a *RewindInputStream* (RIS).

A RIS is an input stream with an *open* method that reads and caches the entire contents of a supplied input stream

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

(such as the input stream associated with a socket). A RIS caches small documents (64 KB or less) entirely in memory, while larger documents are temporarily written to a backing file(limit 1 MB). RIS also provides a method for rewinding its position to the beginning of the stream, and various lexing methods that make it easy to build MIME-type-specific parsers.

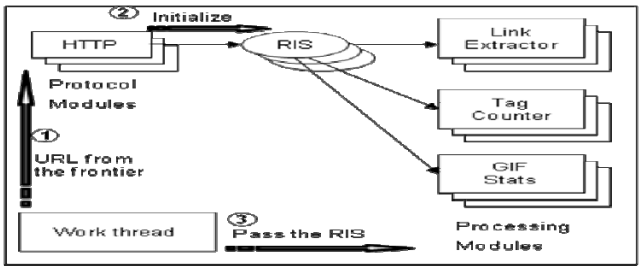


Fig. 4. The Rewind Input Stream

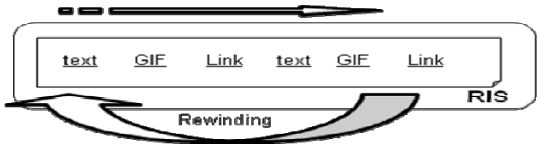


Fig. 5. The Rewind Process

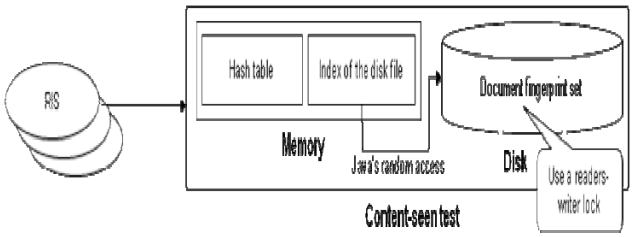
4.4 Content seen test

In web, many documents are available under multiple times with different URLs. There are also many cases in which documents are mirrored on multiple servers show in Fig. 6.

These effects will cause any web crawler to download the same document contents multiple times. To prevent processing a document more than once, a web crawler may wish to perform a *content-seen test* to decide if the document has already been processed. Mercator using a content-seen test makes it possible to suppress link extraction from mirrored pages, which also offers the side benefit of allowing us to keep statistics about the fraction of downloaded documents that are duplicates of pages that have already been downloaded.

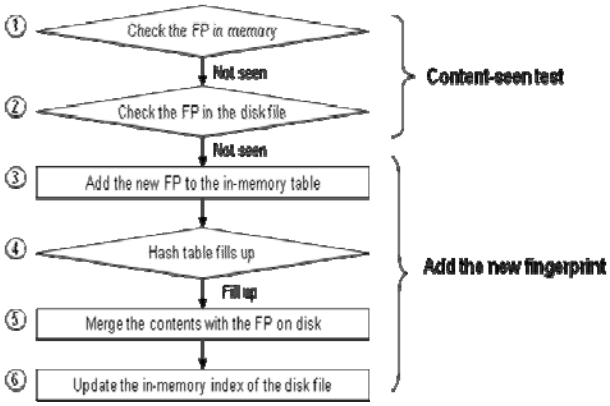
To save space and time, Mercator uses data structure called the *document fingerprint set* that stores a 64-bit checksum of the contents of each downloaded document and also compute the checksum using fingerprinting algorithm. Fingerprints offer provably strong probabilistic guarantees that two different strings will not have the same fingerprint.

Mercator maintains two independent sets of fingerprints: a small hash table kept in memory, and a large sorted list kept in a single disk file. Fig. 7. Content-Seen Test



The content-seen test first checks if the fingerprint is contained in the in-memory table. If not, it has to check if the fingerprint resides in the disk file. To avoid multiple disk seeks and reads per disk search.

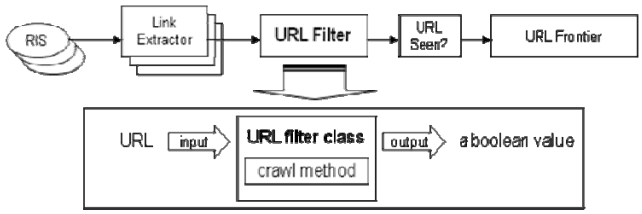
Mercator performs an interpolated binary search of an in-memory index of the disk file to identify the disk block on which the fingerprint would reside if it were present. It then searches the appropriate disk block, again using interpolated binary search. Fig. 8. Steps of Content-Seen Test



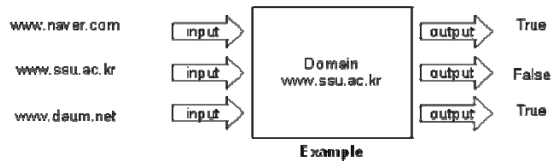
4.5 The url filtering

The URL filtering mechanism provides a customizable way to control the set of URLs that are downloaded.

The URL filter class has a single crawl method that takes a URL and returns a boolean value indicating whether or not to crawl that URL. Mercator includes a collection of different URL filter subclasses that provide facilities for restricting URLs by domain, prefix, or protocol type, and for computing the conjunction, disjunction, or negation of other filters. Fig. 9. The URL Filter with example

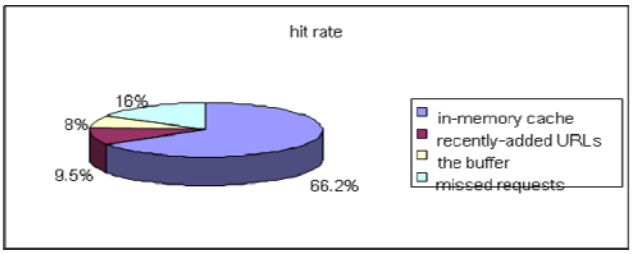


Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

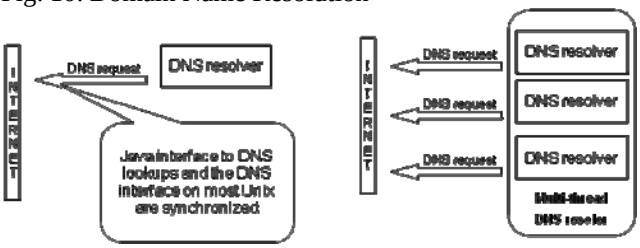


4.6 Domain name service

Before using a web server, a web crawler must use the Domain Name Service (DNS) to map the web server's host name into an IP address. DNS name resolution is a well-documented to avoid bottleneck of most web crawlers.



Mercator tried to alleviate the DNS bottleneck by caching DNS results, but that was only partially effective and also used its own multi-threaded DNS Resolver can resolve host names much more rapidly than either the Java or Unix resolvers. This meant that only one DNS request on an un cached name could be outstanding at once. The cache miss rate is high enough that this limitation causes a bottleneck. Fig. 10. Domain Name Resolution



Perform DNS looking accounted for 87% of each thread elapsed time and reduce that elapsed time to 25%

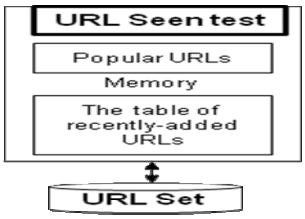
4.7 The url seen test

Any web crawler will encounter multiple links to the same document. To avoid downloading and processing a document multiple times, a URL-seen test must be performed on each extracted link before adding it to the URL frontier to perform the URL-seen test. All of the URLs seen by Mercator in canonical form in a large table called the *URL set*.
To save space, Mercator does not store the textual representation of each URL in the URL set, but rather a fixed sized checksum. To reduce the number of operations on the backing disk file, Mercator therefore keep an in memory cache of popular URLs.

Unlike the fingerprints, the stream of URLs has a non-trivial amount of locality (URL locality). So, each URL set membership test induces one-sixth as many kernel calls as a membership test on the document fingerprint set. Host name locality arises because many links found in web pages are to different documents on the same server. To preserve the locality, we compute the checksum of a URL by merging two independent fingerprints:

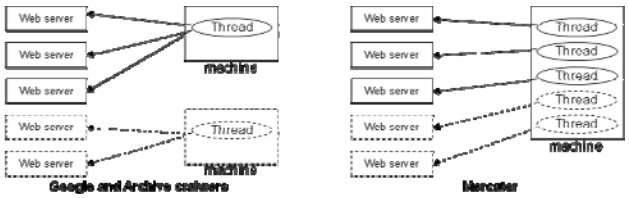
1. The fingerprint of the URL's host name
2. The fingerprint of the complete URL

These two fingerprints are merged so that the high-order bits of the checksum derive from the host name fingerprint. As a result, checksums for URLs with the same host component are numerically close together. So, the host name locality in the stream of URLs translates into access locality on the URL set's backing disk file, thereby allowing the kernel's file system buffers to service read requests from memory more often. On extended crawls, this technique results in a significant reduction in disk load, and hence, in a significant performance improvement. Figure 11.Using an in-memory cache of 2^18 entries and the LRU-like clock replacement policy and Fig. 11. URL SEEN TEST



4.8 Synchronous and asynchronous i/o

Both Google and Internet Archive crawlers use single-threaded crawling processes and asynchronous I/O to perform multiple download in parallel and they are designed from the ground up to scale to multiple machines. Whereas, Mercator uses a multi-threaded process in which each thread performs synchronous I/O (It leads to a much simpler program structure **is** the main advantage of Mercator) and it would not be too difficult to adapt Mercator to run on multiple machines. One strength of the Google and the Internet Archive crawlers is that they are designed from the ground up to scale to multiple machines. Fig.12. Synchronous vs. asynchronous I/O



Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

4.9 Checkpointing

To complete a crawl of the entire web, Mercator writes regular snapshots of its state to disk. An interrupted or aborted crawl can easily be restarted from the latest checkpoint. Mercator's core classes and all user-supplied modules are required to implement the check pointing interface. User-supplied protocol or processing modules are also required to implement the check pointing interface.
Checkpoints are coordinated using a global readers-writer lock. Each worker thread acquires a read
share of the lock while processing a downloaded document. Once a day, Mercator's main thread acquires the write lock, so it is guaranteed to be running in isolation and also it acquired the lock, the main thread arranges for the checkpoint methods to be called on Mercator's core.

5. Extensibility

Mercator is an extensible crawler it means two things.

1. First, Mercator can be extended with new functionality.
2. Second, Mercator can easily be reconfigured to use different versions of most of its major components.

Different versions of the URL frontier, document fingerprint set, URL filter, and URL set may be all are dynamically "plugged into" the crawler means multiple versions of each of these components, which we employ for different crawling tasks.
Making an extensible system, Mercator requires three ingredients:

1. The interface to each of the system's components must be well-specified is defined by an *abstract* class.
2. A mechanism must exist for specifying how the crawler is to be configured from its various components by supplying a *configuration file* which specifies which additional protocol and processing modules should be used, as well as the concrete implementation to use for each of the crawler's "pluggable" components.
3. Sufficient infrastructure must exist to make it easy to write new components such as rich set of utility libraries with a set of existing pluggable components.

To demonstrate Mercator's extensibility, here we describe some of the extensions.

5.1 Protocol and processing module

By default, Mercator will crawl the web by fetching documents using the HTTP protocol, extracting links from documents of type text/html. To fetch documents using additional protocols or to process the documents once they are fetched, new protocol and processing modules must be supplied.
The abstract Protocol class includes two methods.

1. The fetch method downloads the document corresponding to a given URL, and
2. the new URL method parses a given string, returning a structured URL object.

The abstract Analyzer class is the superclass for all processing modules and defines a single process method which is responsible for reading the document and processing it appropriately. Analyzers often keep private state or write data to the disk.
Other processing modules simply write the contents of each downloaded document to disk. As another experiment, WebLinter processing module that runs the weblint program on each downloaded HTML page to
check it for errors, logging all discovered errors to a file.

5.2 Alternative url frontier and implementation

We described one implementation of the URL frontier data structure and that implementation on a crawl of our corporate intranet, with the drawback is that multiple hosts might be assigned to the same worker thread, while other threads were left idle. This situation is occur on an intranet because intranets typically contain a substantially smaller number of hosts than the internet at large.
To restore the parallelism, an alternative URL frontier component that dynamically assigns hosts to worker threads and that at most one worker thread will download documents from any given web server at once. It also maximizes the number of busy worker threads within the limits and all worker threads will be busy so long as the number of different hosts in the frontier is at least the number of worker threads.

5.3 Configuring mercator as random walker

Mercator used to perform random walks of the web in order to gather a sample of web to measure the quality of search engines. A random walk starts at a random page taken from a set of seeds. The next page to fetch is selected by choosing a random link from the current page. The process continues until it arrives at a page with no links, at which time the walk is restarted from a new random seed page. The seed set is dynamically extended by the newly discovered pages, and cycles are broken by performing random restarts.
Performing a random walk of the web is quite different from an ordinary crawl for two reasons. First, a page may be revisited multiple times during the course of a random walk. Second, only one link is followed each time a page is visited.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

6. Conclusions

Scalable web crawlers are an important component of many web services, but they have not been very well documented in the literature. Building a scalable crawler is a non-trivial endeavor because the data manipulated by the crawler is too big to fit entirely in memory, so there are performance issues relating to how to balance the use of disk and memory. This paper has enumerated the main components required in any scalable crawler, and it has discussed design alternatives for those components.

The paper described Mercator; an extensible, scalable crawler is design with the features a crawler core for handling the main crawling tasks, and extensibility through protocol and processing modules. Users may supply new modules for performing customized crawling tasks. We have used Mercator for a variety of purposes, including performing random walks on the web crawling our corporate intranet, and collecting statistics about the web at large. Mercator's scalability design has worked well. It is easy to configure the crawler for varying memory footprints. The ability to configure Mercator for a wide variety of hardware platforms makes it possible to select the most cost-effective platform for any given crawling task.

Mercator's extensibility features have also been successful and able to adapt Mercator to a variety of crawling tasks. Mercator is scheduled to be included in the next version of the *AltaVista Search Intranet* product is mostly to corporate clients who use it to crawl and index their intranets.

7. References

[1] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual Web search engine. In *Proceedings of the Seventh International World Wide Web Conference*, April 1998

[2] Google! Search Engine. http://google.stanford.edu/.

[3] Monika Henzinger, Allan Heydon, Michael Mitzenmacher, and Marc A. Najork. Measuring Index Quality using Random Walks on the Web. In *Proceedings of the Eighth International World Wide Web Conference*, May 1999.

[4] The Internet Archive. http://www.archive.org/,

[5] Z. Smith. The Truth About the Web: Crawling towards Eternity. *Web Techniques Magazine*, 2(5), May 1997.

[6] The Web Robots Pages. http://info.webcrawler.com/mak/projects/robots/robots.html.

[7] The Robots Exclusion Protocol. http://info.webcrawler.com/mak/projects/robots/exclusion.html,

[8] M. O. Rabin. Fingerprinting by Random Polynomials. Report TR-15-81, Center for Research in Computing Technology, Harvard University, 1981.

[9] Oliver A. McBryan. GENVL and WWWW: Tools for Taming the Web. In *Proceedings of the First International World Wide Web Conference*, pages 79–90, 1994.

[10] Robert C. Miller and Krishna Bharat. SPHINX: A framework for creating personal, site-specific Web crawlers. In *Proceedings of the Seventh International World Wide Web Conference*, April 1998.

First Author- I am Priyanka Saxena, presuing M.Tech in Computer Science Engineering from Shobhit University, Meerut Uttar Pradesh, INDIA and got 8 cgpa in every semester. My Thesis is left, which is hopefully completed up to August 2012. I have completed my B.Tech in Computer Science from Uttar Pradesh Technical University in 2009 with 72%. I have one year gap between my studies in which I was complete my Hardware Support Training form HP Delhi with 84% and A+ grade. It is my first paper on Web Crawler and I will find the details of IJCSI through Internet. I am heartily thank full to IJCSI for giving me this big opportunity to show my skills and knowledge which definitely generate positive energy for my next paper.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Noncooperative Iris Segmentation

PROF.Dr. Elsayed Mostafa[1], Dr. Maher Mansour[2] and Eng. Heba Saad[3]

[1]Faculty of engineering in Helwan, Helwan university
Cairo, Egypt

[2]Faculty of engineering in Helwan, Helwan university
Cairo, Egypt

[3]Shorouk academy
Cairo, Egypt

Abstract

In noncooperative iris recognition one should deal with uncontrolled behavior of the subject as well as uncontrolled lighting conditions. That means eyelids and eyelashes occlusion, non uniform intensities, reflections, imperfect focus, and orientation among the others are to be considered. To cope with this situation a noncooperative iris segmentation algorithm based on numerically stable direct least squares fitting of ellipses model and modified Chan-Vese model (local binary fitting energy) with variational level set formulation is to be proposed. The proposed algorithm is tested using CASIA-IrisV3.

Keywords: *Noncooperative Iris Segmentation, Active Contour Model, Modified Chan-Vese Model, Variational Level Set Formulation.*

1. Introduction

The biometric technology deals with recognizing the identity of individuals based on their unique physical or behavioral characteristics. Physical characteristics as finger print, palm print, hand geometry, and iris patterns and behavioral characteristics as typing pattern, gait, voice, and hand-written signature. The biometrics are used in applications where it is required to establish or confirm the identity of individuals as passenger control in airports, access control in restricted area, border control, database access, and financial services.

The iris pattern has several advantages as stability over time, statistically unique, well protected from the external environment, and impossibility of surgically modifying the iris without unacceptable risk to vision.

The typical stages of iris recognition begin with the iris image acquisition. After that the segmentation of the region correspondent to the iris in detecting the iris inner and outer boundaries at the pupil and sclera, detecting the upper and lower eyelid boundaries if they occlude, and detecting and excluding any superimposed eyelashes or reflections. The iris images are highly affected by their distance and angular position with respect to the camera; moreover, the illumination has a direct impact on the pupil size and causes non linear variations of the iris patterns. To compensate these variations, the iris image is transformed (normalized) into a polar coordinate system where fixed-size rectangular block is produced. The feature extraction process produces a set of numeric values known as the iris signature or the iris template. The comparison between the acquired and the enrolled templates (the enrolled template is stored in database previously once) produces a dissimilarity value used to conclude the identity of the subject.

The emerging needs for a safer and quicker access (building, weapons, and restricted areas) requires noncooperative recognition methods, in which the user has no active participation in the image capture process and is not even aware of the recognition method and also dealing with uncontrolled lighting conditions. This means that we process images with heterogeneous characteristics.

Some segmentation methods [1-3] assume that the inner and the outer boundaries and eyelid boundaries are circular or elliptical. This is a source of error, since the iris boundaries are not exactly circles or ellipses even in cooperative recognition. Daugman [1] applied an integro-differential operator which acts as a circular edge detector to detect the inner and outer boundaries of the iris, while

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

the upper and lower eyelids are also detected using the integro-differential operator by adjusting the contour search from circular to a designed arc. Wildes [2] used an edge detector and a circular Hough transform-based method to segment the iris, while the upper and lower boundaries of the eyelid are approximated using parabolic curves. In [3] noncooperative segmentation method is used where the process begins with the image-feature extraction in which three discrete values are extracted for each image pixel, followed by the application of a clustering algorithm that will classify each pixel and produce the intermediate image. This intermediate image is then used by circular Hough transform-based method.

Recently, several researchers proposed different noncooperative iris segmentation methods based on active contours models [4, 5, 7, 8, 10]. Active contour models are curves that evolve within digital images to recover object boundaries (whatever their shapes are) under the influence of both internal and external forces and user defined constraints. These algorithms have been at the heart of one of the most active and successful research areas in edge detection, image segmentation, shape modeling, and visual tracking. Active contour models are broadly classified as either parametric active contour models (Snakes) or geometric active contour models (level sets). Geometric active contour models have several important advantages over parametric models such as computational simplicity and the ability to change the contour topology during the evolution automatically (e.g., splitting and merging) without requiring an elaborate mechanism to handle such changes. However, this advantage of geometric active contour models turns out to be a liability in applications where the object to be segmented has a known topology that must be preserved.

Geometric active contour models are divided into edge and region based active contour models (the Snake is edge based active contour model). Edge based active contour models depend on large image gradient to stop the curve evolution on the object boundary. Region based active contour models depend on partitioning the image domain by fitting statistical models as the intensity, color, texture, or motion in each of a set of regions and therefore have better performance for the image with weak object boundaries.

In [4, 5, 7] edge based active contour models are used. Daugman [4] used Snake model for detection of the inner and outer boundaries of the iris and to correct the off-axis images, the images are converted into orthographic form by using Fourier-based method and used statistical methods for detecting the eyelashes. In [5], the inner boundary is detected by smoothing the eye image using a 2-D median filter and binarization the smoothed image by certain threshold and then 2-D median filter is applied on the binary image to discard the relatively smaller regions. A circle-fitting procedure is executed on all detected regions. Finally, the circle whose circumference contains the maximum number of black pixels is deemed to be the detected pupil. If reflection is detected in the vicinity of the pupil, it is inpainted. The outer boundary of the iris is detected by level set representation of the geodesic active contour model where the pupil boundary is used to initialize the contour. The contour evolves around a small narrow band [6] to accelerate the evolution process. Important feature of GACs is their ability to handle "splitting and merging" boundaries. This is especially important in the case of iris segmentation (the Snake model has not this feature). But if the iris boundary is weak, the contour evolution may not stop at the desired iris boundary leading to an oversegmentation of the iris. In [7], a level set method with area preserving has been used for determining inner and outer iris boundaries, both performed in one step. The contour initialization is determined by one dimensional derivation in the right horizontal axes from the approximate center of inner boundary of the iris (the approximate center is calculated by intersection of the minimum vertical and horizontal histograms). The one step segmentation approach improves the speed of the whole process in comparison with two–step boundary detection methods. The main difference between [7] and [5] is using an area preserving which make this method is robust in case of the blurred images. But the method which is used in calculating the approximate center of inner boundary of the iris does not give right results in some images because of the noise.

In [8, 10] region based active contour models are used. In [8], a two stages iris segmentation process is proposed. In the first stage, approximation of the inner and outer boundaries of the iris using iterative elliptical fitting model. In the second stage, the approximate inner boundary is used as the initial contour and the contour is evolved by Chan-Vese model (global binary fitting energy) [9] in a narrow band of ±5 pixels. Similarly, the approximate outer boundary is used as the initial contour and the contour is evolved by Chan-Vese model in a narrow band of ±10 pixels. Using the approximate iris boundaries as the initial contours reduces the complexity of curve evolution and is suitable for real-time applications. The eyelids and the eyelashes may be present as noise, the eyelids are isolated by fitting lines to the upper and lower eyelids. A mask based on the detected eyelids and eyelashes is then used to extract the iris without noise. But using iterative algorithm for fitting an ellipse is computationally expensive, also using Chan-Vese model is not suitable for images with intensity inhomogeneity.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Chan-Vese model (region based active contour model) has been successful for images with two regions, each having a distinct mean of pixel intensity where this model assumes that the image consists of statistically homogeneous regions, with intensity in each region being a constant up to certain noise. Since the iris region in the noisy images is not statistically homogeneous, the difficulty due to image inhomogeneity is incorrectly label part of background pixels as the iris, while a significant part of the iris is missing.

The iris segmentation methods in [4, 5, 7, 8] depend on image homogeneity. In [10], also the iris segmentation process is divided into two stages. In the first stage, approximation of the inner and outer boundaries of the iris using a Direct Least Square (DLS)-based elliptical fitting model [11]. In the second stage, the approximate inner boundary is used as the initial contour and the contour is evolved by modified Chan-Vese model (local binary fitting energy) [12] with variational level set formulation [13] in a narrow band of ±10 pixels from outside the approximated inner boundary to remove the effect of reflections. Similarly, the approximate iris boundary is used as the initial contour and the contour is evolved by modified Chan-Vese model with variational level set formulation in a narrow band of ±20 pixels from inside the approximated iris boundary to reduce the effect of the eyelids and the eyelashes. The eyelashes detection technique is deployed. The advantages of [10] are using modified Chan-Vese model which extracts the intensity information in local regions at a controllable scale to solve the problem of image inhomogeneity and using variational level set formulation to speed up the curve evolution and to obtain stable evolution of the level set function. The iris segmentation methods in [5, 7, 8] consume huge computational time for using the traditional level set. But although DLS-based elliptical fitting model is the first non-iterative ellipse-specific fitting and it guarantees an ellipse-specific solution even for scattered or noisy data, in some cases produces wrong results. Our contribution is to attempt solving the previous drawbacks by using numerically stable Direct Least Square fitting of ellipses model [14] to approximate the iris boundary and using modified Chan-Vese model with variational level set formulation.

2. The Proposed Method

The iris segmentation process is divided into two stages. In the first stage we use a numerically stable DLS-based elliptical fitting model to approximate the outer boundary of the iris. In the second stage we use a modified Chan-Vese model to estimate exact outer and inner iris boundaries.

2.1 The Numerically Stable DLS-Based Elliptical Fitting Model

We use a numerically stable DLS-based elliptical fitting model (instead of DLS-based elliptical fitting model as in [10]) to approximate the outer boundary of the iris only through five parameters (p_1, p_2, r_1, r_2, θ) where

(p_1, p_2) are horizontal and vertical coordinates of the outer boundary center,

(r_1, r_2) are the length of the major and minor axes, and

(θ) is the orientation of the ellipse.

These five parameters are used to draw an ellipse which is used as initial contour.

The DLS-based elliptical fitting model is the first non-iterative ellipse-specific fitting and it guarantees an ellipse-specific solution even for scattered or noisy data, but in some cases produces wrong results. These wrong results are eliminated by using a numerically stable DLS-based elliptical fitting model as in "Fig. 1".

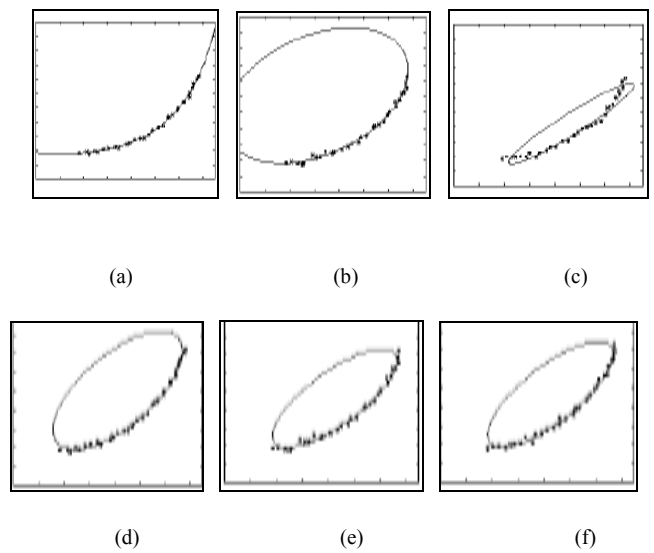


Fig. 1 (a, b, c) illustrate that DLS-based elliptical fitting model, sometimes produces wrong results, (d, e, f) illustrate that numerically DLS-based elliptical fitting model produces good results on the same data points.

2.2 The Modified Chan-Vese Model

We estimate exactly the outer and inner iris boundaries. The initial contour is evolved using a modified Chan-Vese model in a narrow band of ±20 pixels from outside the approximated outer boundary to inside where the iris region is considered one region and the rest of the image and the superimposed eyelids and eyelashes and

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

reflections are the other region. For each point x Є Ω (where Ω is the image domain), the modified Chan-Vese model (local binary fitting energy E_x^{LBF}) is as Eq. (1)

$$E_x^{LBF}(C, f_1(x), f_2(x)) = \lambda_1 \int_{in(C)} K(x-y)|I(y)-f_1(x)|^2 dy + \lambda_2 \int_{out(C)} K(x-y)|I(y)-f_2(x)|^2 dy \quad (1)$$

where

C is a contour in the image domain Ω,

$f_1(x)$ and $f_2(x)$ are the values that approximate image intensities inside and outside the closed contour C,

λ_1 and λ_2 are positive constants,

$I(y)$ is the intensity in a local region centered at the point x,

$K(x - y)$ is a weighting function with a localization property where $K(x - y)$ decreases and approaches zero as $|x - y|$ increases (I.e., at the points y away from the center point x) as Eq. (2)

$$K_\sigma(x - y) = \frac{1}{(2\pi)^{n/2}\sigma^n} e^{-|x-y|^2/2\sigma^2} \quad (2)$$

with a scale parameter σ > 0.
In level set methods, a contour C is represented by the zero level set function Φ. Thus the local binary fitting energy (LBF) in "(1)" can be written as Eq. (3)

$$E_x^{LBF}(\phi, f_1, f_2) = \lambda_1 \int K_\sigma(x-y)|I(y)-f_1(x)|^2 H(\phi(y))dy + \lambda_2 \int K_\sigma(x-y)|I(y)-f_2(x)|^2 (1-H(\phi(y))dy \quad (3)$$

where H is the Heaviside function. Thus, the energy over all the center point x in the image domain Ω is as Eq. (4)

$$E^{LBF}(\phi, f_1, f_2) = \int_\Omega E_x^{LBF} dx = \lambda_1 \int[\int K_\sigma(x-y)|I(y)-f_1(x)|^2 H(\phi(y))dy]dx + \lambda_2 \int[\int K_\sigma(x-y)|I(y)-f_2(x)|^2 (1-H(\phi(y))dy]dx \quad (4)$$

The previous energy is computed from the image data I, so it is called an external energy. Some internal energy terms (as the distance regularizing term $D(\Phi)$ and the length of the zero level contour $L(\Phi)$) are added. Then the total energy function is as Eq. (5)

$$F(\phi, f_1, f_2) = E^{LBF}(\phi, f_1, f_2) + \beta D(\phi) + \nu L(\phi) \quad (5)$$

where β and v are nonnegative constants.

The term $\beta D(\Phi)$ is to penalize the deviation of the level set function Φ from a signed distance function and therefore completely eliminates the need of the reinitialization procedure which is time consuming in traditional level set methods and therefore significantly larger time step can be used which speeds up the curve evolution and stable evolution of the level set function Φ is produced as Eq. (6)

$$D(\phi) = \int_\Omega \frac{1}{2}(|\nabla\phi(x)| - 1)^2 dx \quad (6)$$

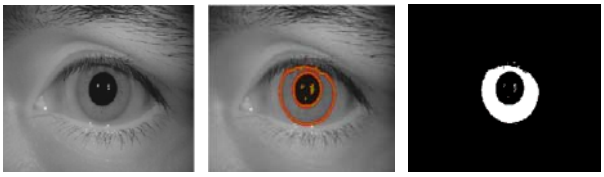
And the length of the zero level contour to regularize the contour can be defined by Eq. (7)

$$L(\phi) = \int_\Omega \delta(\phi(x))|\nabla\phi(x)|dx \quad (7)$$

(a) (b) (c)

(i)

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.



(a) (b) (c)

(ii)

Fig. 2 Illustrates the segmentation process by (i) modified Chan-Vese model (ii) Chan-Vese model. Where (a) original image, (b) the image and the contour, (c) the segmentation.

In our proposed method, we use a single contour instead of two contours [8] and [10]. For using two contours the superimposed reflections on the iris region which are not near the boundaries are not detected as in "Fig. 3".

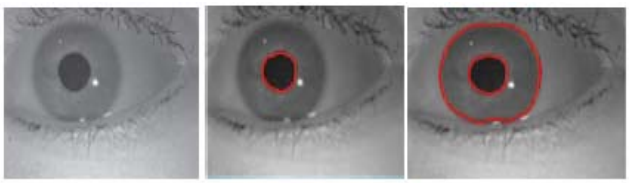


Fig. 3 Illustrates the superimposed reflections on the iris region are not detected.

While using single contour evolves from outside the approximated outer boundary to inside will reduce the effect of eyelids and eyelashes as shown in "Fig. 2". For using a single contour evolves from inside the inner boundary to outside, the contour will detect the pupil boundary and the eyelashes (which have the same intensity) while it can not detect the outer boundary.

3. Experimental Results

3.1 The Database

In our experimentation, we have used CASIA-IrisV3 database [15] which is a noisy database. CASIA-IrisV3 includes three subsets which are labeled as CASIA-IrisV3-Interval, CASIA-IrisV3-Lamp, and CASIA-IrisV3-Twins as in "Fig. 4". CASIA-IrisV3 contains a total of 22051 iris images from more than 700 subjects. All iris images are 8-bit gray level JPEG files, collected under near infrared illumination.

(a) (b) (c)

Fig. 4 (a) image from CASIA-IrisV3-Interval, (b) image from CASIA-IrisV3- Lamp, (c) image from CASIA-IrisV3- Twins.

3.2 Numerically Stable DLS

Numerically stable DLS-based elliptical fitting model is for fitting an ellipse to a set of data points. To determine the data points, we apply circular Hough transform with certain radius. Then applying numerically stable DLS on arc from the resultant circle. The arc interval (in degrees) is 0 - 30.

3.3 The curve evolution parameters

The values of the selected parameters to find the inner and outer boundaries in case of CASIA-IrisV3-Interval are $\sigma = 2$, $\beta = 0.001$, $\nu = 0.2$, $\lambda_1 = 1$, $\lambda_2 = 90$, step size = 90, and the number of iterations = 3 (the maximum number to this database). In case of CASIA- IrisV3-Lamp and CASIA-IrisV3-Twins are $\sigma = 2$, $\beta = 0.001$, $\nu = 0.2$, $\lambda_1 = 1$, $\lambda_2 = 3$, step size = 190, and the number of iterations = 2 (the maximum number).

3.4 Comparison between the proposed method using either Chan-Vese model or modified Chan-Vese model

All the algorithms were implemented in MATLAB 7.0. The 'Time' column from Table 1 contains the average execution time for each of the segmentation tested methods. These time values were obtained by averaging 100 segmentation processes on 100 distinct images.

Table 1 : Contains the average time and segmentation efficiency for each of the segmentation tested methods.

Method	***Time (s)***	***Segmentation Effeciency (%)***

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

The proposed method using Chan-Vese model	.3	96.7
The proposed method using modified Chan-Vese model	.17	99.02

4. Conclusions

The process of segmenting the iris plays a crucial role in iris recognition systems. This paper presents a noncooperative iris segmentation using numerically stable direct least squares fitting of ellipses model and using modified Chan-Vese model (region based active contour model). The irregular shape of the inner and outer boundaries of the iris was the motivation of using the active contour models where the inner and outer boundaries of the iris are not exactly circular or elliptical. Region based active contour models have better performance for the iris images with weak iris boundaries than edge based active contour models. Also, using local region based active contour model provides a better performance than global region based active contour model when the iris images suffer from intensity inhomogeneity.

References

[1] J. G. Daugman, "High confidence visual recognition of persons by a test of statistical independence," IEEE Trans. Pattern Anal. Mach. Intell., vol. 15, no. 11, pp. 1148–1160, Nov. 1993

[2] R. P. Wildes, "Iris recognition: An emerging biometric technology," Proc. IEEE, vol. 85, no. 9, pp. 1348–1363, Sep. 1997.

[3] H. Proenc¸a and L.A. Alexandre, "Iris Segmentation Methodology for Noncooperative Iris Recognition," IEE Proc. Vision, Image, and Signal Processing, vol. 153, no. 2, pp. 199-205, Apr. 2006.

[4] J. Daugman, "New methods in iris recognition," IEEE Trans. on SMC-B, vol. 37, pp. 1167-1175, 2007.

[5] S. Shah, and A. Ross, "Iris Segmentation Using Geodesic Active Contours," IEEE Trans. Info. Forensic and Security, vol. 4, no. 4, 2009.

[6] R. Malladi, J. A. Sethian, and B. C.Vemuri, "Shape modeling with front propagation: A level set approach," IEEE Trans. Pattern Anal. Mach. Intell., vol. 17, no. 2, pp. 125–133, Feb. 1995.

[7] N. Barzegar and M. S. Moin, "A New User Dependent Iris Recognition System Based on an Area Preserving Pointwise Level Set Segmentation Approach," in EURASIP Journal on Advances in Signal Processing , 2009, Article ID 980159, pp. 516–523.

[8] M. Vatsa, R. Singh, and A. Noore, "Improving iris recognition performance using segmentation, quality enhancement, match score fusion, and indexing," IEEE Trans. on SMC-B, vol. 38, pp.1021-1035, 2008.

[9] T. Chan, and L. Vese, "Active contours without edges," IEEE Trans. on Image Process., vol. 10, pp. 266-277, 2001.

[10] Kaushik Roy, Prabir Bhattacharya, Ching Y. Suen and Jane You, "RECOGNITION OF UNIDEAL IRIS IMAGES USING REGION-BASED ACTIVE CONTOUR MODEL AND GAME THEORY," Proceedings of 2010 IEEE 17th International Conference on Image Processing September 26-29, 2010, Hong Kong.

[11] Fitzgibbon, A. W. Pilu, M and Fischer, R. B.: Direct least squares fitting of ellipses. In Proc. of the 13th International Conference on Pattern Recognition, pages 253–257, Vienna, September 1996.

[12] C. Li, C. Y. Kao, J. C. Gore, and Z. Ding, "Implicit active contours driven by local binary fitting energy. Int. Conf. on Comp. Vis. and Pattern Recog., pp. 1-7, 2007.

[13] C. L., C. Xu, C. Gui, and M.D. Fox, "Level set evolution without re-initialization: A new variational formulation," in IEEE Proc. of CVPR, 2005, vol. 1, pp. 430–436.

[14] Halir, R. and Flusser, J. Numerically Stable Direct Least Squares Fitting of Ellipses. Proc. WSCG'98, pp 125-132.

[15][Online].Available:http://www.cbsr.ia.ac.cn/IrisDatabase/irisdatabase.php, January 2011.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

(IPTV) INTERNET PROTOCOL TELEVISION'S KEY BROADCAST BUILDING BLOCKS

Muhammad Asif[1], Zahid Farid[2], Muhammad Lal[3], Junaid Qayyum[4] and Fayyaz Gul[5]

[1]Department of Information Technology and Media (ITM), Mid Sweden University Sundsvall 85170, Sweden.

[2]School of Electrical and Electronics Engineering, Universiti Sains Malaysia (USM) Penang 14300, Malaysia.

[3, 4, 5]Department of Information Technology, Gandhara University of Sciences Peshawar 25000, Pakistan.

Abstract

Telcos, in pursuit of the triple play (voice, video, and data), are investing heavily to compete with the cable companies in providing quality video content to the consumer. However, most telco infrastructures do not currently have the bandwidth to support high-quality video distribution. The telcos are updating their Internet protocol (IP) networks and using new encoding schemes such as MPEG4-Part 10 (also called H.264-AVC) to deliver video content to the consumer. This creates a huge opportunity for supplying telcos with updated equipment including the latest encoders and decoders. This paper will investigate the technology that is fueling this new Internet protocol television (IPTV) infrastructure. The first portion will be looking at the video encoding method and the second portion will focus on the video-over-IP network design that is being used for IPTV.

Keywords Telco, Internet Protocol (IP), MPEG4-Part 10, H.264-AVC, Internet Protocol Television (IPTV), Encoder, Decoder.

1. Introduction

The telcos are on the offensive to gain a big piece of the video market share from the cable TV providers. Cable multiple service operators (MSOs) have made great progress in delivering a "triple play" of voice, video, and data services to the consumer in the last few years. Now the telcos are responding in a big way to provide the same triple play by offering not only voice and data, but also high-quality digital TV video via a new technology called IPTV. IPTV is an emerging technology that allows consumers to watch high-quality digital TV over the Internet via an IPTV set-top box or a PC. The traditional cable companies use a RF signal to carry the digital video by means of QAM. Technology advancements have made it possible for telcos to bring the same quality of video via the Internet. The key building blocks on the transmission side are advanced video encoding and video over IP. Advanced video encoding is the most critical building block. The availability of high-definition (HD) content along with standard definition (SD) content have created a challenge for the telcos, since telcos still rely on bandwidth-limited twisted copper pair of wires and usually do not have the luxury of cable's broadband capability. A typical HD channel requires 20 Mbps and a SD channel requires 4 Mbps. Therefore, a bandwidth-efficient video transport mechanism is needed. The H.264 format of MPEG4 Part 10 and Microsoft's VC-1 encoder can offer 2.5 to 3 times more bandwidth-efficient improvement over the popular MPEG2 encoding. Most broadcasters are adapting the H.264 standard rather than the VC-1 standard. The other building block of IPTV transmission is video over IP, which maps or bridges the encoded video data onto the Internet for delivery.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

2. Building Blocks for IPTV and Broadcasting

2.1 Broadband and Telephone Networks

IPTV is gathering momentum. One main driver is the technical developments taking place in telephone networks. That can be illustrated by the sharply increasing downstream data rates becoming available on twisted-pair copper lines *(see Fig. 1)*. In the past, with an analogue modem or narrow-band ISDN connection, the last mile to the home was the bottleneck for high bandwidth applications. DSL (Digital Subscriber Line) technology has provided the breakthrough for broadband telephone networks in recent years. With ADSL, and its variants ADSL2 and ADSL2+, downstream data rates of several Mbit/s are now common in many households. In the future, with VDSL, the data rates will extend well into the two-digit Mbit/s range.

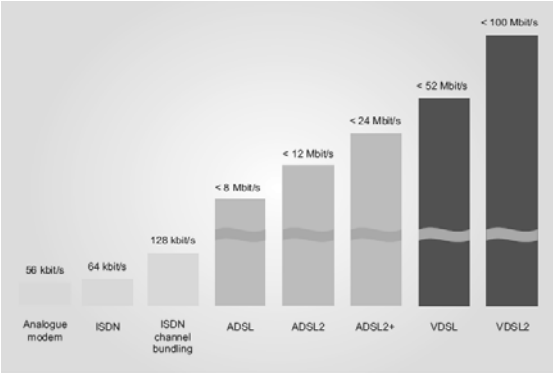


Figure 1: Downstream data rates on twisted-pair copper lines

DSL technology is based on sophisticated multi-carrier modulation schemes for physical transmission. The nature of the traffic at the consumer end suggests an asymmetric split with a broadband downstream data rate into the home and a relatively narrow-band upstream data rate from the home to the DSL provider. The multi-carrier approach of discrete multi-tone (DMT) splits the frequency band in a number of subchannels, each addressed by a carrier frequency. For example, 256 subchannels can be used in ***ADSL*** [2]. Physically, ADSL uses frequencies up to 1.1 MHz on a twisted-pair copper line. The length and the diameter of the twisted-pair copper line determine the attenuation of the ADSL signal. Line-noise and crosstalk between lines can significantly impair the signal-to-noise ratio. Under ideal conditions, the maximum downstream data rate is in the region of 8 Mbit/s and the upstream data rate is up to 1 Mbit/s. However, if the length of the line approaches the limit of about four to five kilometers, for example in rural areas, the downstream data rate may be just 1-2 Mbit/s or even lower. **ADSL2** [3] extends the maximum downstream data rate to about 12 Mbit/s. **ADSL2+** [4] doubles the upper frequency limit from 1.1 to 2.2 MHz. The maximum downstream data rate is in the region of 24 Mbit/s. **VDSL1** [5] uses frequencies up to 12 MHz on a twisted-pair copper line and **VDSL2** [6] up to 30 MHz. The resulting downstream data rates are in the region of 52 and 100 Mbit/s, respectively. However, in order to achieve the maximum data rates, the length of the line has to be reduced from several kilometers down to about 150 - 300 meters. Additionally, in order to maintain the very high data rates to and from the provider backbone, typically a ***fiber optic network*** is needed up to the point where the last mile starts with VDSL transmission. Thus, VDSL requires a major investment. The focus will be on selected urban areas to reach as many households as possible for a given capital investment and operational expenditure. At present it is not clear how fast the telephone companies are going to upgrade their networks to accommodate VDSL and for how long they will stick with ADSL and the variants thereof, for the vast majority of households.

2.2. Advanced Source Coding for Data Compression

Advanced source coding for data compression enables full-quality television on DSL networks. The transition from the ISO-MPEG-2 standard to MPEG-4 Advanced Video Coding (AVC) [7] reduces the data rates required for the picture signal. Using MPEG-4 AVC, the data rates required for high quality television in standard definition (SDTV) are in the region of 2 to 4 Mbit/s depending on the source signal. The data rates of MPEG-4 AVC streams are thus quite commensurate with the data rates available over ADSL. VDSL lends itself even to HDTV. Source coding is not performed using only the MPEG open standards. On the Web, proprietary codecs that are integrated in downloadable software players are also used. In the case of

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

IPTV, there is notably Microsoft IPTV edition as a platform with integrated data compression, based on Windows Media 9 / VC-1.

2.3. IPTV and Physical Transmission

Technically, IPTV is not a well-defined term. In a layered communication model, the Internet Protocol (IP) resides at the network layer and television (TV) at the application layer. Nothing is said about the physical transmission. Three simplified examples are shown in *Fig. 2*. After source coding, the compressed video bit-stream is packetized at the network layer with the Internet Protocol. The resulting IPTV packets are then physically transmitted with a transmission scheme that provides a guaranteed quality of service for digital television – in terms of bandwidth, error rate and delay. For example, as shown on the left in *Fig. 2*, DSL is used for physical transmission over twisted-pair copper lines. In the home, the DSL modem contains a so-called Integrated Access Device (IAD) which is connected to an IPTV set-top-box, using a cat-5 (Ethernet) cable. Alternatively, a wireless LAN (WLAN) can be used for connecting the IPTV set-top-box to the IAD.

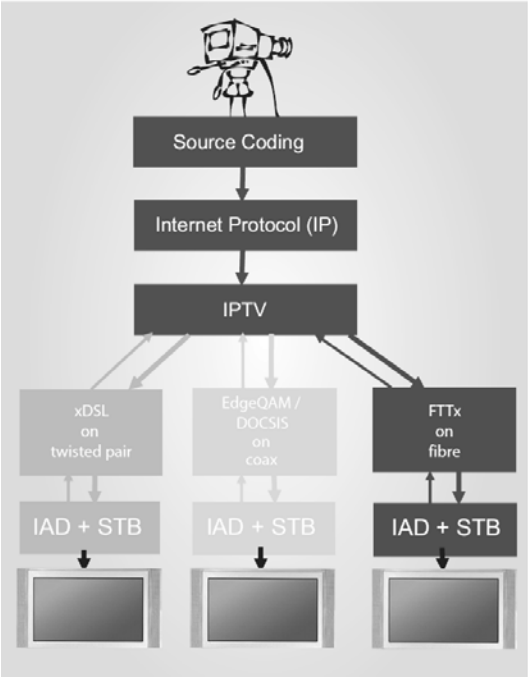


Figure 2: IPTV and physical transmission

IPTV is not bound to telephone networks and can be physically transmitted on broadband cable networks as well. One needs a transmission scheme which matches the physical properties of coax cable, e.g. EdgeQAM/ Docsis [8]. A similar reasoning applies to IPTV transmission on fiber-optic networks, e.g. fibreto- the-curb (FTTC) or fiber-tothe-home (FTTH). Actually, VDSL networks are typically a combination of FTTC to get to the point where the last mile starts and, from there, VDSL over twisted-pair copper lines to the home. For example, Deutsche Telekom is planning a fiber-optic network in Ten German Cities which, in combination with VDSL2, aims to provide downstream data rates of 25 Mbit/s over the last mile. At a national and European regulatory level, there is much debate about whether competitors of Deutsche Telekom should also have bit-stream access to this new network. Why is there IP? It is because IP fits in well with the existing infrastructure of the telecom-munications industry. IP is a bidirectional protocol with source and destination addresses in the IP header. One can address packets using **IP unicast** to a single household for video or television ondemand and, for greater efficiency, use seamless switching to **IP multicast** [9] for addressing packets to a group when live broadcasts attract a large audience. It is not strictly a "must" but typically, as shown in *Fig. 2*, an IPTV system comprises at least a narrow-band return channel to complement the broadband forward channel. Within this asymmetric structure, usually a client server architecture is implemented for IPTV. In that respect, IPTV is interactive from the very start which makes an important difference when compared with one-way digital television such as DTT.

2.4. WEBTV and IPTV

IPTV, as outlined in this article, should not be confused with WebTV. While WebTV is also based on IP and compressed TV signals, it provides for live video streaming or video-on-demand over the worldwide web. Downloadable software players for a PC or Laptop are common. The quality of
service relies on the "best effort" principle. The IP unicast traffic of WebTV can lead to significant distribution costs for the broadcaster, depending on the number of users. In contrast, IPTV is implemented in fully-managed networks similar to broadband cable networks. The quality of service is
guaranteed. The IPTV technology is under the control of the provider. Specific provider-defined

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

settop boxes were common in the early days of IPTV, but open standards should be a requirement and are being developed by DVB. For retransmission of live broadcast signals as IPTV, thanks to IP multicast, the distribution costs on the provider's backbone network do not scale with the number of users. Moving away from fixed networks for a moment, a similar difference between WebTV and IPTV can be found in the case of terrestrial networks for mobile reception.

WebTV over 3G networks is an approach already being followed by the mobile network operators (MNOs). In contrast to mobile WebTV, **mobile IPTV** can be physically transmitted over a broadband terrestrial forward channel. A suitable broadcast transmission scheme, such as DAB-IP, DMB-DXB or DVB-H, provides a guaranteed quality of service that is independent of the number of users receiving the signal. It is that proposition – in combination with a 3G return channel to deliver IPTV services to battery-powered handheld devices with small displays and integrated antennas *(Fig. 3)* – that is attracting the imagination of MNOs for creating new revenue streams.

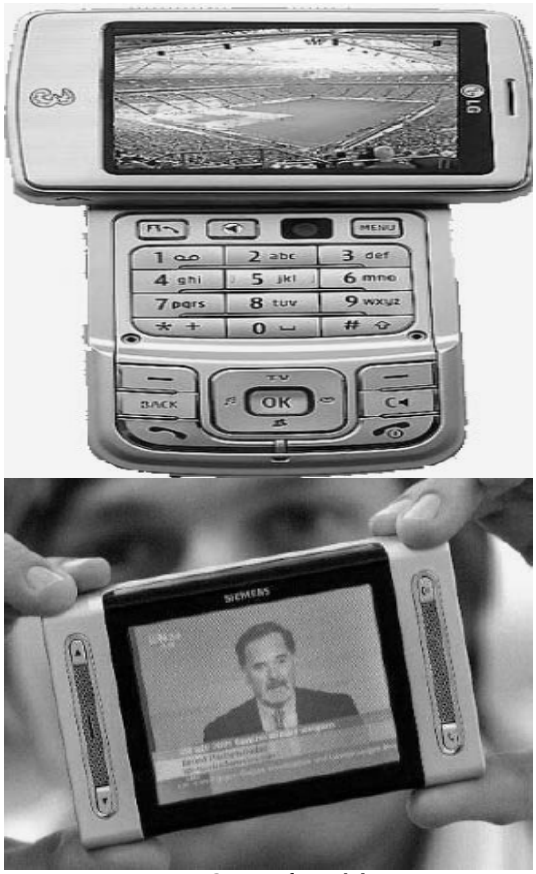


Figure 3: IPTV for mobiles

2.5. IPTV Over DSL Versus DVB Over Satellite and Cable

Returning to IPTV over DSL again, a simplified comparison with the DVB transmission schemes for satellite (DVB-S) and cable (DVB-C) is shown in Fig. 4. The colored lines are the different television programs of a digital bouquet that are bundled together in one or more multiplexes, each carrying a multi-programme transport stream (MPTS). Similar to a broadband cable head-end, a DSL headend picks up the satellite signals and converts them to IPTV signals. The difference here though is the client-server architecture used for IPTV, which sits between the provider backbone and the IPTV home. In contrast to DVB-S and DVB-C, however, not a single programme goes automatically into the IPTV home. Instead, the press of a button on the remote control is communicated back via the return channel to a server of the provider. The server then routes the requested programme, typically as a single programme transport stream (SPTS), over the backbone to the last mile which, in Fig. 4, starts behind the point of presence (PoP) at the DSL access multiplex (DSLAM). In this context, a technical challenge for IPTV is the short time delay during channel hopping (i.e. the zapping time).

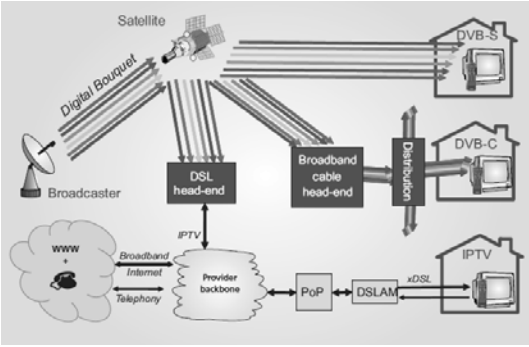


Figure 4: Simplified comparison between IPTV over DSL and DVB-S / DVB-C

Theoretically, every command on the remote control at home can be registered at the back-end system of the provider. This permits the ratings of a television programme for every single home to be measured very accurately and opens up a new dimension for personalized and individualized services on the TV screen. Equally, the client server architecture poses a new challenge to the data protection act and to

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

privacy, quite different from anonymous free-to-air reception in one-way digital television systems. Different IPTV providers may use different DSL head-end structures, different backbones and different client server architectures including the settop-boxes. Under regulatory terms, market competition among telephone companies in Germany is enabled through bit-stream access at the point of presence (PoP) or at the DSLAM *(Fig. 4)*. Thus, a telephone household can choose from several IPTV providers, unlike subscribers to existing coax broadband networks for cable television. The activities of IPTV providers in Germany are, however, still very much in the start-up phase. To summarize this section, IPTV can be regarded as a new variant of interactive cable television which at present, in Germany, is implemented primarily for DSL transmission over fixed line telephone networks in urban areas *(Fig. 5)*.

Figure 5: Examples of xDSL networks for IPTV in Germany

3. H.264 Encoder

The H.264 is also known as MPEG-4 ISO/IEC14496-10 or MPEG-4/AVC. This standard was co-developed by a JVT group composed by MPEG-ISO/IEC members and VCEG-ITU-T members. Three profiles (main, baseline, and high) have been defined, each with several levels. The main profile is required for broadcast video quality, while the baseline profile is typically used for mobile and video conferencing applications. The H.264 encoder system block diagram (Figure 6) includes two dataflow paths, a "forward" path and a "reconstruction" or feedback path.

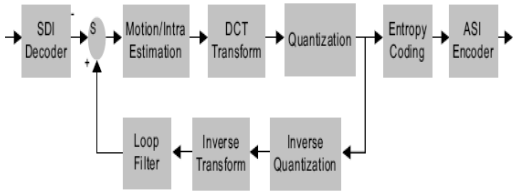


Figure 6: H.264 Encoder Block Diagram

H.264 encoding is ten times more complex than the MPEG2 encoding. For the motion estimation and compensation
section, H.264 employs blocks of different sizes and shapes, multiple reference frame selection, and multiple bi-directional mode selection. For the transform section, H.264 uses an integer-based transform that roughly approximates the discrete cosine transform (DCT) used in previous MPEG standards, but does not have the mismatch problem in the inverse transform. Entropy coding can be performed using either a combination of a single universal variable-length codes (UVLC) table with context adaptive variable-length codes (CAVLC) for the transform coefficients, or using context-based adaptive binary arithmetic coding (CABAC). The H.264 design is very complex, computing-hungry, and requires parallel processing. If a general-purpose processor is used, it will be limited by its internal architecture (i.e., if it has eight internal multipliers, it can perform eight multiplications per cycle). A programmable logic device (PLD) is flexible and highly scalable: if an algorithm needs 100 multiplications per cycle, then the PLD can be programmed to perform the required task. Figure 2 shows a single-chip main profile H.264 encoder PLD implementation using an Altera® Stratix® II device.

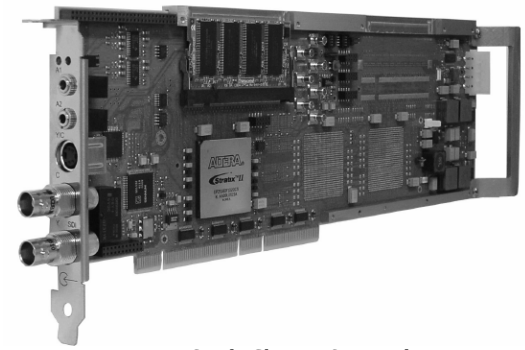


Figure 7: Single Chip H.264 Encoder

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

4. Video Over IP

Video over IP is the transmission of encoded video transport stream (TS) data over IP-based networks. It bridges between one or more encoded video streams and IP packets carried over 100 Mbps or 1 Gbps Ethernet. Video over IP accepts TS data and encapsulates it for transmission over Ethernet. Various standards define video over IP: real-time transport protocol (RTP), RTP payload format for MPEG video, UDP/IP, Pro MPEG code of Practice #3, and DVB-IPI. The TS input to video over IP is either a DVB-ASI or uncompressed SDI video data that will be mapped onto the Ethernet protocol layer. Figure 8 shows a video-over-IP reference design block diagram that receives a DVB-ASI TS and then converts the TS to IP. The design includes the following main blocks: TS input logic, frame buffer, queue system, Ethernet-receive DMA, encapsulator, transmits channel information, receive channel information, timestamp, media access control (MAC) interface, and host processor interface.

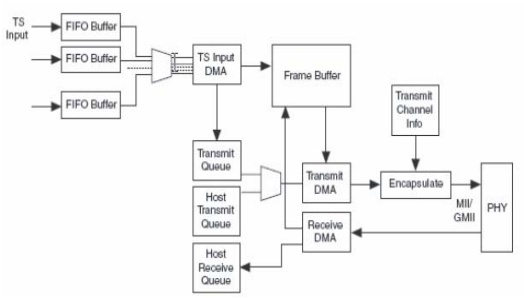


Figure 8: Video-Over-IP Reference Design Block Diagram

Conclusion

In summary, in order to provide quality video over IP, the latest H.264 video encoding technology is used to conserve bandwidth for delivery. Figure 9 shows the overall IPTV transmission system block diagram. The video content can be either SD or HD, uncompressed video, or previous MPEG2 TS. All these formats will be converted to H.264 video format before transmitting. All the key pieces can be implemented efficiently using PLDs for system upgradeability and flexibility.

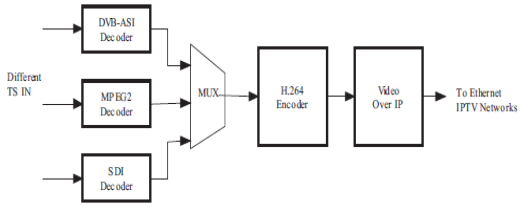


Figure 9: IPTV Core Block Diagram

5. Requirements

For the system outlined in Fig. 4, we will now concentrate on the requirements for IPTV over DSL networks. The business model is key. Obviously, Pay-TV and Free-To-Air (FTA) broadcasting are different models. We will focus on the latter for defining the requirements.

5.1. Retransmission of Live Broadcast Signals

If one regards IPTV as a new variant of interactive cable television, the general requirements for one-way digital cable television can be carried over. For example, broadcasters wish to provide a guaranteed quality of service with short signal delay, high picture and sound quality and lip sync. If transcoding of compressed source signals is involved at the DSL head-end, for example from MPEG-2 video received from a DVB-S signal to MPEG-4 AVC for IPTV retransmission, special attention has to be paid. The transcoding parameters have to be carefully chosen and adjusted in order to maintain a high picture quality and to make sure that the processed video does not lose the lip sync by delaying the unprocessed audio. The client-server architecture of a two-way IPTV system leads to additional requirements. Guidelines for programme listings in the navigator and interactive functions have to be revisited. The syntax and semantic for the related metadata server at the provider's back end system is a new component beyond the traditional broadcasting model. For example, a solution is required for the mapping of DVB service information (DVB SI data) that is carried in the transport stream of a DVB-S signal onto the client-server architecture and to identify what additional data is needed for the metadata server in order to get the electronic programme guide (EPG) or broadband content guide (BCG) fully functional and operational. In order to avoid expensive multi-platform authoring for the IPTV systems of different providers, standard data

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

formats are very helpful, e.g. XML. In particular, if DVB transport streams feed the DSL head-end as shown in Fig. 4, one solution is the mechanism of Service Discovery and Selection (SD&S) defined in the DVB-IP (Internet Protocol) recommendation for the transport of DVB services over IP networks [10]. In that case, every DVB-IP receiver, no matter what brand, can access the services and that can be quite relevant for free-to-air broadcasting. In the IPTV system, the return channel is always on. Therefore, requirements have to address in general what user- and service-related data is or can be collected on the return channel. Broadcasters want access to the interactive network functions, because this adds a new layer of functions for enhancing existing services or creating new ones. For copyright reasons, there should be no redistribution of IPTV content over the worldwide web. On the other hand, if a broadcaster does unicast live streaming over the worldwide web and in parallel the same content is distributed as multicast IPTV, then synergies through clever routing might be possible. An IPTV system can carry IP-Radio as well and therefore the radio requirements have to be negotiated between the broadcaster and the provider where appropriate. Free-to-air broadcasting, in its strict meaning, relies on unencrypted transmission of all service and bouquet components, not just picture signal and stereo sound but also multi-channel sound, teletext, interactive applications, etc. In terms of signalling, free-to-air means open standards for retransmission and coding such as DVB-IP [10] and MPEG-4 AVC [7]. If a digital rights management (DRM) system is used by the IPTV provider, the functions should be switched off by default for FTA services unless specified otherwise by the broadcaster. The EBU requirements, summarized in the content protection and copy management (CPCM) guidelines, provide a yardstick [11]. The users should have non-discriminatory access to all FTA services through the navigator and EPG/BCG. For the FTA services of public service broadcasters, the IPTV provider should not impose an extra fee on the user, related to content that is already covered through the license fee.

5.2. On-Demand IPTV Services

For on-demand services, another set of requirements relates to the time-shifted or time-independent use of content. Clearing of copyrights can be difficult. Broadcasters may not have the relevant rights for each programme or item in the archive because, at the time of acquisition, the television on-demand proposition was unknown and therefore not covered in the copyright contract. This brings up the problem of contemplating new and unknown forms of use for repurposed content that will be available in the future through advances in technology. Not an easy task. For rights acquisition today, the aspect of on-demand should be considered from the start alongside live broadcast. The take-up of hard disk video recording, be it on a server at the provider's back-end system (Network PVR) or at home (PVR) and more recent developments such as pod casting, are merely the tip of the iceberg. For on-demand it can make a difference whether one deals with (i) streaming and one-off time-independent viewing or (ii) file download yielding a permanent digital copy or, more precisely, a digital clone. Sophisticated DRM systems aim at refining the access rights for maximizing the income of commercial broadcasters and content owners. In this setting for an IPTV system, very careful thinking is needed when on-demand services are considered for programs that are originally made for free-to-air public service broadcasting. Nondiscriminatory and fair access to the interactive network functions that are relevant for on-demand services is a clear requirement. The IPTV on-demand system has to comply with the national copyright regime. For fully user-controlled personal video recorders (PVRs) that are combined with the IPTV set-top box, similar requirements apply as for PVRs with an integrated DVB-S, DVB-T or DVBC tuner. For example, a copy for private use must be possible. In the case of a provider-controlled Network PVR or PVR, the terms and conditions have to be negotiated between the broadcaster and the IPTV provider, e.g. unencrypted transmission or the marketing model by which the content is advertised and made available to the user. The best quality on-demand content is useless if it cannot be searched, located and referenced to others by the user. Metadata is the key to enable podcast-like functions and to facilitate sophisticated navigation tools. The syntax and semantic of relevant metadata should be agreed on between the broadcaster and the IPTV provider. The efficient generation of metadata for search engines and navigation tools is getting increasingly important in the media production

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

process. In passing, we should note that synergies between on-demand video services for the worldwide web and for IPTV networks could be an interesting follow-up.

6. Meeting the Requirements

In order to meet the requirements for IPTV, a look back at DVB and at the worldwide web is instructive. The DVB project was created for digital television. It brought together the broadcast industry, i.e. broadcasters, broadcast network operators and manufacturers. Out came mandatory forward transmission schemes for terrestrial, cable and satellite, now commonly used in the European broadcast markets. On the consumer side there is specific hardware, i.e. DVB receivers with mandatory and fixed integrated decoding functions. On top of the mandatory part there are options, e.g. for encryption or the MHP application programming interface. At the birth of the worldwide web, the Internet Engineering Task Force (IETF) defined the real-time streaming protocol (RTSP) without involving the traditional broadcast industry. Building on RTP/ RTSP and other mechanisms defined by the IETF, proprietary audio and video streaming schemes are not uncommon and are even used by broadcasters for WebTV. Proprietary schemes and widespread availability co-habitat through using general-purpose IT consumer hardware (PCs, Laptops) along with proprietary downloadable software players. Recent developments such as podcasting, and related RSS feeds, were originated for user-generated content and again rely on downloadable software. Pod casting is now increasingly used in the broadcast industry, even though it was not code signed by incumbent broadcasters. IPTV stands at the crossroads of these two examples, WWW and DVB. There is the content provider model as followed by WebTV and pod casting, and the broadcaster model of DVB. In the content-provider model, one concentrates on the generation of content. For distribution of the content, the necessary functions are specified at a rather abstract level and one does not really care about the specifics of the transmission scheme and the receiver technology. The content-provider model is agnostic to a specific IPTV technology. The IPTV system is a black box. The IPTV technology is left to the IPTV provider who has to make sure that the requirements of the broadcaster are met. For example, abstract requirements include (i) avoiding multi-platform authoring and (ii) ensuring a high quality of service with a free-to-air experience for the user. But how can a broadcaster be sure that such requirements are really met in cases of dispute, when no parameters are defined that could be metered, measured and quantified. The broadcaster model goes one step further and cares about certain specifics of the transmission scheme and the receiver technology. The business functions are developed into a set of technical requirements. The FTA principle, the avoidance of multi-platform authoring and the benefit of the economies of scale are promoted through technical requirements for open standards such as DVBIP for transmission and MPEG-4 AVC for source coding. However, the policy of intellectual property rights (IPR) and the conditions for obtaining a license to apply open standards is getting more difficult as the example of MHP highlights. The IPR policy and the license conditions for DVB and other standards relevant to IPTV should be clear upfront. Of course, IPR can be an even more important issue in the content-provider model when proprietary systems are involved.

7. Conclusion

IPTV can be regarded as a new variant of interactive cable television. IPTV brings together telephone companies and broadcasters. Partnerships are needed for carrying the concept of free-to-air public service broadcasting forward to IPTV. The implementation of a broadcaster model with broadcast-centric requirements and open technical standards as a base for fair and non-discriminatory access is a worthwhile goal, of benefit of the audience. The high-quality content and the specific know-how of public service broadcasters is a valuable asset for building strategic partnerships with IPTV providers, relevant software and hardware manufacturers and other broadcasters. The short term effects of IPTV should not be overestimated. In Germany, IPTV is still in its infancy. In the long run, IPTV could have a lasting effect and bring profound changes to the broadcast business. IPTV combines the traditional concept of live broadcast with on-demand and interactive viewing. That combination is most likely to prevail because it is an attractive proposition that appeals to a large audience. With ever-increasing bandwidth on the worldwide web, the differences between WebTV and IPTV will eventually

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

vanish. But this will take some time, and the guaranteed quality of service of IPTV is an advantage for the time being.

Abbreviations

ADSL Asynchronous Digital Subscriber Line
AVC (MPEG-4) Advanced Video Coding
BCG (DVB) Broadband Content Guide
CPCM (DVB) Content Protection and Copy Management
DAB Digital Audio Broadcasting (Eureka-147)
http://www.worlddab.org/
DMB Digital Multimedia Broadcasting
http://www.t-dmb.org/
DMT Discrete Multi-Tone
DRM Digital Rights Management
DSL Digital Subscriber Line
DSLAM DSL Access Multiplexer
DTT Digital Terrestrial Television
DVB Digital Video Broadcasting
http://www.dvb.org/
DVB-H DVB - Handheld
DxB Combination of DAB and DVB-T/DVB-H
EPG Electronic Programme Guide
FTA Free-To-Air
FTTC Fiber To The Curb
FTTH Fiber To The Home
IAD Integrated Access Device
IEC International Electro technical Commission
http://www.iec.ch/
IETF Internet Engineering Task Force
http://www.ietf.org/
IP Internet Protocol
IPI (DVB) Internet Protocol Infrastructure
IPR Intellectual Property Rights
IPTV Internet Protocol Television
ISDN Integrated Services Digital Network
ISO International Organization for Standardization
http://www.iso.org
MHP (DVB) Multimedia Home Platform
MPEG Moving Picture Experts Group
http://www.chiariglione.org/mpeg/
MPTS Multi-Programme Transport Stream
MSO Mobile Service Operator
PoP Point Of Presence
PVR Personal Video Recorder
RSS Really Simple Syndication
RTSP Real-Time Streaming Protocol
SD&S (DVB) Service Discovery & Selection
SDTV Standard-Definition Television
SPTS Single Programme Transport Stream
STB Set-Top Box
VDSL Very high bit-rate Digital Subscriber Line
WLAN Wireless Local Area Network
xDSL *(Different variants of)* Digital Subscriber Line

References

[1] Source: http://www.bundesnetzagentur.de/media/archive/7671.pdf
[2] ADSL: ITU Rec. G. 992.1
[3] ADSL2: ITU Rec. G.992.3/4 Annex J
[4] ADSL2+: ITU-T Rec. G.992.5: Asymmetric digital subscriber line (ADSL) transceivers – extended bandwidth ADSL2
[5] VDSL1: ITU-T Rec. G.993.1: Very high data rate digital subscriber line
[6] VDSL2: ITU-T Rec. G.993.2: Very high data rate digital subscriber line 2
[7] MPEG-4 Advanced Video Coding (AVC): MPEG-4 Part 10, ISO/IEC 14496-10
[8] Edge/QAM Docsis, http://www.cablemodem.com/specifications
[9] Internet Group Management Protocol (IGMP): http://www.ietf.org/rfc/rfc3376.txt
[10] DVB Internet Protocol (IP),
ETSI TS 102 034: Digital Video Broadcasting (DVB): Transport of DVB Services over IP based Networks
[11] Content Protection and Copy Management (CPCM) Guidelines for Free-To-Air (FTA), EBU Requirements, Geneva, January 2006

FURTHER READING
[a] F. Kozamernik and L. Vermaele: Will Broadband TV shape the future of broadcasting? EBU Technical Review No. 302, April 2005
http://www.ebu.ch/en/technical/trev/trev_302-kozamernik.pdf
[b] EBU tech. doc. 3308: Broadband Television - Opportunities and Challenges
http://www.ebu.ch/CMSimages/en/tec_doc_t3308_tcm6-40404.pdf

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

A Review of QoS in Mobility Phillic Communication

Mr. Sumit Joshi[1] Mr. Rajeev Kumar[2]

Department of Electronics & Communications
GRD-IMT, Dehradun-248001, India

ABSTRACT: ***This particular outlook represents, in brief the recent & upcoming issues associated with QoS guarantee & mobility in communications. Moreover it also presents a brief context of QoS attributes & handling techniques thereby. Even more, it covers aspects of QoS fundamentals in deployment of mobility led approaches. There are different kinds of systems that facilitate different QoS attributes for variable classes of services. This review represents different types of approaches involved in the process of modern communication networks with respect to the QoS offered by a particular service. Moreover the parameters such as: delay, bandwidth, jitter etc. are system specific along with being service specific too. All the system types are actually a broad classification which involves different service primitives. In this review we focus mainly on mobility aspects of computing in the modern arena.***

Keywords: *Quality aspects, mobility, QoS management.*

I.INTRODUCTION

Many business & technological issues employ good processing, battery life, interface compatibility etc. and hence application based deployment is a concern. Thus a constant feed backing in the form of standards, developments & algorithms are continuously used. Therefore the background engineering must concentrate on issues of QoS& system mobility.

II.DEFINING QoS FOR MOBILITY LEAD COMMUNICATIONS

Quality is the first term of QoS which addresses the behavioral approach of action in any kind of devices or strategies used for communicating. More elaborative, the term actually expresses the characteristic point which focuses on the issue of quality perception. FIG.1 shows distinct parameters on which QoS functions. The predicates used to review it can vary according to the kind of service & the application for which it is being provided. Here, we look in to the model for QoS with a bit deeper insight. Moreover, ITU defines it as the level of satisfaction of a user of any particular or specific service primitive. Another definition relates QoS to systems/networks capability in terms of factors like: bandwidth, transmission rate etc.[1] Broadly, we can sub-divide QoS in to three types: 1.operational 2.perceptability 3.assessebility

Figure.1.Demonstrating core factors in relation to QoS.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

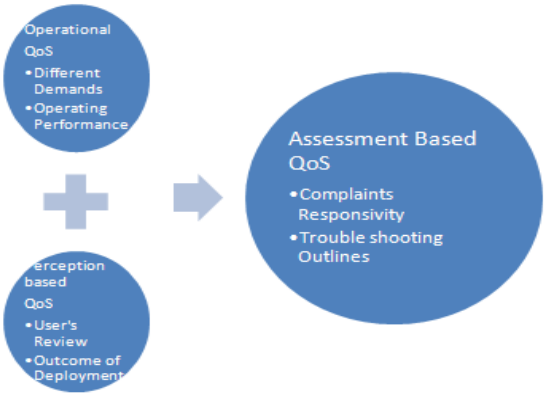


Figure.2.Types of QoS Sub-divisions.

This classification actually is multiple-leveled & works only in accordance with information availability at each respective level, each level is responsible for a particular kind of a service primitive as shown in the functional outlet in Figure.2

While defining QoS it is important to mention & specify these types as described in the section earlier. Operational QoS refers to the network's connection & resources associated to it from design point of view.[5] On contrary the perception based QoS refers to the degree of satisfaction estimates obtained via. Responses from customer side .The most interesting is the assessment based QoS as it depends on the first two; it deals only with after sales review of the product.

The service quality can be expressed as a function of all these three types of QoS. Also the major dependence of service quality is on the first one, as it paves the way for all of the others to play their part. This QoS term can also be defined in terms of degrees of freedom available to the user for better perception.

MOBILE COMPUTING APPLICATION

Mobility along with it accompanies portability as a feature which leads to reduction in product's market price .These type of devices find use in accessing basic services of today[2] .many other devices are being used in conjunction with these devices which has lead to the convolving of communication with computer based mathematics. Most of these applications are less massy and thus are preferred over earlier user preferred services delivery methodologies. These applications are even more favorable in terms of integrity with other similar user-selective applications. A lot many web based applications are presently in run but are bounded only up to personal computers, for a better integrity that could cover recent innovations & growing technologies, a positive development towards a more practically accepted monetary value with respect to the wireless communications is required for better & numerically advanced functioning. However, earliest of the developments like: internet supporting devices, mobility enabled communication platforms etc. are in a process of continuous eruption.

A more sophisticated set of services such as: position aware services are offered via. most of the new devices that have embedded capabilities for the same. These devices can easily locate device coordinates; update weather information & roadside traffic updates etc. on a given map with Global Positioning Systems. Even more it can be used in different prospects of getting details regarding nearby local centers of interest.

Another aspect discussed, focuses on usability part of mobility lead computations in integrity with services offered in serendipitous conditions like fire, earthquake or any situation of natural disasters .In these particular cases far away information can easily be assessed & fetched with it. Portable hand-held devices, cell phones etc. are other features which supports regulation & information exchange simultaneously. Other requirements, such as: safe initiation while any kind of communication is in progress.

It also supports access to digital data warehouse from which academic content is easily distributed to the local as well as far away users being mobile. Further more the concepts of learning in alignment with web access & online resource sharing is addressed. The most happening feature of all these technologies is

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

integrity & portability with respect to any type of data under processing.

CHARACTERSTIC SUPPORTING FACILITIES

The first issue to be fore taken is to sub-divide communications in to its types based on whether the communication platform is static or dynamic (movable).It is also important to consider the factors or parameters associated with respective platforms.

Mobile systems are intended to work under either LAN or line based dialing configurations, moreover the issues such as opaqueness & desired performance while the user is in mobile state need to be addressed, i.e. we need to have broader capability spectrum with better deployment in terms of commonly available resources. The term 'mobility' involves both types of systems i.e. the systems which are actually mobile & the systems which are not mobile but uses the support of mobile devices. It is observed that all such sorts of services are categorized under "Location-Aware Services".

Still, there are certain services that are called 'Context-Aware Services', these services do require information regarding characteristics of particular computation tool being used along with locating information. In order to provide the required quality associated with communication services computation facilitation concept is employed with resource allocation as the core.

AN INSIGHT IN TO QoS MANAGEMENT

QoS deals fundamentally with those parameters of a system which are related on service basis but independent of functionality of a system. The flow-charts below shows the core networking parameters defined in relation to QoS and also the user-perception lead QoS characteristics, as shown in Figure.3

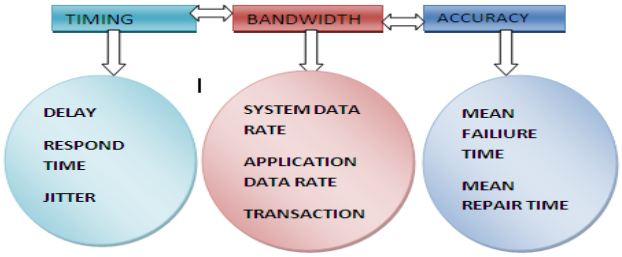


Figure.3.QoS AS BASIS OF TECHNOLOGY

Broadly QoS management can be classified under two heads, namely: static QoS management & dynamic QoS management.

NON-ADAPTIVE QoS MANAGEMENT

The static QoS management deals with those parameters which seldom change during the entire period of transaction or operation within any kind of network in action. Moreover user intervention here, in any operation is must as it is the user who perceives the service standards. Its main objective is to focus only on those services which are non-varying or where variability is unacceptable example: frame transfer rate in multimedia processing.

ADAPTIVE QoS MANAGEMENT

Dynamic QoS aspects focus on changing environment under which service is being provided via. Any kind of network & it also allows the network services to adjust themselves as per the requirements & environment in which network is functioning, whether it be any. This mainly involves variable parameters like data flow rate, resource utilization index etc.[3] Table 1.1 displays a comparison between the two management aspects as discussed in this review.

NON-ADAPTIVE QoS	ADAPTIVE QoS
1. Involves features that remain constant throughout some activity.	**1. Involves changing features with respect to environment.**

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

2.Activities like resource reservation, admission control are the key functions	**2. Activities like adaptation, synchronization are the key functions with respect to changing conditions.**
3. Other functions involved are specification at various abstraction levels.	**3. Other functions here are monitoring & policing to keep a check over QoS actually being offered.**
4. Negotiations for QoS of an acceptable liability between two entities are also a feature of importance.	**4. Maintenance is another very important aspect which if required also modifies required parameter as per needs.**

Table 1.1.Comparison between two types of QoS managment.

EFFECT OF MOBILE ENVIRONMENT ON QoS

This covers those aspects of QoS which are effected via. Mobile conditions proportionally .The major focus will be on mobility impact on QoS for a given network under work.

PERFORMANCE WITH DIFFERENT LINKS

Mobile systems usually are dynamic by nature and thus connect through a local network via. a LAN or WI-FI access point in a network. Basically, wireless links are included as sub-centers of support for mobility prone environments.

These sorts of systems have certain essentially required considerations like:

- Range of coverage.
- Infrastructural & interfacial costs
- Required bandwidth support.

In the present time lots of new technologies supporting greater quality services as compared to earlier one's are budding out & this will lead to improved standards in terms of customer satisfaction. Still, the issue of wire phobic applications involves environmental susceptibility as a key. Other majority multimedia applications have some common issues like throughput, data rate etc.[6] as the fundamental problems with respect to mobility. In designing & analyzing such systems the sole responsibility is to make these systems more adaptive to these kinds of variations normally encountered which are a much better step in this direction.

ISSUES WITH MOBILITY

These fundamentally include issues like coverage area, handing off etc. which are of greater concern in applications such as image & video transmission. This requires utilizing QoS based estimation of available resources & processing this information for seizing of the same, this can be done using knowledge of available routes in any network located in any known & well-mapped boundary. This knowledge in convolution with information with respect to operating environment facilitates better measures based on statistical predictions & thus makes the resource seizing more efficient & reliable. Moreover it leads to better analysis possibilities for further deployment later on.

In all these techniques are good enough but not the best to compensate for all sorts of connectivity issues because there are still some loop holes left which may lead to degradation in quality & non-permanent falling of signal strength. The cause of these kinds of degradations can be many & thus an efficient mechanism is required in curing this out in form of adaptive QoS.

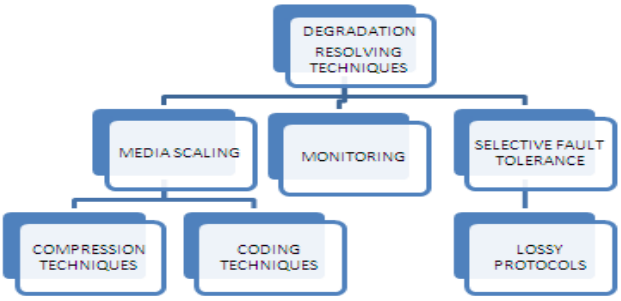


Figure.4. Speedy Degradation Recovery Methodologies

For speedy degradation recovery the major methodologies used are summarized as in the block diagram of Figure.4 this clarifies the fact that speedy recovery & quality are the issues which require resolute focus. Even more the system should be such, that it has enough time to

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

adjust according to the transient changes encountered.

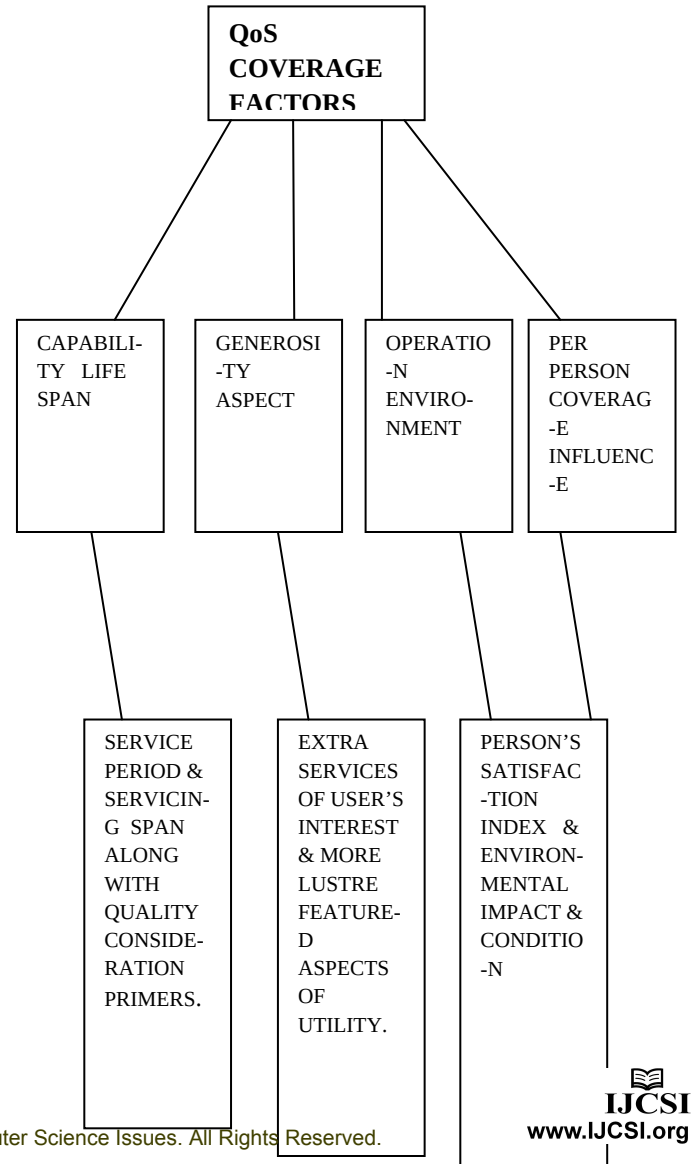
Thus, this clarifies the fact that for better service quality some standards in terms of certain major well defined & structured parameters like: capability life, generosity aspects, operational environment etc. are the key concerns. Moreover another very delightful parameter of interest is the 'user health'; this parameter is a key concern nowadays when many upcoming technologies mostly inclined towards radio propagation are in huge & populous eruption all over the world.

These concerned parameters really have an impact on user behavior & market too. Even some special core challenging & more managerial then technical parameters of concern along with QoS enabled service infrastructure are as: recreational outlook, per person popularity index etc. these are some of the core parameters & challenges overlooked here in this context in a brief summarization.[4] Issues of network availability & reliability are also major concerns associated with proper design technicalities in reference to a networks basic functional outlook, which have been foreseen with impact.

We now describe each of these QoS coverage factors in detailed aspects. The major coverage aspects involve following:

1. **Capability life span –**

This under covers various aspects of exact service primitives like period for which the services are available along with aspects of quality improvement & enhancement features. Specifically, for a wireless mobile channel where channel parameters are always variable & ever-changing the QoS aspects like these are of due interest. When mobility as a whole is considered, it even becomes more important to measure or represent QoS in terms of connectivity life span which actually signifies the time for which the transmission is smooth & of the estimated quality. Another aspect integrated with the capability term is the estimated handling capacity for a link in terms of packets delivered at receiving end via. a network of links, that means QoS is not only associated with link failures or packet delivery ratio only but also with the aspects of data being transmitted over kinds of links. A comparative assessment of different kinds of data over the mobile channel under research under varying conditions facilitates the study of possible performance profiles that could be obtained while real time situation is met. On, the basis of this kind of study more adaptive & less quality compromising algorithms for better performance can be structured & thus it also includes aspects of link life or link failures at same.



Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Figure.5.Total reviewed factors binding QoS

2. **Generosity Aspects-**
   Special facilities provided along with usual QoS concerns are a part of these aspects; these sorts of features in today's market are core issues & are a definite feature for good QoS providence in eyes of users of service. These aspects sometimes are available as specific extra facilities wit already available ongoing scenario; these not only enhance the user base but at the same time impinge on psychological framework of a user in terms of trust & goodwill for that service. That means QoS is indirectly associated to these aspects too.

3. **Operational Environment-**
   An important feature of any QoS architecture is the environment for which it is designed & considered. More over the factors like terrain, mobility etc. also encompasses actual QoS measurement for a network.
4. **Per Person Coverage Influence-**
   This is one of the cause-effect aspect which actually expresses the satisfaction of a customer or a user in a particular domain of influence, for example- a user if feels more satisfied & at ease with the services provided by (say) network A in comparison to network B ,definitely in his understanding the QoS offered by the same is of an acceptable level, although the same network in some another domain of interest may not be of equivalent state. So, this means it is very important to categorize the services before actually evaluating them under required domains of interest.[8] Here, we propose a domain based approach in this review for better deployment of QoS.

QoS Domains of Interest- Under this we categorize different kinds of mobility prone networks under different domains that are of required interest**.**

- Wireless adhoc networks domain.
- Cellular networks domain.
- Wireless sensor networks domain.
- Wireless multimedia networks domain

These domains along with earlier proposed QoS models will be reviewed in the continuing context as:

Wireless Adhoc Networks Domain

Here, in adhoc networks usually the topology is ubiquitous & configurable that means for these kinds of applications there must be issues related to a definite level of quality regarding the services offered by these networks & hence QoS is to be considered seriously. Moreover, here in these kinds of networks we have some major issues as: unicast routing, multicast routing, battery management & TCP transport issues.[7] These all major issues influence the QoS in real aspects as depicted in Figure.6

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

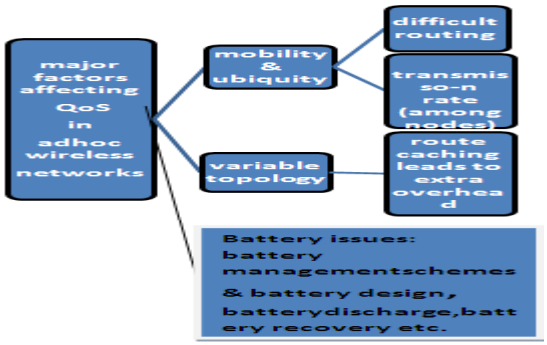


Figure.6.QoS effecting factors in wireless adhoc networks

Cellular networks domain

In cellular networks the majority factors effecting QoS are: mobility, handoff, population density, coverage area etc. these are the parameters which do impinge QoS in a certain manner. Although, apart from these, there are many other factors also like: network accessibility, offered service cost, special offers etc.

Thus we can classify these in to 2 specific classes for the sake of simplicity as: macro-impact factors & micro-impact factors, as Figure.7.

Wireless Sensor Networks Domain

Although, these are a special class of adhoc wireless networks but on basis of the issues associated, they need a separate addressing.[2] They consist of miniature devices that are called sensor nodes which are capably designed to sense & monitor different physical entities of due interest with respect to the application for which these are being used. As, we know sensing can be either of the two types: periodic & sporadic depending on the application type. Thus, the issues associated are: mobility among nodes, network dimensions, deployable coverage, power limitations etc. depicted as Figure.8

Wireless multimedia network domain

This domain focuses on multimedia data transmission over mobility led wireless networks, the major causal factors effecting QoS here are network load capability, actual load processed, reconfiguration actions etc. these are depicted as Figure.9

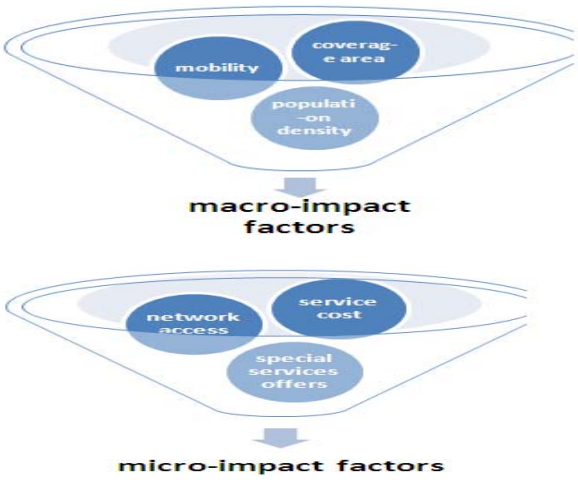


Figure.7. Micro & Macro Impact factors.

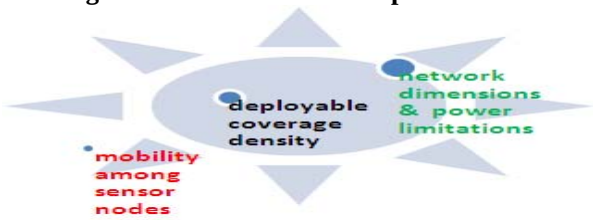


Figure.8. QoS effecting factors in sensor network domain

Figure.9.Factors effecting QoS in wireless multimedia networks.

Now, some of wireless QoS schemes which are based on 5 principles as:

1) **PRINCIPLE OF UNITEDNESS→**According to this principle quality of service for any network based application should include features like: reliability, maintenance capability, reconfiguring ability over different designs in order to achieve better user-to-user service quality.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

2) **PRINCIPLE OF SAGREGATION→**It illustrates that functions like: data transfer of any kind, flow control & resource management etc. are the activities which are distinct from each other in terms of operation & design aspects.[7] Therefore, these should be handled in a segregating manner. The actual segregation is between the two core aspects of data transfer & information control.

3) **PRINCIPLE OF CLARITY→**This principle can also be termed as principle of abstraction that means different variety of associated or non-associated applications should be abstracted i.e they should be kept apart from influence of issues such as: QoS control & handling. Advantages of abstraction are depicted in form of a block as Figure.10

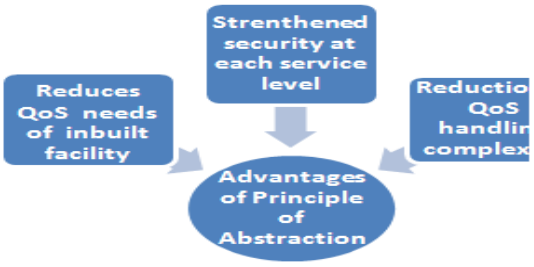


Figure.10.Advantages of Abstraction

4) **PRINCIPLE OF RANDOM ENTITY DISTRIBUTION→**
This principle coordinates the distribution of operation between different structural designs, in all it is responsible for design of the entire control, scheduling, flow & operational management aspects.

5) **PRINCIPLE OF RELIABILITY→**This specifies the aspects of reliable performance by different structured communication protocols for enhanced performance with respect to some or the other of design principles.

Different QoS specifications are as:

1) **Tuning in flow control:**
This is the feature that characterizes different related traffics on the basis of tuning which exists between them. In many applications like: multimedia applications data tightness or tuning is an important centralized feature, without which information has no logical meaning. Therefore, precise tuning is the key in such cases & certain other applications may not require this level of tuning; there we don't require such controlled tightness.

2) **Flow action primitives:** These primitives define the requirements of flow activity from user end. Specifically, for multimedia applications it becomes important to ensure that the factors like: throughput, jitter, delay etc. are guaranteed from network traffic point of view. These flow-action primitives may vary with the application on demand i.e. QoS infrastructure must have pre-information regarding traffic volume.

3) **Service hierarchy:** It specifies the performance indices in a qualitative fashion of distinction which classifies them into classes such as: micro, macro & sublime guarantees of performance.

4) **Policy based QoS control:**___It demonstrates the tolerable flow based QoS adaptability. It also covers aspects of QoS management at every level of service being provided.

5) **Monetary aspects of service:** This includes the monetary value being charged as per user interest in a particular

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

level of service to be offered to the users as per their liking. This certainly enables user to select the services in accordance with the cost of service as per QoS specifications.

Some of the popular QoS models as:

1) Heidelberg QoS model:

This model is designed specifically for heterogeneous networks of today. Moreover, it provides better services for end-systems & networks. Its design is

composed of regular data transport mechanism which is responsible for providing QoS provisioning functionalities to the design along with compression & coding aspects as depicted below as Figure.11.

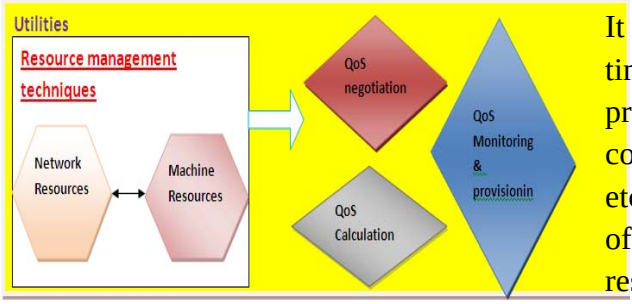


Figure.11.Simplified Heidelberg'sQoS Model

2) The OMEGA Architecture:

This architecture provides QoS guarantees at different imaginary levels in a network. These kinds of resource management by this model are facilitated via. a middleman model which includes features like: QoS based resource allocation, QoS middleman based translation at network-node interface. The OMEGA Architecture is developed by University of Pennsylvania with end-point architecture in view. Its simplified outlook has been presented as fig.3 below. This architecture classifies OMEGA communication prototype in to two fundamental units namely: transport unit & application unit, where each unit consists of two functional sub units. Moreover the core feature of the design is a middle-man which serves both the fundamental units by performing translational operations at different imaginary network levels.

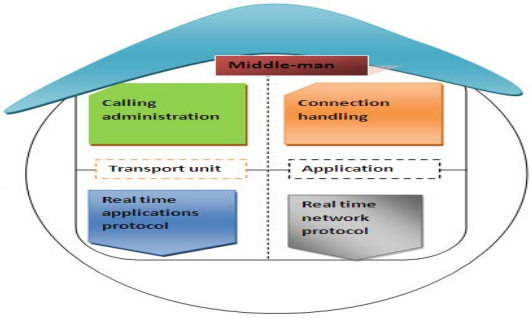


Figure.12.Simplified OMEGA Communication prototype

3) The IMAC architecture:

It is based on ANSA QoS outlines & is a real-time operation facilitating design. Thus, it provides features such as: QoS details, real-time coding based on QoS facilitated encapsulation etc. this architecture in all covers 3 major aspects of real-time network services ,which are resource demand , resource assigning & resource timing.

Apart from these architectures discussed above, there are many more designs which are present or are in their initial phase of development & deployment. This discussion focuses only on reviewing the fundamental aspects.

Heterogeneity & mobility prone networks

Still a lot is to be done & a lot is being done by different groups & people at different places in the world for a better QoS guarantee in modern ever mobile & heterogeneous operational environment. While discussing different QoS parameters we observed that each parameter needs special attention with regard to some or

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

the other applications ,that means QoS is now no more a general aspect but has become application specific by its sole nature in modern times. We also noticed that all the designs & strategies developed & deployed till now are more focused on adaptively of the network parameters with the flow control & resource allocation as the core. The most important observation is that, the QoS guarantee depends not only on network parameters only but also on the nature of these parameters with respect to the deploying circumstances in which the network will be actually functioning. One of the key aspect that is to be thought of is that we need to develop more stringent & abstract designs for an appreciable QoS demand to be satisfied.

Comparative outlook of different modern networks

It is clear that modern networks are more or less similar in terms of their operating environments & aspects of functioning .For example the adhoc networks & sensor networks are nearly same functionally & moreover the parameters for QoS guarantee to be looked for are not much different .we observe that these sort of networks are definitely going to be a big boom if the issues of ubiquity & mobility associated with them are properly addressed via. Better QoS provisioning & management schemes. Further, the issues of QoS routing & energy efficient network handling are the core issues to be considered while any design is to be undertaken.

CONCLUSION

This review is in context of effect of mobility on performance of any network in communication & by the review of QoS for it, we came to certain important derivatives, that there are many factors capable of affecting the performance of a network in mobile conditions & also that each of the factor needs attention singularly for better QoS design. Generally, a desired QoS is never offered but a average or near is offered by means of better network handling & resource allocation techniques.

REFERENCES

[1] R.Timusset al., Mobility and Multimedia Information Software Agents and Soft Computing, Towards Enhancing Machine Intelligence Concepts and Applications, Springer-Verlag,1997,pp.146-59.

[2]. KimayaSanzgiri, Elizabeth M. Belding-Royer, Leveraging Mobility to Improve Quality of Service in Mobility Networks, Deptt. of Computer Science, University of California.

[3]. AngelosAmanatiadias, KonstantinosDrakatos, Lukas Tsironis, VassilisMoustakis, Defining the Main Factors of Quality of Service in Mobile Telephony, IEEE-2006.

[4].Dan Chalmers, Morris Sloman, A Survey of Quality of Service In Mobile Computing Environments, IEEE Communication surveys, second quarter-1999.

[5]. LilianaDiazOlavarrieta, Wireless Communications Education Guide to Important Topics, November, 2005.

[6].Diedrich, Simple Quality of Service for Wireless Mobile Networks, Verlag, 2003.

[7]. YigalBejerano, Randeep S Bhatia, MiFi: A Framework for Fairness and QoS Assurance in Current IEEE 802.11 Networks with Multiple Access Points, IEEE-2004.

[8].ManojB.S, Shiva Rama Murthy C, Adhoc Wireless Networks, Prentice Hall, 2004.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Rajeev Kumar
Presently, Assistant professor at Dehradun Institute of Technology. He has many years of experience in Teaching and is an IEEE reviewer.

Sumit Joshi
Mr. Sumit Joshi is working as an assistant professor in Deptt. Of Electronics & Communications at GRD-IMT, Dehradun. He is also a member of IAENG & IJOE.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Conceptualizing of social Networking Sites

Dr. J. S. Sodhi [1] Shilpi Sharma [2]

[1] AKC Data Systems, New Delhi, 110024, India

[2] Computer Science and Engineering, Amity University, Noida –UP, 201303,India

Abstract:

People often move to their friends, families and colleagues when they feel urge and having doubts or queries to solve. Participation in social networking site has dramatically increased in recent years. Many social networking sites boost with million of members using their network on regular basis to communicate, share , create and collaborate with others. In this paper we explore the phenomenon of using social networking site to trace a link of the search from the community of users for better exploration of link. If a person wishes to visit a place that belongs to the community of his friend list. Similarly it will search upto nth level and filter to get best option by implementing the concept of deterministic agents.

Keyword: hcard, FOAF, RDF

I. Introduction:

The social network site will implement this concept of exploration of optimal link as a search engine, that benefits the user to search friend of friend list in order to visit any place which is new for him and he wants to explore the new place [Refer Fig. 1]. It will act as an open ended API. It will provide a great opportunity for increased search ranking and more increased relevance because the search engine can better understand what the content of your site is about. Web Standards such as hCard and FOAF as a part of semantic Uniform Resource identifier (URI) scheme.We will use the Web technology FOAF to mark up the friend list of users.

II. Concept

FOAF provides a unique concept that, no other single social network Web site has a distributed friend list. Because our friends are referenced within FOAF by a URI, and the URI that references them can itself return FOAF, it can follow the path through the FOAF files of our friends, to the friends of your friends and so on. The URIs can refer to any location on the web, so your network can expand across the Internet.

When a user browse to the page of another user, they will see and search for the place into friend link. Then, on their profile page, we present the list of the user's friends marked up with FOAF using RDF. Their data is available in a standard, portable format for direct use in other FOAF-enabled Web sites they come across.

In general , we must use either Micoformats or a RDFa depending on the case. Microformats, from the start , are designed to work with existing markup to add semantic value for specific use

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Figure 1.

III. Challenges & Threats:

In any social networking site there exists several challenges while expressing the idea of a distributed friends list to get use of optimum link usage and obviously due to open nature of connectivity mechanism it may face hurdles that can be overcome, but first, let's identify them.

III.1 From others

Difficult to Monetize

Even Google says it's having a hard time monetizing social networks, The use case is completely different. Members aren't hunting for information like they do on a search, instead they are communicating with each other, and self-expressing.

Excess of Players

In the case of the many white label social networks, there are too many players in the space.

As Marketers Move In, Users Move Out

Remember Friendster? Tribe, or way back and eCircles? Nothing is new, as communities form, marketers will move in, and in some cases bastardize the experience and the hip, cool, influencers will leave to the next network.

Untrustworthy Member Data

In many cases users submit inaccurate information on their profile. As a result,

marketing efforts will not be aimed at the right audiences, members continuing to be an elusive target.

III.2 From Systems

Lack of Metrics Makes Success Hard to Measure

For many marketers who want to deploy a campaign on a social network, access to server metrics isn't always available. As a result, they have to often visually monitor the interaction on the site, or measure click through their site. In some of the more sophisticated platforms, a crude dashboard is provided.

Successful Networks have hard time scaling

Facebook and Twitter are suffering from scaling issues, as a result, their sites have a great deal of downtime or latency. The complicated applications will only increase in intricacy as more users are added.

Privacy Concerns Mount as Developers Move In

The great community push back from the recent Beacon experiment, launching of newsfeeds, and social networks sharing too much information with third party widget developers puts members at risk, and visibly makes them uncomfortable

Strings Attached to Membership: Difficult to Leave

leaving Facebook is difficult, there are hooks, saved accounts, and ways to continue to reconnect to the site, even after you've left.

Stalkers and Other Unwanted Activity Ruins Lives

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Child stalkers in MySpace continues to be a problem, and in some cases, masking oneself as someone else is easy, and to readily fool others. As a result, one young teen committed suicide from the deception, rejection, and embarrasment from a peer's mother.

Loss in workplace productivity "Social *not*-working"

Companies, organizations, and individuals are concerned about the time wasted in managing social network profiles, in some cases, companies have banned Facebook from their employees, often using Firewalls.

IV. Characteristics OF Social Site

1. It needs to be able to attract people whom we care about, so it has to be easy and convenient. Because there are so many social networking sites out there, it needs to be able to publish to or from other formats so that people don't feel like they have to choose one or the other. It should incorporate the following concept even if exists must be implemented or used in a trusted manner so that optimality using the concept of friends path can be achieved.

1. No ads. Can not take them.
2. A personal blog like page
3. Tags that automatically link to other people with similar interests.
4. Granular privacy permissions.
5. A front page with public calendar info on it, and all the publicly available voted on posts.
6. A calendar, again, with granular permissions.
7. The front page should be somewhat like a portal with parts of it customizable by people who log in,

A page with all posts I have permission to view on it.

↓

The ability to have your own domain name and still be a part of the network.

↓

The ability to vote on blog posts, and music, and comments.

↓

Real names and pictures to increase integrity among posters.

V. Operational Concept

5.1 A user posts a message to their page, with a "blast radius" option. The system then sends the message that many levels outward from the original person: A radius of 1 would go to their friends, a radius of 2 would go to their friends' friends, 3 goes to grand-friends, etc. It shows that 3 levels would be the reasonable cut-off point-- I'm figuring that anything beyond that would involve sending the thing to the entire Internet.

The first option would be to store every interpersonal relationship up to three levels deep in a table: The person's ID, The relation's ID, and the distance between the two (all integer identifiers). This would allow faster searching, since the distances are pre-calculated, and it's just a matter of getting anything relating to that particular person with a distance less than the blast radius.

5.2 The second option would be to only store immediate (distance=1) relationships, and make multiple queries, iterating through each friends' relationships, and each friends'-friends' relationships. This just sounds like a great way to choke a server, especially if it's being done often.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

VI. Optimization

Social media optimization and search engine optimization are the methods used for web optimization. Social media optimization is one of the very popular methods and the reason for this is the increasingly large number of people who are active in different social media networking platforms like MySpace, Facebook, Orkut and several others.

Social media optimization is being undertaken extensively and if anyone want his site to rank at the top of all major search engine results, this is what he must do extensively and regularly.

- Optimize their connections with the social networking community,
- Continually enhance the on-site social experiences on their websites,
- Analyse the results and adapt their approach to improve their social business strategy and how it is implemented. This strategy is often called Social Optimization of Websites.
- Develop enhancement strategies by offering incentives for referrals and packages for customers to tell their friends about it.

- Social network linkages tend to multply and interconnect with third party sites. Friends tend to go rapidly to other networks via these expanding linkages. The use of social networking sites like Facebook, Twitter and Linkedin exploded because it initially provided a fantastic tool for groups with existing offline personal networks to connect with each other online. It quickly developed as a system rather than a tool for making new online social networks. Now these networks systems are more widely available to online businesses – through Facebook Connect, and Sign-in with Twitter, among others. These services now offer tools for online businesses to register users using established logon identities.

- Users who sign up for these services can easily share content and activities on the site back to their friends on social networks via these feeds. Profile information can be gathered and enhance to personalize the site experience for the customers. If you know who your users are connecting to, you can adapt the site accordingly. You can even tap into the networks yourself to learn about them and so provide a more highly developed and personalized social experience on your site. Other friends will want to explore your site with their friend and groups, rather than by themselves.

VII. Generate Referrals

The concepts of general referral will proactively enabling social participation by users on site. This should be tested by:

- counting the number of users who use the tools to register their social network identification on site,

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

- assessing the amount of website content and activity on site that can be shared via social networks
- seeing the amount of time users spend exploring site while sharing with in friends on the social networks.

The more we optimize your site for social networking, the more will be number of users, and their friends will use your site and the better will be the business outcomes. An effective website that is socially optimized can generate:

increases in traffic from social networks,

foster greater sales revenue,

increase search engine rankings,

promote better brand or product awareness

greatly reduce your costs in finding new customers, and

provided more focused users and visitors, meaning that you site can be better designed for a specific target group

Social Optimization for a website will require attention on three key components:

Social Connectivity

Social Analytics

Connected Experience

First Step:

Generating Social Connectivity

The first step in evolving and applying a scheme is to connect the website to the social networks in a way that maximizes the number of active participants. Fortunately, most of the social networks are aware the worth in this as well, and most have supplied methods that enable websites to directly connect with them. The APIs they provide and the various routines and programs allow hosting websites to get data from, and interchange information with the social network. These routines are provided for Facebook Connect and there are other programs for Facebook, Myspace, Twitter, LinkedIn, Yahoo, Google, and many others.

Second Step:

Developing or Improving the Experience when Connected

One should greatly enhance the connected experience offered to users of your website by re-designing it for heightened socially-connected participation. This means encouraging your users to register with their social network identification, to share content and experience with their social networks, and to interact with friends while on your website. It is all about how to enable sharing by users.

The key points are:

- Keep users on your site for the entire sharing process, by including a sharing dialogue box on each page that can be shared, or the home page. The newest APIs

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

allow users to make the connection and start sharing without having to leave your site.

- Build a sharing experience into the user activity scheme on your site.
- Allow users to sign-in on the home page with their social network identities. Make the sharing part of the normal log-in or registration process, not a separate extra requirement.
- Make it easy for users to share to multiple networks, simultaneously. This can greatly increase the network audience.

Step Three. Develop advanced social metrics for your site

Businesses need to be able to test and measure changes in connected user activity by social networks, and make changes based on the monitoring information. Some of the major performance measures that you can use are:

- Yearly and Monthly growth in traffic figures derived from social networks overall, and by from each specific social network site.
- Statistics on the numbers of users that sharing content on your website
- Average number of completed shares and responses to shares via messages, status updates etc. per connected user.
- The mean number of referred visits and other referrals
- The specific parts and features of your site which drive the highest volume of sharing activity

Conclusion:

This research paper trying to convey the concept of Social Networking Site, Its optimization by overcoming barrier may be showing the way how one can be benefitted from using friends social net link and put lights on concept like metrics for its performance point of view.

Reference:

[1]http://clickforseo.com/blog/2011/05/search-engine-optimization-and-social-networking-%E2%80%93-friends-with-benefits/
[2]http://www.socialmediadfw.com/services.php
[3] Imagined communities Awareness, information Sharing and Privacy on the Facebook by Alessandro Acquisti and Ralph Gross.
[4] All friends are Not Equal: using weights in Social Graphs to Improve Search by Sudheendra Hangal, Diana MacLean, Monica S. lam, Jeffery Heer
[5] Finding Experts using Internet Based Discussions in online Communities and associated social networks by John G. Breslin, Uldis Bojars, Boanerges, Almen Meza, Harold Boley , Malgorzata Mochol, Lyndon, Axel Pollers and Anna v. Zhdanova
[6] www.bukisa.com
[7] www.yoyouneeda**website**.com
[8] http://www.foafrealm.org
[9] Modelling relationship strength in online social networks. R.Xiang,J.Neville and M.Rogati.

First Author:

Dr JS Sodhi is the CIO, Head-IT designated as Assistant Vice President at AKC Data Systems Pvt. Ltd.. He is leading a team of 150 IT professionals and responsible for the overall IT management of Amity Institutions throughout 25 locations in India and AKC Group of Companies.He had his early schooling in

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Lucknow and thereafter took his Bachelor's Engineering Degree from the Pantnagar University at Nainital. Subsequently, he obtained his Post Graduate Degree in Management from the IIMS, Calcutta and Doctorate from Amity University in Information Security. He is the first Certified Security Compliance Specialist in India. He is Honorary member of Expert Group of Green IT of Council of Power Utilities. He is also Honorary member of National Advisory Board of OWASP, India besides many more committees. Dr JS Sodhi has over 20 years of Experience in Engineering and IT Project Management in organizations like NDDB, BenXL Technologies and Amity Group. He has spent over 10 years in Amity Group and managed team for implementation of various latest technological solutions like Vmware Virtualization of Servers, Blade Servers, Managed Wireless Implementation, Live Broadcast of Class Rooms, IP Surveillance Camera Network etc. He led the team for getting ISO-27001 Certification for company. His comprehensive Problem Solving Abilities, Team Building and Team Management skills, strong IT Strategies, IT Policies & Monitoring Process brings in the cutting edge experience and expertise. He has participated as distinguished Speaker at various National & International Conferences like Convergence India-2007 at New Delhi, Renvoi-2008 at New Delhi, Fortinet Seminar-2007, International Conference at Manglore in October,2008, SecurityByte Asia Pacific Conference in 2009, Intel Conference-2009 and given guest lecture to Management students. The Solution Implementation Project at Amity Network led by Dr JS Sodhi has been won for Best IT Implementation Award-2009by PCQuest and another project won the award of $95639 from HP Innovations Grant Program.

Second Author:

Ms. Shilpi Sharma is designated as Associate Professor in Computer Science and Engineering Department of Amity University from last three years. She has 13 years of experience in teaching. She had early schooling in Jaipur and Bachelor's Degree from Rajasthan University and Post Graduate degree in Computer Science from Rajasthan Vidhyapeeth. She has been All India Topper in 1995, in subject Data Structure using C and C++, in DOEACC. Published research paper in International Conference of ICCACN'11 titled security threats in Social networking Sites. She is a member of UACEE.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Email Mining: A Review

Mrs. Pranjal S. Bogawar[1], Dr. Kishor. K. Bhoyar[2]

[1]Information Technology, R.T.M. Nagpur University, Priyadarshini College of Engineering, Nagpur, Maharashtra, India

[2] Information Technology, R.T.M. Nagpur University, Yashavantrao Chavhan College of Engineering, Nagpur, Maharashtra, India

Abstract

E-mail is one of the most widely used ways of written communication over the internet, and its traffic has increased exponentially with the advent of World Wide Web. The increase in email traffic comes also with an increase in the use of emails for illegitimate purpose. Phishing, Spamming, email bombing, threatening, cyber bullying, racial vilification, terrorist activities, child pornography and sexual harassment are common examples of e-mail abuses. So, there is a need for e-mail mining. Various methods and approaches were used by the scientists for classification of email messages in above categories. In this paper we are presenting various techniques and approaches used by researchers for email mining and subsequent classification.

Keywords: *email, email mining, spam, header body, MIME, SMTP, POP*

1. Introduction

According to the survey of Radicati group from April 2010, there are about 1.9 billion users of email worldwide [1]. As the popularity of email increased, it becomes an important form of communication for many computer users, for both legitimate and illegitimate activities. Legitimate activities are like messages and document exchanges which are also misused eg., distribution of junk mails, unauthorized conveyance of sensitive information, mailing of offensive or threatening material.

Email system is inherently vulnerable to misuse for three main reasons. First, an email can be spoofed and metadata contained in its header about the sender and the path along which the message has travelled can be forged or anonymzed. An email can be routed through anonymous e-mail servers to hide the information about its origin. Second, e-mail systems are capable of transporting executables, hyperlinks, Trojan horses, and scripts. Third, the internet including email services is accessible through public places, such as net cafe and libraries. Hence there is a need of email mining. Many authors worked on emails. This paper is an overview of them.

2. E-MAIL

Electronic mail, commonly known as email or e-mail, is a method of exchanging digital messages from an author to one or more recipients. Ray Tomlinson is generally credited as having sent the first email across a network, initiating the use of the "@" sign to separate the names of the user and the user's machine in 1971, when he sent a message from one Digital Equipment Corporation DEC-10 computer to another DEC-10. The two machines were placed next to each other. Tomlinson's work was quickly adopted across the ARPANET, which significantly increased the popularity of email.

Modern email operates across the internet or other computer networks. Today's email systems are based on a store-and-forward model. Email servers accept, forward, deliver and store messages [16, 17].

Email's have specific format which is defined in RFC(Request for Comment) 5322, with multi-media content attachments being defined in RFC 2045 through RFC 2049, collectively called Multipurpose Internet Mail Extensions or MIME.

2.1. Message Format

Internet email messages consist of two major sections:

- **Header** — the message header contains control information, an originator's email address and one or more recipient addresses and descriptive information, such as a subject header field and a message submission date/time stamp which is structured into fields such as From, To, CC, Subject, Date, and other information about the email.
- **Body** — body content is unstructured text which sometimes contains a signature block at the end. This is exactly the same as the body of a regular letter.

The header is separated from the body by a blank line.

2.1.1. Message Header:

Each message has exactly one header, which is structured into fields. Each field has a name and a value. RFC 5322 specifies the precise syntax. Informally, each line of text in the header that begins with a printable character begins a separate field. The field name starts with the first character of the line and ends before the separator character ":". The separator is then followed by the field value (the "body" of the field).

Header fields: The message header must include at least the following fields:

- *From:* It is the email address, and optionally the name of the author(s).

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

- *Date:* The local time and date when the message was written.
- *Message-ID:* Also an automatically generated field; used to prevent multiple deliveries and for reference in In-Reply-To: (see below).
- *In-Reply-To:* Message-ID of the message which is a reply to. It is used to link related messages together. This field only applies for reply messages.

Along with above header fields some common header fields of email which every person is using are:

- *To*: The email addresses, and optionally name(s) of the message's recipient(s). Indicates primary recipients (multiple allowed), for secondary recipients see Cc: and Bcc: below.
- *Bcc:* Blind Carbon Copy; addresses added to the SMTP delivery list but not (usually) listed in the message data, remaining invisible to other recipients.
- *Cc:* Carbon copy; many email clients will mark email in your inbox differently depending on whether you are in the To: or Cc: list.
- *Subject:* A brief summary of the topic of the message. Certain abbreviations are commonly used in the subject, including "RE:" and "FW:"
- *Content-Type:* Information about how the message is to be displayed, usually a MIME type.
- *Precedence:* commonly with values "bulk", "junk", or "list".
- *Received:* Tracking information generated by mail servers that have previously handled a message, in reverse order (last handler first).
- *References:* Message-ID of the message that this is a reply to, and the message-id of the message the previous reply were a reply to, etc.
- *Reply-To:* Address that should be used to reply to the message.
- *Sender:* Address of the actual sender acting on behalf of the author listed in the From: field.
- *Archived-At:* A direct link to the archived form of an individual email message.[18]

2.1.2 Message Body

In the SMTP (Simple mail Transfer Protocol) standard, the body is the full email message. Most modern graphic email clients allow the use of either plain text or HTML for the message body at the option of the user. HTML email messages often include an automatically generated plain text copy as well, for compatibility reasons.

In order to ensure that HTML sent in an email is rendered properly by the recipient's client software, an additional header must be specified while sending: "Content-type: text/html". Most email programs send this header automatically.

2.2. Working of E-Mail

Email relies on two basic communications protocols: SMTP (Simple Mail Transfer Protocol), which is used to send messages and POP3 (Post Office Protocol), which is used to receive messages. A simplified version of the email life cycle can be seen in Figure 1.

The most important logical elements of the Internet Mail System are:

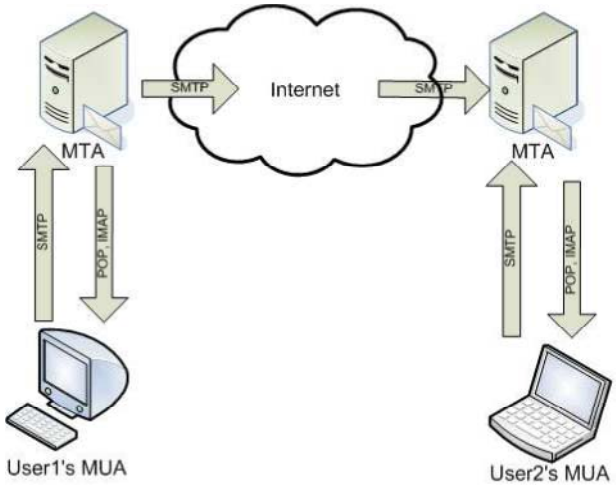


Fig 1: Life Cycle of an email [19]

1) Mail User Agent (MUA*)* – It is responsible for helping the user to read and write email messages. The MUA is usually implemented in software usually referred to as "email client". Popular email clients are Microsoft Outlook2 and Mozilla Thunderbird3, claws mail, Zimbra Collaboration Suite etc. These programs transform a text message into the appropriate internet format in order for the message to reach its destination.

2) Mail Transfer Agent (MTA**):** It accepts a message passed to it by either an MUA or another MTA and then decides for the appropriate delivery method and the route that the mail should follow. It uses the SMTP protocol to send the message to another MTA or an MDA.

3) Mail Delivery Agent (MDA): It receives messages from MTAs and delivers them to the user's mailbox in the user's mail server

4) Mail Retrieval Agents (MRA): It fetches mail messages from the user's mail server to the user's local inbox. MRAs are often embedded in email clients [19].

2.3. Filename Extensions

Upon reception of email messages, email client applications save messages in operating system files in the file system. Some clients save individual messages as separate files, while others use various database formats, often proprietary, for collective storage. A historical standard of storage is the mbox format. The specific filename extensions are in table1.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Table 1: Specific formats used for filename extension

Email Extension	***Using Agency***
eml	Used by many email clients including Microsoft Outlook Express, Windows Mail and Mozilla Thunderbird. The files are plain text in MIME format, containing the email header as well as the message contents and attachments in one or more of several formats.
emlx	Used by Apple Mail.
msg	Used by Microsoft Office Outlook and Office Logic Groupware.
mbx	Used by Opera Mail, KMail, and Apple Mail based on the mbox format.

3. Email mining

Email mining can be considered as mining of data embedded in header and or body of the email message. Various text mining techniques which extract unknown and useful information from huge set of emails can be employed to achieve email mining. Email Mining can be considered as an application of the upcoming research area of Text Mining (TM or also known as Knowledge Discovery from Textual Data) on email data.

However, there are some specific characteristics of email data that set a distinctive separating line between Email and Text Mining:

1. Email includes additional information in the headers of email that can be exploited for various email mining tasks.
2. Text in email is significantly shorter and, therefore, some Text Mining techniques might be inefficient in email data.
3. Email is often cursorily written and, thus, linguistic well-formedness is not guaranteed [19]. Spelling and grammar mistakes as well as nonstandard user acronyms also appear frequently.
4. Email is personal and therefore generic techniques are difficult to be effective to individuals.
5. Email is a data stream targeted to a particular user and concepts or distributions of target classes of the messages may change over time, with respect to the messages received by that user.
6. Email will probably have noise. HTML tags and attachments must be removed in order to apply a text mining technique. In some other cases, noise is intensively inserted. In spam filtering for example, noisy words and phrases are inserted, in order to mislead machine learning algorithms.
7. It is rather difficult to have public email data for experiments, due to privacy issues. This is a drawback especially for research since comparative studies cannot be conducted without public available datasets. An exception to the above statement is the Enron Corpus (Klimt & Yang, 2004), which was made public after a legal investigation concerning the Enron Corporation [19, 26].

Email mining is done by various researchers to extract different information from email. Some topics for which invention is done are discussed below.

3.1. Authorship Attribution

Email authorship attribution means identify the most plausible author of an anonymous email from a group of potential suspects. For author attribution various techniques used by various authors. The various topics on which work was done are gender, language, various writing styles.

Oliver de Vel et.al. used combination of stylometric, structural, gender preferential features, and language preferential features together with support vector machine algorithm to classify author's gender and language. The author also classified the language of the person as EFL (English as first language) and ESL (English as second language). Researchers also classified English and Arabic language [10]. Farkhund Iqbal et.al. gave a data mining technique to capture the write prints of every suspect and model it as combination of features that occur frequently in the suspect's email called frequent patterns. Every person has unique identity, features and writing styles. Writing patterns usually contain the characteristics of word usage, word sequence, compositions, layouts, common spelling and grammatical mistakes, vocabulary richness, hyphenation and punctuations.

3.2. Content Analysis

"Content analysis is a summarising, quantitative analysis of messages that relies on the scientific method (including attention to objectivity, intersubjectivity, a priori design, reliability, validity, generalisability, replicability, and hypothesis testing) and is not limited as to the types of variables that may be measured or the context in which the messages are created or presented."

As communication on e-mail get increased, marketing, sales, customer services and dedicated call centres required to process high volumes of emails with many messages having repetitive enquiries. Indrajit Mukerjee et.al created the automatic email answering mechanism by finding the keywords of product in email's of client.

In another research of content analysis relationships of email files were found by using keyword based matching techniques. They created a tool called Email Mining Set (EMS) for analyzing email archives which includes a graphical display to explore the relationship between users and groups of email users [14].

Appavu alias Balamurugan et.al., introduced the new Ad Infinitum algorithm to classify the threatening messages. Ad infinitum would be the extension of the decision tree induction algorithm.

Vatcharaporn Esichaikul et.al. proposed a simple and intuitive model to locate significant messages and users from an analysis of email message.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

3.3. Phishing

Phishing can be defined as a scam by which email users are duped into surrendering private information that will be used for identity theft. Phishing attacks use both social engineering and technical subterfuge to steal personal identity data and financial account credentials. It is one of the fastest growing scams on the Internet. The exclusive motivation of phishers is financial gain.
John Yearwood et. al. used the structural characteristics of the emails received by persons and the information derived on hyperlinks from 'Whois Database' for profiling of phishing emails. While generating profiles author used structural characteristics such as text content, vlinks, html content, script, table, images/logos, hyperlinks, form tag, fake tags etc.

3.4. Spam Filtering

Spam, also known as unsolicited bulk email (UBE), is becoming increasingly harmful for email traffics. Filtering is a simple and efficient way to combat against spam. Machine-learning-based classification algorithms are of excellent performance in filtering spam.

An email naturally has no label indicating whether it is a spam or a legitimate message. The label can only be set by humans for accuracy, which needs heavy manual labor. This challenge is especially significant for email service providers (ESPs) because the spam filtering system deployed on the email servers need to be retrained frequently to meet the rapidly changing spam. Figure 4 shows the email behavior model and its applications.

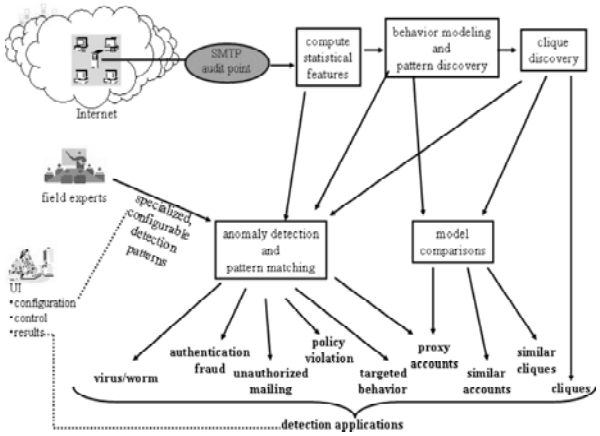


Fig. 2: Overview of email behaviour modelling, architecture and applications [20].

Yong Hu et.al suggested the fuzzy clustering method for spam and legitimate emails. They proposed spam filter consists of four components, namely, "feature extractor", "fuzzy clustering algorithm", "labeling algorithm" and "adjusting algorithm"[11].

Chun Wei et.al. is concentrating on the advanced analysis of spam emails, by considering eleven attributes from the message: message id, sender's IP address, sender email,subject, body length, word count, attachment filename, attachment_MD5, attachment size, body_URL, body_URL_domain. Some attributes were again broken. E.g. body_URL into machine name and path[13].

Salvatore J. Stolfo et.al.; gave a data mining system called EMT(Email Mining tool kit) which is used for core security applications to detect virus propagations, "spambot" activity and security policy violation. They used behaviour based analysis rather than content analysis[20, 21, 22].

4. Algorithms used in email mining

There are various algorithms used for email mining which are given below.

4.1 Support Vector Machine Algorithm:

The original SVM algorithm was invented by Vladimir Vapnik. SVM concept is based on the idea of structural risk minimisation which minimizes the generalization error. The advantage of SVM is that they do not require a reduction in number of features in order to avoid the problem of over fitting, which is useful when dealing with large dimensions as encountered in the area of text mining. It is a learning machine that classifies an input Vector X using decision function:
$f(X)= \langle X,W \rangle + b$ ……(1)
SVMs are hyper plane classifiers and work by determining which side of hyper plane classifiers and work by determining which side of the hyper plane X lies. In the above formula given in eq. no. 1 the hyper plane is perpendicular to W and at a distance b/|| W|| from the origin.
SVM maximize the margin around the separating hyper plane. The decision function is fully specified by a subset of training samples [7, 8, 23, 24, and 25].

4.2 Naive Bayes Algorithm:

The Bayesian Classification represents a supervised learning method as well as a statistical method for classification. This Classification is named after Thomas Bayes (1702-1761), who proposed the Bayes Theorem. Bayesian reasoning is applied to decision making and inferential statistics that deals with probability inference. It is used the knowledge of prior events to predict future events. Baye's Theorem says that

$$P\left(\frac{h}{D}\right) = \frac{P(\frac{h}{D})}{P(D)} \text{ .....(2)}$$

where
P(h) : Prior probability of hypothesis h
P(D) : Prior probability of training data D
P(h/D) : Probability of h given D
P(D/h) : Probability of D given h[27]

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

4.3 Clustering Algorithms:

Researchers used some clustering algorithms they are

4.3.1. Expectation Maximization (EM):

The EM algorithm was explained and given its name in a classic 1977 paper by Arthur Dempster, Nan Laird, and Donald Rubin [31].
EM is an iterative optimization method to estimate some unknown parameters Θ, given measurement data U. However, we are not given some "hidden" nuisance variables J, which need to be integrated out. In particular, we want to maximize the posterior probability of the parameters Θ given the data U, marginalizing over J:

$$\Theta = \text{argmax}_{\Theta} \sum_{J \in \mathcal{J}^n} P(\Theta, J|U)$$

The intuition behind EM is an old one: alternate between estimating the unknowns Θ and the hidden variables J. This idea has been around for a long time. However, instead of finding the best J Є $\mathcal{J}$ given an estimate Θ at each iteration, EM computes a distribution over the space $\mathcal{J}$[28, 29, 30 ,31].

4.3.2. K-Means:

K-means clustering (MacQueen, 1967) is a method commonly used to automatically partition a data set into k groups. It proceeds by selecting k initial cluster centers and then iteratively refining them as follows:
1. Each instance di is assigned to its closest cluster center.
2. Each cluster center Cj is updated to be the mean of its constituent instances [32].

4.3.3. Bisecting K-means:

It produces the clusters of the similar sizes and with smaller entropy than K-means [33].

4.3.4. Agglomerative Algorithm:

This algorithm starts with each individual item in its own cluster and iteratively merges clusters until all items belong in one cluster.

4.4 Behaviour Based Models:

1) The *user cliques* model profiles a user's communication groups that naturally occur in her or his email communication history. These cliques models provide important information to rank order the relative importance of individuals in an organization.
2) The *Hellinger distance* model profiles the distribution of the frequency of communication of users, and the variability of frequency, between a user and his/her correspondents. The recipient frequency analysis also identifies the relative importance of various email users. By extending the analysis to compute the response rates to a user's typical recipients, one can learn relative rank ordering of various people.
3) The *cumulative distribution* model profiles the rate at which a user sends emails to distinct parties in sequential order. A virus would generally not know this statistics and so would violate the user's typical behaviour while propagating itself to new victim

5. Conclusions

Email is very popular and necessity of many users. The paper gives an overview of email, its message format and working. Now a day's criminal are also using emails. So, email is considered as powerful evidence. In this paper, have presented the research carried out in the field of email mining. Researchers worked on the email body and found gender, language, writing styles of author. They found the relationships between users, threatening messages. Researchers also found the profiles of phishers. In this paper we have also presented the overview of spam messages and spam filters. Some spam filters used contents of emails; some used link information, while others used behavioural information.

6. References

[1] http://www.radicati.com
[2] Farkhund Iqbal, Rachid Hadjidj, Benjamin C.M. Fung, Mourad Debbabi, "A novel Approach of mining write prints for authorship attribution in email forensic", Digital Investigation(Elsevier Journal)2008,pp42-51
[3] Farkhund Iqbal, Hamad binsalleeh, Benjamin C.M. Fung, Mourad Debbabi, "Mining writeprints from anonymous emails for forensic investigation", Digital Investigation(Elsvier Journal)2010, pp1-9
[4] Robert Layton, Paul Watters, Richard Dazeley, "Authorship Attribution for Twitter in 140 characters or less", second Cybercrime and Trustworthy Computing workshop, 2010 IEEE, pp-1-8
[5] Nouf Al Fe'ar, Einas Al Turki, Asma Al Zaid, Mashael Ai Duwais, "E-Classifier: A Bi-Lingual Email Classification System", Information technology, ITSim2008, IEEE,Vol-2, pp-1-4
[6] Indrajit Mukerjee, Mohammad Al-Fayoumi, P.K. Mahanti, "Content Analysis based on text Mining using genetic algorithm", Second internation conference on Computer technology and Development(ICCTD 2010),IEEE, pp -432-436
[7] Oliver de Vel, Malcolm Corney, Alison Anderson, and George Mohaly, " Language and Gender Author Cohort Analysis of Email for Computer Forensics", In Proc. Of digital forensic workshop(2002),pp-1-16
[8] O.De Vel, A. Anderson, M. Corney, G. Mohaly, "Muti-Topic E-mail Authorship Attribution Forensic", ACM Conference on Computer Security- Workshop on data Mining for Security Applications, November 8,2001,pp: 1-7
[9] John Yearwood, Musa Mammadov, Arunava Banerjee, " Profiling Phishing Emails Based on Hyperlink Information", International conference on Advances in Social Network Analysis and Mining, 2010 IEEE,pp-120-126
[10] Nouf Al Fe'ar, Einas Al Turki, Asma Al Zaid, Mashael Ai Duwais, "E-Classifier: A Bi-Lingual Email Classification System", Information technology, ITSim2008, IEEE,Vol-2, pp-1-4
[11] Yong Hu, Ce Guo, Xiangzhou Zhang Zhihui Guo, Jing Zhang, Xiaojuan He, "An Intelligent Spam Filtering System Based on Fuzzy Clustering", 2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery, pp515-519
[12] Appavu alias Balamurugan, Rajaram,Muthupandian and Athiappan, " Automatic mining of threatening e-mail using

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Ad Infinitum algorithm, In International Journal of Information Technology,Vol14. No.2,2008

[13] Chun Wei, Alan Spragur, Gary Warner, Anthony Skjellum, "Mining spam Email to identify Common Origins for Forensic Applications", ACM 2008, pp: 1433-1437

[14] Hongjun Li, Jiangang Zhang, Haibo Wang, Shaoming Huang," A Mining Algorithm For Email's Relationships Based on Neural Networks",International Conference on computer science and Software Engineering ,2008 IEEE, pp- 1122-1125

[15] Appavu alias Balamurugan, Rajaram,Muthupandian and Athiappan, " Automatic mining of threatening e-mail using Ad Infinitum algorithm, In International Journal of Information Technology,Vol14. No.2,2008

[16] "The Technical Development of Internet Email" Craig Partridge, April–June 2008, pp.5

[17] "The First network Email", http://openmap.bbn.com/~tomlinso/ray/firstemailframe.html

[18] P. Resnick, "Internet Message Format", Qualcomm Incorporated, October 2008, http://tools.ietf.org/html/rfc5322

[19] Ioannis Katakis, Grigorios Tsoumakas, Ioannis Vlahavas, "Email Mining: Emerging Techniques for Email Management", 2006, pp- 1-32

[20] Salvatore J. Stolfo, Sholomo Hershkop, Chia-Wei Hu, Wei-Jen Li, Olivier Nimeskern, and Ke Wang, " Behavior Based Modeling and its Application to Email Analysis ",ACM transactions on Internet technology, Vol6, No.2 May 2006, pp187-221

[21] Salvatore J. Stolfo, Sholomo Hershkop,Ke Wang, Oliver NImeskern, Chia Wei Hu, " Behavior Profiling of Email", ISI 2003,LNCS2665, pp74-90

[22] Salvatore J. Stolfo, Shlomo Hershkop, "Email Mining Toolkit Supporting Law enforcement Forensic Analyses", ACM international conference Proceeding series(2005) , pp -221-222

[23] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin, "A Practical Guide to Support Vector Classification", http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf, April 2010

[24] Steve Gunn, "Support Vector Machines for Classification and Regression", ISIS Technical Report http://www.svms.org/tutorials/Gunn1998.pdf, 14 May 1998, pp:1-52

[25] V. Vapnik. The Nature of Statistical Learning Theory. Springer-Verlag, New York. 1995.

[26] Enron Email Dataset, http://www.cs.cmu.edu/~enron/

[27] Naive Baye's Classification Algorithm, http:// software.ucv.ro /~cmihaescu/ro/teaching/AIR/docs/Lab4-NaiveBayes.pdf ,pp1-17

[28] Frank Dellaert, "The Expectation Maximization Algorithm", http://www.cc.gatech.edu/~dellaert/em-paper.pdf,pp1-7

[29] Chuong B Do & Serafim Batzoglou, "What is the expectation maximization algorithm?", Nature Publishing Group http://www.nature.com/naturebiotechnology , 2008,pp1-3

[30] Sean Borman, "The Expectation Maximization Algorithm A short tutorial", http://www.seanborman.com/ publications/ EM_algorithm.pdf, January 09, 2009 ,pp1-9

[31] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm", Journal of the Royal Statistical Society. Series B (Methodological) Vol. 39, No. 1 (1977), pp. 1-38

[32] Kiri Wagstaff,Claire Cardie, Seth Rogers Stefan Schroedl, "Constrained K-means Clustering with Background Knowledge", Proceedings of the Eighteenth International Conference on Machine Learning, 2001, p. 577-584.

[33] Sergio M. Salveresi and Daniel Boley, "A comparative analysis on the bisecting K-means and PDPP clustering algorithms", http://www-users.cs.umn.edu/~boley/ publications / papers / savaresi04.pdf, pp1-18

[34] Devi Prasad Bhukya and S. Ramachandram, "Decision Tree Induction: An Approach for Data Classification Using AVL-Tree", International Journal of Computer and Electrical Engineering, Vol. 2, No. 4, August, 2010, pp 660-665

Mrs. Pranjal S. Bogawar has completed her B.E. (Computer technology) in 2000 from Rashtrasant Tukadoji Maharaj Nagpur University and M.E.(Computer science and Engineering) in 2009 from Sant Gadge Baba Amravati University of India. She is assistant professor in the department of Information Technology of Priyadarshini College of Engineering, R.T. M. Nagpur University at Maharashtra, India. She published 3 national, 3 international and one journal paper. Her research interests are Database Management System, Data mining and Email mining. She is Life member of Indian Society for Technical Education.

Prof. Kishor Bhoyar has completed his B.E. (Computer Science & Engineering) and M.E.(Computer Tech.) Degrees from Dr. Babasaheb Ambedkar Marathwada University and Swami Ramanad Teertha Marathwada University of India respectively in the years 1990 and 2001 respectively. He has achieved his Ph.D. degree in Computer Science and Engineering from Vishweswarayya National Institute of Technology, Nagpur, India in the year 2010. He is a professional member of ACM, Associate member of Computer Society of India and Life member of Indian Society for Technical Education. His areas of interest include Image Processing, Soft Computing and Data Mining.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Data Optimization from a Multiple Species Network Using Modified SFLA(Shuffled Frog Leaping Algorithm)

N.Kannaiya Raja, M.E., (P.hd) .,A.P[1], Dr. K.Arulanandam, Prof & Head[2], S.K. Sugunedham, (M.E)[3]

1CSE Dept. Arulmigu Meenakshi Amman College of Engg
Thiruvannamalai Dt, Near Kanchipuram ,India

[2]CSE Department Ganadipathy Tulsi's Jain Engineering
College, Vellore , India

[3]CSE Dept.Arulmigu Meenakshi Amman College of Engg ,
Thiruvannamalai Dt, Near Kanchipuram ,India

Abstract--In this work we present a novel approach that uses interspecies sequences homology to connect the networks of multi species and possible more species and possible more species together with gene ontology dependencies in order to improve protein classification for research work. Proteins are involved in many for all biological process such energy metabolism, signal transduction and translation initiation. Even though for a large portion of proteins and their biological function are unknown or incomplete, therefore constructing efficient and reliable models for predicting the protein function has to be used in research work. Our method readily extends to multi species food and produce the improvements similar to them multi species. In the presence of multi interacting networks are using data mining for integration of a data from various sources and contributing increased accuracy of the function prediction of the multiple species for research work. We further enhance our model to account for the gene ontology dependencies by linking multiple related ontology categories such as, we have selected the food items from various countries such as from America the famous food items of yoghurt and Australia food items of oats and Indian food items of soya bean. The data sets are highly desirable for this use from various countries using logical networks from center for bioinformatics research institute (Chennai) and stored in the mining.SFLA aims to set a generic paradism of the efficient mining that acquire the data set of proteins for these food items and promotes predictions of protein functions with gene ontology for research work.

Keywords— Biology and genetics, machine learning, bioinformatics (genome or protein) databases.

1. Introduction

Protein-protein interactions (PPIs) can reveal insights on biological regulatory path ways and metabolic processes. a complete and reliable protein interaction map provides us with an opportunity to understand the basic biological processes within a cell. Every node have protein-protein networks, which represents protein between nodes should represent a protein and edges between nodes represents different types of function has been assigned, such as protein-protein interactions, sequence similarity, co-expression patterns and others. in biological research, various experiments have been developed to map the interactions among the proteins, such as mass spectrometry, yeast two-hybrid tandem affinity purification and co-immuno prediction and chain graph probabilistic approach [1][2][3]. The computational methods for protein classification depend on nearest neighbors in a protein-protein network, can share related functions. These methods are appointed the functions to a protein of interest based on the annotation of its neighbors. Those approaches have accomplishment in cases where protein have multiple, [4][5][6]mostly annotated neighbors. On the other hand the method shows much less success on protein with inadequate neighborhoods those proteins isolated in their own network or the ones surrounded by weakly annotated neighbors. In this paper, we propose a new approach to protein function prediction which trounce the drawbacks of these method and illustrates interspecies sequences homology to connect the networks of multi species and possible more species together with multifunctional gene ontology (GO) dependencies. The primary conceptual improvement of our method is to connect protein-protein networks of two or more different food items,

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

but related multiple species, into a single-computational model. During the edges of high homology, proteins can increase their learning neighborhood and acquire extra functional information from their neighbors-homolog's of a different food items networks.

Our new and promising technique, the modified shuffled frog leaping algorithm for classifying the protein functions from various sources with the use of interspecies sequence similarity and integrate multiple sources of information. In connecting networks we rely on the fact that proteins of different food items from various countries, which share high sequence similarity, are likely to share similar protein classification. Thus, high similarity of sequences between species is likely to lead to shared functions. Even though the resulting shuffled frog leaping algorithm used the data mining with leave-one-out cross validation is used to evaluate classification accuracy when applied to the datasets, compared all multi species protein functions. We demonstrate that the combined models often lead to efficient implementations and significant improvements in predictive accuracy not experimental in isolated networks or other competing approaches.
Additionally we propose an efficient evolutionary approach for selecting protein subsets from stored mining data's, that effectively achieves higher accuracy for classification problems. The purpose of classification is to build an efficient model for predicting protein functions [6]. We embedded a genetic algorithm in a shuffled frog leaping algorithm and to serve as a local optimizer for each generation in implementing feature selection. To evaluate the SFLA and GA results showed that our proposal achieves superior classification accuracy when applied to the data sets. The SFLA is derived from a virtual population of protein functions in which individual proteins are equivalent to the GA chromosomes, and represent a set of solutions. Each protein is distributed to a different subset of the whole population called a memeplex. An independent local search is conducted for each protein memplex, in what is called memeplex evolution. After a defined number of memetic evolutionary steps, protein functions are shuffled among memeplexes [10][11], enabling proteins to interchange messages among different memplexes and ensure that they move to an optimal position. Local search and shuffling continue until defined convergence criteria are met. SFLA have demonstrated effectiveness in a number of global optimization problems difficult to solve using other methods.

The remaining portion of the paper as follows: in section 2, we first present a general view of closely related network approaches to protein function prediction. then we goes in section 3, the modified shuffled frog leaping algorithm that combines both go structure and the information from protein-protein networks of multiple species. section4 illustrates the effectiveness of the proposed approach when implemented to yoghurt, oats and soya beans networks, at different granularities of the go(gene ontology). We discuss the new results and analysis in section5. Finally in section 6 conclusion and future enhancement.

2. Related Work

Proteins and are used in many biological process such as energy gains and DNA-rna metabolism, signal transduction and translation, initialization. However the large portion of protein and their biological function cannot be used in maximum in biological research work. Therefore we have proposed constructing efficient and reliable model for predicting protein function which has the task of our model. For this purpose we have selected three food items for prediction of protein functions[1][2]. The computational biology described a enormous methods for understanding the importance of protein-protein interactions. In these methods, supervised learning is a leading approach. The state-of-the-art supervised learning methods include k-nearest neighbor, support vector machines (SVMS),[4] Random forest and so on. Using learning classifier positive examples of truly interacting protein pairs and negative examples of non-interacting protein pairs, to predict a not noticed relationship between two proteins. Each protein pair is fixed as a feature vector in the data. This attempt has been tried on developing informative and effective feature representation methods for PPI prediction. Feature vectors may be extracted based on protein sequences directly or may involve indirect evidences, including domain compositions, motif pairs and related mRNA expression. Compositions of amino acids and physiochemical descriptors based SVM method is used by bock and gouge. Urquiza et al. extracted 26 genomic or proteomic features of yeast from various databases for each pair, such as information of protein domains, domain-domain interactions in proteins whose 3D-structures are known and high quality annotation of gene ontology. Espadaleret al. considered protein structural similarities among domains found in the databases of interacting proteins, combining maintenance of pairs of sequence patches based on the observation that structural evidence has shown that usually interacting pairs of close homolog's physically interact in the same way. Numerous methods are facilitate to concluded protein interactions based on the conservation of gene neighborhood, conservation of gene order, gene fusion events, and the co-evolution of interacting protein pair sequences. The valuable facts of protein interaction networks across organisms. A sensitive idea is to utilize the interaction map of one organism as a pattern to predict interactions in another . The 'interaction domain profile pairs' method (IDPP) applied to protein interaction datasets of Escherichia coli and Helicobater pylori by Wojcik and Schachter. The IDPP method required a high quality protein interaction map to be real world, protein interaction datasets are frequently sparse,

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

incomplete and deafening. The researchers in machine learning develop a matrix factorization based methods for reassign learning. An important application is recommendation systems, where collaborative filtering is modeled using matrix factorization, moreover countenance the problem of network sparsity. We are interested in predicting links between nodes, which corresponds to binary connectivity, where the proteins have different semantics from users and products. The family probabilistic graphical models utilized by computational biology for protein function prediction, such as hopeful networks, to conclude the functions datasets of incomplete annotated proteins. With the use of partial knowledge of functional observation, probabilistic assumption is to determine other proteins unknown functions of momentary on and accumulating undecided information on a huge sets of related proteins. The preference of functional association between proteins of the network models performance impacts by a critical factor. The most traditional methods are based on sequence similarly using blast. The large set of method relies on the truth that similar proteins are likely to share common functions, sub cellular location, or protein- protein interactions. The same kind of similarity based methods take account of sequence homology similarity in short signaling motifs, amino acid composition and expression data. The PPI data to established protein function within a own network. For example Markov random filed method is to define the complete set of proteins involved in PPI[17][18]. These methods are based on the concept that interacting neighbors in networks might also share function. The approach of incorporating gene ontology structure into probabilistic graphical models has gave the possible results for predicting protein functions. The approach considers multiple functional discipline in the gene ontology (GO) simultaneously. In this model, each protein is represented by its own annotation space - the go structure. The information is distributed within the ontology structure as well as between neighboring proteins, leading to an added ability of the model to explain potentially tentative single term predictions. Markov random field (MRF)[12] model for protein function prediction using protein-protein interaction data. Two main features differentiate the MRF based methods from other guilt-by-association methods. One is that the MRF model uses global information on the whole interaction network as a replacement for of local interaction network. the method was applied to predict protein function based on "cellular role" using protein functions defined in yeast proteome database (YPD)[14][15]. The results showed that the MRF based method outperforms the two guilt-by-association based methods. Features of individual proteins have long been used for protein function prediction. A feature here refers to an surveillance about a protein. It can be the presence or absence of a motif signal, its isoelectric point, its absolute mRNA expression level, or mutant phenotypes from experiments about the sensitivity or resistance of disruption mutants under various growth conditions. features were used for protein function prediction as pattern recognition problems. Drawid and Gerstein developed a general Bayesian approach to predict protein localization based on a large number of features of individual protein. we broaden the MRF[10][11] based method to an integrated approach including other protein pair wise relationship such as correlations of gene expression patterns, genetic interactions, and features of individual proteins such as domain information. The model is bendable that other protein pair wise relationship information such as pair wise protein sequence similarities and features of individual proteins can be easily incorporated. The integrated approach to predict functions of yeast proteins based on MIPS protein functions and the interaction networks based on MIPS physical and genetic interactions, gene expression profiles, tandem affinity purification (tap) protein complex data, and protein domain information. We study the sensitivity and specificity of the integrated approach using different sources of information by the leave-one-out approach. Compared to using MIPS physical interactions only, the integrated approach combining all the information increases the sensitivity from 57% to 87% when the specificity is set at 57%, an increase of 30%. it should also be noted that by enlarging the interaction network, the number of proteins whose functions can be predicted is also greatly increased. Many learning approaches rely on information available from neighbors in a protein network. however, two largest protein networks of yeast and fly as well as joint yeast-fly network. predictive performance[2][3][4] of network models is evaluated in a five-cross-validation setting. the experiment set consists of a 20 percent same proportion of negatively and positively annotated proteins, as the left behind 80 percent of the data used for training the model. For each arbitrarily chosen test protein, all of its annotations are left out—the gene ontology structure remains in place but the functions at all terms now listed as unknown. in the case of a joint fly-yeast network, we eradicate annotations of 20 percent of annotated proteins from each network. in the testing phase, upon junction of the message-passing process, predictions at terms whose annotations were left out are tested against the known eliminated annotations. We conducted a 10 experimental rounds using the random splitting process for each tested network, and compared results of runs on single networks (without joining) to that of the joint network. individual and joint networks are skilled and evaluated on the same training/testing data. normalized BLAST scores measures the intra and interspecies similarity, and its divided by self score of query (i.e., blast score of the homolog divided by the blast score of the protein against itself), ranging from 0 to 1.From the saccharomyces Genome database for yeast and fly base for fly we got sequence and annotation data. Same as protein-protein interaction data were obtained from biogrid database. This evidence resulted in a combined se of fly and yeast proteins that were used to construct the joint hopeful networks.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Gene ontology structure downloaded from gene ontology databases. while analysis gene ontology annotation, we consider two basic go assumptions: GO hierarchy is expanded up for positively annotated proteins (if a protein is positively annotated to a term, then it is also positively annotated to all of its parents/ancestors) and is expanded down for negatively annotated proteins (if a protein is negatively annotated to a term, then it is negatively annotated to all of its children/ descendants). We construct a negative set relying on co-annotation (co-occurence) statistics of GO annotations in the data (further maintaining two fundamental go assumptions). In exacting, a protein is considered negatively annotated to a definite GO term if this term has never been observed to co-occur with a known function for this protein in the training data. This method can be applied to the entire gene ontology, at the cost of time and space complexity. Though, in practice, biologists and clinicians are interested in precise, relatively small, subontologies, targeted in our study.

Data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining technique is one of a number of analytical tools for analyzing data. it allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. While large-scale information technology has been evolving separate[14] transaction and analytical systems, data mining provides the link between the two. Data mining software analyzes relationships and patterns in stored transaction data based on open-ended user queries.

Data mining consists of five major elements:
1. Extract, transform, and load transaction data onto the data warehouse system.
2. Store and manage the data in a multidimensional database system.
3. Provide data access to business analysts and information technology professionals.
4. Analyze the data by application software.
5. Present the data in a useful format, such as a graph or table.

TECHNIQUES OF DATA MINING

Data mining involves following four techniques:

Clustering - is the task of discovering groups and structures in the data that are in some way or another "similar", without using known structures in the data.[14][15]

Classification - is the task of generalizing known structure to apply to new data. for example, an email program might attempt to classify an email as legitimate or spam. Common algorithms include decision tree learning, nearest neighbor, naïve bayesian classification,neural networks and support vector machines.

Regression - attempts to find a function which models the data with the least error.

Association rule learning - searches for relationships between variables. For example a supermarket might gather data on customer purchasing habits. Using association rule learning, the supermarket can determine which products are frequently bought together and use this information for marketing purposes. This is sometimes referred to as market basket Analysis

3. Method

Multi species method is to join networks of two (or more) species by edges of high sequence similarity into one computational model, as shown in fig. 1. In exacting, an edge is introduced between homologous proteins in two species if their normalized blast score is above 0.5 (the similarity is high). On the other hand, interspecies edges are not introduced [1][2]when the score is below 0.5 (the similarity is low), since dissimilar proteins may or may not be involved in the same biological process. Furthermore, most of the protein pairs would share some low similarity, which would obscure the network with potentially irrelevant low similarity edges. We take a three-species setting, and we classify a similarity measure between protein i in yoghurt network and protein y in oats network, and protein z in soya bean network at term c, as $s_{i,y,c}^{between}$ $s_{i,z,c}^{between}$ a normalized pair wise blast score. Consequently, the potential function for homolog's between different species is defined as if the relationships between children and parents are directional, the protein function is positively annotated to a child as well as positively annotated to a parent. The reverse relationships does not hold but the parent have negatively annotated protein function, it will be negatively annotated to all children terms. By the definition clearly known where the presence of multiple parents in negative state,it immediately yields a negative state of child.

$$\omega(+,+) = \omega(-,-) = s_{i,y,c}^{between}$$

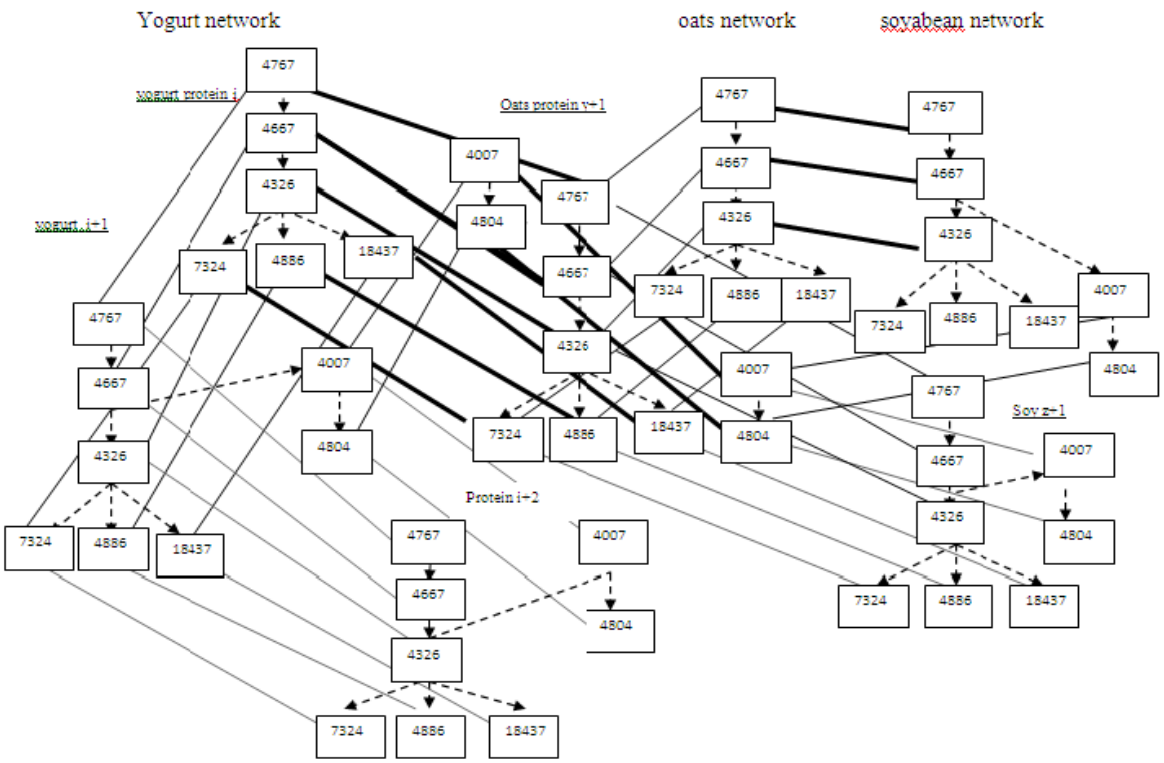
$$\omega(+,-) = \omega(-,+) = 1 - s_{i,y,c}^{between}$$

$$\omega(+,+) = \omega(-,-) = s_{i,z,c}^{between}$$

$$\omega(+,-) = \omega(-,+) = 1 - s_{i,z,c}^{between}$$

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

bond angles ω, the creation of these structures neutralizes the polar groups on each amino acid. the secondary structures are tightly packed in the protein core in a hydrophobic environment. each amino acid side group has a some degree of volume to occupy and a limited number of possible interactions with other near- by side chains, a situation that must be taken into account in molecular modeling and alignments. While this assumption maybe open to debate, it is shown to lead to improved annotation performance. Considering heterogeneous values of similarity $s_{i,y}^{between}$ at each term c may lead to additional improvements, at a cost of a more complex and demanding parameter estimation process. The combined model for joint yoghurt-oats-soya bean (referred to as species 1 and 2) network now defines a joint Gibbs distribution of functional term annotations over a set of all proteins in the shuffled frog leaping, detailed in (2). Here, z is the normalizing constant, ω_{within} is similarity measure within one species network, $\omega_{between}$ is a similarity measure between multiple species network. After the joint network is built, the confidence broadcast is used to compose predictions at all ontology terms in both species. We consider a state of convergence and decision thresholds. Adding interspecies homology information into the learning model has exclusive advantages and shows noteworthy Improvements in protein function prediction.The model is specifically beneficial for proteins isolated in their own networks (having no interaction neighbors) or for proteins which are surrounded by poorly annotated neighbors. in a multispecies setting, the neighborhood of such proteins is expanded so that they can learn their functional annotations from their homologs in the different species



Basic SFLA

The SFLA has four main stages: generating initial population, partition, local search and shuffling.

Generating initial population

Generate p frogs as initial population arbitrarily. Each Frog (correctly, the position of a frog) represents a sufficient solution of the problem, Frog i is denoted as $x_{i=}x_{i1}, x_{i2}, \ldots \ldots . . x_{is}$ where S is the space dimension of a solution. later than that, calculate the fitness for each frog.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Partition

The Frogs are sorted in descending order according to their fitness. then partition the entire population into m memeplexes (subsets), each one containing n frogs. in this scheme, put the first memeplex, the mth frog into the mth memeplex, and the (m +1) th frog back into the first memeplex, etc. finally, all the frogs are put into m different memeplexes and p (m×n) . Here, the different memeplexes symbolize different cultures of the specified frogs.

Local search

The local search process improves the average fitness of the memeplexes by serving the frogs with the worst fitness in the subsets communicate with the frogs by way of local, even global best fitness. the upgrading process is described as follows. to a given memeplex, find out the frogs with the best and the worst fitness and categorize them as x_b and x_w respectively. Also find out the frog with the global best fitness within the whole population and identify it as x_g . Then apply a process similar to particle swarm optimization to improve only the frog with the worst fitness (i.e. x_w) in each cycle. Accordingly, the position of the frog with the worst fitness is adjusted as follows:

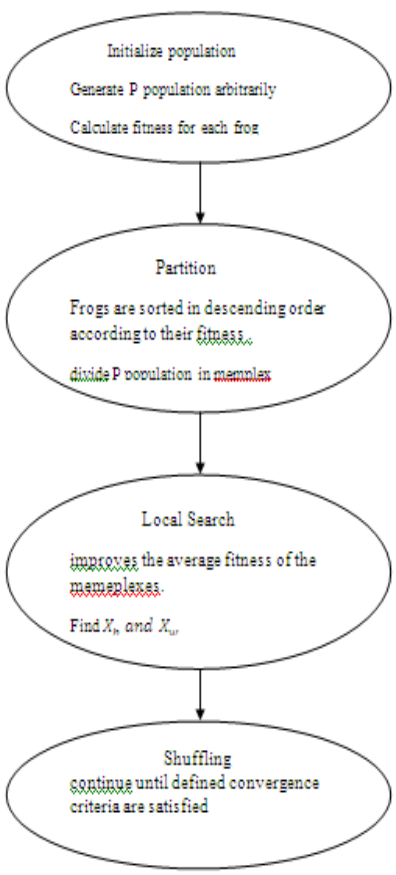


Figure2 Stages in SFLA

where rand() is a random number between 0 and 1; d_{max} is the maximum allowed change in a frog's position. If the fitness of new x_w is better than that of current x_w , then replace current x_w with new x_w . Otherwise, replace x_b with x_g and repeat the calculation according to (1) and (2). If no improvement is made in this case yet, then randomly generate a new frog to replace the current x_w . After the memetic change for the current frog in the memeplex, sort the frogs of the memeplex in order of decreasing performance value and repeat local iteration shuffling.

The next step of memetic evolution steps within each memeplex, ideas are agreed among memeplexes in a shuffling process which helps to increase the quality of the memes after being infected by frogs from different cultures[6]. The partition, the local search and the shuffling processes continue until defined convergence criteria are satisfied. Usually, the convergence criteria can be defined as follows:

The qualified change in the fitness of the best frog within a number of successive shuffling iterations is less than a pre-specified tolerance. The main parameters of SFLA are: number of total frogs, number of memeplexes, number of generation for each memeplex before shuffling, number of iteration, and maximum step size.

procedure msfl

Randomly initialize population of *p* frogs;
for *g*=1 to *gshuff*
begin
calculate the fitness of all frogs;
sort all frogs in order of descending fitness and do the memeplex
division according to (1);
for k=1 to m do
begin
for *j*=1 to *gmeme* do
begin
set *xbk* ,*xwk* and *xg*;
do local evolution according to (4)(5);
if the evolution can't produce a better solution,
(4)(5) are repeated
but with *xbk* replaced by *xg* ;
else if still no improvement, randomly generate a new solution;
resort the *k*-th memeplex in order of decreasing fitness;
end;
end;
re-shuffling the entire frogs;
end

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

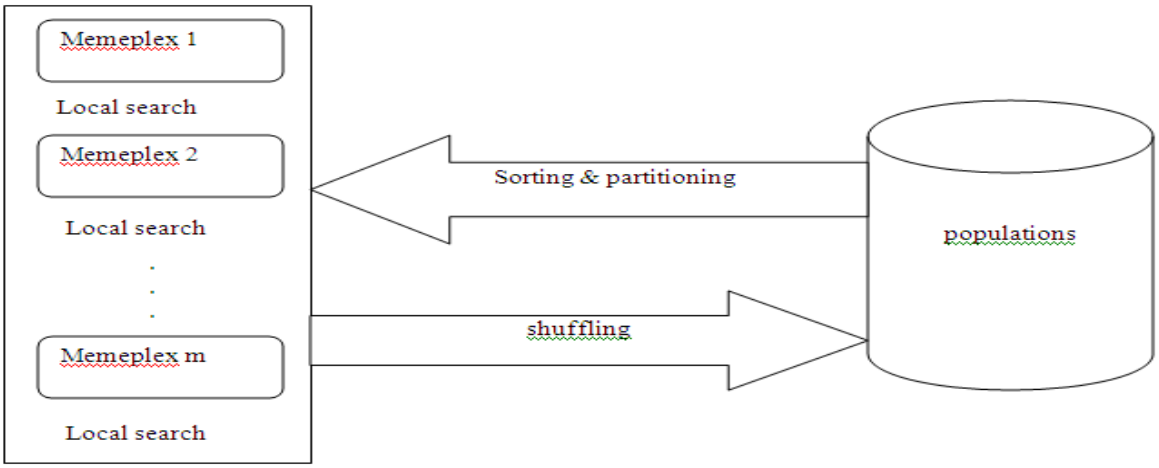


Figure 3 Construction of MSFLA

4. Experiments and Results

As a new and hopeful technique, the Modified shuffled frog leaping algorithm for multi species networks . We apply our method to three largest joint protein networks of yogurt, oats and soya bean , for optimizing the data from multispecies networks. Initially we discussed the two search techniques: the local search of the "particle swarm optimization" technique and the competitiveness mixing of information of the "shuffled complex evolution" technique . with each other. So the frogs stand for the protein functions here. Frog i can be written as P_i =($p_{i1}, p_{i2}, \dots \dots \dots p_{is}$) where s is total number of food items. The decision variable F_{ij} (j=1,2,…….s) is the number equivalent to a food items. Here we took three food items so, s=3 then calculating the fitness function ,it evaluate the protein's position and returns a single numerical value, and higher return value, the better the protein The SFLA is a heuristic search algorithm.Experiments was carried out by using different numbers of food items from various countries from America the food item of yogurt. Australia food item of oats and Indian food item of soya bean. In the joint yogurt-oats-soya bean networks the best solution of protein function is to be selected for protein through interaction together function.

Now a protein function starts to find the nearest neighbor for making strong link between each network. According to this simple near neighborhood rule, we can generate S food items as part of intial population, in which the first resolution variable is 1,or 2 or S. The second variable is the number which represents the closest protein function to the first protein. The third variable is the number which represents the nearest protein function among the protein networks has not been connected and next to second protein node.

The protein functions are sorted in descending order according to their fitness and then partitioned into subsets called as memeplexes (m). Within each memeplex, the frog with worst and best fitness are identified as x_b , x_w protein function with global best fitnees is identified as x_g the protein function with worst fitness is improved according to above equation. We want to calculate intra- and interspecies similarity, used BLAST scores, and defined as a BLAST score divided by character score of protein functions, its score of the protein against itself ranging from 0 to 1. We obtained sequence and annotation data from NCBI genome database [3] for yogurt, oats and soya-bean. NCBI provides various levels of computation, analysis, and Curation as needed per organism. Blast results, as well as the sequence features, are readily displayed on NCBI'S map viewer. Protein-Protein interaction data were obtained from string [7] database. Network visualization of the protein interactome where each point represents a protein function and each line between them is an interaction. This resulted in a combined set of 13,200 yogurt,5,004 oats and 6,008 soyabean proteins that were used to construct our joint belief networks. To generate the database to mine we formed a single deductive database of genes and their known functional assignments.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

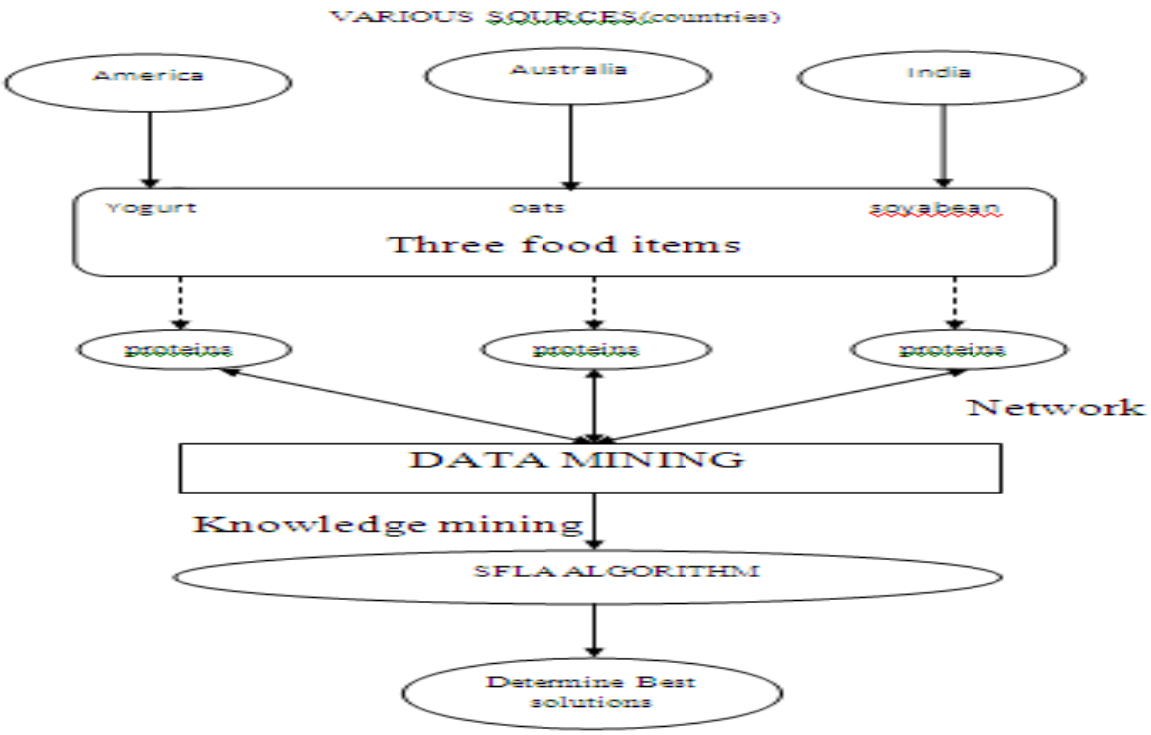


Figure 4 Architecture Diagram

We then processed this data to form sequence descriptions of the genes. the most commonly used technique to gain information about a sequence is to run a sequence similarity search, and this was used as the starting point in forming descriptions. The most commonly used technique to gain information about a sequence is to run a sequence similarity search, and this was used as the starting point in forming descriptions. Such rules provide a way of predicting function in the absence of recognizable sequence homology. The data mining approach described is extendable to analysis of other forms of bio informatics data, such as expression profiles, pathway analysis,structural studies.

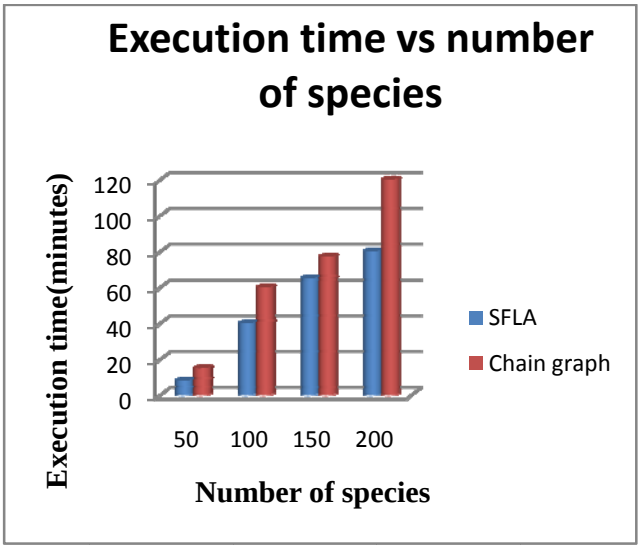


Figure5 Execution time vs Number of species

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

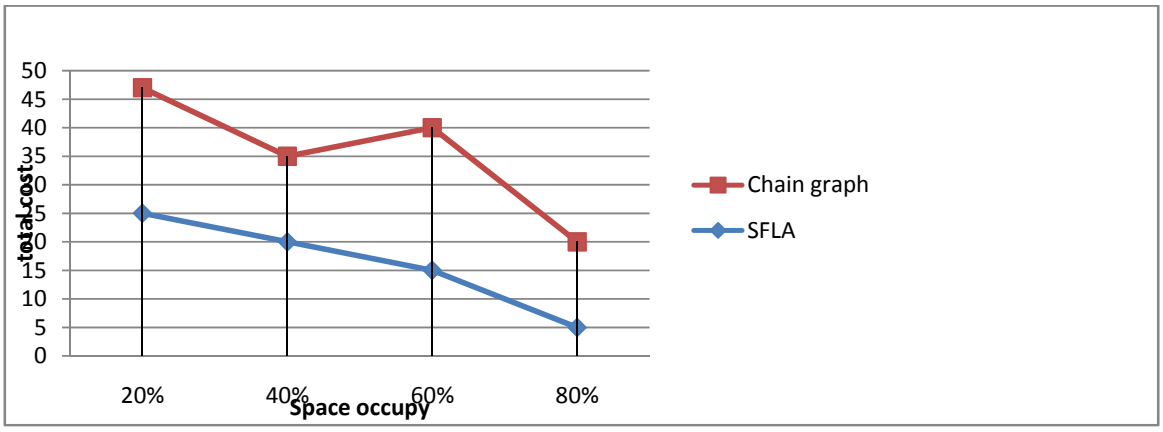
Table1 The number of protein function found are those selected on the validation set. A rule predictsmore than one homology class if there is more than one sequence similarity. A rule predicts a new homology class if there is a sequence similarity cluster in the test predictions that has no members in the training data. average test accuracy is the accuracy of the predictions on the test proteins of assigned function (if conflicts occur, the prediction with the highest *a priori* probability is chosen). Default test accuracy is the accuracy that could be achieved by always selecting the most populous class.

Networks	Yogurt	oats	soya-bean
number of protein functions found	27	32	21
rules predictiong more than one homology class	19	18	8
rules predicting a new homology class	14	16	2
average test accuracy	66%	65%	73%
default test accurach	55%	20%	6%

Table2.Number of species Execution time comparison

Number of Species	Execution Time of Chaingraph	Execution Time of SFLA
50	12.22 min	1.34 min
100	25.34 min	3.02 min
150	52.77 min	7.33min
200	1.65 hour	12.2 min
above 200	infinite	Finite

Figure6 Total cost with respect to space occupy



Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Table3 classified protein function values in SFLA and

Func-tion	Dimen-sion	Best		Mean Best		Std. Dev	
		SFLA	MSFLA	SFLA	MSFLA	SFLA	MSFLA
f_1	10	8.24E-18	9.68E-24	3.44E-16	8.65E-23	3.91E-16	6.31E-23
	20	2.74E-10	1.28E-20	2.41E-09	6.20E-20	1.70E-09	3.19E-20
	30	7.54E-07	4.45E-19	3.33E-06	1.22E-18	2.16E-06	5.78E-19
f_2	10	0	0	1.01	0	9.65E+01	0
	20	3.98	0	7.90	0	2.08	0
	30	7.96	0	1.60E+01	0	4.25	0
f_3	10	7.27E-10	0	4.0E-09	0	2.31E-09	0
	20	2.63E-06	0	1.10E-05	0	5.33E-06	0
	30	1.91E-04	0	4.18E-04	0	2.07E-04	0
f_4	10	1.99E-02	0	7.30E-02	1.02E-02	3.47E-02	5.17E-02
	20	1.47E-08	0	2.74E-02	0	2.59E-02	0
	30	5.70E-06	0	9.32E-03	0	1.20E-02	0
f_5	2	1.60E-16	0	9.09E-16	0	5.82E-16	0

5. Conclusion

In this paper, we proposed a modified SFLA techniques to classified the protein functions with joining networks of three different species. Experimental results on the standard test data sets show that the proposed algorithm is an effective and efficiency algorithm for optimizing data from multispecies networks. The use of the Gene Ontology enables synchronized deliberation of multiple but related functional categories,forfurther improvements to the model's predictive ability.

Our method readily extends to multiple, but not related species setting . The presence of multiple interacting networks may further enable integration of additional sources of evidence, thus contributing to increased accuracy of functional predictions.

References

[1] http://www.geneontology.org/, 2010.

[2] Antonina Mitrofanova, Vladimir Pavlovic, And Bud Mishra"Prediction Of Protein Functions With Gene Ontology And Interspecies Protein Homology Data" IEEE/ACM Transactions On Computational Biology And Bioinformatics, Vol. 8, No. 3, May/June 2011.

[3] U. Karaoz et al., "Whole-Genome Annotation by Using Evidence Integration in Functional-Linkage Networks," Proc. Nat'l Academy of Sciences USA, vol. 101, pp. 2888-2893, 2004.

[4] A. Vinayagam, R. Konig, J. Moormann, F. Schubert, R. Eils, K.-H. Glatting, and S. Suhai, "Applying Support Vector Machines for Gene Ontology Based Gene Function Prediction," BMC Bioinformatics, vol. 5, p. 116, 2004.

[5] J. Liu and B. Rost, "Comparing Function and Structure between Entire Proteomes," Protein Science, vol. 10, pp. 1970-1979, 2001.

[6] Alireza R.V., Ali Hossein M, "Solving a bi-criteria permutation flowshop problem using shuffled frog-leaping algorithm," Soft Computing, vol. 12(5), pp. 435-452, 2008.

[7] Eusuff M. M, Lansey K.E, "Shuffled frog-leaping algorithm: A memetic meta-heuristic for discrete optimization," Engineering Optimization, vol. 38 (2), pp. 129-154, 2006.

[8] N. Nariai, E. Kolaczyk, and S. Kasif. Probabilistic protein function prediction from heterogeneous genome-wide data. PLoS ONE, 2(3), 2007.

[9] Enright, A.J., Iliopoulos, I., Kyrpides, N.C. & Ouzounis, C.A. Protein interaction maps for complete genomes based on gene fusion events. *Nature* 402, 86–90 (1999).

[10] Altschul,S.F., Gish,W. *et al.* (1990) Basic local alignment search tool. *J. Mol. Biol.*, 215, 403–410

[11] Vazquez, A. Flammini, A. Maritan, and A. Vespignani. Global protein functionprediction from protein-protein interaction networks. *Nature Biotechnology*, 21(6): 697–700, 2003

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

[12] von Mering, R. Krause, B. Snel, M. Cornell, S. G. Olivier, S. Fields, and P. Bork. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature*, 417:399–403, 2002.

[13] G. Yona, N. Linial, and M. Linial. Protomap: Automatic classification of protein sequences, a hierarchy of protein families, and local maps of the protein space. *Proteins: Structure, Function, and Genetics*, 37:360–678, 1999.

[14] Han, J., Kamber M. "Data Mining: Concepts and Techniques", Morgan Kaufmann Publishers, (2006).

[15] Jensen, L., et al. "Prediction of Human Protein Function from Post - Translational Modifications and Localization Features", Journal of Molecular Biology, vol. 319(5), (2002) pp 1257-65.

[16] E.Elbeltagi, T.Hegazy, And D.Grierson, "comparison among five evolutionary-based optimization algorithms," advanced engineering informatics,vol.19,pp.43-53, january 2005.

[17] A. R. Rahimi-Vahed And A. H. Mirzaei, " solving a bi-criteria permutation flowshop problem using shuffled frog-leaping algorithm," soft computing, vol.12(5), pp. 435–452, december 2008.

[1]Mr.N.Kannaiya Raja received degree MCA from Alagappa University and ME from Anna University Chennai in 2007 joined assistant professor in various engineering colleges in Tamil Nadu affiliated to Anna University and has eight years teaching experience. His research work in deep packet inspection. He has been session chair in major conference and workshops in computer vision on algorithm, network, mobile communication, image processing papers and pattern recognition. His current primary areas of research are packet inspection and network. He is interested to conduct guest lecturer in various engineering in Tamil Nadu.

[2]Dr.Mr..KArulanandam received PhD doctorate degree in 2010 from Vinayaka Missions University. He has twelve years teaching experience in various engineering colleges in Tamil Nadu which are affiliated to Anna University and his research experience network, mobile communication works, image processing papers and algorithm papers. Currently working in Ganadipathy Tulasi's Jain Engineering College Vellore .

[3]S.K.Sugunedham received degree B.Tech Computer Science Engineering from Pondicherry University Puducherry in 2010. Now pursuing ME Computer Science and Engineering in Arulmigu Meenakshi Amman College of Engineering Kanchipuram affiliated to Anna University Chennai.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Fuzzy ID3 Decision Tree Approach for Network Reliability Estimation

Ak. Ashakumar Singh[1], Momtaz Thingujam[2]

[1]Department of Computer Science, Thoubal College, Manipur University,
Manipur-795138, India.

[2] Department of Computer Science, Manipur University, Imphal,
Manipur-795001 India.

Abstract

The Computer communication networks have been rapidly increasing recently to share expensive hardware and software resources, and provide access to main system from distant locations. The reliability and the cost of these systems are important considerations that are largely determined by placement of the nodes and the links between nodes. Network Reliability Estimation is a process of predicting the efforts and cost in terms of money, schedule and staff for any network system. In this study, we investigate the use of Fuzzy ID3 decision tree for network reliability estimation; it is designed by integrating the principles of ID3 decision tree and the fuzzy set-theoretic concepts, enabling the model to handle uncertain and imprecise data when describing the reliabilities of the networks, which can improve greatly the accuracy of obtained estimates. MMRE is used as measures of prediction accuracy for this study. The results are compared with those produced by the crisp version of the ID3 decision tree.

Keywords: *Network reliability, network design, Decision Tree, Fuzzy ID3, Fuzzy Entropy.*

1. Introduction

The design of reliable communication networks is a significant problem in the telecommunications industry. Improving the accuracy of the reliability estimation models available to project managers would facilitate more effective control of time and budgets during network design. An important stage of network design is to find the best layout of components to minimize cost while meeting a performance criterion such as transmission delay, throughput or reliability [1]. Generally, a large scale network has a multilevel and hierarchical structure consisting of a backbone network and several local access networks [2]. This paper is focused on large scale backbone communication network design where the relevant reliability metric is all terminal network reliability [3].

In this paper, we are concerned with network cost estimation models based on fuzzy decision trees especially Fuzzy Interactive Dichotomizer 3.

There are three major advantages when using estimation by decision trees (DT). First, decision trees approach may be considered as "white boxes", it is simple to understand and easy to explain its process to the users, contrary to other learning methods. Second, it allows the learning from previous situations and outcomes. The learning criterion is very important for cost estimation models because network design technology is supposed to be continuously evolving. Third, it may be used to feature subset selection to avoid the problem of cost driver selection in network cost estimation model.

On the other hand, fuzzy logic has been used in network cost estimation. It's based on fuzzy set theory, which was introduced by Zadeh in 1965 [4]. Attempts have been made to rehabilitate some of the existing models in order to handle uncertainties and imprecision problems. The aim of this study is to evaluate and to discuss the use of fuzzy decision trees, especially the fuzzy ID3 algorithm in designing DT for network reliability estimation. Instead of crisp DT, fuzzy DT may allow to exploit complementary advantages of fuzzy logic theory which is the ability to deal with inexact and uncertain information when describing the network designing.

The paper is organized as follows: In section 2, we present the fuzzy ID3 decision tree for network reliability estimation. The description of datasets used to perform the empirical studies and the evaluation criteria adopted to measure the predictive accuracy of the designed models are given in section 3. Section 4 focuses on the experimental design. In Section V, we present and discuss the obtained results when the fuzzy ID3 is used to estimate the network designing. A comparison of the estimation results produced by means of the fuzzy ID3 Models and the crisp ID3 model are also provided in section 5. A conclusion and an overview of future work conclude the paper in the last section 6.

2.Fuzzy ID3 for Network Reliability Estimation

Based on the Concept Learning System algorithm, Quinlan proposed a decision tree called the Interactive Dichotomizer 3 (ID3). The ID3 technique is based on information theory and attempts to minimize the expected number of comparisons.

The fuzzy ID3 is based on a fuzzy implementation of the ID3 algorithm [5,6]. It's formed of one root node, which is the tree top, or starting point, and a series of other nodes. Terminal nodes are leaves (effort). Each node corresponds to a split on the values of one input variable (cost drivers). This variable is chosen in order to reach a maximum of homogeneity amongst the examples that belong to the node, relatively to the output variable.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

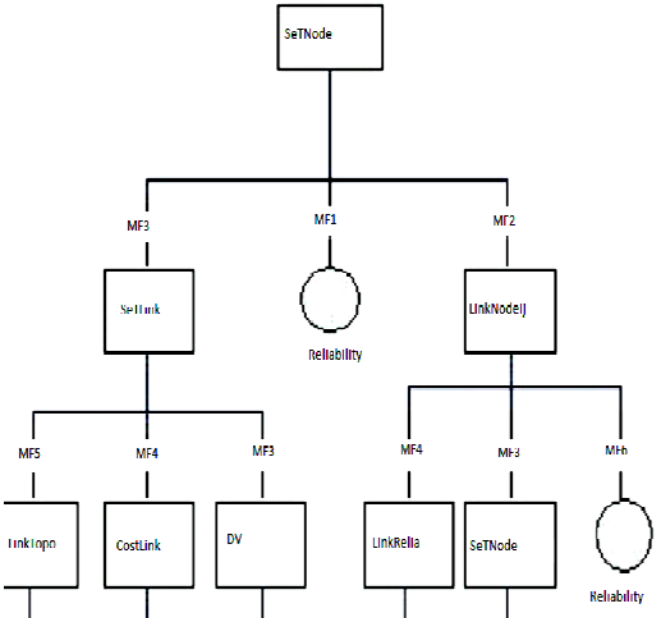


Fig. 1 An example of fuzzy ID3 decision tree for network reliability estimation

.

The major characteristic of fuzzy ID3 is that an example belongs to a node to a certain degree. The proportion of P_k^n of examples with k classification at node n is calculated using the membership degrees as follows:

$$P_n^k = \frac{\sum_{i=1}^{N} u_k(y_i) \wedge u_n(x_i)}{\sum_{c=1}^{k}\sum_{i=1}^{N} u_c(y_i) \wedge u_n(x_i)} \quad (1)$$

Where K represents the classes and N is the number of examples in the subset. $u_k(y_i)$ is the membership degree on the attribute i that belongs to the class k and $u_n(x_i)$ is the membership degree of the attribute i at node n.

Λ represents the conjunction operator. T-norm, which generalizes intersection in the domain of fuzzy sets, is usually used for fuzzy conjunction. The most popular T- norms are minimum and product. The fuzzy entropy uses the membership degree of examples at a particular node and contributes to enhance the discriminative power of an attribute, is computed as:

$$H_n = -\sum_k p_k^n * \log(p_k^n) \quad (2)$$

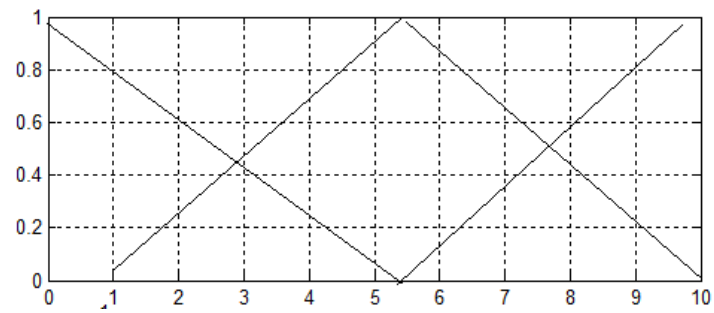
The growth of the fuzzy ID3 is realized by expanding a node of tree characterized by the highest information gain. The information gain is calculated as follows:

$$G_n^j = H_n - \sum_{l=1}^{M} w_l H_l \quad (3)$$

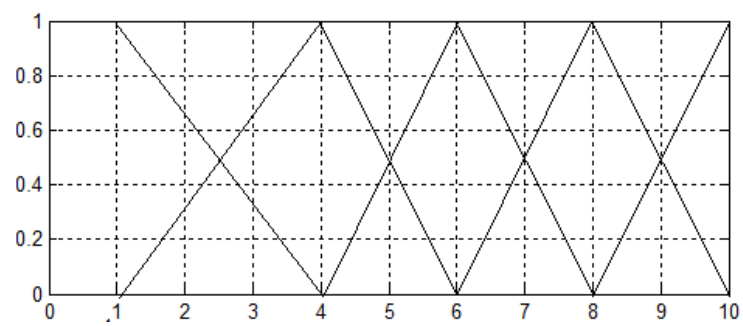
Where H_n is the entropy in the node n. H_l is the entropy of the node that belongs to the fuzzy set L of the j variable. W_l is the fuzzy relative weight.

The node *n* is split into as many sub-nodes as there are attributes. The algorithm terminated when all attributes are used for splits, or when all examples at a node have the same classification.

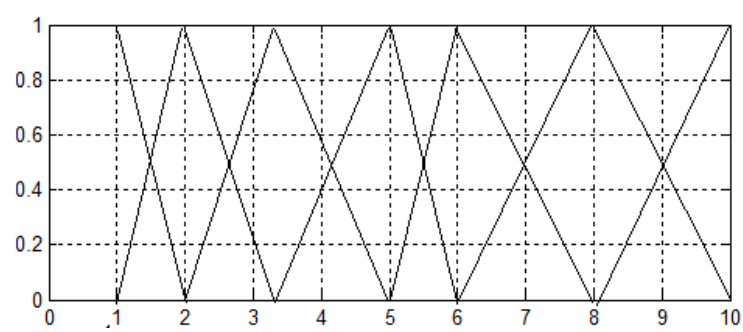
The fuzzification of the network cost drivers converts crisp cost drivers into membership degrees to the different fuzzy sets of the partition. Many algorithms can be found in the specialized literature for generating partitions from data, we choose the hierarchical fuzzy partitioning [7]. It corresponds to an ascending procedure. At each step, for each given variable, two fuzzy sets are merged. This method combines two different clustering techniques, hierarchical clustering and fuzzy clustering techniques. The triangular membership functions are used to represent the fuzzy sets because of its simplicity, easy comprehension and computational efficiency. Figure 2 illustrates the membership function associated to the fuzzy sets of the SeTNode attribute.



(a) Membership function of 3 fuzzy sets defined for the SeTNode cost driver



(b) Membership function of 5 fuzzy sets defined for the SeTNode cost driver



Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

(c) Membership function of 7 fuzzy sets defined for the SeTNode cost driver

Fig. 2 Membership function associated to the fuzzy of the SeTNode attribute.

The fuzzy decision tree is interpreted by rules, each path of the branches from root to leaf can be converted into a rule with condition part represents the attributes on the passing branches from the root to the leaf and condition part represents the class at the leaf of the form: IF (condition 1 and condition 2….. and condition n) THEN C, where the conditions are extracted from the nodes and C is the leaf.

3. Data Description and Evaluation Criteria

3.1 Data Description

This section describes the dataset used to perform this empirical study and the evaluation criteria adopted to measure the estimates accuracy of the designed network cost estimation model based on fuzzy ID3 method.

The dataset is described using 9 numerical attributes as given in table 1.

Table 1: Attributes for network cost estimation

Attributes	Description
SeTNode	Set of nodes (terminal)
SeTLink	Set of links (edge, arcs)
LinkNodeij	Link between nodes i and j
LinkRelia	Link Reliability
DV	Decision variables
LinkTopo	Link Topology
TerRelia	All terminal Reliability
ReliaReq	Network Reliability Requirement
CostLink	Cost of Link

3.2 Evaluation criteria

We employ the following criteria to measure the accuracy of the estimates generated by the fuzzy ID3. A common criterion for the evaluation of reliability estimation models is the magnitude of relative error (MRE), which is defined as

$$MRE = \left| \frac{\mathrm{Re}liability_{actual} - \mathrm{Re}liability_{estimated}}{\mathrm{Re}liability_{actual}} \right| \quad (4)$$

Where $\mathrm{Reliability}_{actual}$ is the actual reliability of the network in the dataset and $\mathrm{Reliability}_{estimated}$ is the estimated Reliability that was obtained using a model or a technique.

The MRE value are calculated for each attribute in the dataset, while mean magnitude of relative error (MMRE) computes the average over N items

$$MMRE = \frac{1}{N}\sum_{I=1}^{N} \left| \frac{\mathrm{Re}liability_{actual,i} - \mathrm{Re}liability_{estimated,j}}{\mathrm{Re}liability_{actual,i}} \right| \times 100 \quad (5)$$

The acceptable target values for MMRE are MMRE ≤ 25. This indicates that on the average, the accuracy of the established estimation model would be less than 25%.

4. Experiment Design

This section describes the experiment design of the fuzzy ID3 decision tree on the given datasets.

The use of fuzzy ID3 to estimate network design effort requires the determination of the parameters, namely the number of input variables, the maximum number of fuzzy sets for each input variable, the significant level value and the conjunction operator. The last two parameters play an essential role in the generation of Fuzzy Decision trees. It greatly affects the calculation of fuzzy entropy and classification results of Fuzzy Decision trees.

The number of input variables is the number of the attributes describing in the used dataset. Therefore, when applying fuzzy ID3 to the given dataset, the number of input variables is equal to 9. Concerning the maximum number of fuzzy sets is the maximum partition size for each variable, is fixed to 7 for all experiments.

In the present paper we are interested in studying the impact of the fuzzy conjunction operators (t-norms) and the significant level parameter (β) on the accuracy of fuzzy ID3. The significant level is the membership degree for an example to be considered as belonging to the node.

For each dataset, two models of fuzzy ID3 were generated. The first Fuzzy ID3 effort estimation model uses the product entropy conjunction operator to measure the fuzzy entropy (t-norm=product), and the second model uses the minimum entropy conjunction operator to calculate the fuzzy entropy t-norm=min). These conjunction operators are the two commonly used t-norm operators because of their well behavior and their computational simplicity [8]

The minimum entropy conjunction operator is defined as:

$$u_k(y_i) \wedge u_n(x_i) = \min[u_k(y_i), u_n(x_i)] \quad (6)$$

Concerning, the product entropy conjunction operator is given as:

$$u_k(y_i) \wedge u_n(x_i) = u_k(y_i) * u_n(x_i) \quad (7)$$

For each model, a series of experiments is conducted with the fuzzy ID3 algorithm each time using a different value of the significant level parameter (β). The significant level is varied within the interval [0, 1].

5. Results

This section presents and discusses the results obtained when applying the fuzzy ID3 to the given datasets. The calculations were made using MatLab software. We conducted several experiments

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

using different configurations of fuzzy ID3. For these experiments, a holdout validation on the entire datasets was performed. Datasets were randomly split into two groups: training set and test set.

Two models of fuzzy ID3 were designed. The first Fuzzy ID3 effort estimation model (Model 1) uses the formula of the conjunction operator given in Eq. (7) to compute fuzzy entropy, and the second model (Model 2) uses the formula of the conjunction operator given in Eq. (6). For each model, different configurations have been obtained by varying the significant level (β). The aim is to determine which configuration improves the estimates.

We have trained and tested the two models using the given dataset. The results for the different configurations have been compared. Figure 3 show the accuracy of the two fuzzy ID3 models, measured in terms of MMRE.

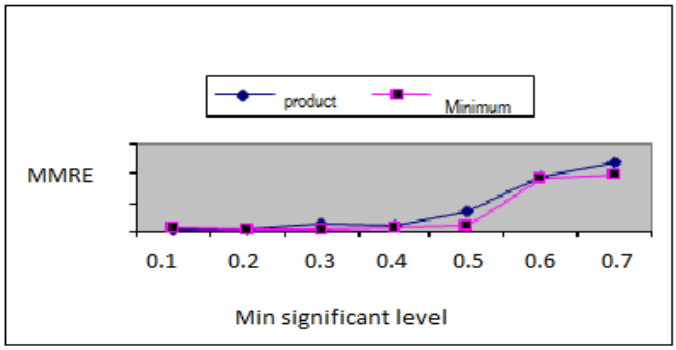


Fig. 3: Relationship between the accuracy of fuzzy ID3(MMRE), the used conjunction operator and the SL value.

Figure 3 compares the accuracy of the two models, in terms of MMRE, when varying the significant level. We note that the fuzzy ID3 model using the product conjunction operator generates a lower MMRE that the other model using the minimum conjunction operator for significant level value less than 0.2.

For example, for β=0.1 the model 1 generates a lower prediction error (MMRE=2.45) than the model 2 (MMRE=5.31). By against, model 2 generates a lower MMRE than the model 1 for significant level value greater than or equal to 0.2. For example, for β=0.5 the model 2 generate a lower prediction error (MMRE=9.09) than the model 1 (MMRE=34.48).

Table 2 summarizes the results obtained using different configurations of fuzzy ID3 for the given dataset. It shows the variation of the accuracy according to the significant level value and to the used conjunction operator.

Table 2: MMRE results of different fuzzy ID3 configurations for the given dataset

Significant level (β)	Accuracy of Fuzzy ID3	
	t-norm=Product	t-norm=Minimum
	MMRE	MMRE
0.1	2.45	5.31
0.2	4.09	1.82
0.3	11.7	3.87
0.4	8.49	5.82
0.5	34.48	9.09
0.6	93.08	90
0.7	119.3	97.41
0.8	210	111.83
0.9	176.99	176.99

Table 3: Result of the different models used on the given dataset

Performance criteria	Crisp ID3	Model 1	Model 2
MMRE	24	2.45	1.82

The experimental results show that the fuzzy ID3 models show better estimation accuracy than the crisp ID3 model in terms of MMRE.

6. Conclusions

In the paper, we have empirically studied two fuzzy ID3 models for network reliability estimation. Each one used a different formula to compute the fuzzy entropy. These fuzzy ID3 models were trained and tested using the given datasets. The results show that the use of an optimal significant level value and an adequate conjunction operator for computing the fuzzy entropy improves greatly the estimates generated by fuzzy ID3 model. The comparison with the crisp version of ID3 decision tree shows encouraging results.

References

[1] R.H.Jan, F.J.Hwang, and S.T.Chen, "Topological optimization of a communication network subject to a reliability constraint", *IEEE Transactions on Reliability*, Vol.42(1993), 63-70

[2] R.R.Boorstyn and H.Frank, "Large scale network topological optimization", IEEE Transactions on communication, Com-25(1977), 29-7

[3] C.J.Colbourn, *The Combinatories of Network Reliability*, Oxford University Press(1987).

[4] L.A.Zadeh, "Fuzzy sets", Information and Control, Vol. 8, 1965,338-353.

[5] M.Umano, H.Okamoto, I.Hatono, H.Tamura, F.Kawachi, S.Umedzu, J.Kinoshita, "Fuzzy Decision Trees by Fuzzy ID3 algorithm and its Applications to Diagnosis System", In

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

proceedings of the third IEEE Conference on Fuzzy Systems, vol. 3, Orlando, 1994, 2113-2118.

[6] R.Weber, Fuzzy ID3: a class of methods for automatic knowledge acquisition, proceedings of the 2nd International Conference on Fuzzy Logic and Neural Networks, Iizuka, Japan, July, 17-22, 1992, 265-268

[7] S.Guillaume and B.Charnomordic, "Generating an interpretable family of Fuzzy partitions", IEEE Transactions on Fuzzy Systems, 12(3), June 2004. 324-335.

[8] E.H.Mamdani, "Applications of Fuzzy Set Theory to Control Systems: A survey", in Fuzzy Automata and Decision Processes, M.M.Gupta, G.N.Saridis and B.R.Gaines, eds., North-Holland, New York, 1977, 1-13.

Biography:

Ak. Ashakumar Singh graduated in Mathematics from Manipur University, Imphal and passed MCA in the year 2000 from the same varsity. He was awarded Ph.D. in the area of Computer Science from the Dept. of Mathematics of the same varsity in the year 2008. Then produced eight M.Phil scholars in Computer Science and now supervising three scholars leading to Ph.D. in *Computer* Science. The area of research is on Soft computing and related applications of computer science.

Ms. Momtaz Thingujam had received her B.Tech degree (C.Sc.) from IET , Lucknow U.P(INDIA) and M.Tech degree from AAIDU, Allahabad U.P (INDIA) respectively.
She is currently an Assistant Professor (Senior Scale) in Computer Science Department, Manipur University. Previously she was a Programmer in Department of Science and Technology, Manipur and Computer Center, Manipur University. Her primary research interests are Natural Language Processing and Computer Network. She also worked as a Co-Investigator for a project (Development of Manipuri language Technology) sponsored by Department of Information Technology, New Delhi.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Implementation & Throughput Analysis of Perfect Difference Network (PDN) in Wired Environment

Yogesh Golhar[1] , R.K.Krishna[2] and Mahendra A. Gaikwad[3]

[1] Rajiv Gandhi College Of Engineering,
Research and Technology, Chandrapur, M.S.442403

[2] Rajiv Gandhi College Of Engineering,
Research and Technology, Chandrapur, M.S.442403

[3] Mahendra A. Gaikwad
Bapurao Deshmukh College of Engineering, Sevagram

Abstract

This paper is about Perfect Difference Network (PDN) which is an asymptotically optimal method for connecting a set of nodes into a Perfect Difference Network with diameter 2, so that any node is reachable from any other node in one or two hops.
We aim to show a simulated result of the data transmission in PDN for δ value 2 & 3 for wired network and later analysis of trace file to find the Throughput. During implementation we have tried to make disturbance in the network to better analysis of PDN implementation in wired network.

Keywords: *Perfect Difference Network*, *Perfect Difference Network*, *throughput, Wired Network, Wireless network.*

1. Introduction

Network throughput is measured as the average rate of successful message delivery over a communication channel. The data may be delivered over a link, or pass through a certain network node and is usually measured in bits per second, data packets per second or data packets per time slot. Throughput varies from wireless to wired network.

The n-node complete graph and the n-node ring represent the two extremes of network connectivity patterns. Intermediate architectures between n-node complete graph connectivity and n-node ring connectivity can be obtained in a variety of ways, providing tradeoffs in cost and performance. Network cost is affected, among other things, by the (maximum) node degree d, while indicators of network performance and network cost. Perfect difference network is one such network which is based on the concept of perfect difference sets. Fig.1. the spectrum of networks in terms of node degree.

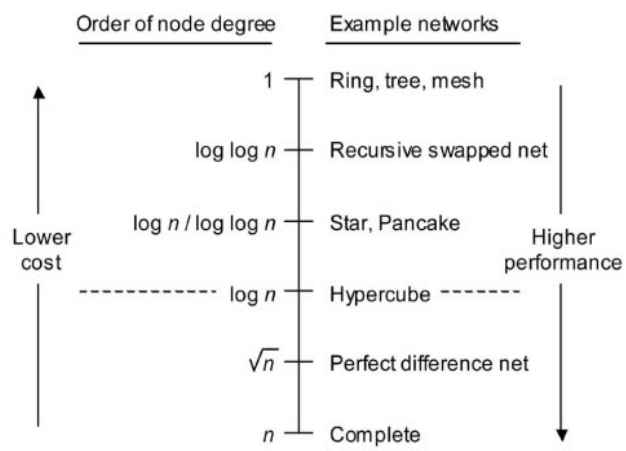


Fig 1. Various network with node degree & cost

2. Perfect Difference Set

Given that the complete graph Kn (with diameter D ¼ 1) is impractical for large n, it is quite natural to consider the best topology for D ¼ 2, the next most desirable network diameter. Based on Moore bounds, a degree-d digraph with D ¼ 2 can have no more than n ¼ d2 þ d þ 1 nodes. The corresponding upper bound n ¼ d2 þ 1 for undirected graphs is not much different. Examples of diameter-2 networks of small sizes include the Petersen and Hoffman-Singleton networks [7]. Perfect difference sets provide the mathematical tools for achieving this optimum number of nodes, in an asymptotic manner, within the framework perfect difference networks or PDNs.

TABLE 1
Perfect Difference Sets (PDS) in their Normal forms (0, 1) for δ =16.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

δ	*n*	Example PDS of order δ in normal form
2	7	0, 1, 3
3	13	0, 1, 3, 9
4	21	0, 1, 4, 14, 16
5	31	0, 1, 3, 8, 12, 18
7	57	0, 1, 3, 13, 32, 36, 43, 52
8	73	0, 1, 3, 7, 15, 31, 36, 54, 63
9	91	0, 1, 3, 9, 27, 49, 56, 61, 77, 81
11	133	0, 1, 3, 12, 20, 34, 38, 81, 88, 94, 104, 109
13	183	0, 1, 3, 16, 23, 28, 42, 76, 82, 86, 119, 137, 154, 175
16	273	0, 1, 3, 7, 15, 31, 63, 90, 116, 127, 136, 181, 194, 204, 233, 238, 255

Theorem 1: A sufficient condition that there exist δ + 1 integers $s0, s1, \ldots, s\delta$ having the property that their δ2 + δ differences $si - sj$, $0 \le i \ne j \le \delta$, are congruent, modulo (δ2 + δ + 1), to the integers 1, 2. . . δ2 + δ in some order is that δ be a power of a prime.

Definition 1: Perfect difference set (PDS) – A set $\{ s0, s1, \ldots, s\delta \}$ of δ + 1 integers having the property that their δ2 + δ differences $si - sj$, $0 \le i \ne j \le \delta$, are congruent, modulo (δ2 + δ + 1), to the integers 1, 2, . . . , δ2 + δ in some order is a *perfect difference set* of order δ. Perfect difference sets are sometimes called simple difference sets. PDS need not contain an integer outside the interval [0, δ2 + δ], because any integer outside the interval can be replaced by another integer in the interval without affecting the defining property of the PDS.

Theorem 2: Given a PDS $\{s0, s1, \ldots, s\delta\}$ of order δ, the set $\{as0 + b, as1 + b, \ldots, as\delta + b\}$, where a is prime to δ2 + δ + 1, also forms a perfect difference set. By definition, any perfect difference set contains a pair of integers su and sv such that su –sv ≡ 1 mod (δ2 + δ + 1). By Theorem 2 and the observation that preceded it, subtracting su from all integers in such a PDS yields another PDS that contain 0 and 1

Definition 2: Normal PDS – A PDS $\{s0, s1. \ . \ . \ s\delta\}$ is reduced if it contains the integers 0 and 1. A reduced PDS is in normal form if it satisfies $si < si{+}1 \le \delta 2 + \delta$, $0 \le i < \delta$.

Definition 3: Equivalent PDSs – Two different PDSs are equivalent iff they have the same normal form $\{0, 1, s2, s\delta\}$.

Property 1: Multiplicity- For any order δ, there exist more than one PDS. Perfect Difference sets with order δ as a power of prime number and number of nodes, n= δ2 + δ + 1 are

3. Perfect Difference Networks

Consider the normal-form PDS $\{0, 1, s2. \ . \ . \ s\delta\}$ of order δ. We can construct a direct interconnection network with $n = \delta 2 + \delta + 1$ nodes based on this PDS as follows

Definition 4: Perfect difference network (PDN) based on the PDS $\{0, 1, s2, \ldots, s\delta\}$ – There are $n = \delta 2 + \delta + 1$ nodes, numbered 0 to $n \pm 1$. Node i is connected via directed links to nodes $i \pm 1$ and $i \pm si$ (mod n), for $2 \le i \le \delta$. Because all index expressions in this paper are evaluated modulo n, henceforth we will delete the qualifier "mod n." The preceding connectivity leads to a chordal ring of in- and out-degree $d = 2\delta$ and diameter $D = 2$. Because for each link from node i to node j, the reverse link (j, i) also exists, the network corresponds to an undirected graph. Every normal-form PDS contains 1 as a member. Therefore, PDNs based on normal-form PDSs are chordal rings. In the terminology of chordal rings, the links connecting consecutive nodes i and $i + 1$ are *ring links*, while those that connect nonconsecutive nodes i and $i + si$, $2 \le i \le \delta$, are *skip links* or *chords*. The link connecting nodes i and $i + si$ is referred to as the *forward skip link* of node i and *backward skip link* of node $i + si$. Similarly, the ring link between nodes i and $i + 1$ is a *forward (backward) ring link* for i $(i + 1)$.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

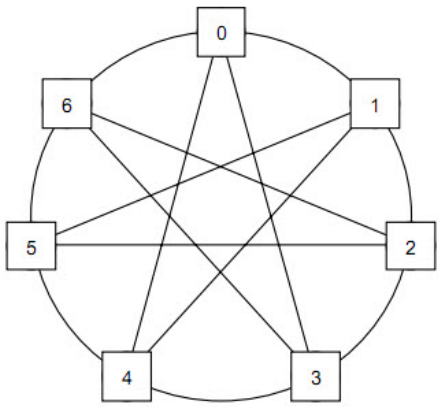


Fig. 2. PDN with $n = 7$ nodes based on the perfect difference set {0, 1, 3}.

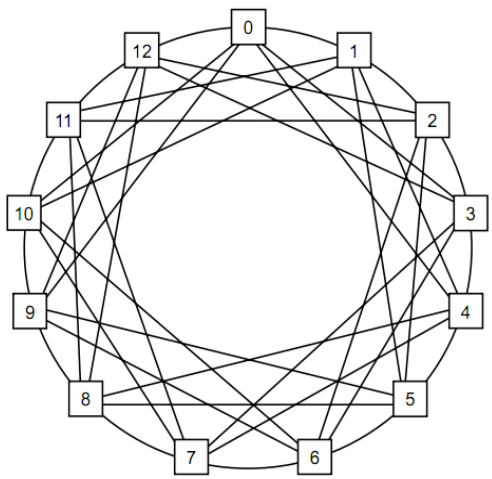


Fig. 3. PDN with $n = 13$ nodes based on the perfect difference set {0, 1, 3, 9}.

4. PDN generation and analysis from PDS

In this section we will look at the procedures for obtaining PDN from PDS for δ having values 2, 3 and 4 respectively. Then we will create a connectivity table for each node based on result of findings obtained from above step. We will also find out optimal lookup table for the PDNs by using the property that a PDN has diameter D of 2.

PDN generation for PDS.
It is given under the following headings:
Steps of implementation of PDN for δ

- Here value of δ is 2 & 3 as is evident from the given PDS.
- Number of nodes n is given by equation $n = \delta^2 + \delta + 1$.
- Go through Table 1 and note the corresponding PDS for the chosen value of δ.
- Calculate the ring links and skip links for each node by using the formulae specified in the steps for PDS to PDN conversion.
- Construct the PDN from above obtained data.
- Find out the lookup table for the PDN, keeping in mind that the diameter of the network is 2.

a. Calculations for PDN with δ = 2 and number of nodes n = 7

TABLE 2
Node connectivity table for δ = 2

Node i	Skip Links	Ring Links	Connecting Links
0	3,4	1,6	1,3,4,6
1	4,5	0,2	0,2,4,5
2	5,6	1,3	1,3,5,6
3	0,6	2,4	0,2,4,6
4	0,1	3,5	0,1,3,5
5	1,2	4,6	1,2,4,6
6	2,3	0,5	0,2,3,5

b. Calculations for PDN with δ = 3 and number of nodes n = 13

TABLE 2
Node connectivity table for δ = 3

Node i	Skip Links	Ring Links	Connecting Links
0	3,4,9,10	1,12	1,3,4,9,10,12
1	4,5,10,11	0,2	0,2,4,5,10,11
2	5,6,11,12	1,3	1,3,5,6,11,12
3	0,6,7,12	2,4	0,2,4,6,7,12
4	0,1,7,8	3,5	0,1,3,5,7,8
5	1,2,8,9	4,6	1,2,4,6,8,9
6	2,3,9,10	5,7	2,3,5,7,9,10
7	3,4,10,11	6,8	3,4,6,8,10,11
8	4,5,11,12	7,9	4,5,7,9,11,12
9	0,5,6,12	8,10	0,5,6,8,10,12
10	0,1,6,7	9,11	0,1,6,7,9,11
11	1,2,7,8	10,12	1,2,7,8,10,12
12	2,3,8,9	0,11	0,2,3,8,9,11

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

5. Experimental Evaluation

Perfect Difference Network in Wired Network

We have implemented PDN in wired environment using the simulator NS2.It primarily shows the communication from one node to the other node. Some conditions are so arising in which the links are made down. So in this case it finds the other way which is again optimum and reaching the destination in which we find it tries to find simple 2 hop path but if more number of links are failed then at most it find 4-hop path. We have taken it for 7 nodes to implement PDN as shown in the figure below:

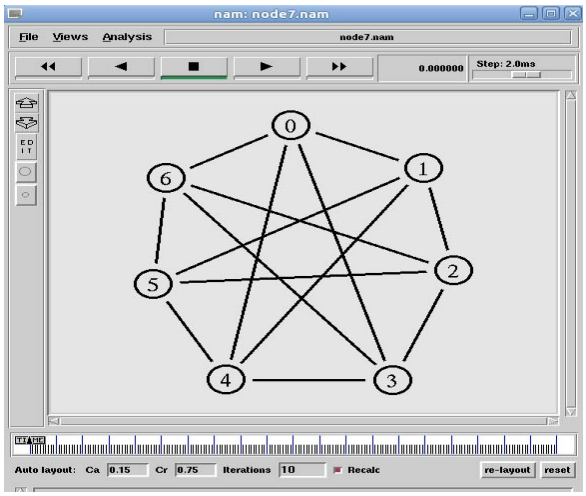


Fig. 4 PDN with 7 nodes

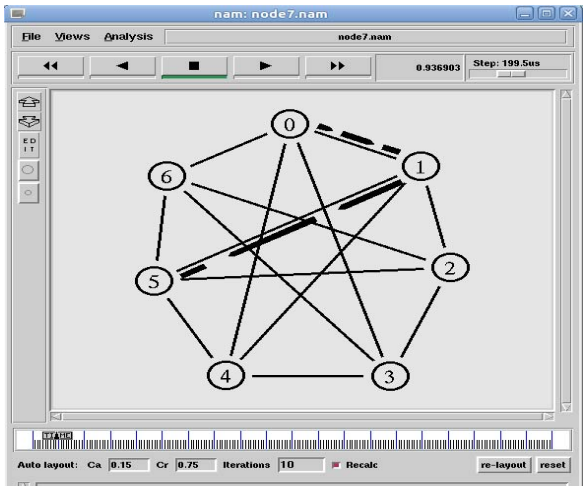


Fig. 5 Data Transmission from node 0 to node 5 via node1

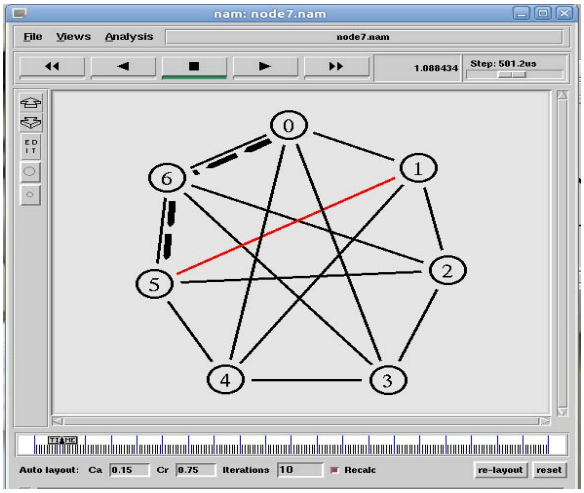


Fig. 6 Data Transmission from node 0 to node 5 via node 6

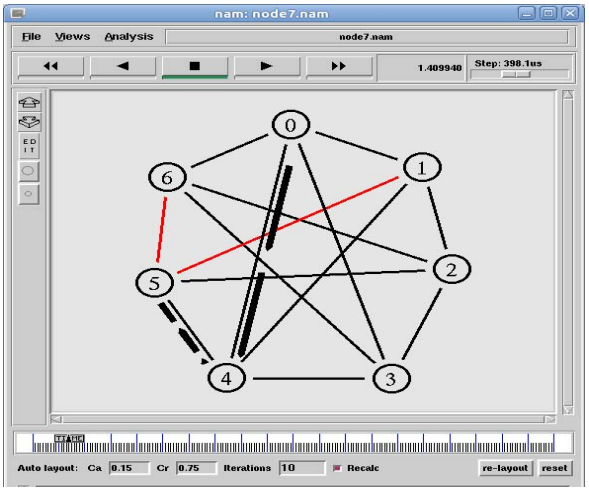


Fig. 7 Data Transmission from node 0 to node 5 via node 4

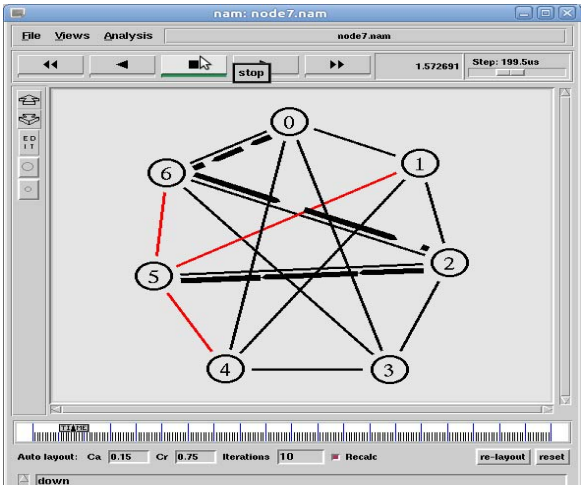


Fig. 8 Data Transmission from node 0 to node 5 via node 6 & node 2

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

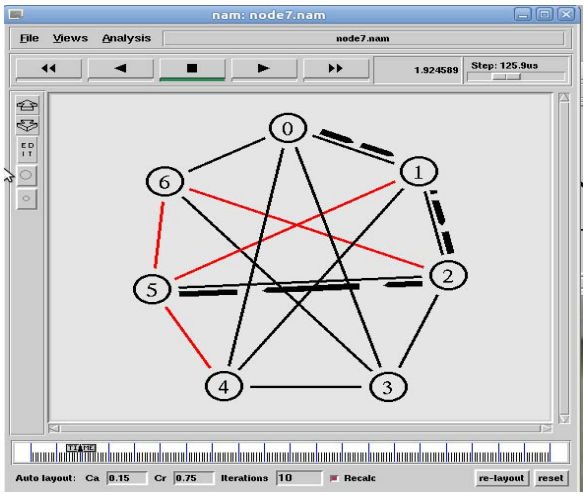


Fig. 9 Data Transmission from node 0 to node 5 via node1

This module shows the connectivity of PDN for δ=2 and n=7 nodes. And other links connected. The data is route from one end to the other end with only two hops. In case of any link failure the route is changed through other nodes. As shown in the figure below.

In the next module we shows the connectivity of PDN for δ=3 and n=13 nodes, other links connected. As shown in the figure below:

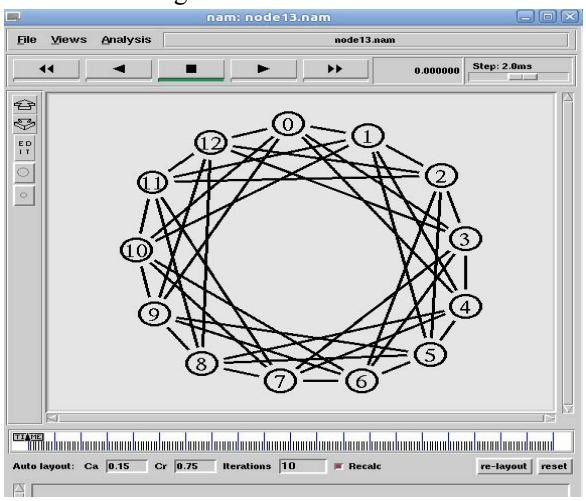


Fig. 10 PDN with 13 nodes

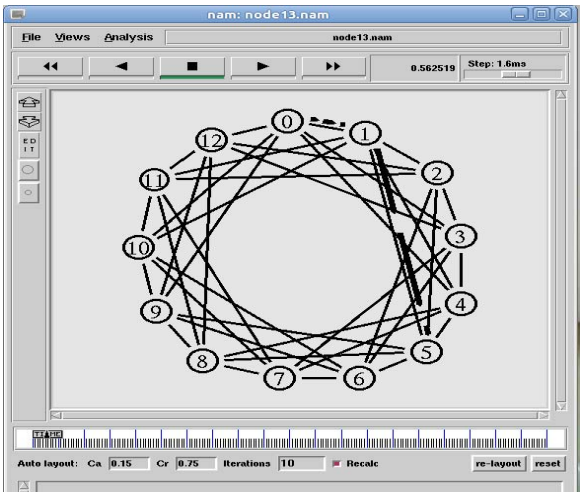


Fig. 11 Data Transmission from node 0 to node 5 via node1

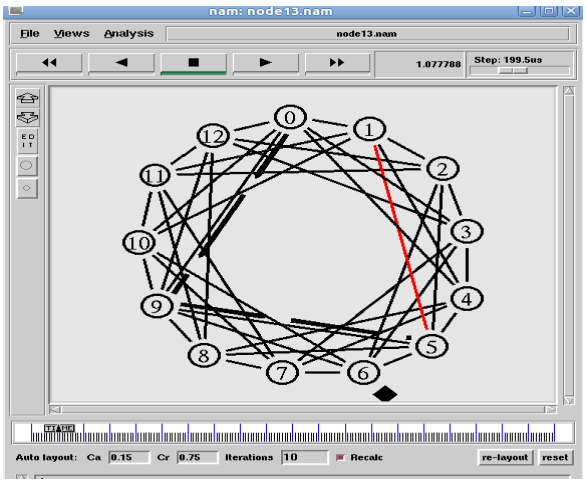


Fig. 12 Data Transmission from node 0 to node 5 via node 9

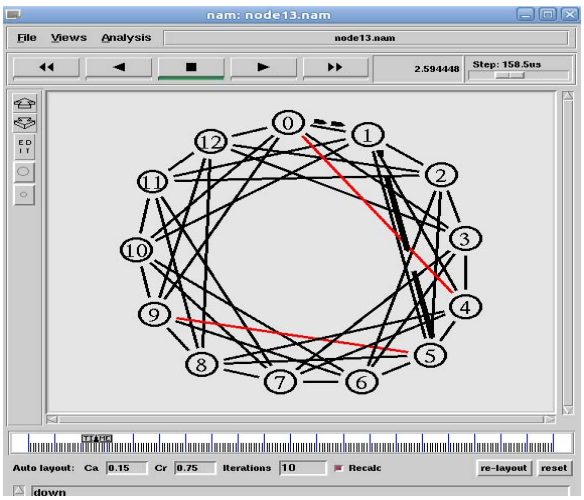


Fig. 13 Data Transmission from node 0 to node 5 via node1

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

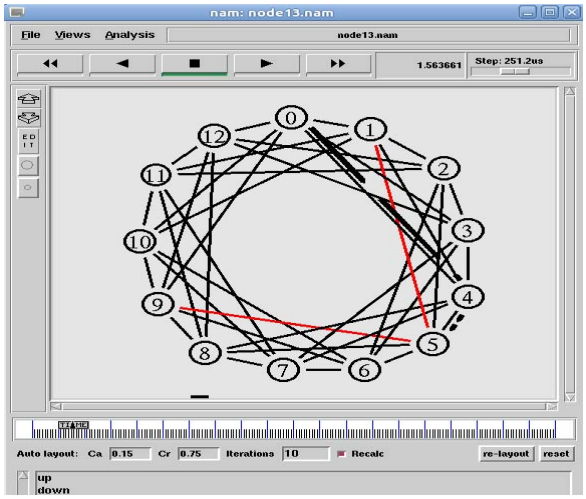


Fig. 14 Data Transmission from node 0 to node 5 via node 4

6. Simulated Results

This is the throughput calculation for two values of 'δ', i.e. 2 and 3.Tis is implemented in wired network as shown above.

TABLE 3
Simulation Result

Time	Node 7 δ=2	Node 13 δ=3
1	94	98
2	169	147
3	190	157
4	203	170
5	203	189
6	210	191
7	204	186
8	209	187
9	204	182
10	208	188
11	205	186
12	208	187
13	205	183
14	208	184
15	203	185
16	190	185
17	179	186
18	169	187
19	160	187
20	160	187

7. Conclusion

In the paper , the results are very different than what expected and though there are link failures during the transition of data but the throughput graph shows the steady output for both the values of 'δ' in wired network.

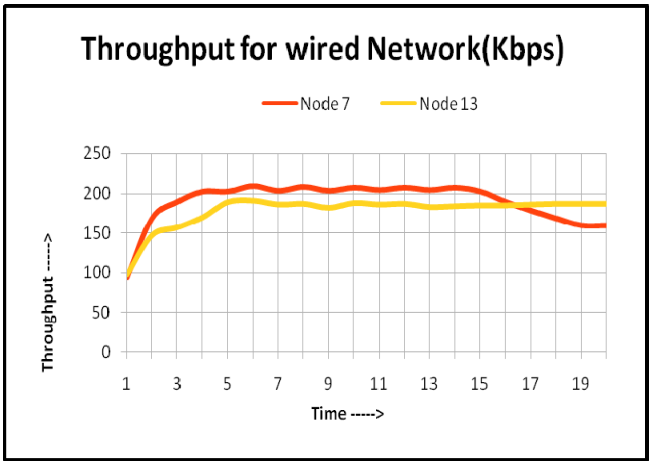


Fig. 15 Throughput analysis for node 7 and node13

8. References

[1] Your 802.11 Wireless Network has No Clothes by William A. Arbaugh, Narendar Shankar, and Y.C. Justin Wan Department of Computer Science University of Maryland College Park, Maryland 20742

[2] Airness for extend DCCP congestion control in Wireless Sensor Networks by Liu Yong-Min; Jiang Xin-Hua; Nian Xiao-Hong; Coll. of Vocational Technol., Central South Univ. of Forestry & Technol., Changsha, China

[3] Behrooz Parhami, Fellow, IEEE, and Mikhail Rakov, "Perfect Difference Networks and Related Interconnection Structures for Parallel and Distributed Systems," IEEE transactions on parallel and distributed systems, vol. 16, no. 8, august 2005, pp714-724

[4] Behrooz Parhami, Fellow, IEEE, and Mikhail Rakov, "Performance, Algorithmic, and Robustness Attributes of Perfect Difference Networks", IEEE transactions on parallel and distributed systems, vol. 16, no. 8,august 2005, pp 725-736.2

[5] Behrooz Parhami, Fellow, IEEE, and Ding-Ming Kwai "Periodically Regular Chordal Rings", IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 10, NO. 6, JUNE 1999

[6] Perfect Difference Network for Network-on-Chip Architecture IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.12, December 2009 by Mahendra Gaikwad & Rajendra Patrikar

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

[7] Sudhir G. Akojwar & Rajendra. M. Patrikar, "Perfect Difference Set – Network for Wireless Sensor Networks" in IEEE 2008

[8] Virtual Inter-Network Test Bed (VINT) Project and National Science Foundation. http://www.isi.edu/nsnam/ns

[9] Object oriented TCL script and NS Manual http://www.isi.edu/nsnam/ns/ns-documentation.html

Yogesh Golhar did his B.E. in Computer Technology in 2008 from Nagpur University. Currently he is pursuing his MTech from Nagpur University at Rajiv Gandhi College of Engineering, Research & Technology Chandrapur. Teaching Experience from last 3years. Currently working in Priyadarshini College of Engineering Nagpur. He has completed his CISCO certification in CCNA (Cisco Certified Network Associate) in 2010. He is the life member of Professional bodies like Indian Society for Technical Education, Institution of Engineers (India), till now 1 International Conference and 5 National Conference & publication and awarded as the best short paper in 2 days workshop on "Emerging Trends in Computer Simulation" at OM COLLEGE OF ENGINEERING, Wardha.

Prof. R.K.Krishna has completed is BE from Magpur University in 1991,MBA from Nagpur University in 1995 and MTech from NIT Bhopal in 1998.He is presently working as Faculty In Rajiv Gandhi College of Engg.,Research and Technology ,Chandrapur and has an Experience of 19 years. He has published three books and 20 papers in various journals and conferences. He is Secretary ISTE RCERT chapter and member of IE. His Areas of interest are communications and image processing.

Mahendra A. Gaikwad did his BE in Electronics Engineering in 1991 from Nagpur University. He did his MBA in Marketing Management from Nagpur University and he has completed his MCM from Nagpur University. He did his Master's Degree in Personal Administration from Nagpur University. He did his M.Tech in Communication Engineering from Indian Institute of Technology, Bombay in 1998. He has submitted his PhD research work on "Network on Chip Architecture using Perfect Difference Network Topologye" at VNIT, Nagpur. Currently he is working as Professor and Dean (R&D) at Bapurao Deshmukh College of Engineering, Sewagram (Wardha), India. He is the life member of professional bodies like Indian Society for Technical Education, Institution of Engineers (India), Indian Society for Telecommunication Engineers and Computer Society of India.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

The application of methods of nonlinear dynamics for ECG in Normal Sinus Rhythm

Nisar Hundewale
College of Computers and Info. Tech., Taif University, Saudi Arabia

Abstract
The ECG signals were processed in three steps: First by reconstructing their phase portraits, second by calculating their spectra and finally by computing their Largest Lyapunov exponents. This study presents a framework to assess nonlinear parameters of ECG signals that may be useful for further exploration of physiological and path-physiological significance of dynamics in electrocardiograms. This paper devotes to identify the nonlinear properties of human heart beat dynamics for 18 characteristic examples of normal subjects with normal sinus rhythm. A new phase space characterization of a class of ECG signals with normal sinus rhythm through is presented. Such approach led to find out a common visual feature namely an anchor-shaped Poincar´e plot resulting from an appropriate choice of the delay time. Any value of the delay time is acceptable, but the shape of the embedded time series depends critically on the choice of its value. The results reported in this study may be useful not only for the classification of ECG states, but can serve as a benchmark to which pathological cases can be compared.

Keywords: ECG, Normal sinus rhythm, nonlinear dynamics, Largest Lyapunov exponent

1. Introduction

In recent literature, approaches to understanding and intervening in the cardiovascular system are being developed using the new methods from nonlinear system theory. The normal development of heart rate controlling system was associated with characteristic changes in linear and nonlinear measures. Non stationarity in the Electrocardiogram (ECG) signals manifest both in RR interval (the time elapsing between two consecutive R waves in the ECG) timing changes and in morphological changes Although the former changes are often thought of as rhythm disturbances and the latter as beat abnormalities. Many scientists turn to the analysis of chaotic dynamics of various phenomena including functioning of alive organisms. It is often difficult to decide whether the dynamics of a biological system is chaotic or not, the question whether the human cardiac system is chaotic or not has been an open one. In chaotic behavior, the initial close orbits of the strange attractor will diverge exponentially with time, with extreme sensitivity to initial conditions. Much evidence shows that the human heart is not a perfect oscillator but it is chaotic system, so ECG signals have been considered as chaotic signals and this has been successfully proved by computing Lyapunov exponents [1][2]. In former studies, strictly periodic cardiac rhythm is not accompanied by healthy condition but, on the contrary, by pathological states. Besides, investigation of continuous ECG with methods of nonlinear system theory is originated from the hypothesis that cardiac rhythms are associated with chaotic dynamics [3][4]. A characteristic feature of a dynamical system exhibiting deterministic chaos is that it is possible to predict the behavior of the system with some degree of confidence in the short-term even though long-term prediction is impossible. Profound inspection of abnormalities in the form of ECG signals allows the medical doctors to identify a larger number of distinct diseases. Some methods of recognition of possible diseases rely on a statistical analysis of a large number of heart beats [5]. The complexity of cardiac dynamics decreases significantly in the morning and this may contribute to the ominously increased rate of cardiac death in the morning hours [6]. The main metrics used to characterize the nonlinear behavior of heart beat signal (Heart Rate Variability and continuous ECG signals) are correlation dimension, Lyapunov Exponent, Kolmogorov entropy, fractal parameters [7]. Largest Lyapunov exponent can be a diagnostic criterion allowing to distinguish between different groups of patients [8]. For instance, The Largest Lyapunov Exponent (LLE) of coronary artery disease patients with sinus rhythms is less than that of healthy people [9]. Use of Lyapunov exponents in studying biological systems has been increasingly frequent. Its have been employed particularly in an application of Adaptive neuro-fuzzy inference system for classification of ECG signals [10], to characterize the states of cardiovascular system in [11] and, more generally, to analyze ECG time series in [12]. As shown in [9], the investigation of the Lyapunov exponents of synchronous 12-lead ECG signals leads to conclude that the Lyapunov exponents of human heart rate are not the same for different lead ECG signals for all the tests for both Coronary Artery Disease patients and healthy people. But more importantly, the Lyapunov exponents have the same sign for all the synchronous eads. It is worth noting that the LLE is a nonlinear measure that reflects the 'overall' properties of the

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

instantaneous heart rhythm[13].A typical chaotic system has at least one positive Lyapunov exponent. Numerous methods for alculating the Lyapunov exponents have been developed during the past decade, among the approaches for calculation of the Lyapunov exponents may be mentioned particularly the Jacobian and the direct method [14]. In general, phenomena observed in finite time intervals are called transient phenomena. Transient chaos is one of transient phenomena that can be found in a class of systems whose asymptotic behavior is regular. Lyapunov exponents is among the most useful tools applied for measuring the sensitivity to initial conditions in the case of
asymptotic chaos. Finite-time Lyapunov exponents were used for characterizing sensitivity to initial conditions in the case of transient chaos[15]. The embedding dimension plays an essential role for the identification and the prediction of nonlinear dynamical systems including chaos. Analysis of heart rate variability focused on the RR-interval fluctuations data. Measurements of heart rate variability by conventional and new nonlinear methods is a useful research tool for documenting changes in neural regulation in relation to arrhythmia events in various clinical settings [16]. In [17] it was reported that heart rate variability can be used as a reliable indicator of the heart state. It becomes less random with aging, it has also been shown that the parameter values fall into different ranges for different health conditions. A heart rate variability analyzer was designed in [18] to assess nonlinear metrics like, LLE, fractal dimension, correlation dimension, approximate entropy and slope of Poincar´e plots. Many attempts to find a relationship between some nonlinear indices and the biological aspects of heart rate regulation have been reported in[19]. Phase space plots facilitate the visualization of beat-to-beat dynamics in the heart rate. The measures of nonlinear dynamics through Poincar´e maps, symbolic dynamics and renormalized entropy have been investigated in [20]. In this study, an approach based on phase space plots is used to determine a new characterization of a certain class of 18 human heart beat recordings of subjects who have had no significant arrhythmias. Such approach led to find out a common visual feature in phase space of the so called recordings. Simultaneously with the phase space plots, Spectral analysis of data strings of the ECG signals was performed in order to provide additional details related to such class of recordings. Investigation of chaotic nature of the ECG recordings was then done by computing their corresponding LLE.

2. Materials and Methods

2.1. ECG Description

An ECG in normal sinus rhythm, P-waves are followed after a brief pause by a QRS complex, then a T-wave. The time intervals between heartbeats usually vary in a complex and irregular manner. However, even without arrythmia, the variability of sinus rhythm in a healthy individual is very complex. Normal sinus rhythm not only indicates that the rhythm is normally generated by the sinus node and traveling in a normal manner in the heart, but also that the heart rate is within normal limits. In a normal heart rhythm, the sinus node generates an electrical impulse which travels through the right and left atrial muscles producing electrical changes which is represented on the electrocardiogram (ECG) by the P-wave. The ventricular contraction is represented electrically on the ECG by the QRS complex of waves. cardiac cycle after a short pause repeats itself, and so on. There is no one normal heart rate, but this varies by age. It is normal for a newborn to have a heart rate up to 150 beats per minute, while a child of five years of age may have a heart rate of 100 beats per minute. The adult's heart rate is even slower at about 60-80 beats per minute. The database of MIT-BIH normal sinus rhythm, includes 18 long-term ECG recordings (sampling frequency is 128 Hz) of subjects referred to the Physiobank databases[21]. Subjects included in this database were found to have had no significant arrhythmias; they include 4 men, aged 26 to 45, and 14 women, aged 20 to 50. From such recordings we calculate an estimate value of the heart beat rate expressed in beats per minute (BPM) see table 1.

Table 1: MIT-BIH Normal Sinus Rhythm Database

Record	Gender	Age	BPM
nsrdb16265	M	32	98
nsrdb16272	F	20	62
nsrdb16273	F	28	85
nsrdb16420	F	38	93
nsrdb16483	M	42	88
nsrdb16539	F	35	78
nsrdb16773	M	26	74
nsrdb16786	F	32	72
nsrdb16795	F	20	61
nsrdb17052	F	45	67
nsrdb17453	F	32	81
nsrdb18177	F	26	92
nsrdb18184	F	34	82
nsrdb19088	F	41	83
nsrdb19090	M	45	84
nsrdb19093	F	34	70
nsrdb19140	F	38	87
nsrdb19830	F	50	102

2.2. Poincar´e plot analysis

The Poincar´e plot is a tool developed by Henri Poincar´e for analyzing complex systems. Such technique taken from nonlinear dynamics, also named return map, is the

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

simplest technique to describe the nonlinear dynamics of a phenomenon.

Table 2: Spectral parameters of ECG signals; *a*0: Fourier constant term; *fM*: the predominant frequency; *AM*: The amplitude of the predominant higher harmonic; *LLE*: Largest Lyapunov Exponent

Record	a_0	$f_M(Hz)$	A_M	*LLE*
nsrdb16265	14.020	6.406	11.840	all < 0
nsrdb16272	9.147	1.063	5.595	all < 0
nsrdb16273	9.870	4.531	14.910	all < 0
nsrdb16420	10.430	7.906	1.802	some *LLE* > 0
nsrdb16483	8.436	1.500	17.090	all < 0
nsrdb16539	6.971	4.125	1.838	all < 0
nsrdb16773	5.779	1.188	21.450	all < 0
nsrdb16786	9.989	3.594	15.653	all < 0
nsrdb16795	7.099	2.281	2.289	some *LLE* > 0
nsrdb17052	6.975	1.094	1.410	all < 0
nsrdb17453	10.570	1.344	3.427	some *LLE* > 0
nsrdb18177	9.427	3.313	2.153	all < 0
nsrdb18184	7.392	0.250	5.344	all < 0
nsrdb19088	5.526	1.500	0.852	all < 0
nsrdb19090	4.917	2.969	1.869	all < 0
nsrdb19093	23.440	1.156	50.360	some *LLE* > 0
nsrdb19140	3.008	4.625	6.420	some *LLE* > 0
nsrdb19830	24.330	0.250	1.164	all < 0

The Poincar´e plot had been used to portray the heart rate fluctuations because it provides summary information as well as detailed beat-to-beat information on the behavior of the heart [18]. Poincar´e plot was used as a quantitative visual technique whereby the shape of the phase plane plot was categorized into functional classes that indicate the degree of the heart failure in a subject[17]. Since the phase space representation is a convenient method for graphical characterization of ECG signals, it is used in this study as a visual technique to recognize the hidden correlation patterns of a time series signal. In fact, a three dimensional phase space will be described by plots in which the temporal samples of ECG are plotted as a function of the previous ones. A phase plane plot is obtained by plotting the value at time (t+τ) vs. the value at time t where τ is an arbitrary delay which should be sufficient to reconstruct the attractor, efficiency with a limited amount of data is enhanced by particular choices of the time delay τ. In the following, we use three dimensional representations to distinguish between different ECG signals. A phase space plot is obtained by plotting the values of the continuous
ECG with three axes (e(t), e(t+2τ), e(t+τ)).

2.3. Frequency domain measures

The spectral analysis of heart rate variability had been applied to populations with cardiovascular diseases. 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. Spectral analysis of ECG signals aims to put into evidence features that may lead to categorize certain particular physiological signals through their spectral composition. An ECG frequency spectrum can be thick or sparse, so the role of higher harmonics can be of great interest to explain and interpret particular healthy or diseased situations. 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. Aiming to analyze Electrocardiogram (ECG) signals given as discrete time signals, one can investigate their frequency content and other particular parameters such as the Fourier constant term, the frequency and the magnitude of the predominant harmonic. Thus, in the following, spectral analysis is performed by computing the FFT of a data string of 512 sample points lasting 4ms.

2.4. Largest Lyapunov exponent

Spectral and correlation analysis have been widely used for
the processing of time series far before the concept of dynamical chaos has been settled[8]. Lyapunov exponents are a quantitative measure of the sensitive dependence of nonlinear system variables on the initial conditions. Its define the average rate of divergence of initially nearby trajectories in phase plan e for distinguishing among the various types of orbits based on their sensitive dependence on the initial conditions, and are used to determine the stability of any steady-state behavior, including chaotic solutions. Starting from two neighbor points $X0$ and $\Delta X0$ which will generate two different orbits in the phase space, the divergence between such orbits is measured by the Lyapunov exponents. In chaotic states the separation between the orbits, which is a function of time will behave erratically. In the following The Largest Lyapunov exponents will be estimated from the observed time series in order to investigate the 'chaocity' of the studied class of ECG signals. If the Largest Lyapunov exponent of an ECG signal is positive, then it is said to have chaotic characteristics, this may mean that the electrical activity of the human heart is not periodic but chaotic. The Largest Lyapunov exponent equals zero for strict periodic motion and infinite for pure noise.

2.5. Embedding dimension and delay time

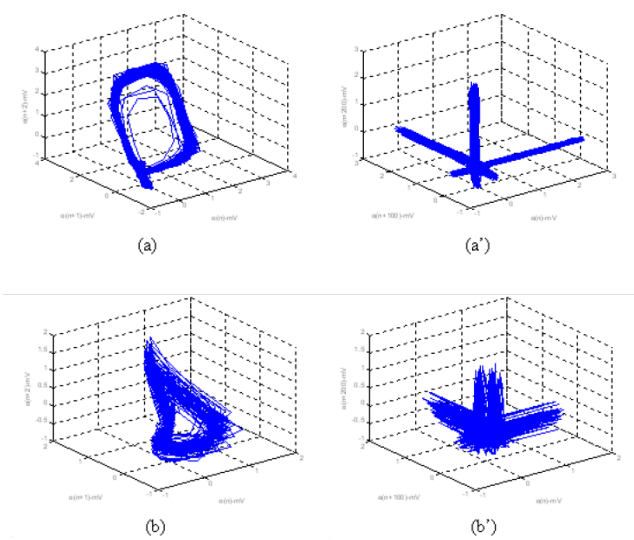
Aiming to perform the phase space reconstruction, it is worth noting that the selection of a suitable pair of embedding dimension and time delay is based on two

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

different points of view. One is that such parameters are not correlated with each other and the second viewpoint is that these parameters are closely related, for most researchers the latter is more practical and reasonable than the first one in the engineering practice[23]. To perform an adequate embedding, the first parameter to be estimated is the embedding dimension d. Practically, the method used to determine the embedding dimension consists in guessing suitable value for d, by successively embedding in higher dimensions and looking for consistency of results. Selecting an embedding dimension devotes to make sufficiently many observations of the system state so that the deterministic state of the system can be determined unambiguously. In former studies, autoregressive modeling
was used to extract features in order to classify certain cardiac arrhythmias. A nonlinear autoregressive model is described by:

$$Xt = Fd(Xt-1, Xt-2, ..., Xt-d)$$

The embedding dimension is defined as follows[22]: A time series Xt is said to have the embedding dimension $d0$ with the delay time $\tau 0$ if and only if there exist non-negative integers $d0 < \infty$, and any $\tau 0 < \infty$ such that: $Fd(Xt-\tau, Xt-2\tau, ..., Xt-d\tau)$, $Fd0\ (Xt-\tau 0\ , Xt-2\tau 0\ , ..., Xt-d0\tau 0\ )$ *a.e* for any $d < d0$, and any $\tau > 0$, and $Fd(Xt-\tau, Xt-2\tau, ..., Xt-d\tau) = Fd0\ (Xt-\tau 0\ , Xt-2\tau 0\ , ..., Xt-d0\tau 0\ )$ *a.e*, for any $(d, \tau) \in B(d0, \tau 0)$ where $B(d0, \tau 0) = \{(d, \tau) \mid \{\tau 0, 2\tau 0, ..., d0\tau 0\} \subset \{\tau, 2\tau, ..., d\tau\}\}$.



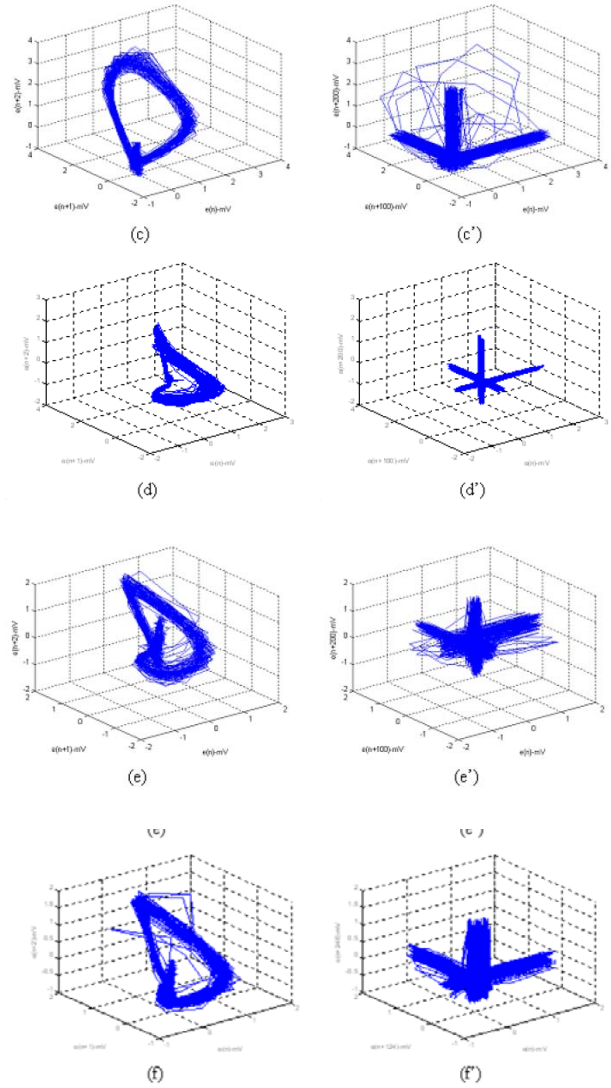
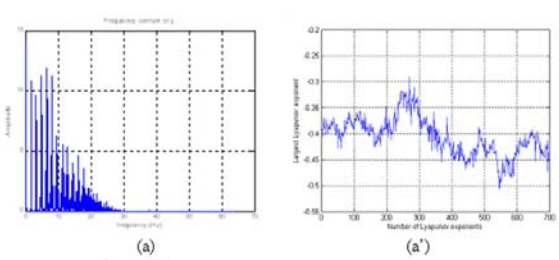


Figure 1: phase plots for the ECG signals: attractor and reconstruction of the attractor. The left panel is a single 7680 point trajectory plotted in e(t), e(t+1) and e(t+2) coordinates (embedding lag of 1), the right hand panel is a delay embedding in three dimension (embedding lag of 100 in figures a',b',c',d' and e' and 124 in figure f') of the e(t) coordinate



Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

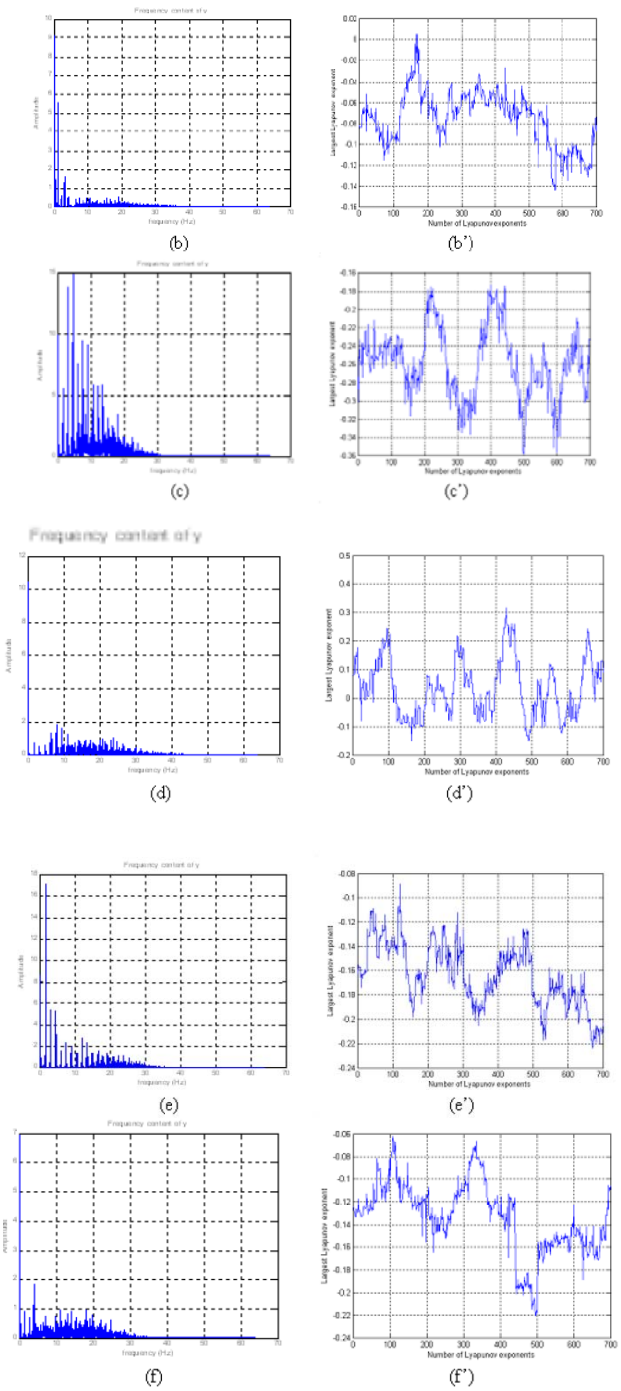
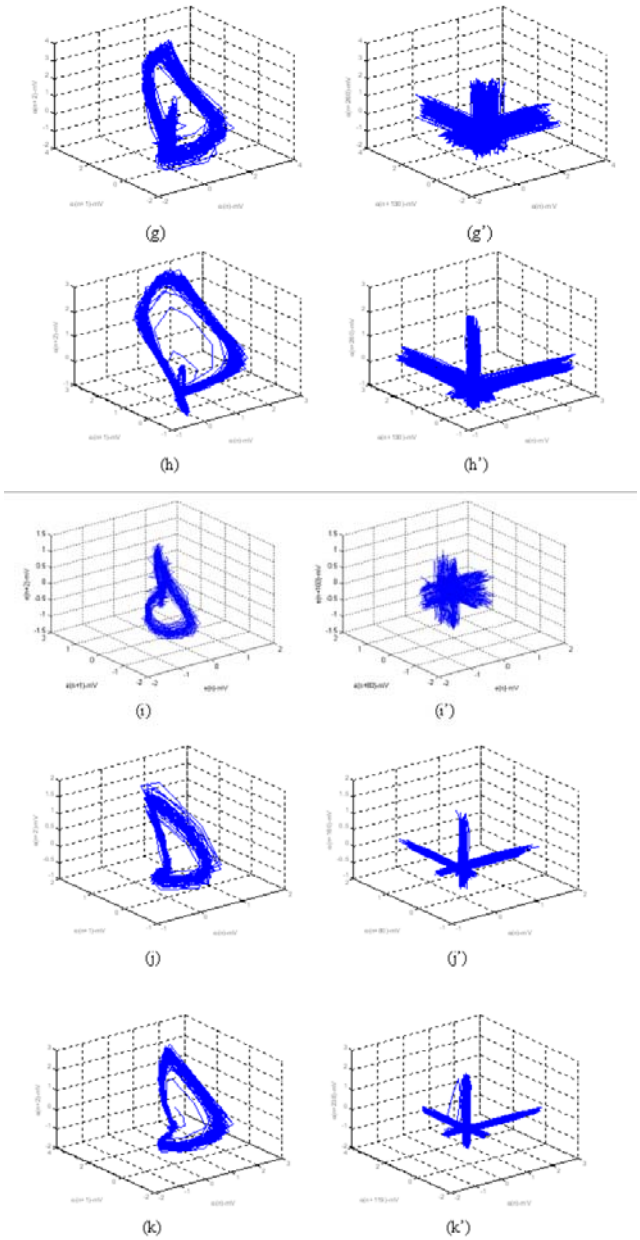


Figure 2: The left panel is the power spectrum distribution (FFT of a data string of 512 sample points lasting 4ms), the right hand panel is the Lyapunov exponents of the data string of the ECG (sliding window of 610 samples)

3. Results and Evaluation

Despite that the signals belong to the same class of ECG namely, the normal sinus rhythm, their phase space plots are quite different. The three dimensional phase space display the attractors of the heart beat dynamics. of the ECG signals was reconstructed using the delay time embedding method, where the delay can be chosen one or more samples. Figures 1,2 and3 show the reconstructed phase space graphs of 18 data sets using one minute string data of ECG signals. A three dimensional phase space graph was constructed using e(t), e(t+τ) and e(t+2τ). The choice of τ is important to find out a visual characterization of human ECG classes. For τ = 1, several different shapes of phase portraits were determined see Figures 1-(a-f),2-(g-l),3-(m-r). An appropriate choice of τ leads to put into evidence a common visual shape of the three dimensional representation.



Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

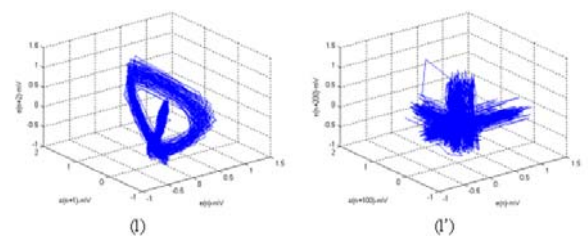
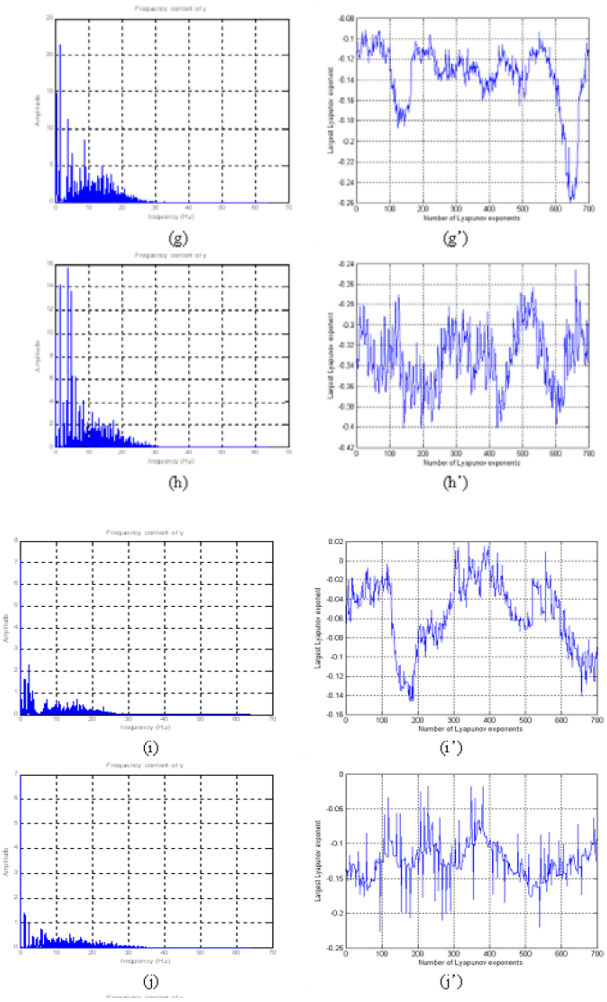


Figure 3: phase plots for the ECG signals: attractor and reconstruction of the attractor. The left panel is a single 7680 point trajectory plotted in e(t), e(t+1) and e(t+2) coordinates (embedding lag of 1), the right hand panel is delay embedding in three dimension (embedding lag of 100 in figures g',h' and l', 80 in figures i' and j', and 119 in figure k') of the e(t) coordinate.



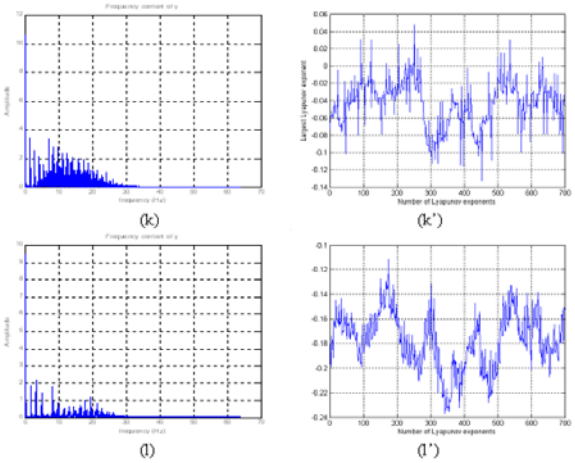
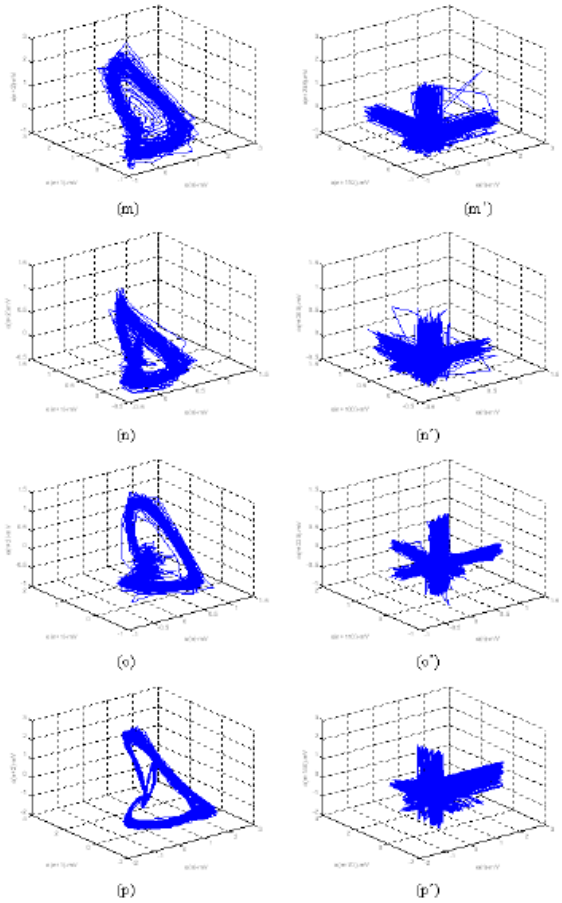


Figure 4: The left panel is the power spectrum distribution (FFT of a data string of 512 sample points lasting 4ms)), the right hand panel is the Lyapunov exponents of the data string of the ECG (sliding window of 610 samples).



Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

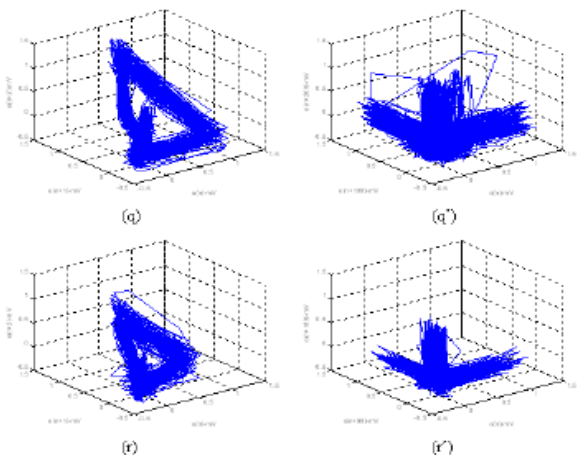
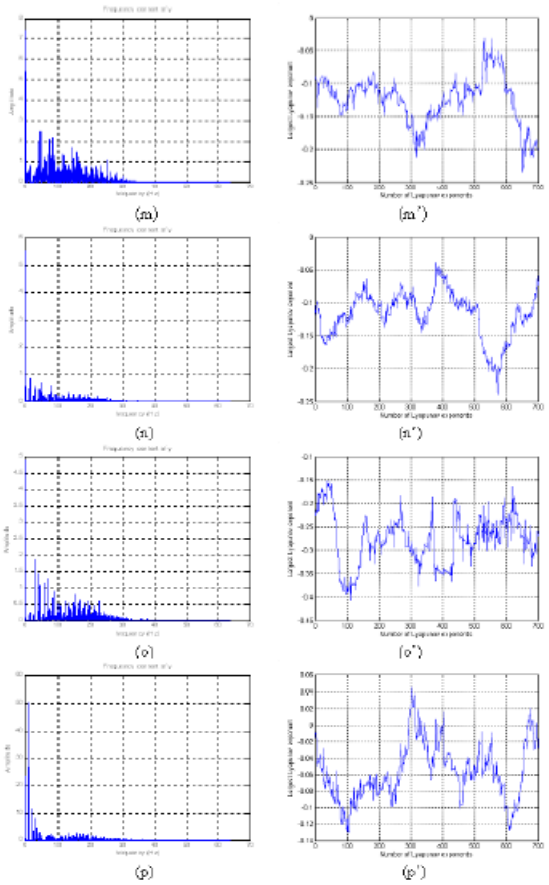


Figure 5: phase plots for the ECG signals: attractor and reconstruction of the attractor. The left panel is a single 7680 point trajectory plotted in e(t), e(t+1) and e(t+2) coordinates (embedding lag of 1), the right hand panel is a delay embedding in three dimension (embedding lag of 100 in figures n' and q', 70 in figures p', 90 in figure r', 110 in figure o' and 119 in figure m') of the e(t) coordinate.



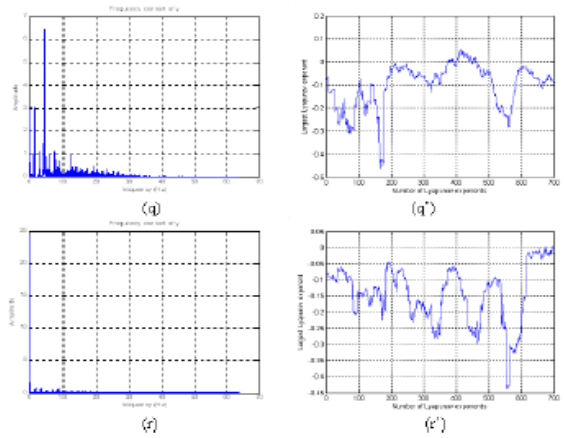


Figure 6: The left panel is the power spectrum distribution (FFT of a data string of 512 sample points lasting 4ms), the right hand panel is the Lyapunov exponents of the data string of the ECG (sliding window of 610 samples)

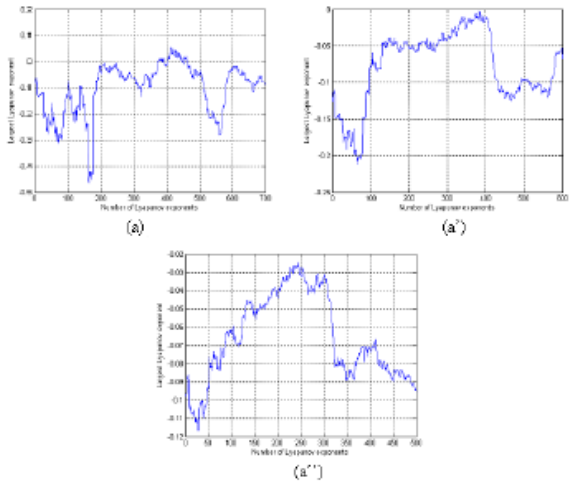


Figure 7: The Lyapunov exponents of the ECG beats (record. nsrd19140)(a) sliding window of 610 samples, a') sliding window of 1610 samples, (a") sliding window of 2610 samples

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

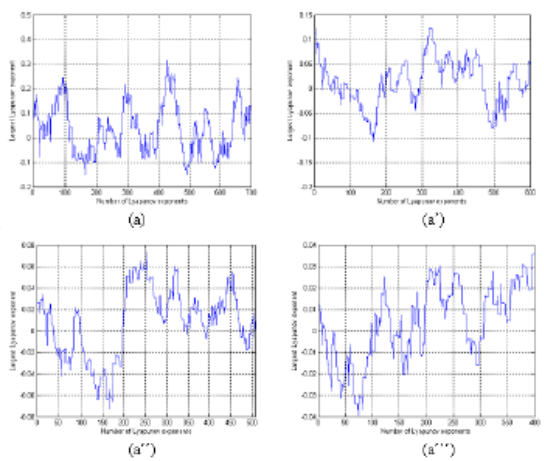


Figure 8: The Lyapunov exponents of the ECG beats(record. nsrd16420) (a) sliding window of 610 samples, (a') sliding window of 1610 samples, (a") sliding window of 2610 samples,(a"') sliding window of 3610 samples.

Such visual phase plots resemble to three invariant cross axes parallel to the phase space axes see Figures 1-(a'-f'),2-(g'-l'),3-(m'-r'). Such particular geometric shape of phase space may related to a certain class of healthy people ECG dynamics, nevertheless , we should not neglect the fact that ECG signals may be noise corrupted.

We develop the spectral point of view, through the parameters defined in table2, namely the Fourier constant term *a*0, the predominant frequency *fM* and the corresponding amplitude *AM*. Each 1-minute ECG segment includes 7680 samples, the FFT is applied to the 4096 initial samples (32*s*). The Largest Lyapunov exponents were computed by using a sliding window of 610 discrete data (contains nearly 8 heart beats). For each string of data(610 samples), 700 Lyapunov exponents of the ECG signal samples are computed and plotted in Figures 4-(a'-f'),5-(g'-l'),6-(m'-r'). For most cases treated in this study, the Largest Lyapunov Exponent are negatives, meaning that the ECG signals with normal sinus rhythm are generally not chaotic. Nevertheless, few cases exhibit positive Largest Lyapunov exponents thereby confirming the chaotic nature of the associated data strings see Figures 2-d',4-(i',k'),6-(p',q').
Regarding the power spectra distribution of the considered ECG signals, all of the power is contained in frequencies below 30 Hz with peak power occurring in the range of 0.25 to 8 Hz. The ratio of *a*0 to *AM* is in the range of 0.27 to 21.(see table 2)
Plots from Figure 7.(a) show that some narrow width intervals of 610 samples having positive LLEs may possess a chaotic behavior, it is seemingly a 'local chaocity'. whereas intervals of larger widths such as 1610 and 2610 samples (see Figure 7.(a') Figure 8: The Lyapunov exponents of the ECG beats(record. nsrd16420) (a) sliding window of 610 samples, (a') sliding window of 1610 samples, (a") sliding window of 2610 samples,(a"') sliding window of 3610 samples and (a")) give rise to negative LLEs only. Certainly a signal having a larger LLE is supposed to be 'more chaotic' than a signal having a lower one. Thereby, the chaotic behavior put into evidence in few data strings is vanishing by enlarging the interval widths, because initially the LLEs have positive low values. whereas when the LLEs are initially large, increasing the width of the intervals for calculating these values only weaken the latter which remain positives see Figure 8 (a.610 samples, a'.1610 samples,a".2610 samples, and a"'.3610 samples).

4. Discussion

It was the aim of this study to find out a common feature that may be a specific characteristic of a certain category of ECG signals namely that corresponding to human heart beats with normal sinus rhythm. In a typical QRS of normal duration,virtually all of the power is contained in frequencies below 30 Hz with peak power occurring in the range of 4 to 12 Hz, in our study a peak power occurred at 0.25Hz.
Results show, in certain cases in Figures 2-d',4-(i',k'),6-(p',q'), that the LLEs were above zero, and this validated that the non linear dynamics really exist in cardiovascular system.
When the LLE decreases the complexity of the attractors decrease and the heart electrical activity becomes more regular. For the majority of the signals examined in this study, the Lyapunov exponents are negatives, except for few cases which possess certain intervals with positive Lyapunov exponents. The results of this study demonstrate that every subject in the studied group had an anchor-shaped Poincar´e plot resulting from an appropriate choice of the delay time. The inter-beat time can be considered as a 'period' of the ECG signal although the latter cannot be perfectly periodic. In this study, it has been obvious to note that the delay time is close to the value of the so-called ECG 'period'. Thus, an open question in future researches, is to find out a relationship between the delay time and the heart beat 'period'. The chaotic behavior may characterize ECGs of healthy people as well as those of diseased ones. Furthermore, certain long-term ECG recordings may include local distortions, disturbances or noise of any origin. The problem that arises is how wide the range to choose to calculate the Largest Lyapunov Exponent? In fact, it must be pointed out that Low distortions occurring on the ECG, not necessarily of pathological origin, can give rise to chaotic behavior in a narrow range (positive LLE).The extension of the width of

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

such interval leads to having a negative LLE meaning that the overall behavior is not chaotic. The Fourier constant term may be too high compared to the amplitude of the predominant harmonics; in figure 6.(r), the ratio of *a*0 to *AM* equals 20.94, the frequency of the predominant harmonic is 0.25 Hz. The estimation of the heart beat rate expressed in beats per minute (102 BPM) of such subject is the highest among all the others (see table 1).

5. Summary and Conclusion

Applying nonlinear signal processing techniques to signals like ECG provides very useful information for detection of cardiac abnormalities. The proposed technique have been shown to be effective for the characterization of a certain class of ECG signals namely those of normal sinus rhythm. In this study we are concerned with the evolution of the dynamics in phase space and the related properties in temporal and frequential domains. The main findings highlighted are essentially the following: - characterization of a class of ECG signals with Normal Sinus Rhythm through anchor-shaped Poincar´e plot resulting from an appropriate choice of the delay time. - chaotic behavior is not a specific property of the ECG in healthy people. – spectral analysis can provide additional information to characterize the dynamics of the ECG. given the small number of cases studied, these results need to be confirmed by larger and prospective clinical investigation. The obtained results are compatible with a non linear complex behavior in heart rate variability of human being. The different structure of the return map of the chosen group could express some information that we are not able to define otherwise.

The methods described in this study may be useful not only for the characterization of the particular class of ECGs studied, but can serve as a benchmark to which pathological cases can be compared. Future researches should be devoted to investigate more profoundly eventual relationships between nonlinear indices and the biological aspects of heart beat dynamics.

Table .3: List of figures

Record	Figures
nsrdb16265	Figure1-(a),(a'),Figure2-(a),(a')
nsrdb16272	Figure1-(b),(b'),Figure2-(b),(b')
nsrdb16273	Figure1-(c),(c'),Figure2-(c),(c')
nsrdb16420	Figure1-(d),(d'),Figure2-(d),(d')
nsrdb16483	Figure1-(e),(e'),Figure2-(e),(e')
nsrdb16539	Figure1-(f),(f'),Figure2-(f),(f')
nsrdb16773	Figure3-(g),(g'),Figure4-(g),(g')
nsrdb16786	Figure3-(h),(h'),Figure4-(h),(h')
nsrdb16795	Figure3-(i),(i'),Figure4-(i),(i')
nsrdb17052	Figure3-(j),(j'),Figure4-(j),(j')
nsrdb17453	Figure3-(k),(k'),Figure4-(k),(k')
nsrdb18177	Figure3-(l),(l'),Figure4-(l),(l')
nsrdb18184	Figure5-(m),(m'),Figure6-(m),(m')
nsrdb19088	Figure5-(n),(n'),Figure6-(n),(n')
nsrdb19090	Figure5-(o),(o'),Figure6-(o),(o')
nsrdb19093	Figure5-(p),(p'),Figure6-(p),(p')
nsrdb19140	Figure5-(q),(q'),Figure6-(q),(q')
nsrdb19830	Figure5-(r),(r'),Figure6-(r),(r')

References

[1] Ubeyli E.D., 'Detecting variabilities of ECG signals by Lyapunov exponents', Neural Computing and Applications 18, pp 653-662.

[2] Babloyantz A., and Destexhe A., 'Is the normal heart a periodic oscillator' Biological Cybernetics, 1988, 58:203.

[3] Fell J., Mann K., Roschke J., and Gopinathan S., 'Nonlinear analysis of continuous ECG during sleep I. Reconstruction', Biological Cybernetics

[4] Fell J., Mann K., Roschke J., and Gopinathan S., 'Nonlinear analysis of continuous ECG during sleep II. Reconstruction', Biological Cybernetics 82, pp. 485-491, 2000.

[5] Bartsch R., Hennig T., Heinen A., Heinrichs S., and Maass P., 'Statistical analysis of fluctuations in the ECG morphology', Physica A 354, pp 415- 431, 2005.

[6] Yum M.K.,'Non-linear cardiac dynamics and morning dip: an unsound circadian rhythm', Clin Physiol. 1999 Jan;19(1):56-67

[7] Xinbao N., Chunhua B., Jun W., and Ying C., 'Research progress in nonlinear analysis of heart electric activities', Chinese Science Bulletin, Volume 51, N 4 pp 385-393, 2006.

[8] Pavlov A.N., Janson N.B., Anishchenko V.S., Gridnev V.I., and Dovgalevsky P.Y. 'Diagnostic of cardio-vascular disease with help of largest Lyapunov exponent of RR-sequences', Chaos, Solitons and Fractals, volume 11, Issue 5, pp. 807-814, April 2000.

[9] Zhenzhou W., Zheng L., Yixiang W., Xinbao N., and Yuzheng L., 'Lyapunov exponents for synchronous

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

12-lead ECG signals', Chinese Science Bulletin Vol. 47 N 21 November 2002.

[10] Ubeyli E.D., 'Adaptive neuro-fuzzy inference system for classification of ECG signals using Lyapunov exponents', Computer Methods and Programs in Biomedicine 93, pp 313-321.

[11] Xia L., Tianliang K., Jinghua L., and Xin T., 'Largest Lyapunov index of EHRV for analyzing status of cardiovascular system', IFMBE Proceedings 19, pp. 361-362, 2008.

[12] Wang N. and Jiong R., 'Principal components cluster analysis of ECG time series based on Lyapunov exponents spectrum', Chinese Science Bulletin, Volume 49, N 18 pp 1980-1985, 2004.

[13] Mrowka R., Patzak A., Schubert E., and Persson P. B., 'Linear and nonlinear properties of heart rate in postnatal maturation', Cardiovascular Research 31, pp. 447-454, 1996.

[14] Govindan R.B., Narayanan K., and Gopinathan M.S., 'On the evidence of deterministic chaos in ECG: Surrogate and predictability analysis', Chaos, volume 8, N2, june 1998.

[15] Klonowski W., Olejarczyk E., and Stepien R.,'Chaocity and Dimensional Complexity of EEG-Signal, International Symposium on Nonlinear Theory and its Applications, Special Session 'Nonlinear Dynamics in Brain and Nervous System', NOLTA'2001, Miyagi, Japan, October 28 - November 1, 2001.

[16] Lombardi F., Makikallio T.H., Myerburg R.J., and Huikuri H.V., 'Sudden cardiac death: role of heart rate variability to identify patients at risk', Cardiovascular Research 50, pp. 210-217, 2001.

[17] Acharya R.U., Kannathal N., Sing O.W., Ping L.Y., and Chua T., 'Heart rate analysis in normal subjects of various age groups', Biomedical Engineering Online 2004, 3:24.

[18] Mohan A., James F., and Fazil S., 'Design and Development of a Heart Rate Variability Analyzer', Journal of Medical Systems, Springer Science and Business Media, LLC 2010.

[19] Guzzetti S, Signorini MG., Cogliati C., Mezzetti S., Porta A., Cerutti S., and Malliani A., 'Non-linear dynamics and chaotic indices in heart rate variability of normal subjects and heart-transplanted patients. Cardiovasc Res 1996;31:441446.

[20] Voss A., Kurths J., Kleiner H.J., Witt A., Wessel N., Saparin P., Osterziel K.J., Schurath R., and Dietz R., 'The application of methods of nonlinear dynamics for the improved and predictive recognition of patients threatened by sudden cardiac death', Cardiovascular Research 31, pp. 419-433, 1996.

[21] Goldberger A. L. et al.,'PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals', Circulation,2000 June 13(23), Vol. 101,pp.215-220.

[22] Kaoru F., and Takashi Y., ' Estimating the embedding dimension and the delay time from chaotic time series with dynamic noise', Journal of the Japan Statistical Society, Vol. 31, No. 1, pp 27-38, 2001.

[23] Hong-guang M., and Chong-zhao H., 'Selection of Embedding Dimension and Delay Time in Phase Space Reconstruction', Front. Electr. Electron. Eng. China (2006) 1: pp 111-114.

Dr. Nisar Hundewale received his Ph.D. in Computer Science from Georgia State University, USA. He has worked at National Institutes of Health (NIH), USA, as post-doctoral Fellow. Currently, he is an Assistant Professor at Taif University. His research interests are Algorithms, Machine Learning, Bioinformatics, Distributed Computing, and Networking.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Performance evaluation of Statistical Approaches for Automatic Text-Independent Speaker Recognition using Robust Features

R. Rajeswara Rao[1] , A. Prasad[2] and Ch. Kedari Rao[3]

[1] Professor of CSE, MGIT
Hyderabad, India

[2]Professor of CSE,Vignan University, Guntur
Guntur, India

[3] Asst. Prof, Dept of CSE, DVRCET
Hyderabad, India

Abstract

This paper introduces the performance evaluation of statististical approaches for Automatic-text-independent Speaker Recognition system. Automatic-text-independent Speaker Recognition system is to quickly and accurately identify the person from his/her voice. The study on the effect of feature vector size for good speaker recognition demonstrates that the feature vector size in the range of 18-22 can capture speaker related information effectively for a speech signal sampled at 16 kHz. it is demonstrated that the timing varying speaker related information can be effectively captured using hidden Markov models (HMMs) than GMM. It is established that the HMM based speaker recognition system requires significantly less amount of data during both during training as well as in testing than GMM based Speaker Recognition System. The performance evaluation of speaker recognition study using robust features for HMM based method and GMM based method is exploited for different mixtures components, training and test durations We demonstrate the speaker recognition studies on TIMIT database.

Keywords: hidden Markov models (HMMs), Gaussian Mixture Model (GMM)), MFCC, Robust Features, Speaker

1. Introduction

Speaker recognition refers to recognizing persons from their voice. No two individuals sound identical because their vocal tract shapes, larynx sizes, and other parts of their voice production organs are different. In addition to these physical differences, each speaker has his or her characteristic manner of speaking, including the use of a particular accent, rhythm, intonation style, pronunciation pattern, choice of vocabulary and so on. State-of-the-art speaker recognition systems use a number of these features in parallel, attempting to cover these different aspects and employing them in a complementary way to achieve more accurate recognition.

An important application of speaker recognition technology is forensics. Much of information is exchanged between two parties in telephone conversations, including between criminals, and in recent years there has been increasing interest to integrate automatic speaker recognition to supplement auditory and semi-automatic analysis methods.

Automatic speaker recognition is an application of pattern recognition. Speaker recognition system, like any other pattern recognition system, can be represented as shown in Fig. 1. This task involves three phases, feature extraction phase, training phase and testing phase [1]. Training is the process of familiarizing the system with the voice characteristics of a speaker, whereas testing is the actual recognition task.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

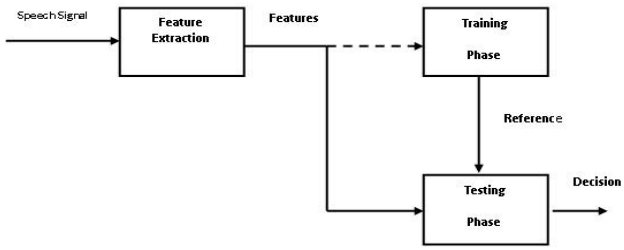


Fig. 1: A typical Block diagram representation of a speaker recognition task.

1.1 Feature extraction

For any pattern recognition task like Automatic Speaker Recognition (ASR), the relevant information has to be captured in terms of suitable feature vectors. In speaker recognition, the feature vectors are in general some parameter vectors extracted from frames of the speech signal. Most of the present day ASR systems are developed using parameters that are derived based on spectral analysis, and the speaker variability is captured in terms of the distribution of these feature vectors. But, it is a fact that the spectrum of a signal is prone to channel characteristics and noise. Channel characteristics and noise play a prominent role in the performance of spectral feature-based systems [2]. Another drawback with the existing techniques is the way in which speaker-specific information is being captured. Mostly, they are statistical techniques, capturing the variability in terms of distribution of the feature vectors and hence large amount of data is required for a better estimate.
Since all the real world services have to deal with speech coming over telephone channel, the ASR systems have to be robust to environmental variations. Also, the requirement of large amount of data has to overcome, as in the real world applications we may not have large amount of data to recognize a person. Hence, in order to make the ASR work in noisy conditions, and with less amount of data, features other than those derived based on spectral analysis also need to be explored.

1.2 Selection of Features

Speech signal includes many features of which not all are important for speaker discrimination. An ideal feature would

- have large between-speaker variability and small within-speaker variability
- be robust against noise and distortion
- occur frequently and naturally in speech
- be easy to measure from speech signal
- be difficult to impersonate/mimic
- not be affected by speaker's health or long-term variations in voice.

1.3 Motivation to use Mel frequency cepstral coefficients (MFCC)

Since our interest is in capturing global features which corresponds the low frequency or pitch components are to be emphasized. To fulfil this requirement it is felt that MFCC are most suitable as they emphasize low frequency and de-emphasize high frequencies

1.4 Mel frequency cepstral coefficients (MFCC)

In this phase the digital speech signal is partitioning into segments (frames) with fixed length 10-30 ms from which the features are extracted due to their spectral qualities. Spectrum is achieved with fast Fourier transformation [3]. Then an arrangement of frequency range to mel scale follows according to relation

$$f_{mel} = 2595 \log\left(1 + \frac{f_{Hz}}{700}\right) \quad (1)$$

By logarithm of amplitude of mel spectrum and applying reverse Fourier transformation we achieve frame cepstrum:

$$mel - cepstrum(frame) = FFT^{-1}\left[mel(\log|FFT(frame)|)\right]$$

The FFT-base cepstral coefficients are computed by taking IFFT of the log magnitude spectrum of the Speech signal. The mel-warped cepstrum is obtained by inserting a intermediate step of transforming the frequency scale to place less emphasis on higher frequencies before taking the IFFT [4][5][6].

1.5 High-Level Features

Speakers differ not only in their voice timbre and accent/pronunciation, but also in their lexicon – the kind of words the speakers tend to use in their conversations. The work on such "high-level" conversational features was initiated in [7] where a speaker's characteristic vocabulary, the so-called idiolect, was used to characterize speakers. The idea in "high-level" modeling is to convert each utterance into a sequence of tokens where the co-occurrence patterns of tokens characterize speaker differences. The information being modeled is hence in categorical (discrete) rather than in numeric (continuous) form.
The tokens considered have included words [7], phones [8][9], prosodic gestures (rising/failing pitch/energy) [10][11][12], and even articulatory tokens (manner and place of articulation) [13]. The top-1 scoring Gaussian mixture component indices have also been used as tokens [14][15][16].

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Sometimes several parallel tokenizers are utilized [9][14][17]. This is partly motivated by the success of parallel phone recognizers in state-of-the-art spoken language recognition [18][19]. This direction is driven by the hope that different tokenizers (e.g. phone recognizers trained on different languages or with different phone models) would capture complementary aspects of the utterance. As an example, in [14] a set of parallel GMM tokenizers [15[[16] were used. Each tokenizer was trained from a different group of speakers obtained by clustering.

One of the issues in speaker recognition is how to represent utterances that, in general, have a varying number of feature vectors. In early studies [20] speaker models were generated by time-averaging features so that each utterance could be represented as a single vector. The average vectors would then be compared using a distance measure [21] which is computationally very efficient but gives poor recognition accuracy. Since the 1980's, the predominant trend has been creating a model of the training utterances followed by "data-to-model" type of matching at run-time (e.g. likelihood of an utterance with respect to a GMM). This is computationally more demanding but gives good recognition accuracy.
Interestingly, the speaker recognition community has recently re-discovered a robust way to present utterances using a single vector, a so-called super vector [22].

2. Exploring Robust Features for Speaker Recognition

Here, the GMM is used as front-end to extract features vectors from speech signal. For the ASR task, the basic requirement is to obtain the feature vectors form the speech signal. Recently, some attempts are made to explore the alternative representation of feature vectors based on GMM feature extraction.
For Speaker Recognition task, robust features are derived from the speech signal based on estimating a Gaussian mixture model. The underlying speaker discrimination information is represented by Gaussians. The estimated GMM parameters means, co-variance and component weight can be related to the formant locations, bandwidths and magnitudes.
For the proposed new feature vectors, from the speech signal of a speaker S_i, a 12 dimensional MFCC feature vectors are obtained with a window size of 20ms and window shift of 5 ms. These MFCC feature vectors are distributed into 'R' Gaussians mixtures as shown in Fig. 2.

Fig. 2: R Gaussians for Speaker S_i .

The feature vector X=(X1, X2,……, X12) is passed through a Gaussian G1 by calculating a Gaussian probability P1 using Gaussian probability density function. This P1 is first coefficient in the new feature vector. In the same way feature vector X is passed through R Gaussians by creating R feature vector coefficients namely P1,P2,….,PR, as shown in Fig. 3. These R coefficients create a new R dimensional feature vector. The newly created R dimensional feature vector is shown in the Fig. 4.
Experiments are carried to find the dimension new feature vector for good speaker recognition performance. This is done by varying the number of Gaussians from 12 to 30, i.e number of coefficients in the new feature vectors. When the numbers of coefficients are 20, the good identification performance is achieved.

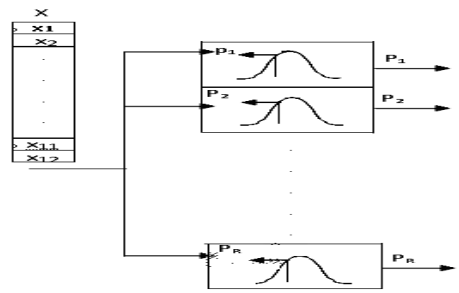


Fig. 3: Parameter estimation for new vector P. When R=15, the optimal recognition performance has been achieved.

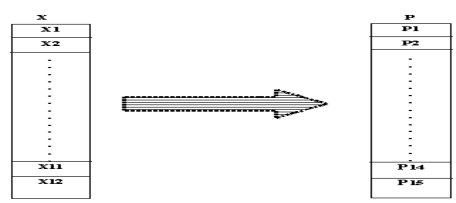


Fig. 4: Transforming from 12 dimensional MFCC feature vector to R dimensional feature vector.

3. Performance evaluation of Parametric approaches

Parametric approaches are model-based approaches. The parameters of the model are estimated using the training feature vectors. It is assumed that the model is adequate to represent the distribution. The most widely used parametric approaches are GMM and HMM based approaches.

3.1 Gaussian Mixture Models for LID

GMM is a classic parametric method best used to model speaker identities due to the fact that Gaussian components have the capability of representing speaker discrimination

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

information effectively. Gaussian classifier has been successfully employed in several text-independent speaker recognition applications. As shown in Fig. 5 in a GMM model, the probability distribution of the observed data takes the form given by the following equation [23][24].

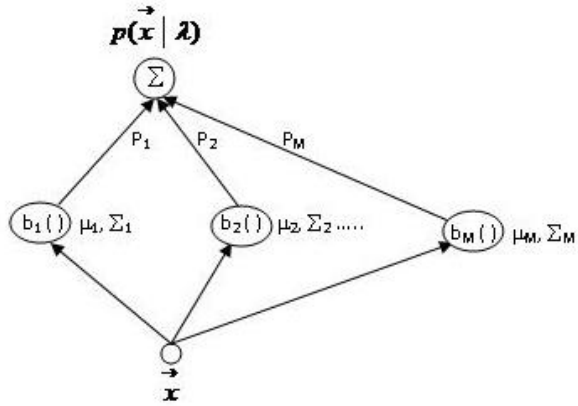


Fig. 5: Gaussian Mixture Model

$$p(\bar{x} \mid \lambda) = \sum_{i=1}^{M} p_i b_i(\bar{x})$$

Where M is the number of component densities, $\bar{x}$ is a D dimensional observed data (random vector), $b_i(\bar{x})$ are the component densities and p_i are the mixture weights for i = 1, .., M.

$$b_i(\bar{x}) = \frac{1}{(2\pi)^{D/2} |\sum_i|^{1/2}} \exp\left\{-\frac{1}{2}(\bar{x}-\bar{\mu}_i)^T \sum_i^{-1} (\bar{x}-\bar{\mu}_i)\right\}$$

Each component density $b_i(\bar{x})$ denotes a D-dimensional normal distribution with mean vector $\bar{\mu}_i$ and covariance matrix $\sum_i$. The mixture weights satisfy the condition $\sum_{i=1}^{M} p_i = 1$ and therefore represent positive scalar values. These parameters can be collectively represented as $\lambda = \{p_i, \bar{\mu}_i, \sum_i\}$ for i = 1 … M. Each speaker in a language system can be represented by a GMM and is referred by the language respective model λ.

The parameters of a GMM model can be estimated using maximum likelihood (ML) [25] estimation. The main objective of the ML estimation is to derive the optimum model Parameters that can maximize the likelihood of GMM. Unfortunately direct maximization using ML

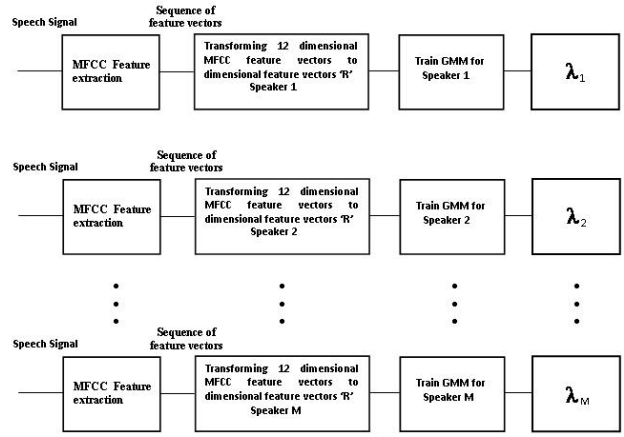


Fig. 6: Training GMM for Speaker Recognition Task

Parameter estimation is not possible and therefore a special case of ML estimation known as Expectation-Maximization (EM) [25] algorithm is used to extract the model parameters.

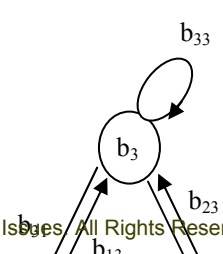
The GMM likelihood of a sequence of T training vectors $X = \{\bar{x}_1, .. \bar{x}_T\}$ can be given as [25]

$$p(X \mid \lambda) = \prod_{t=1}^{T} p(\bar{x}_t \mid \lambda)$$

The EM algorithm begins with an initial model λ and tends to estimate a new model $\bar{\lambda}$ such that $p(X \mid \bar{\lambda}) \geq p(X \mid \lambda)$ [19]. This is an iterative process where the new model is considered to be an initial model in the next iteration and the entire process is repeated until a certain convergence threshold is obtained.

3.2 Continuous Ergodic Hidden Markov model for speaker recognition

The HMM is a doubly embedded stochastic process where the underlying stochastic process is not directly observable. HMMs have the capability of effectively modeling statistical variations in spectral features. In a variety of ways, HMMs can be used as probabilistic speaker models for both text-dependent and text-independent speaker recognition [27][28][29]. HMM not only models the underlying speech patterns but also the temporal sequencing among the sounds. This temporal modeling is advantageous for text-dependent speaker recognition system. Left Right HMM can model temporal sequence of patterns only, where as to capture the patterns of different type ergodic HMM is used [21].



Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

FIG. 4:

Fig. 7: Three-state ergodic HMM.

As shown in the Fig. 8 in the training phase, one HMM for each speaker is obtained (i.e., parameters of model are estimated) using training feature vectors. The parameters of HMM are [19] State-transition probability distribution: It is represented by $A = \lfloor a_{ij} \rfloor$

Where

$$a_{ij} = P(q_{t+1} = j \mid q_t = i) \quad 1 \le i, j \le N \qquad (2)$$

defines the probability of transition from state i to j at time t.

For a three state left-right model the state transition matrix is given as

$$A = \{a_{ij}\} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{bmatrix} \qquad (3)$$

The state transition matrix of three state ergodic model is given by

$$A = \{a_{ij}\} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \qquad (4)$$

Observation symbol probability distribution: It is given by $B = [b_j(k)]$ in which

$$b_j(k) = P(O_t = V_k \mid q_t = j) \quad 1 \le k \le M \qquad (5)$$

defines the symbol distribution in state $j = 1,2,3.....N$. The initial state distribution is given by $\pi = P(q_1 = i)$ where

$$\pi_i = P(q_1 = i) \quad 1 \le i \le N \qquad (6)$$

Here, N is the total number of states, and q_t is the state at time t, M is the number of distinct observation symbols per state, and O_t is the observation symbol at time t. In testing phase, $P(O/\lambda)$ for each model is calculated, where $O = (O_1 O_2 O_3 .... O_T)$ Here the goal is to find out the probability for a given model to which the test utterance belongs to. The speaker whose model gives the highest score is declared as the identified speaker. GMM corresponds to a single-state continuous ergodic HMM.

The model parameters can be collectively represented as $\lambda = (A_i, B_i, \pi_i)$ for $i = 1.....M$. Each speaker in a speaker identification system can be represented by a HMM and is referred to by the speaker's respective models λ.

In the testing phase, p (O/λ) for each model is calculated [29]. Where O=(o1o2o3…OT) is the sequence of the test feature vectors. The goal is to find the probability, given the model that the test utterance belongs to that particular model. The speaker model that gives the highest score is declared as the identified speaker.

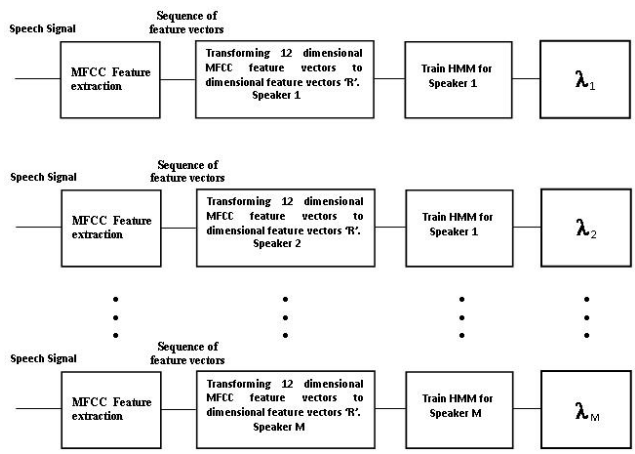


Fig. 8: Training HMM for Speaker Recognition Task

The parameter estimation is not possible and therefore a special case of ML estimation known as Expectation-Maximization (EM) [30] algorithm is used to extract the model parameters.

The GMM likelihood of a sequence of T training vectors $X = \{\bar{x}_1, .. \bar{x}_T\}$ can be given as [30]

$$p(X \mid \lambda) = \prod_{t=1}^{T} p(\bar{x}_t \mid \lambda)$$

The EM algorithm begins with an initial model λ and tends to estimate a new model $\bar{\lambda}$ such that $p(X \mid \bar{\lambda}) \ge p(X \mid \lambda)$ [29]. This is an iterative process where the new model is considered to be an initial model in the next iteration and the entire process is repeated until a certain convergence threshold is obtained.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

4. Experimental Evaluation

4.1 Database used for the study

Speaker Recognition is the task of identifying the speaker from the registered set of speakers. In this paper we consider identification task for TIMIT Speaker database [30].

The TIMIT corpus of read speech has been designed to provide speaker data for the acquisition of acoustic-phonetic knowledge and for the development and evaluation of automatic speaker recognition systems. TIMIT contains a total of 6300 sentences, 10 sentences spoken by each of 630 speakers from 8 major dialect regions of the United States. We consider 100 male speakers and 100 female out of 630 speakers for speaker recognition. Maximum of 30 sec. of speech data is used for training and minimum of 1 sec. of data for testing. In all the cases the speech signal was sampled at 16 kHz sampling frequency. Through out this study, closed set identification experiments are done to demonstrate the feasibility of capturing the speaker-discrimination information from the speech signal. Requirement of significantly less amount data for speaker-discrimination information and Gaussian mixture models is also demonstrated.

4.2 Experimental setup

The system has been implemented in Matlab 7 on Windows XP platform. We have trained the GMM model using Gaussian Components as 4, 8, 16, 32 and 64 and HMM with 2, 3 and 4 states with 4, 8, 16, 32 and 64 components at each state for training speech duration of 10, 20 and 30 sec. testing is performed using different test speech durations such as 1 sec., 3 sec., and 5 sec..

5. Performance Evaluation

The system has been implemented in Matlab7 on windows XP platform. The result of the study has been presented in Table 3. We have used coefficient order of 20 for all experiments. We have trained the GMM and HMM models using Gaussian components as 4, 8, 16, 32 and 64 components at each state for different training speech lengths as 10 sec., 20 sec., and 30 sec. by varying HMM states such as 2, 3 and 4. Testing is performed using different test speech lengths such as 1 sec, 3 sec, and 5 sec.. Here, recognition rate is defined as the ratio of the number of speaker identified to the total number of speakers tested. As shown in Fig. 9, the speaker recognition performance for varying co-efficient \order is depicated. It is established that the co-efficient order range of of 18-22 is found to be optimal for good speaker recognition. As shown in Fig. 10, Fig. 11, Fig. 12 for a 3-state HMM, the recognition rate for testing length for 5 sec. outperformed, where as for testing length of 3 sec. is also on par with 5 sec. testing length.

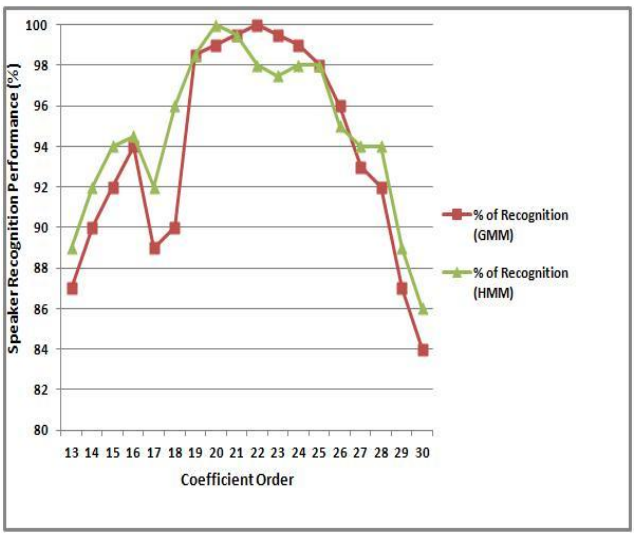


Fig. 9: Speaker Recognition Performance for varying Co-efficient order

Table 1: Speaker Recognition Performance for varying Co-efficient order

Coefficients Order	Speaker Recognition (%) of GMM	Speaker Recognition (%) of HMM
13	87	89
14	90	92
15	92	94
16	94	94.5
17	89	92
18	90	96
19	98.5	98.5
20	99	100
21	99.5	99.5
22	100	98
23	99.5	97.5
24	99	98
25	98	98
26	96	95
27	93	94
28	92	94
29	87	89
30	84	86

Table 2: Avg. Speaker Recognition Performance for varying test duration.

As shown in Table 3,, for a 3-state HMM the percentage (%) recognition for Gaussian Components such as 4, 8, 16, 32 and 64 seems to be uniformly increasing. The minimum number of Gaussian components to achieve good speaker recognition performance seems to be 16 and thereafter the recognition performance is minimal. As shown in Fig. 12.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

No. of Mixture Components	Avg. Speaker Recognition (%)					
	Test Duration					
	1 Sec.		3 Sec.		5 Sec.	
	GMM	HMM	GMM	HMM	GMM	HMM
4	54.66	86.66	86	92.83	92.66	96.16
8	67.83	90.5	95	96.33	97.33	98.66
16	74.66	92.16	97.5	98.16	99	99
32	82.85	94	97.83	97.66	99	98.83
64	79.83	92	96	96.33	98.16	98.16

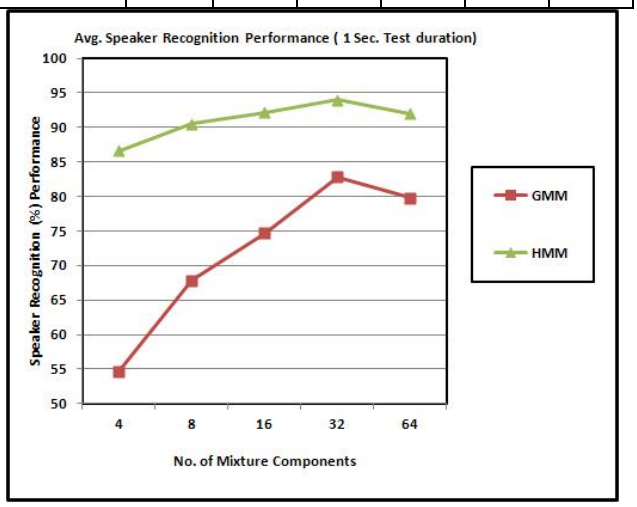


Fig. 10: Avg. Speaker Recognition Performance of GMM/HMM for varying test duration of 1Sec.

As shown in Fig. 10, the average speaker recognition performance for 10 sec., 20 sec. and 30 sec. training duration for varying mixture components as 4, 8, 16, 32 and 64 tested with 1 sec., 3 sec., and 5 sec., test durations for 2, 3 and 4 states indicate that for 20 sec., of training speech duration with 16 mixture components test duration of 3 sec. And for 3-state HMM gives good speaker recognition performance.

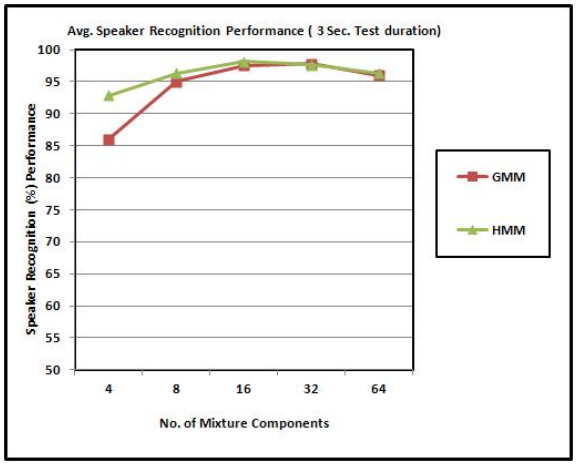


Fig. 11: Avg. Speaker Recognition Performance of GMM/HMM for varying test duration of 3 Sec.

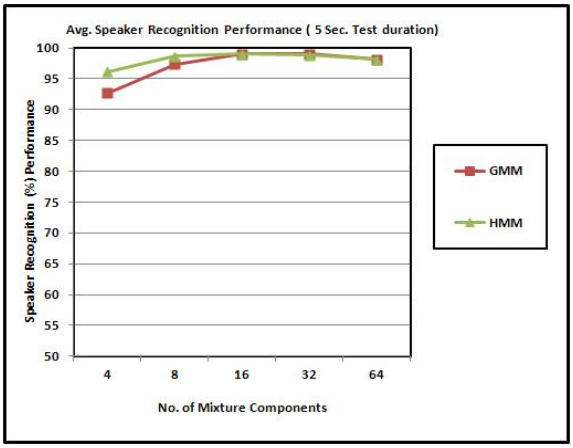


Fig. 11: Avg. Speaker Recognition Performance of GMM/HMM for varying test duration of 5 Sec.

Table 3: Avg. Speaker Recognition Performance of GMM & HMM for 20 Sec. Training duration.

No. of Mixture Components	Avg. Speaker Recognition Performance of GMM	Avg. Speaker Recognition Performance of HMM
4	78.33	96.83
8	87.66	98.33
16	90.83	99.5
32	93.33	99.16
64	94.16	97.83

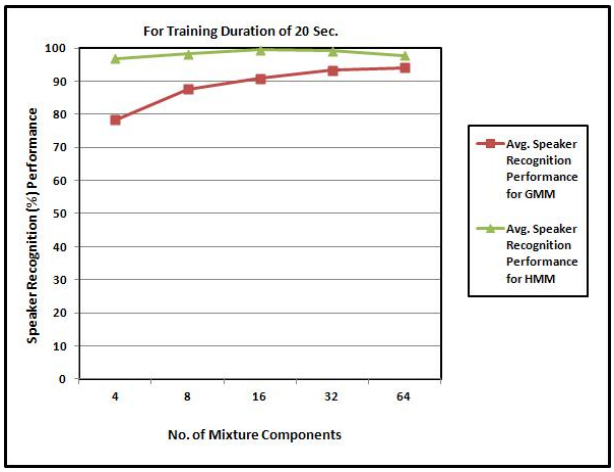


Fig. 11: Avg. Speaker Recognition Performance of GMM & HMM for 20 Sec. training duration

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

1 CONCLUSION

In this work we have demonstrated the importance of coefficient order for speaker recognition task. Speaker discrimination information is effectively captured for coefficient order 20 using a HMM .The recognition performance depends on the training speech length selected for training to capture the speaker-discrimination information. Larger the training length, the better is the performance, although smaller number reduces computational complexity.

Effectiveness of the HMM for speaker recognition task using the time varying speech signal is demonstrated. GMM based approaches do not capture prosodic information and acoustic variations, which vary with time. Hence, in order to capture time varying properties in the speech signal effectively, HMM based system is well suited for speaker recognition task. Continuous HMM embeds benefits associated with GMM approaches. We have not made any attempt to optimize the parameters of the model used for feature extraction, and also the decision making stage. Therefore the performance of language identification may be improved by optimizing the various design parameters

The objective in this paper was mainly to demonstrate the significance of the speaker-discrimination information present in the speech signal for speaker recognition. We have not made any attempt to optimize the parameters of the model used for feature extraction, and also the decision making stage. Therefore the performance of speaker recognition may be improved by optimizing the various design parameters.

.

References

[1]. Furui, "An overview of speaker recognition technology", in Automatic Speech and Speaker Recognition (C.-H. Lee, F. K. Soong, and K. K. Paliwal, eds.), ch. 2, pp. 31-56, Boston: Kluwer Academic, 1996.

[2] D. A. Reynolds, "The effects of handset variability on speaker recognition performance: Experiments on switch board corpus," in Proceedings of IEEE Int. Conf. Acoust., Speech, and Signal Processing, pp. 113-116, 1996.

[3] Dempster, A., Laird, N., and Rubin, D., "Maximum likelihood from incomplete data via the EM algorithm," Journal of the Royal Statistical Society, vol. 39, pp. 1-38, 1977.

[4] Makhoul, J., 1975. Linear prediction: a tutorial review. Proc. IEEE 63, 561-580.

[5] Molau, S., Pitz, M., Schluter, R., and Ney, H., "Computing Mel-frequency cepstral coefficients on the power spectrum," Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 1, pp. 73-76, May. 2001.

[6] Picone, J. W., "Signal modeling techniques in speech recognition," Proceedings of IEEE, vol. 81, no.9, pp. 1215-1247, Sep. 1993.

[7] Doddington.G., Speaker Recognition based on idiolectal differences between speakers. In Proc. 7th European Conference on Speech Communication and Technology (Eurospeech 2001) (Aalborg, Denmark, September 2001), pp. 2521 – 2524.

[8] Andrews.W., Kohler.M., Campbell.J., Godfrey.J., and Hernandez –Cordero.J. Gender-dependent phonetic refraction for speaker recognition. In Proc. Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP 2002) (Orlando, Florida, USA, May 2002), vol. 1, pp. 149-152.

[9] Campbell.W., Campbell.J., Reynolds.D., Jones.D., and Leek.T., Phonetic speaker recognition with support vector machines. In Advances in Neural Information Processing Systems 16, S.Thrun, L.Saul, and B.Scholkopf, Eds. MIT Press, Cambridge, MA, 2004.

[10] Adami.A, Mihaescu.R, Reynolds.D, and Godfrey.J., Modelling Prosodic dynamics for speaker recognition. In Proc. Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP 2003) (Hong Kong, China, April 2003), pp. 788-791.

[11] Chen, Z.-H., Liao, Y.-F., and Juang, Y.-T. Eigen-Prosody analysis for robust speaker recognition under mismatch handset environment. In Proc. Int. Conf. on Spoken language processing (ICSLP 2004) (Jeju, South Korea, October 2004), pp. 1421-1424.

[12] Shriberg, E., Ferrer, L., Kajarekar, S., Venkataraman, A., and Stolcke, A. Modeling Prosodic feature sequences for speaker recognition. Speech Communication 46, 3-4 (July 2005), 455-472.

[13] Leung, K., Mak, M., Siu, M., and Kung, S. Adaptive articulatory feature – based conditional pronunciation modeling for speaker verification. Speech Communication 48, 1 (January 2006), 71-84.

[14] MA, B., Zhu, D., Tong, R., and Li, H. Speaker cluster based GMM tokenization for Speaker recognition. In Proc. Interspeech 2006 (ICSLP) (Pittsburgh, Pennsylvania, USA, September 2006), pp. 505-508.

[15] Torres-Carrasquillo, P., Reynolds, D., and Jr., J. D. Language identification using Gaussian mixture model tokenization. In Proc. Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP 2002) (Orlando, Florida, USA, May 2002), vol. 1, pp. 757-760.

[16] Xiang, B. Text-independent speaker verification with dynamic trajectory model. IEEE Signal Processing Letters 10 (May 2003), 141 -143.

[17] Jin, Q., Schultz , T., and Waibel, A. Speaker identification using multilingual phone strings. In Proc. Int. Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP 2002) (Orlando, Florida, USA, May 2002), vol. 1, pp. 145-148.

[18] Zissman, M. Comparison of four approaches to automatic language identification of telephone speech. IEEE Trans. on Speech and Audio Processing 4, 1 (January 1996), 31-44.

[19] MA, B., Li, H., and Tong, R. Spoken language recognition with ensemble classifier. IEEE Trans. Audio, Speech and Language Processing 15, 7 (September 2007), 2053-2062.)

[20] Markel, J., Oshika, B., and A.H. Gray, J. Long-term feature averaging for speaker recognition. IEEE Trans. Acoustics, Speech, and Signal Processing 25, 4 (August 1977), 330-337.

[21] Kinnunen, T., Hautamaki, V., and Franti, P. On the use of long-term average spectrum in automatic speaker recognition. In 5th Int. Symposium on Chinese Spoken Language Processing (ISCSLP'06) (Singapore, December 2006), pp.559-567.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

[22] Tomi Kinnunen., and Haizhou Li., An overview of Text-Independent Speaker Recognition: from Features to Supervectors. Speech Communication, July 1, 2009.

[23] Gish, H., Krasner, M., Russell, W., and Wolf, J., " Methods and Experiments for text-independent speaker recognition over telephone channels," Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 11, pp. 865-868, Apr. 1986.

[24] Reynolds, D.A., and Rose, R.C., "Robust Text-Independent Speaker Identification using Gaussian Mixture Models," IEEE-Transactions on Speech and Audio Processing, vol. 3, no.1, pp. 72-83,1995.

[25] A. P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm", J. Royal Statist. Soc. Ser. B. (methodologies), vol. 39, pp. 1-38, 1977.

[26] Gish, H., Krasner, M., Russell, W., and Wolf, J., " Methods and Experiments for text-independent speaker recognition over telephone channels," Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 11, pp. 865-868, Apr. 1986.

[27] Reynolds, D.A., and Rose, R.C., "Robust Text-Independent Speaker Identification using Gaussian Mixture Models," IEEE-Transactions on Speech and Audio Processing, vol. 3, no.1, pp. 72-83,1995.

[28] A. P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm", J. Royal Statist. Soc. Ser. B. (methodologies), vol. 39, pp. 1-38, 1977.

[29] K.N. Stevens, Acoustic Phonetics. Cambridge, England: The MIT Press, 1999.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Secured Distributed Cognitive MAC and Complexity Reduction in Channel Estimation for the Cross Layer based Cognitive Radio Networks

Niraj Shakhakarmi

Department of Electrical & Computer Engineering, Prairie View A&M University (Texas A&M University System)
Prairie View, Houston, Texas, 77446, USA

Abstract

Secured opportunistic Medium Access Control (MAC) and complexity reduction in channel estimation are proposed in the Cross layer design Cognitive Radio Networks deploying the secured dynamic channel allocation from the endorsed channel reservation. Channel Endorsement and Transmission policy is deployed to optimize the free channel selection as well as channel utilization to cognitive radio users. This strategy provide the secured and reliable link to secondary users as well as the collision free link to primary users between the physical and MAC layers which yields the better network performance. On the other hand, Complexity Reduction in Minimum Mean Square Errror (CR-MMSE) and Maximum Likelihood (CR-ML) algorithm on Decision Directed Channel Estimation (DDCE) is deployed significantly to achieve computational complexity as Least Square (LS) method. Rigorously, CR-MMSE in sample spaced channel impulse response (SS-CIR) is implemented by allowing the computationally inspired matrix inversion. Regarding CR-ML, Pilot Symbol Assisted Modulation (PSAM) with DDCE is implemented such the pilot symbol sequence provides the significant performance gain in frequency correlation using the finite delay spread. It is found that CR-MMSE demonstrates outstanding Symbol Error Rate (SER) performance over MMSE and LS, and CR-ML over MMSE and ML.

Keywords: *Secured, Distributed, Medium Access Control, Low Complexity, Channel Estimation, Cross Layer, Cognitive Radio Networks.*

1. Introduction

The concept of cross-layer design is to optimize the overall performance (the routing efficiency, throughput, fairness and delay variance) of Cognitive radio networks with the exchange of information between the different layers of among different applications. Physical layer information exchange with the MAC and network layer has also exhibited superior network performance. The CRN cross layer design involves the MAC and PHY layers, and the central decisions are taken by the CR Manager (CRM) which also has the capability of interaction between different modules. The other two modules in the MAC layer are responsible for efficient dynamic channel allocation among mobile nodes with limiting power constraints. The Channel Sense and Allocation scans spectrum in periodic intervals for possible free unoccupied channels. The next module is called the Power-aware Scheduling aims at a multi-objective joint power control and link scheduling of data frames. Additionally, it also performs the hybrid queuing strategy to achieve fairness among requesting applications. The noise and interference measurement, coded-OFDM transceiver and receiver modules and channel estimation are associated with the PHY layer of each node in the network. Noise and Interference measures the interfering power sensed in each sub-channel due to users in adjacent sub-bands. The coded OFDM transceiver and receiver deals with the PHY layer transceiver and receiver, it can use CR technology to select optimal channel for maximum signal power and hence maximum Signal-to-Noise Ratio (SNR) at minimal Bit Error Rate (BER).

The channel estimation approximates the fading condition of the channel as well as the channel error rates. The CE shares the cross-layer information with the CRM to select the best link for data transmission among the adjacent one hop neighbors. The Decision Directed Channel Estimation (DDCE) scheme is preferred for CR in both information symbols and the pilot symbols for channel estimation. Least Square method has low complexity but the energy-fluctuations associated with the near-Gaussian distributed subcarrier-related samples render the poor performance for decision directed channel in OFDM. On the other hand, MMSE in DDCE provide significant performance over LS but complexity need to be mitigated the intensive matrix inversion operation. Similarly, the Maximum Likelihood (ML) DDCE for CR needs novel time-frequency interpretation for complexity reduction.

The spectrum mobility is the ability of a distributed cognitive radio to scan the spectrum and adaptively switch channels to mitigate secondary user's link degradation. Channel reservation mechanism to reduce the collision when forced termination occurs for the cognitive user's communication. The protocol is employed in decentralized multi channel network with continuous spectrum sensing and scanning capability. The channel status table and reservation information is embedded in the control messages and exchanged with communicating nodes and then announced to the neighbor's nodes. A novel idea is to include the reservation information for the backup channel into the control messages of the current transmission instead of sending an extra control packet, and to use the

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

duration identification (ID) to decide when the currently occupied channel will be released. Thus, the reservation information can be heard by all nodes within the transmission range of the sender with little extra overhead.

Cognitive radio with the spread spectrum modulation techniques, provide a highly secure communication format resistant to deliberate narrowband jamming and other obstruction tactics. Spread spectrum technique, has unique feature to make the data look like noise, and secure in the sense that the jamming and the interfering elements are unable to distinguish mixed data with noise being sent over the channel and prevent from eavesdropping. Different encryption techniques such as integrated key encryption public key encryption (RSA, Elliptic, SHA) and private key encryption (DES, Triple DES, AES) algorithms, and integrated trust computing with public cryptography are useful with cognitive radio to provide a form of secure communication. These encryption algorithms make sure that the key that is used at the transmitter side should be provided by the receiver for correct information retrieval and hence ensures the security and also prevent the malicious users from taking control over the system, blocking the access to other secondary users.

2. Related Work

Opportunistic MAC for CR network coexistence with any primary network is a discrete protocol with the advantages of fast-employment and flexibility [1]. The major issue is MAC cannot realize the physical layer status. Furthermore, Multiple channel cognitive radio with backup channel reservation MAC protocol facilitate quick and smoother shift to a reserved channel before the secondary link degrades due to primary user arrival [2] . The protocol significantly improves the network throughput; keep the privilege to the licensee and utilizes the available free channel perfectly with less packet exchange overhead. The simplicity feature makes OC-MAC a more feasible solution than other schemes even without synchronization. But, the major drawback is the physical layer performance degradation due to the lack of co-operation between PHY and MAC layers and lack of security check during the communication [3]. Additionally, OFDM Based Framework for cognitive radio network (OFDM-Based CRN) is deployed in cross-layer design in CRN [4]. The optimization of joint power control and link scheduling is done on the cross-layer architecture of OFDM-based CRN. The dynamic channel assignment within power constraints and the transmission power reduce the interference and improve performance. But, the channel collision is not addressed specifically to optimize the instantly available channel bandwidth. For security, secured and reliable CR communication solutions include the integration of the merits of spread spectrum modulation, using integrated key encryption algorithms as well as trust computing with encryption, and its potential to switch over various frequency bands [5] ,[6].

A fundamental component of the DDCE scheme for complexity reduction in channel estimation for cross layer based CR is a posteriori Least Squares (LS) temporal estimator of the CR-OFDM subcarrier-related Frequency-Domain Channel Transfer Function (FD-CTF) coefficients [7]. The precision of the consequential secular estimates is normally improved using either one or two-dimensional interpolation using both the time and the frequency-domain correlation between the adjacent FD-CTF coefficients. The LS-based secular FD-CTF estimator, with the energy of the transmitted subcarrier related symbols varies as a function of both the modulated sequence and the alternative of the potentially non-constant modulus modulation scheme itself. Minimum Mean Square Error (MMSE) Channel Impulse Response (CIR) estimator-aided DDCE method causes to the complexity reduction in channel estimation which is comparable with LS.

The drawbacks of high SNR performance limitation for MMSE, limitation on the spacing/number of pilots can be mitigated by the finite delay spread of the channel and developing a low complexity algorithm capable of estimating the channel from part of the carriers only [8]. ML estimator can be interpreted as a transformation from frequency domain to time domain and back to frequency. The actual estimation can be done in the time domain, where the number of parameters such as the channel length is small. The estimator can be achieved by minimizing a quadratic criterion, which, combined with the small number of parameters, for lower complexity algorithm with Pilot Symbol Assisted Modulation (PSAM) and Constrained Least Squares (CLS).

3. Secured Distributed Cognitive MAC

The implementation of the secured distributed cognitive MAC and complexity reduction in channel estimation is depicted in Fig. 1. When the user application received by the CR manager (CRM) for the link request and processing delay constraint, it performs initial step of authentication using security policy and then forwards the authenticated link request to the Channel Sense and Allocation (CSA) for available free channels. The CSA consists of direct status check of the noise and interference measurement module for probable channel interference power to

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

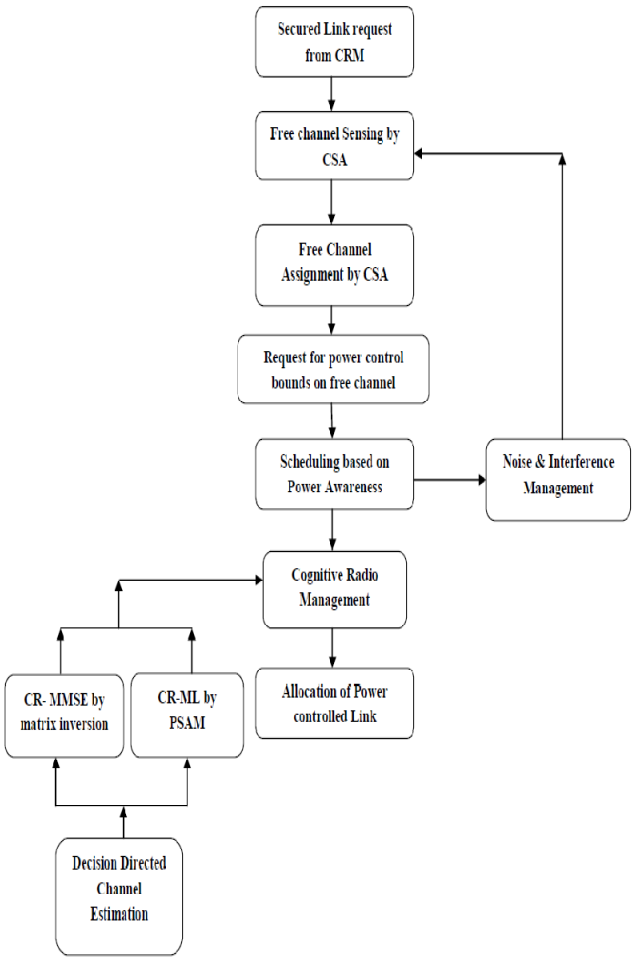


Fig. 1 Distributed Cognitive MAC and Complexity Reduction in Channel Estimation

to detect the free or idle channels. From the response of the noise and interference sensing module located in the PHY layer, the channel sensing and allocation module selects entirely free channel for both authenticated transmitter and receiver by using Channel Endorsement and Transmission Policy (CETP) and sends the detected free channels with their respective identifications to the CRM. This results that the CRM has now the complete information regarding the available free channels for the authenticated requesting applications users. The CRM requests the power aware scheduling module for transmit power limits on each of available free channels. The CRM also sends information about the delay constraints for the requesting application users. The scheduling module recommends to the interference measuring for signal-to-interference and noise ratio needed for joint power control and link scheduling. The scheduling module divides the MAC layer frames into sub-frames depending on delay constraints of authenticated requesting applications and assigns a group of links to each sub-frame. The module also allocates a group of transmit powers based on the delay constraints. The scheduling module sends the information to the sub-frame and probable group of transmit power allocation to each sub-frame. Finally, the CRM checks with the channel estimation for probable error rates and channel fading conditions based on the information received from the power scheduling module. The channel estimation is described in section -4 in details. Then, the CSA allocates the power constrained links to the authenticated application determined by the CRM and update security authentication at some periodic time interval for robustly secured communication.

3.1 Security Policy

The security scheme is including the authentication of CR before giving an exclusive access to spectrum or link and the authentication of users before communication and interval after communication. This addresses the mis-detections and false alarms issues. Security is deployed in three steps to make robustly secured in MAC layer and PHY layer listed as follows:

Step-1: Digital signatures are used in authenticating a CR,

Step-2: Integrated Trust computing and Public Cryptography technique is used in authenticating CRs in communication [6].

Step-3: Ensure security during the interval of communication between users.

In step-3, CR user estimates a maximum probable number of free sub-bands available for its communication using preliminary knowledge. The central idea of computing this probability is to compute the probable number of free sub-bands over which a cognitive radio switches during its communications. A cognitive user uses this information along with the information obtained from the dedicated database to comprehend the sequence of switching pattern for its communication. Once the two cognitive users have authenticated themselves to be genuine, the transmitting user conveys the switching pattern to its intended receiver. This procedure ensures that only the intended users are aware of the switching pattern of frequencies.

3.2 Channel Endorsement and Transmission Policy

For channel endorsement and transmission policy, when a CR node A wants to transmit data to CR node B, A updates its Channel State Table at first. If a free channel y becomes busy after update, A resets the timer of y which is represented in equation (1) .

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

$$\text{Timer}(y) = \text{pkt_size}(y)/\ \text{rate}(y) \quad (1)$$

This sensing makes sure that there is not unexpected present of PUs because PUs don't send any message on control channel where the messages are used for recording data channels' status. Generally, a free channel turns to be busy after sensing, when it is occupied by PUs. It is necessary to initiate the available timer as packet transmission duration.

When *B* received RTS from *A*, it updates the Channel State Table at first. Then *B* finds the overlapping channel set between *A*'s and *B*'s available lists and uses the evaluating function to score channels. For each channel *x*, the most appropriate channel is the one over which CR pairs can send most packets cumulatively as follows:

$$\text{Max}(\text{Min}x\{\text{pkt_ num}_{tran}(x), \text{pkt_ num}_{recv}(x)\}) \quad (2)$$

$$\text{pkt_num}(x) = [\ \beta \times \{\ \text{slot}(x) - \text{LC}\ /\text{pkt_size}(x)\}] \quad (3)$$

For the maximum slot number, Equation [2] calculates the maximum cumulatively packet number for distributed MAC. The maximum slot number of channel *x, slot(x),* is given as:

$$1 - (1 - \text{Ut}(x))^{\text{slot}(x)} < \text{threshold}(x) \quad (4)$$

$$\text{Ut} = \alpha \text{Ut}{-1} + (1 - \alpha)\hat{\text{U}} \quad (5)$$

where, Ut−1 is the exact utilization of last time slot $t - 1$ acquiring by physical layer.

$\hat{\text{U}}$ is the average experienced utilization of the past.

If there are more than one channel having the same pkt _num, it will select the one with maximum transmission rate. If there still are more than one channel have the same pkt _num with the same rate, randomization is used for decision. After making choice, *B* sends CTS to *A* with selected channel index and maximum packet number.

Under the condition that there is no available channel for *B* or maximal packet number is zero, *B* do nothing, and *A*'s timer for CTS will expire. After receiving CTS from *B*, *A* sends CRTS to *B* for confirmation; at the same time, all CR neighbors of *A* are also informed of the upcoming transmission over channel *x* and update their Channel State Tables for channel.

4. Complexity Reduction in Channel Estimation

The Decision Directed Channel Estimation (DDCE) scheme is preferred for CR-OFDM rather than purely pilot assisted channel estimation methods because it deploys both information symbols and the pilot symbols for channel estimation. The favorable case for DDCE is in the absence of transmission errors where the full pilot information can be used to detect subcarrier symbols as a posteriori reference signal which drastically reduce the number of pilot symbols required.

Least Square method has low complexity but the performance is degraded due to the energy-fluctuations associated with the near-Gaussian distributed subcarrier-related samples which renders the LS estimator is the worst for decision directed channel in CR-OFDM. On the other hand, MMSE in DDCE has good performance but high complexity over LS which can be mitigated by passing the computationally intensive matrix inversion operation. In addition, the Maximum Likelihood (ML) DDCE for CR-OFDM is optimized for low complexity by novel time-frequency interpretation.

4.1 Complexity Reduction in MMSE SS-CIR Estimator

The block diagram of the channel estimation method consists of a posteriori decision-directed CIR estimator, followed by a priori CIR predictor as shown in Fig. 2. The posteriori SS-CIR estimator inputs are the frequency-domain signal y[n] and the decision-based transmitted symbol estimate x[n]. The transformation from the frequency to time domain is performed within the CIR

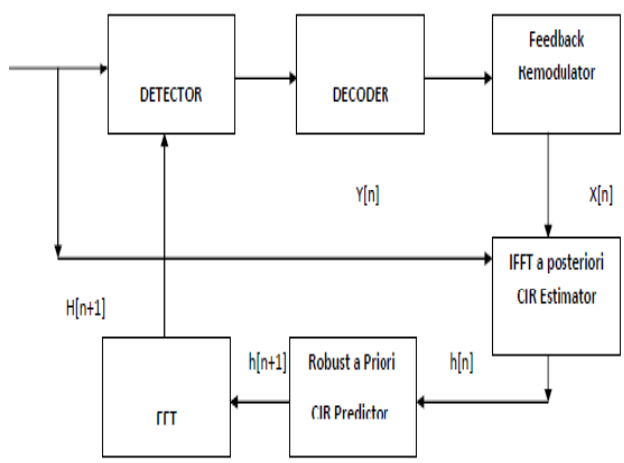


Fig. 2 Channel estimator constituted by an a posteriori decision-directed CIR with priori CIR

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

estimator and its output is a posteriori estimate x[n, k] of the SS-CIR taps , which is fed into the low-rank time-domain SS-CIR tap predictor to produce *a* priori estimate h[n+1, l], l = 0, 1, … ,$K_0 - 1$ of the next SS-CIR on a SS-CIR tap-by-tap basis. Finally, the predicted SS-CIR is converted to the Frequency Domain-CTF at the FFT block. The resultant FD-CTF is employed by the receiver for detecting and decoding the next OFDM symbol.The discrete baseband model of the OFDM system can be given by:

$$y[n, k] = H[n, k]x[n, k] + w[n, k] \quad (6)$$

for k = 0, . . . ,K − 1 and all n, and y[n, k], x[n, k] and w[n, k] are the received symbol, the transmitted symbol and the Gaussian noise sample respectively, corresponding to the k_{th} subcarrier of the n_{th} OFDM symbol. Moreover, H[n, k] is the complex channel transfer function (CTF) coefficient associated with k_{th} subcarrier and time instant n. For an M-QAM modulated OFDM system, x[n, k] corresponds to the M-QAM symbol accommodated by the kth subcarrier. In OFDM systems using a sufficiently long cyclic prefix and adequate synchronisation, the discrete CTF can be formed as follows. For k = 0, 1, …., K − 1, where h[n, l] = h(nT, lT/K) is the sample-spaced CIR having significant nonzero tap values only at sample-spaced raster positions and $W_K = \exp(-j2\pi/K)$.

$$H\left[n,k\right] = H\left(nT, k\Delta f\right) = \sum_{l=0}^{k_0-1} W_K^{kl}\, h\left[n,l\right], \quad (7)$$

By substituting $\hat{h} = \sigma_w^2 I + diag\left(\sigma_l^2\right) \overset{H}{W} diag$ from Frequency Domain CTF Equation (1) into (2);

$$y\left[n,k\right] = \sum_{l=0}^{K0-1} W_k^{kl}\; h\left[n,l\right] x\left[n,k\right] + w\left[n,k\right]. \quad (8)$$

The SS-CIR taps h[l] are assumed to be uncorrelated complex- Gaussian distributed variables having a zero mean and a covariance matrix $C_h = \text{diag}\,(\sigma^2)$. The MMSE estimator of the SS-CIR taps h[n, l] of the linear vector model described by equation (8) is given as follows:

$$\hat{h} = \left(\sigma_w^2 I + diag\left(\sigma_l^2\right) \overset{H}{W} diag\left(\left|\hat{x}[k]\right|^2\right) W\right)^{-1} \times diag\left(\sigma_l^2\right) \overset{H}{W} diag\left(\overset{(h*)}{x}[k]\right) y \quad (9)$$

The direct MMSE for estimation of the SS-CIR taps h[n, l] involves an inversion of a time-variant matrix which imposes a relatively high computational complexity as compared to the least square. In order to reduce the allied computational complexity, a two-step reduced complexity SS-CIR estimator invoking a method, which bypasses the computationally intensive matrix inversion operation. The channel Transfer function of MMSE can be given as follows:

$$H^{n}_{MMSE}\left[n,k\right] = \frac{\overset{\times}{x}\left[n,k\right] \times y\left[n,k\right]}{\left|x\left[n,k\right]\right|^2 + \frac{\sigma_w^2}{\sigma_H^2}} \quad (10)$$

MMSE CTF from equation (10) may be modelled as complex Gaussian-distributed variables having a mean identical to that of H[n, k], which represents the actual FD-CTF coefficients encountered and a variance of $\sigma_2^v = \sigma_2^{H_{MMSE}}$.

$$H^{n}_{MMSE}\left[n,k\right] = H\left[n,k\right] + v\left[n,k\right], \quad (11)$$

where, v[n, k] represents the i.i.d. complex-Gaussian noise samples having a zero mean and a variance, and substituting equation (7) into (11) yields as follows:

$$H^{n}_{MMSE}\left[n,k\right] = \sum_{l=0}^{K0-1} W_k^{kl}\; h\left[n,k\right] + v\left[n,k\right], \quad (12)$$

Furthermore, the MMSE estimator of the SS-CIR taps h[n, l] of the linear vector model described by equation (12) is given as follows:

$$h^{\wedge} - \overset{H}{W} \frac{H^{n}_{MMSE}}{\left[C_v\left(\frac{1}{C_h}\right) \quad \overset{H}{W}\left(\frac{1}{C_v}\right) W\right]} \quad (13)$$

where, we define C_h and C_v as the covariance matrices of the SS-CIR vector h and the scalar-MMSE FD-CTF estimator's noise vector v, respectively. The elements of the noise vector v are assumed to be complex-Gaussian i.i.d. samples and therefore we have $C_v = \sigma^2_v I$. On the other hand, as follows from the assumption of having uncorrelated SS-CIR taps, the SS-CIR taps' covariance matrix is a diagonal matrix $C_h = \text{diag}\,(\sigma^2)$ where $\sigma^2_l = E\{h[n, l]2\}$. Substituting C_h and C_v into Equation (8) yields;

$$\hat{h} = \left(diag\left(\frac{1}{\sigma_l^2}\right) + \left(\frac{1}{\sigma_v^2}\right) \overset{H}{W} W\right)^{-1} \overset{H}{W} \left(\frac{1}{\sigma_v^2}\right) H^{n}_{MMSE}$$

$$\hat{h} = diag\left(\frac{\sigma_l^2}{\sigma_v^2 + K\sigma_l^2}\right) \overset{H}{W} H^{n}_{MMSE} \quad (14)$$

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

Finally, upon substituting Equation (10) into Equation (12), it gives a scalar expression for the Reduced-Complexity (RC) a posteriori MMSE SS-CIR estimator in the form of;

$$\hat{h}[n,l] = \left(\frac{\sigma_l^2}{\sigma_v^2 + K\sigma_l^2}\right)\left(\sum_{k=0}^{k-1} W_k^H \frac{\hat{x}[n,k] \times y[n,k]}{|x[n,k]|^2 - \frac{\sigma_v^2}{\sigma_H^2}}\right) \quad (15)$$

The Complexity Reduced MMSE (CR-MMSE) estimator of equation (15) does not require any matrix inversion and therefore it has outstanding performance as the allied complexity is equivalent to the general LS estimator.

4.2 Complexity Reduction in Maximum Likelihood

The complexity of ML solution can be applied to a combination of pilot symbol assisted modulation (PSAM) and decision-directed channel estimation (DDCE) OFDM system as Complexity Reduced Maximum Likelihood (CR-ML) The pilot symbols along with decisions taken on the other carriers, then r = h + v remains valid, with a given C_{vv}, which leads to H =P_{Fh} y. The combination of PSAM and decision-feedback is desirable as the increment in complexity can be dealt. The complexity of the ML estimator can be significantly minimized both for spectral shaping and PSAM systems. This low complexity depends on the time-frequency interpretation and pruning of the FFT/IFFTs.

Regarding spectral shaping, manipulating P_{Fuh} as a low rank matrix of rank $N_{\hat{h}}$ considering its hermiticity, $P_{Fuh} = VV^H$ where V is a matrix of size N_c x $N_{\hat{h}}$ and can be computed easily. The complexity of computing the estimator is 2 N_u x $N_{\hat{h}}$ complex multiplications for the general ML estimator. This complexity can be reduced by using the time-frequency interpretation. The projection operation can be expressed by the cascade of two partial FFTs, weighted by a $N_{\hat{h}}$ x $N_{\hat{h}}$ matrix assuming that all carriers are used as pilots, called an identity matrix.

When using Pilot Symbol Assisted Modulation, a comb spectrum has to be measured, and only the teeth of the comb are used for the FFTs. The DFT can be computed with $N_c/4 + N_{\hat{h}}/2 + \log 2\, N_{\hat{h}} - N_{\hat{h}}$ complex multiplications, which refers a large gain for a large number of carriers [8]. The complexity for FFT-based solutions is much lower than for the singular value decomposition (SVD) based approach, both for spectral shaping and PSAM systems. Futhermore, the ML algorithm can efficiently work with a significantly smaller $N_{\hat{h}}$ than the LMMSE and LS, which yields to comparatively larger gain.

The major inspiration to the complexity is due to the weighting matrix for a relatively large number of pilot carriers. Rigorously, pure PSAM with regularly spaced pilot carriers can easily be depicted that the weighting matrix $\left(F_{uk}^{H^{-1}} F_{uk}\right)$ is proportional to the identity matrix, and complexity is even lower. This is the special case of frequency correlation interpretation.

5. Simulation and Performance Evaluation

The Simulation is conducted setting the total licensed frequency 600 MHz and 100 allocated channels at the rate of 6MHz. The maximum licensed primary users are 100 and spectrum hole is generated randomly for secondary users in the absence of incumbents. Fig. 3 represents WRAN spectrum considering video signal spectrum for instant 58 users including both incumbents as well as secured CR users in ad hoc manner without distributed MAC. Fig. 4 illustrates 58 active primary users SNR, whereas Fig. 5 refers the dropped or unsecured 22 secondary users and Fig. 6 refers authenticated 20 secondary user SNR at corresponding base stations. The power controlled link without using Transmission policy and security policy is represented in Fig. 7 and found adverse power assigned in links non-uniformly which degrades the network performance.

Furthermore, the Fig. 8 Fig. 9 and Fig. 10 show that the transmitted packets, received packets and their corresponding packet ratio at different channel sub-bands deploying Channel Endorsement and Transmission Policy. The channels are selected based on the joint information of the utilization function and slots to determine the cumulative packet distribution in distributed MAC and compared to threshold. The best channels are having maximum good put or higher packet transmission, in which even the minimal data transmitted without loss. The Power Controlled Link is assigned to best channel application by CSA after channel estimation in Fig.11 and their corresponding optimized throughput up to 75% in Fig. 12. This concurs that the CETP and SP with cross-layer based Power Controlled Links (PCL) provide the optimal number of maximum transmission oriented channels and provide higher throughput as compared to without PCL.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

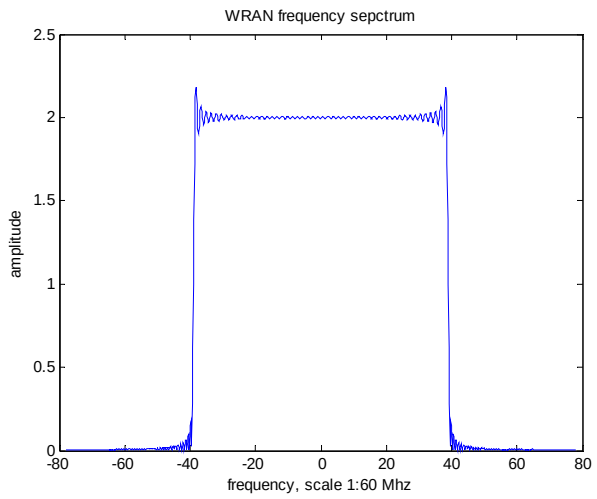


Fig. 3 Ad hoc WRAN Spectrum 600 MHz (scale 1:60)

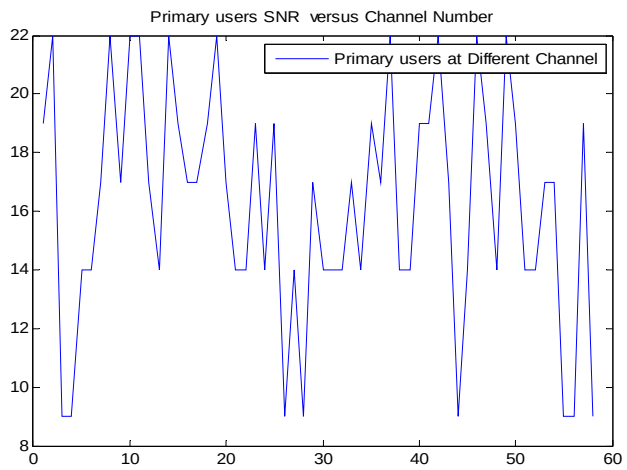


Fig. 4 SNR in Primary user channels

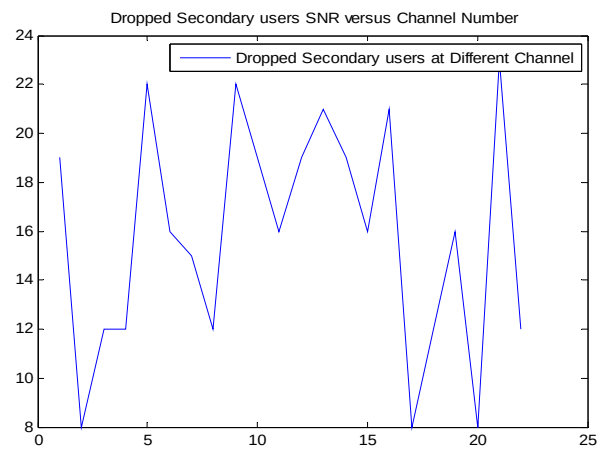


Fig. 5 SNR in Dropped (unsecured) Secondary user channels

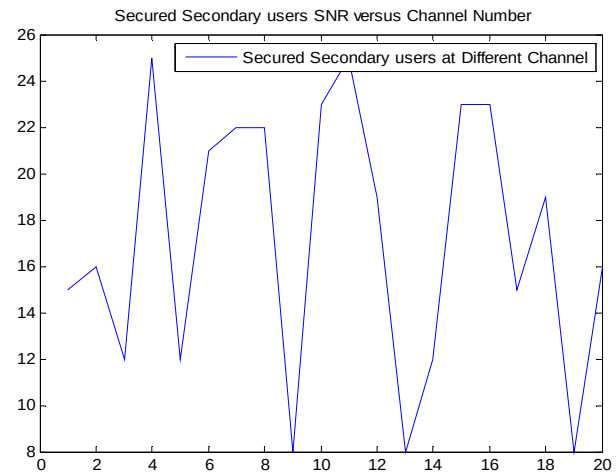


Fig. 6 SNR in authenticated Secondary user channels

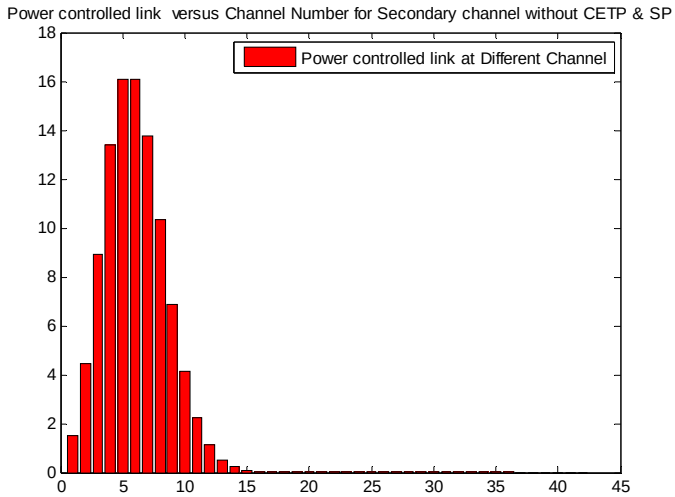


Fig. 7 Power controlled links without Transmission and Security policy

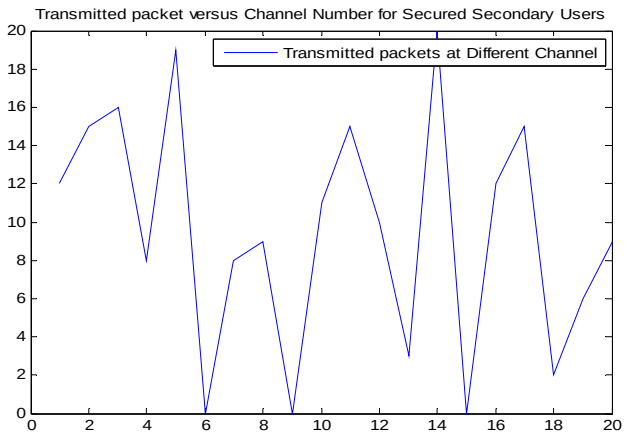


Fig. 8 Transmitted packets in Secondary channels

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

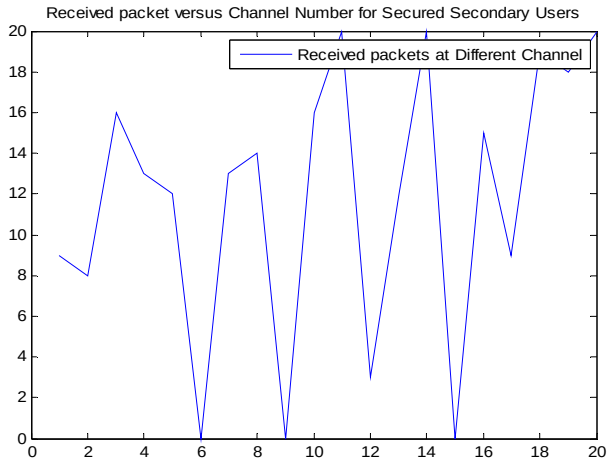


Fig. 9 Received packets in Secondary channels

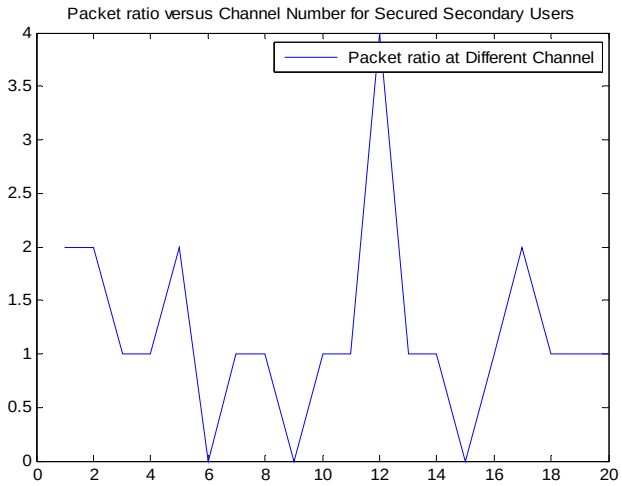


Fig. 10 Packets Ratio of Transmitted and Received packets in Secondary channels using CET

Regarding complexity reduction in channel estimation, CR based on OFDM with DDCE specifications including the various parameters are as given below in Table-1. From simulation, SER is optimized at $10^{-1.5}$ for CR-MMSE, $10^{-1.45}$ for MMSE and $10^{-1.35}$ for LS as shown in Fig. 13. In other words, SER is minimized in CR-MMSE and outperforms over MMSE and LS. In conclusion, MMSE has comparatively higher complexity over Least Square but the performance is better. The LS estimator is the worst option for decision directed channel in OFDM. Complexity reduced MMSE in DDCE has better performance than MMSE and low complexity as LS which is achieved by passing the computationally intensive matrix inversion operation.

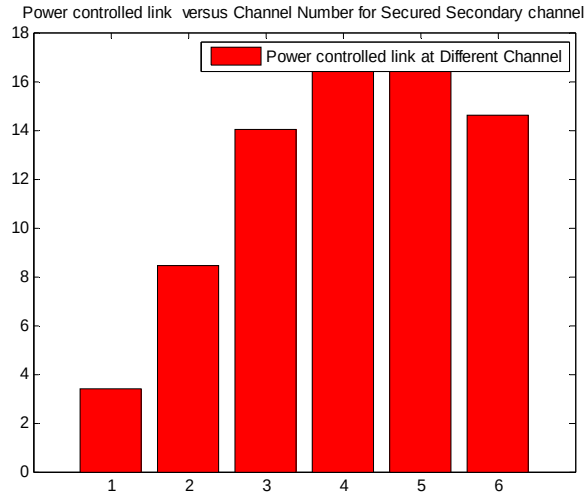


Fig. 11 Power Controlled Secondary Channels deploying UTEP and SP

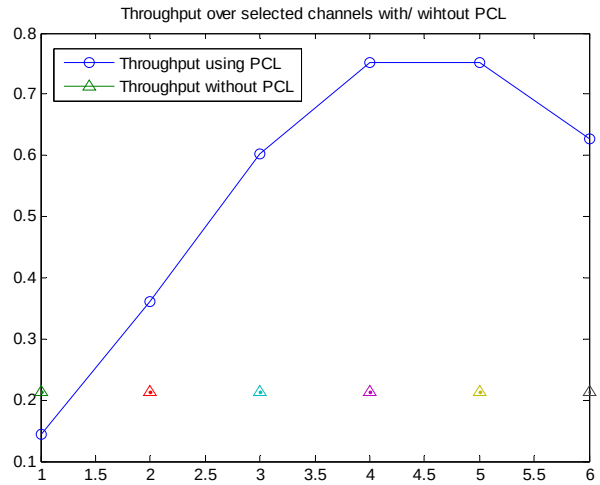


Fig. 12 Throughput over selected channels with/without power Constrained links

The behavior of the ML estimator at low SNR gives some insight on the influence of Nˆh. As moving from higher to lower Nˆh tends to better performance. The values of the channel impulse response beyond the lower Nˆh are below the noise level, so that their estimation introduces more noise than relevant information about the channel. As for the spectral shaping system, the flooring effect of the LMMSE estimator is essentially limiting its effectiveness at high SNR's. Simulation results demonstrate that the LMMSE suffers from a threshold effect at high SNR and gives SER converged to $10^{-1.4}$ whereas improving performance $10^{-1.47}$ and $10^{-1.49}$ in Complexity Reduced ML (CR-ML) and ML as shown in Fig. 14. Nˆh must be 2 to 4

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

times larger for LMMSE to trade off SER and SNR than for ML.

Table- 1 Simulation Parameters for Channel Estimation

Simulation Parameters	Value
Modulation	BPSK
Number of used subcarriers (nDSC)	64
Channel Model	AWGN
Channel B/W	1000 kHz
CIR taps	4
Max Delay Spread	30 µs
Total Symbol duration (Ts)	120 µs
Twiddle Matrix size	64 x 64

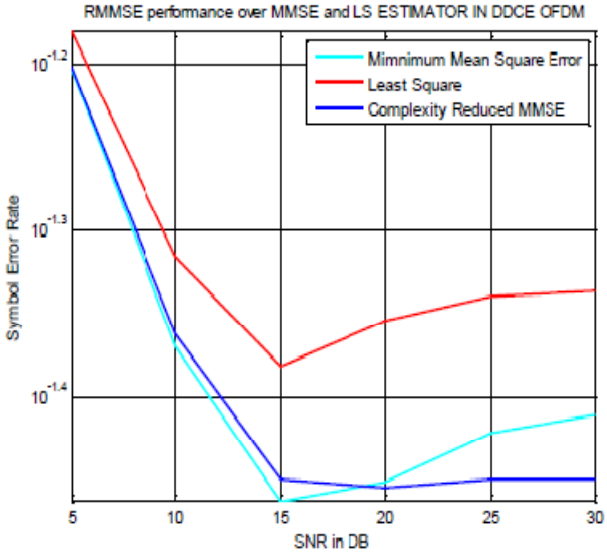


Fig. 13 SNR versus SER plot for MMSE, LS, CR-MMSE

6. Conclusion

Secured Distributed MAC and complexity reduction in channel estimation for Cross-layer based Cognitive radios provide outstanding performance in terms of robustness, SER, joint power control and link scheduling with optimum opportunistic MAC. This may slightly increase the processing overheads and resource consumption as compared to without cross layer approach but it has significant achievement in terms of throughput, quality of service, interference mitigation and better information rate. On the other hand, CR-MMSE has features of complexity reduction by allowing to pass through the computationally intensive matrix inversion operation which demonstrates outstanding SER performance over MMSE and LS in sample spaced channel impulse response (SS-CIR) of DDCE OFDM. Furthermore, CR-ML featured by partial IFFT/FFT with Pilot Symbol Assisted Modulation (PSAM) provides the significant SER performance over MMSE and ML in Cross-layer based Cognitive radio Networks.

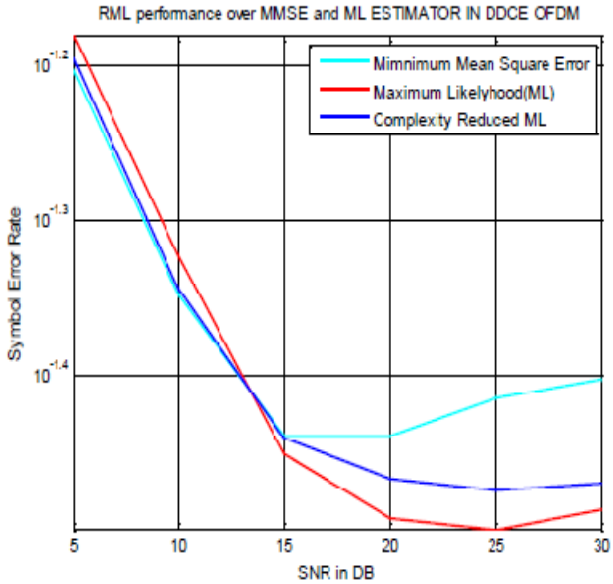


Fig. 14 SNR versus SER plot for MMSE, ML, CR-ML

References

[1] J. Mitola, "Cognitive Radio: an integrated agent architecure for software defined radio", PhD Dissertation, Royal Institute of Technology (KTH), Stockholm, 2000.

[2] T. R. Newman, B. A. Barker, A. M. Wyglinski, A. Agah, J. B. Evans, and G. J. Minden, "Cognitive engine implementation for wireless multicarrier transceivers," Wireless Commun and Mobile Computing, vol. 6, 2006.

[3] V. Srivastava and M. Motani, "Cross-layer design: a survey and the road ahead," IEEE Comm. Mag, vol. 43, no. 12, Dec. 2005, pp: 112-119.

[4] Yi Peng, Zengli Liu, Hua Long, Jie Peng, "The Research of Cross-layer design between the PHY and MAC layer for OFDM Based Cognitive Radio Network", 2009 International Conference on Networks Security, Wireless Communications and Trusted Computing.

[5] D. R. Vaman, N. Shakhakarmi, "Integrated Key based Strict Friendliness Verification of Neighbors in MANET," International Conference on Security Science and Technology, ICSST2011, January 2011.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

[6] N. Shakhakarmi, D. R. Vaman, "Distributed Position Localization and Tracking (DPLT) of Malicious Nodes in Cluster Based Mobile Ad hoc Networks (MANET)", WSEAS Transactions in Communications, ISSN: 1109-2742, Issue 11, Volume 9, November 2010.

[7] P. Vandenameele, L. Van der Perre, M. Engels, Gyselinckx, and H. De Man, "A combined ofdm/sdma approach," IEEE Journal on Special Areas in Communications, 2000.

[8] Y. Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems," IEEE Transactions on Vehicular Technology, vol. 49, pp.1207–1215, July 2000.

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 Cognitive and Cooperative Radio Networks, Wavelets applications, Digital Image Processing, Secured Position Location & Tracking (PL&T), Mobile Ad hoc Networks, 4G Networks, Satellite Networks, QoS & Network Security, and Color Technology. He has published WSEAS journal, IJCSI journal, ICSST conference, Elsevier conference, and several other papers are under review in IEEE and WSEAS journals and conference papers.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.

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)

© IJCSI PUBLICATION 2012

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

© IJCSI PUBLICATION
www.IJCSI.org