

International Journal of Computer Science Issues

**Volume 8, Issue 6, No 1, November 2011
ISSN (Online): 1694-0814**

IJCSI proceedings are currently indexed by:



Cogprints Google scholar



IJCSI Publicity Board 2011

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 Papataxiaris

Department of Informatics and Telecommunications
National and Kapodistrian University of Athens, Athens, Greece

IJCSI Editorial Board 2011

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
Computer Engineering Department
Azad University of Najafabad
Iran

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

N. Jaisankar
Assistant Professor
School of Computing Sciences,
VIT University
Vellore, Tamilnadu, India

EDITORIAL

In this sixth and last edition of 2011, 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 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 8, Issue 6, No 1, November 2011 (IJCSI Vol. 8, Issue 6, No 1). The acceptance rate for this issue is 32.8%.

IJCSI Editorial Board
November 2011 Issue
ISSN (Online): 1694-0814
© IJCSI Publications
www.IJCSI.org

IJCSI Reviewers Committee 2011

- Mr. Markus Schatten, University of Zagreb, Faculty of Organization and Informatics, Croatia
- Mr. Vassilis Papataxiaris, 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 University - 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 Mougiaakou, 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 Egídio 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, Concast, 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. Supheakmungkol SARIN, Waseda University, Japan
- 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 Technology, 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
- Mr. Deo Prakash, SMVD University, Kakryal(J&K), 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. Nadeem Mahmood, Department of Computer Science, University of Karachi, Pakistan
- Mr. Hafeez Ullah Amin, Kohat University of Science & Technology, Pakistan
- Dr. Professor Vikram Singh, Ch. Devi Lal University, Sirsa (Haryana), India
- Mr. M. Azath, Calicut/Mets School of Enginerring, India
- Dr. J. Hanumanthappa, DoS in CS, University of Mysore, 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 Tunnis, 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
- Dr. Mokhled S. AlTarawneh, Computer Engineering Dept., Faculty of Engineering, Mutah University, Jordan, Jordan
- 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
- Dr. Eva Volna, University of Ostrava, Czech Republic
- 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, Ethiopia
- 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 Supérieur 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 Kiourmourtzis, Researcher, University of Patras, Greece
- Mr. Amol D.Potgartwar, 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 - Nahrain University, Iraq
- Dr. Mohd Nazri Ismail, University of Kuala Lumpur, Malaysia
- 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 Phougot, 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
- Dr. Lokesh Kumar Sharma, Chhattisgarh Swami Vivekanand Technical University, Bhilai, 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
- Dr. M. Sreedhar Reddy, JNTU, Hyderabad, SSIETW, India
- Mr. Ojesanmi Olusegun, Ajayi Crowther University, Oyo, Nigeria
- Ms. Mamta Juneja, RBIEBT, PTU, India
- Prof. Chandra Mohan, John Bosco Engineering College, India
- Mr. Nitin A. Naik, Yeshwant Mahavidyalaya, Nanded, India
- Mr. Sunil Kashibarao Nayak, Bahirji Smarak Mahavidyalaya, Basmathnagar Dist-Hingoli., India
- Prof. Rakesh.L, Vijetha Institute of Technology, Bangalore, India
- Mr B. M. Patil, Indian Institute of Technology, Roorkee, Uttarakhand, India

- Mr. Thipendra Pal Singh, Sharda University, K.P. III, Greater Noida, Uttar Pradesh, India
- Prof. Chandra Mohan, John Bosco Engg College, India
- Mr. Hadi Saboohi, University of Malaya - Faculty of Computer Science and Information Technology, Malaysia
- 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. Iqbaldeep Kaur, PTU / RBIEBT, India
- Mrs. Vasudha Bahl, Maharaja Agrasen Institute of Technology, Delhi, India
- Prof. Vinay Uttamrao Kale, P.R.M. Institute of Technology & Research, Badnera, Amravati, Maharashtra, India
- Mr. Suhas J Manangi, Microsoft, 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
- 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
- Mr. Akhter Raza Syed, University of Karachi, Pakistan
- Mrs. Manjula K A, Kannur University, 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
- Dr. V. Nagarajan, SMVEC, Pondicherry university, 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
- Prof. Jyoti Prakash Singh, Academy of Technology, India
- Mr. Peyman Taher, Oklahoma State University, USA
- Dr. S Srinivasan, PDM College of Engineering, India
- Mr. Muhammad Zakarya, CIIT, Pakistan
- Mr. Williamjeet Singh, Chitkara Institute of Engineering and Technology, India
- Mr. G.Jeyakumar, Amrita School of Engineering, India
- Mr. Harmunish Taneja, Maharishi Markandeshwar University, Mullana, Ambala, Haryana, India
- Dr. Sin-Ban Ho, Faculty of IT, Multimedia University, Malaysia
- Mrs. Doreen Hepzibah Miriam, Anna University, Chennai, India
- Mrs. Mitu Dhull, GNKITMS Yamuna Nagar Haryana, India
- Dr. D.I. George Amalarethinam, Jamal Mohamed College, Bharathidasan University, India

- Mr. Neetesh Gupta, Technocrats Inst. of Technology, Bhopal, India
- Ms. A. Lavanya, Manipal University, Karnataka, India
- Ms. D. Pravallika, Manipal University, Karnataka, India
- Prof. Vuda Sreenivasarao, St. Mary's college of Engg & Tech, India
- Prof. Ashutosh Kumar Dubey, Assistant Professor, India
- Mr. Ranjit Singh, Apeejay Institute of Management, Jalandhar, India
- Mr. Prasad S.Halgaonkar, MIT, Pune University, India
- Mr. Anand Sharma, MITS, Lakshmangarh, Sikar (Rajasthan), India
- Mr. Amit Kumar, Jaypee University of Engineering and Technology, India
- Prof. Vasavi Bande, Computer Science and Engineering, Hyderabad Institute of Technology and Management, 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
- Dr. Arti Arya, PES School of Engineering, Bangalore (under VTU, Belgaum, Karnataka), India
- Mr. Pawan Jindal, J.U.E.T. Guna, M.P., India
- Prof. Santhosh.P.Mathew, Saintgits College of Engineering, Kottayam, India
- Dr. P. K. Suri, Department of Computer Science & Applications, Kurukshetra University, Kurukshetra, India
- Dr. Syed Akhter Hossain, Daffodil International University, Bangladesh
- Mr. Nasim Qaisar, Federal Urdu Univertrsity of Arts , Science and Technology, Pakistan
- 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
- Prof. T Venkat Narayana Rao, Department of CSE, Hyderabad Institute of Technology and Management , 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
- Prof. Manpreet Singh, M. M. Engg. College, M. M. University, Haryana, India
- Mr. Syed Imran, University College Cork, Ireland
- Dr. Namfon Assawamekin, University of the Thai Chamber of Commerce, Thailand
- 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
- 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 Institute of Engineering & Technology, India
- Dr. Anuraag Awasthi, JV Womens University, Jaipur, India
- Dr. Mathura Prasad Thapliyal, Department of Computer Science, HNB Garhwal University (Central 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. Yousef Naeemi, Mehr Alborz University, Iran
- 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 Engineering, (Affiliated 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 Pavanasm, Adhiparasakthi Engineering College, Melmaruvathur, India
- Dr. Panagiotis Michailidis, University of Western Macedonia, Greece
- Mr. Amir Seyed Danesh, University of Malaya, Malaysia
- 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
- 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
- Mr. Rupesh Nasre., Indian Institute of Science, Bangalore., India.
- Mrs. Dimpi Srivastava, Dept of Computer science, Information Technology and Computer Application, MIET, Meerut, India
- 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

TABLE OF CONTENTS

1. Using Cryptography in Trust Computing for Networked Communications Dang-Quan Nguyen and Louise Lamont	1-7
2. Modelling and simulation of complex systems: an approach based on multi-level agents Alain-Jerome Fougeres	8-17
3. SMARTNotes: Semantic annotation system for a collaborative learning Driss Bouzidi, Rachid Elouahabi and Noreddine Abghour	18-25
4. An Ensemble Method for Validation of Cluster Analysis Sunghae Jun	26-30
5. An Efficient Method for Mining Event-Related Potential Patterns Seyed Aliakbar Mousavi, Muhammad Rafie Hj Arshad, Hasimah Hj Mohamed and Saleh Ali Alomari	31-38
6. LR-WPAN Formation Topologies using IEEE 802.15.4 Salman Naseer and S R Chaudhary	39-45
7. Dewy Index Based Arabic Document Classification with Synonyms Merge Feature Reduction Amal Alajmi, Elsayed M Saad and Medhat H Awadalla	46-54
8. A Novel Architecture for Quantum-Dot Cellular Automata Multiplexer Arman Roohi, Hossein Khademolhosseini, Samira Sayedsalehi and Keivan Navi	55-60
9. A Collaborative Algorithm for Ontological Matching in E-Learning Courseware Domain Knowledge Network Akanbi C. Olufisoye, Adagunodo E. Rotimi, Omotosho Lawrence O and Adigun Adepeju A	61-67
10. Blind Recognition Algorithm of Turbo Codes for Communication Intelligence Systems Ali Naseri, Omid Azmoon and Samad Fazeli	68-72
11. BPISG: A Batching Heuristic Scheduling Algorithm With Taking Index Parameters for Mapping Independent Tasks on Heterogenous Computing Environment Arash Ghorbannia Delavar, Ali Reza Kalili Boroujeni and Javad Bayrampoor	73-83
12. Design Approaches to Enhance Usability for E-Commerce Sites Mohiuddin Ahmed, Jonayed Kaysar and Nafis Rahman	84-88
13. DTN Routing Protocols for VANETs: Issues and Approaches Ramin Karimi, Norafida Ithnin, Shukor Abd Razak and Sara Najafzadeh	89-93
14. Theoretical analysis of VoIP transmission through different Wireless access technologies Emna Charfi, Lamia Chaari and Lotfi Kamoun	94-103
15. Comparing Methods for segmentation of Microcalcification Clusters in Digitized Mammograms Hajar Moradmand, Saeed Setayeshi and Hossein Khazaei Targhi	104-108
16. Grayscale Image Compression Based on Min Max Block Truncating Coding Hilal Almara'beh, Khalil Barhoum and Majdi Ahed	109-112
17. The Conception and the Study of a Triple-band Planar Inverted-F Antenna Ibnyaich Saida, Ghammaz Abdelilah and Hassani Moha M'rabet	113-117

18. HASoC for Developing a Software System Salah Eldin Abdelrahman and Mohammed Badawy	118-126
19. Nurture IDR Segmentation and Multiple Instruction Queues in Superscalar Pipelining Processor J.Nandini meeraa, N.Indhuja, S.Devi Abirami and K.Rathinakumar	127-133
20. Design of a High Performance Reversible Multiplier Md.Belayet Ali, Hosna Ara Rahman and Md. Mizanur Rahman	134-141
21. A Block Cipher using Rotation and Logical XOR Operations D. Sravan Kumar, CH. Suneetha and A. Chandra Sekhar	142-147
22. Framework for Ethernet Network Functionality Testing Mirza Aamir Mehmood, Ahthasham Sajid and Amir Shehzad	149-155
23. Incorporating Security in Embedded System - A critical analysis Mirza Aamir Mehmood, Amir Shahzad and Mazhar Ali	156-160
24. Implementing VoIP over Fatima Jinnah Women University Ammara Asif, Ammara Asif, and Tahira Mahboob	161-165
25. Analysis of Image Denoising using Wavelet Coefficient and Adaptive Subband Thresholding Technique S.Kalavathy and R.M.Suresh	166-172
26. Strategical Modelling with Virtual Competition for Analyzing Behavior of Malicious Node in Mobile Adhoc Network to Prevent Decamping Anil G.N and A. Venugopal Reddy	173-180
27. Authentication of DICOM Medical Images using Multiple fragile watermarking Techniques in Wavelet Transform Domain A Kannammal, A. Kannammal, S. Subha Rani and S. Subha Rani	181-189
28. E-Business: Application of software and technology in selected Ethiopian banks - Issues and challenges Bhaskar Reddy Muvva Vijay and Tewdros Sisay Asefa	190-198
29. Design and Implementation of Memory-less Forbidden Transition Free Crosstalk Avoidance CODECs for On-Chip Buses J.Venkateswara Rao and P.Sudhakara Rao	199-203
30. Extending XACML to support Credential Based Hybrid Access Control Nirmal Dagdee and Ruchi Vijaywargiya	204-211
31. New Channel Assignment Method for Access Points in Wireless LANs Mohammed Fawzi Al-Hunaity	213-218
32. User Navigation Pattern Discovery using Fast Adaptive Neuro-Fuzzy Inference System J Vellingiri and S. Chenthur Pandian	219-224
33. Approach to a conceptual model of indexation in a weak ontology described in RDFS Traore Issa, Souleymane Oumtanaga, Babri Michel and Claude Lischou	225-234
34. A Novel Approach to Fast Image Filtering Algorithm of Infrared Images based on Intro Sort Algorithm Kapil Kumar Gupta, Kapil Kumar Gupta, M. Rizwan Beg, M. Rizwan Beg, Jitendra Kumar Niranjan and Jitendra Kumar Niranjan	235-241

35. A New Factorization Method to Factorize RSA Public Key Encryption Bhagvant Ram Ambedkar and Sarabjeet Singh Bedi	242-247
36. A Classical Fuzzy Approach for Software Effort Estimation on Machine Learning Technique S.Malathi and S. Sridhar	249-253
37. Intelligent Paging Strategy for Multi-Carrier CDMA System Sheikh Shanawaz Mostafa, Khondker Jahid Reza, Md. Ziaul Amin and Mohiuddin Ahmad	254-260
38. An Efficient and Secure Protocol for Ensuring Data Storage Security in Cloud Computing Syam Kumar P and Subramanian R	261-274
39. Performance Evaluation of Fingerprint Identification Based on DCT and DWT using Multiple Matching Techniques Lavanya B. N. and Raja K.B.	275-283
40. FLoMSqueezer: An Effective Approach For Clustering Categorical Data Stream Marpe Sora, Swarup Roy and S I Singh	284-291
41. iAgile: A Tool for Database Generation Guided by Graphical User Interface Shaimaa Galal and Ehab Hassanein	292-298
42. Enhancing E-Learning with VRML Techniques Sangeetha Senthilkumar and E. Kirubakaran	299-303
43. Knowledge Acquisition and Knowledge Management through E-Learning V.Thennarasu and E. Kirubakaran	304-308
44. A Survey of Web Crawler Algorithms Pavalam S M, S V Kashmir Raja, Felix K Akorli and Jawahar M	309-313
45. Construction of Learning Path Using Ant Colony Optimization from a Frequent Pattern Graph Souvik Sengupta, Sandipan Sahu and Ranjan Dasgupta	314-321
46. Developing an Intelligent User Interaction Development Engine Ashit Kumar Dutta	322-329
47. An Integrative Study on Bioinformatics Computing Concepts, Issues and Problems Muhammad Zakarya, Izaz Ur Rahman, Nadia Dilawar and Reshma Sadaf	330-339
48. Automatic Fracture Detection Using Classifiers- A Review S.K.Mahendran and S. Santhosh Baboo	340-345
49. Automatic Histogram Threshold with Fuzzy Measures using C-means K. Srinivas and V. Srikanth	347-351
50. Modified LRU Algorithm To Implement Proxy Server With Caching Policies Jitendra Singh Kushwah and Brijesh Patel	352-358
51. Classification of Speech for Clinical Data using Artificial Neural Network C.R.Bharathi and V. Shanthi	359-365
52. Efficient Retrieval of Text for Biomedical Domain using Expectation Maximization Algorithm Sumit Vashishtha and Yogendra Kumar Jain	366-370
53. Symmetrical Dispersion Compensation For High Speed Optical Links Ojuswini Arora, Amit Kumar Garg and Savita Punia	371-376

54. Improvement the Flatness, Gain and Bandwidth of Cascaded Raman Amplifiers for Long-Haul UW-WDM Optical Communications Systems	377-384
Fathy M. Mustafa, Ashraf A. Khalaf and F. A. El-Geldawy	
55. Fuzzy Priority CPU Scheduling Algorithm	386-390
Bashir Alam, M.N. Doja, R. Biswas and Mansaf Alam	
56. Dynamic Gesture Recognition Using Hidden Markov Model in Static Background	391-398
Malvika Bansal, Shivin Saxena, Devendra Desale and Dnyaneshwar Jadhav	
57. Image Compression Using Partitioning Around Medoids Clustering Algorithm	399-401
Valmik B Nikam, Vinod J Kadam and B.B.Meshram	
58. Lanes and Road Signs Recognition for Driver Assistance System	402-408
Ahmed Hechri and Abdellatif Mtibaa	

Using Cryptography in Trust Computing for Networked Communications

Dang-Quan Nguyen, Louise Lamont

Communications Research Centre
Ottawa, Ontario, K2H 8S2, Canada

Abstract

We highlight some major difficulties encountered by current approaches that try to model trust computing in a realistic networked communications system. We characterize these approaches as *top-down* since they assume that trust is universal and readily quantifiable. Our main concern with these approaches is that their quest to define a universal trust often ends up with a loose, context-dependent definition of trust value. Based on this shortcoming of the *top-down* approaches, we propose another consideration of trust for networked communications. Namely, we underline the fact that trust is first and foremost a security constraint that exists in each specific network's security operation (*i.e.*, attack and defense). Hence, its definition depends on the attack and defense being involved. We call this approach *bottom-up* and discuss the close relationship between trust and cryptography through some examples of network's attack and defense.

Keywords: Trust computing, Cryptography, Security, Networked communications.

1. Introduction

We present our position regarding a fundamental question about *trust*, considered as a component of security in networked communications: ``*how to efficiently and creatively capture the concept of trust computing in networked communications?*''

Throughout this paper, our objective is two-fold. First, we highlight some major difficulties encountered by the current approaches that try to model trust computing in a realistic networked communications system. We characterize these approaches as *top-down* since they assume that trust is universal and readily quantifiable. Hence they often rely on applications such as Intrusion Detection Systems as a determinant to provide a trust value for various devices in the network. Also, they tend to tailor the underlying networking mechanisms (such as admission control, routing protocols, etc.) to accommodate the constraints induced by trust. In this case, trust is often considered as a network parameter, along with the end-to-

end throughput and delay, and is accounted for as such. Our main concern with these approaches is that their quest to define a universal trust often ends up with a loose, context-dependent definition of trust value. Moreover, from a practical point of view, it is quite unclear what kinds of network attack these approaches aim to resolve.

Based on this shortcoming of the *top-down* approaches, we propose another consideration of trust for networked communications. Namely, we reverse the above process and underline the fact that trust is first and foremost a security constraint that exists in each specific network's security operation (*i.e.*, attack and defense). Hence, trust does not necessarily have a universal definition, but its definition depends on the attack and defense being involved. Therefore, the implementation of trust is automatically embedded within each basic component of the security operations. Trust in this case adds a new dimension of consideration to the existing security operations and also offers guidelines to the design of a new paradigm. We call this approach *bottom-up*.

As an example, consider a large MANET whose topology is not guaranteed to be continuously connected (such as in sparse networks). The nodes are soldiers on the ground. When soldiers from a coalition force meet and they have no connectivity to the authentication server, and if they need to share some valuable information then soldiers from both sides will have to decide whether they can trust each other based on a shared passphrase. However, revealing a secret passphrase to an unknown soldier is dangerous. Therefore, the soldiers may need an exchange protocol that allows for the mutual verification that both sides actually know the secret but without revealing anything on the secret itself, not even a hash of the passphrase since an attacker can reuse the hash values to fool other soldiers. *Top-down* approaches cannot solve this problem other than waiting for the connection with the authentication server to come up. Reputation and recommendation cannot be used since both soldiers are unknown to each other prior to the meeting. We will see in the detailed discussion of Section

3.3 that this problem can be solved at the low level networking mechanisms using a cryptography protocol called *zero knowledge proof*.

To narrow down the scope of security operations that we can afford to work on, while maintaining a subject rich in illustrations and promising in exciting research topics, we limit our security operations to cryptography and cryptanalysis mechanisms that are used to secure and to break the networked communications. We also discuss how the *bottom-up* approach releases the dependence of trust on the network components and blends itself well to distributed systems such as mobile ad hoc networks. We also provide concrete examples and highlight the link with trust in each example.

The rest of the paper is organized as follows. In Section 2, we formulate the issue of trust computing and its importance in secured communications. Section 3 presents the existing proposals of trust computing schemes that we call *top-down* approaches. We argue, by presenting some counter-examples, that *top-down* approaches may not be pertinent. We then present our position on trust, based on the observation that its importance must be granted to solve low-level problems (via some practical attacks on networked communications). Section 3 also shows the close connection between cryptography and trust computing. We conclude this paper in Section 4.

2. Issue and its Importance

Future land operations will be undertaken by geographically dispersed teams in order to gain a better understanding of the battle-space through information gathered by the dispersed teams. In certain situations, a dispersed force must be capable of rapid aggregation in order to conduct operations as a larger aggregated force. The constantly changing nature of the battle-space necessitates adaptive forces equally capable of operating in a dispersed or aggregated posture. Additionally in order to gain greater mission effectiveness, information sharing will extend beyond the land force to allies and joint forces. More than ever we need to develop a framework that can be used to secure network communications against attack.

The trust component is an important concept in network security, as it is the set of relations among agents participating in the network activities. Trust becomes even more challenging in wireless multi-hop and distributed networks where soldiers join and leave the network to support dispersion and rapid aggregation. Trust management is often interpreted as a multifunctional control mechanism, in which the most important aspect is

to establish trust from a small set of agents who are known to be trustworthy. Another important aspect in trust management is trust revocation, which reverses previous trust opinions of agents based on newly obtained evidence with regard to those agents. Trust propagation can also be seen as another component of trust management where a source node must rely on other nodes to forward its packets on multi-hop routes to the destination.

We argue that the fundamentals techniques of cryptography and cryptanalysis need to be refined to ensure that nodes can exchange information once they are trustworthy without starting as a premise that a set of nodes are trustworthy from the first encounter. Furthermore, security in computer networks usually relies on central authorities, certificates directories, or some preinstalled keys and procedures. However, the past decade has witnessed the emergence of Mobile Ad hoc Networks (MANETs) which are characterized as self-organizing systems. These centralized services may not always be accessible in self-organized systems. To address this problem, we propose to develop a trust management scheme where key distribution, key revocation and enforcement are tackled by using light-weight cryptography techniques and key distribution protocols that use broadcasting with selective encryption to guarantee that only a subgroup of members can decrypt the message.

3. Trust Computing and Cryptography

In this section, we first present some existing proposals of trust computing schemes. We characterize these approaches as *top-down* since they consider trust as universal and readily quantifiable by an existing system such as intrusion detection system (IDS). Counter-examples are presented to show that *top-down* approaches may not be pertinent to solve practical security problems involving trust. Based on this observation, we then introduce our *bottom-up* approaches consisting of considering trust as a security constraint proper to each networked communication attack or defense mechanism. Thus, we show that cryptography can be employed to solve practical problems of trust computing. At a lower level, we also discuss how this combination of cryptography and trust computing can provide inputs to other security mechanisms such as admission control, key management and secure routing.

3.1 Existing proposals on the issue

Currently, the existing approaches to trust computing in networked communications consider *trust* as a quantifiable

network metric. We classify existing trust computing proposals into two categories: low-level trust computing focussed on the evaluation of trust metrics and how to propagate such metrics in the network; and high-level trust computing focussed on the creation of frameworks and policies for trust computing.

3.1.1 Low-level trust computing

Low-level trust computing proposals include schemes designed for trust evaluation and trust propagation.

For trust evaluation, the proposals rely on an existing monitoring entity having the ability to observe and to compare nodes' behaviors to report any wrong doings or policy infringement, Cf. [1, 4, 13, 19, 20, 15, 9].

The proposals usually assume that the monitoring entity is an intrusion detection system (IDS) or a watchdog. While these systems are available in the resourceful and centralized networks such as the Internet, it may not be possible to implement them in distributed systems such as MANETs which have very strong constraints on the network resources. For example, IDSes are centralized and heavily rely on the pre-established behavioral rules, the efficiency of IDSes are yet to be proven in highly dynamic networks. In order to prevent serious attacks at low level networking, policy-based IDSes might not be adaptable or sufficiently responsive.

For trust propagation (Cf. [8, 7, 21, 6]), distributed networks such as MANETs require trust metrics to be forwarded on a multi-hop basis. The distant nodes then make assumptions on the degree of the trustworthiness of intermediate recommenders. For example: if node B observes that node C is only 30% trustworthy and node A to node B is 40% then many trust propagation protocols assume that node C is $30\% * 40\% = 12\%$ trustworthy to A. Notice that this multiplicative rule over the trust values assumes that the trustworthiness is independent for the intermediate nodes. This assumption is most likely unrealistic since nodes interact with each other to evaluate their mutual trust values.

To sum up, probabilistic trust evaluation based on behavioral observations relies on many assumptions that may not be true and also on external components that are yet to be proven practical.

3.1.2 High-level trust computing

The existing proposals characterized as high-level trust computing include frameworks or architectures that incorporate trust evaluation, propagation, IDS policies and security components such as secured routing, admission

control, etc. (Cf. [10, 11, 12, 18]). High-level trust computing has many advantages since it offers a unified view on trust throughout the network, and thus facilitates security collaborations of nodes from different domains.

However, in this context, trust is often considered as an issue of quality of service rather than one of security. It also becomes less obvious from the high-level point of view how the frameworks can practically be used to secure communications against attacks that happen at the low-level of the network, since these attacks are specific to many networking properties: wired/wireless, fixed/mobile, centralized/distributed, etc.

To illustrate that the existing *top-down* proposals may not be best suitable for trust computing, we present in the next section some counter-examples in which the network may need other considerations of trust computing to defend itself from some practical attacks.

3.2 Counter-examples for *top-down* trust computing

In the literature of trust computing for networked communications, *top-down* approaches extensively rely on behavioral monitoring to evaluate the trustworthiness of a node. Typical malicious behavior that has been discussed includes *selfishness* and *packet dropping*.

The example that has often been used to illustrate node selfishness is the following. Assuming a distributed communication system in which nodes use cognitive radios to advertise their transmission load and to make reservation of the medium access. A node may behave selfishly by falsely declaring that it has a high amount of priority data to transmit or to forward. Since there is no centralized entity to coordinate the reservation, there will be no easy way for each node to verify the veracity of all the claims of bandwidth prior to the actual transmissions. A trust system would then rely on the monitoring of the actual transmissions for verification. This means each node has to put its network interface into promiscuous mode in order to capture all the transmission around it, decode all the packets and analyze them. This monitoring is usually referred to as *watch dog*. There are many inconveniences of doing so in a distributed, dynamic network.

Firstly, there is no doubt that keeping the network interface in a permanent wake-up mode, in order to decode and process all packets in the transmission range is a very costly operation. In particular, doing so will consume a great deal of the node's battery power. It is worth pointing out that every practical MAC design seeks to put the network interface into idle mode whenever possible to save energy. Therefore, keeping the network interface in

permanent wake-up mode will most certainly not gain wide acceptance.

Secondly, every peer-to-peer transmission in a secured network should be encrypted by some encryption algorithms to ensure that an attacker cannot analyze the network's traffic. Allowing nodes in the network to analyze transmissions around them creates a breach in security if we have not already established trust between all nodes. Even a small amount of information on a data packet such as the source, destination addresses and the data priority, if made available, can be exploitable by the attacker.

Thirdly, an inconsistent behavior of a node may accurately be detectable in a wired network, for example: by monitoring the amount of data that actually came out of a link and by comparing this amount to the bandwidth that the node had reserved on this same link. It is not easy to have such detection in a wireless, error-prone, dynamic network. At the beginning, a node has an actual amount of data to transmit to its peer. After the reservation process, this node may not be able to transmit the data because the peer could have moved away, or some environmental factors (such as interference, physical obstacle) could have blocked the signal. Even if this node has transmitted the data correctly as reserved, the detection by measuring the amount of transmitted data can face the same inaccuracies due to the dynamic environmental factors as the transmission itself.

Finally, using the node selfishness to illustrate a security threat that appeals to trust computing could be inadequate, since the node selfishness is more a protocol design problem (in this case a problem of the bandwidth reservation protocol) than a security problem. Fixing the reservation protocol is deemed to be sufficient to address this issue. Note that a similar problem has been found in some implementations of the CSMA/CA protocol that allow users to manually change the value of the contention window (CW) so that they were always the first ones who occupied the medium.

Similar to the node's selfishness example, *packet dropping* is commonly cited as a relevant example of an attack that can be prevented by trust computing. In this attack, malicious nodes drop packets to be forwarded at some specific rates. All nodes monitor the packet loss rates and evaluate probabilistically the trustworthiness of their peers. We argue that this attack is more a case study of quality of service (with delivery rate acting as the principal metric to optimize) than a problem of security. Many QoS routing protocols can deal efficiently with this issue by avoiding the congested area of the network. Also, a serious attacker is more inclined to conceal his misbehavior by performing

his routing task well and, at the same time, he tries to copy and to decrypt the information, rather than trigger suspicion by dropping packets for no reasons. There are other more efficient ways to willingly disrupt the network, such as distributed denial-of-service (DDoS), than just dropping packets arbitrarily. Furthermore, the environmental factors explained above indicate that relying on the loss rates to estimate the maliciousness of nodes can lead to wrong trust evaluation. Hence, doing so potentially damages the network even more because the trust component may exclude good nodes from the network.

Since we have shown that trust in these two examples cannot be monitored easily and consistently, the trust system, if it relies on this monitoring, cannot produce an accurate trust evaluation that is consistent with the security expectation.

3.3 Our proposal of *bottom-up* trust computing

Based on the shortcomings of the *top-down* approaches, we argue that trust must be considered with all the characteristics of each security defense and attack mechanism. There should not be one single definition of trust throughout the network, but the definition of trust must lie in the context of each security scenario. We call this approach *bottom-up* since it allows for the consideration of the role of trust in each practical, specific scenario. These scenarios or examples will then serve as building blocks for any collection of security solutions. In particular, cryptography and cryptanalysis contain some excellent examples of the strong connection between trust and communications security.

To clarify *bottom-up* trust computing, we present some security examples that appeal to the concept of trust. We acknowledge that a significant amount of work needs to be done in order to determine if (and how) existing cryptography solutions can be implemented in large networks made of low-capacity (and possibly cheap) communicating devices. This should be the starting point of research in trust computing for networked communications.

3.3.1 The requirement for trust and cryptography solution

Starting from the practical security attack and defense mechanisms, trust computing is required in the following examples:

- **Key management problems (Cf. [2, 3, 16]):**

A key management is composed of mechanisms for key distribution, key revocation and must enforce *forward* and *backward security*.

With *forward security*, when an existing node leaves, either voluntarily or by being forced, a group with a shared group key, the key management protocol needs to securely and efficiently change the group key such that all future communications of the group are kept secret to this node. It is natural to distrust a leaving node.

With *backward security*, when a new node joins a group, even if we can trust this node with all future exchanges, we may not want it to be able to decrypt past communications of the group. Hence a new group key must be generated and distributed.

In this context, the key management protocol must implement broadcasting with selective encryption (Cf. [14]), in order to guarantee that only an authorized subgroup of members can decrypt the broadcast message containing the new group key.

- **Zero-knowledge proofs (Cf. [5]):**

As discussed at the beginning of the paper, this issue arises when two parties wish to establish their trust relationship prior to the exchanges of some valuable information, but neither authentication server nor third-party recommender is immediately available. Then both parties must prove to each other that they actually know a shared passphrase without revealing any other information on the passphrase itself. Until trust is established, revealing any information on the passphrase (such as its hash values) could lead to a breach in security since an attacker can collect the information and replay it later. The zero-knowledge proof protocols allow both parties to exchange the proof of their knowledge without revealing it.

- **Good/bad mouthing, data corruption or, more generally, message integrity (Cf. [17]):**

Good and bad mouthing are typical attacks that threaten the security protocols which use some form of reputation and recommendation to establish communications between distant nodes. Malicious nodes may consistently send good reports about other malicious nodes and bad reports about good nodes. They may also modify reports that they forward on behalf of other nodes.

To deal with this problem, nodes may be required to digitally sign their reports, that is, to identify in an unambiguous way the source of the reports. Also, a digital

signature ensures that any modification of the report can be properly detected or invalidate the report.

3.3.2 Practical consideration when using cryptography for trust computing

Difficulties of practical cryptography implementation arise when we place ourselves in the context of working on resource constrained networks. The main obstacles include the overhead in computational power of the device and in bandwidth for the transmissions.

Some constraints also apply for each specific cryptography mechanism, for example: the message to be signed must be large enough to prevent brute-force attacks from recovering the secret key; the ciphertext should be split into several parts and transmitted separately one after another to avoid man-in-the-middle attack. Taking into consideration all these constraints in the real implementation also means demanding a certain level of trust in the low level security mechanisms such as the key length or the strength of the hash functions.

3.4 Inputs to other security mechanisms

As our primary objective is to stay close to practical security scenarios, while aiming for strong impacts at the low level of networking mechanisms, our proposal of using cryptography for trust computing can provide many inputs to the networking mechanisms with security requirements, such as: admission control, routing in multi-hop networks, key management, etc. A few examples can be cited below.

- Zero-knowledge proofs can provide trust collaboration in dynamic, distributed environments where centralized authentication is not always available. Therefore, it offers a possibility to perform admission control in this context.
- Cryptography for selective broadcasting can assist key management systems in renewing keys.
- Digital signature can be used to enforce message integrity and to detect any modifications made to a message. Thus, it provides means to secure routing in multi-hop networks.

4. Conclusions

Trust is an important concept in network security. It becomes even more challenging in wireless multi-hop and distributed networks where dismounted soldiers join and leave the network to support dispersion and rapid aggregation. Trust management is often interpreted as a

multipurpose control mechanism that establishes trust among the participants (or nodes). Traditionally, trust has been considered as a universal metric throughout the network. Nodes try to determine the degree of trustworthiness of other nodes based on the observables. This approach naturally leads to the notion of behavior monitoring and probabilistic estimation based on predefined rules. For example, intrusion detection systems are assumed to play a major role in trust evaluation. Another component of trust management is trust propagation where the nodes relay trust metrics and disseminate them into the network so that distant nodes can also estimate indirectly the trustworthiness of their distant partners.

This view of trust as a single metric, measurable and universal, is simple. However, it has a shortcoming with regards to its applicability. For example, it is not always clearly indicated the kinds of attacks on the network this approach of trust aims at, or is defined for. In practice, most IDSes are centralized and rely on pre-established behavioral rules. Therefore, the efficiency of IDSes in highly dynamic, distributed systems such as MANETs is still an open question. Policy-based IDSes might not be adaptable or sufficiently responsive to the attacks that take place at the low level of the network. Other trust management mechanisms, such as trust propagation, also need to be examined more carefully since their assumptions rely more on qualitative than on quantitative arguments.

Based on the shortcomings of the traditional, *top-down* approaches, we argue that the definition of trust must be found in the context of each networking security scenario. There should not be one single, universal definition of trust in the network but trust should be embedded in each practical, specific networking attack and defense mechanism. These defense scenarios then serve as the building blocks for any collection of security solutions. In particular, quantitative and provable security mechanisms should be the backbone on which trust is built. We show in this paper some typical examples of the strong connection between trust and cryptology. Therefore, we consider trust principally as a requirement, specific to each low level security mechanism. We call this approach *bottom-up*. As our primary objective is to address practical security scenarios and to seek strong impacts at the low level networking mechanisms, our proposal of using cryptography for trust computing can provide many inputs to any collection of security solutions such as admission control, routing in multi-hop networks, key management, etc.

As a promising approach that is positioned half-way between applied and theoretical research on trust computing for communications security, this proposal will need to address the practical challenges posed to the cryptography methods by some resource constrained networks. The main difficulties include a shifting from a centralized infrastructure with heavy key management servers to distributed systems, as well as the overhead in computational power of the devices and in bandwidth for the transmission. Other challenges that are more research-oriented also need to be solved, such as designing new cryptography methods to address the specific requirements of trust in some particular attacks on the network.

Acknowledgments

This work is funded by Defence Research & Development Canada (DRDC).

References

- [1] T. Beth, M. Borcherding, and B. Klein, "Valuation of Trust in Open Networks." In proceedings of the 3rd European Symposium on Research in Computer Security, ESORICS '94, pp. 3–18, London, UK. Springer-Verlag, 1994.
- [2] G. Dini and I. Savino, "S2RP: a Secure and Scalable Rekeying Protocol for Wireless Sensor Networks." In proceedings of IEEE International Conference on Mobile Adhoc and Sensor Systems Conference, pp. 457–466, 2006.
- [3] L. Eschenauer and V.-D. Gligor, "A Key-Management Scheme for Distributed Sensor Networks." In proceedings of the 9th ACM Conference on Computer and Communications Security, CCS '02, pp. 41–47, New York, NY, USA. ACM, 2002.
- [4] L. Eschenauer, V.-D. Gligor, and J. Baras, "On Trust Establishment in Mobile Ad-hoc Networks." In proceedings of Security Protocols Workshop, pp. 47–66. Springer-Verlag, 2002.
- [5] O. Goldreich, S. Micali, and A. Wigderson, "Proofs That Yield Nothing But Their Validity Or All Languages in NP Have Zero-Knowledge Proof Systems." J. ACM, pp. 690–728, July 1991.
- [6] E. Gray, J.-M. Seigneur, Y. Chen, and C. Jensen, "Trust Propagation in Small Worlds." In proceedings of the 1st International Conference on Trust Management, iTrust'03, pp. 239–254, Berlin, Heidelberg. Springer-Verlag, 2003.
- [7] R. Guha, R. Kumar, P. Raghavan, and A. Tomkins, "Propagation of Trust and Distrust." ACM Press, pp. 403–412. ACM, 2004.
- [8] A. Josang, E. Gray, and M. Kinateder, "Analysing Topologies of Transitive Trust." In proceedings of the 1st International Workshop on Formal Aspects in Security and Trust FAST'03, pp. 9–22, September 2003.
- [9] S. D. Kamvar, M.-T. Schlosser, and H. Garcia-Molina, "The Eigentrust Algorithm for Reputation Management in P2P Networks." In proceedings of the 12th International

- Conference on World Wide Web, WWW '03, pp. 640–651, New York, NY, USA. ACM, 2003.
- [10] M. Kinateder and K. Rothermel, “Architecture and Algorithms for a Distributed Reputation System.” In proceedings of the 1st International Conference on Trust Management, iTrust'03, Berlin, Heidelberg. Springer-Verlag, 2003.
- [11] Y. Lacharite, D.-Q. Nguyen, M. Wang, and L. Lamont, “A Trust-Based Security Architecture for Tactical Manets.” In proceedings of IEEE Military Communications Conference. MILCOM 2008.
- [12] F. Martinelli and M. Petrocchi, “A Uniform Framework for Security and Trust Modeling and Analysis with Crypto-CCS.” *Journal of Electron. Notes Theor. Comput. Sci.*, pp. 85–99, July 2007.
- [13] L. Mui, M. Mohtashemi, and A. Halberstadt, “A Computational Model of Trust and Reputation for E-Businesses.” In proceedings of IEEE Hawaii International Conference on System Sciences, pp. 188–193, Los Alamitos, CA, USA. IEEE Computer Society, 2002.
- [14] D.-Q. Nguyen and L. Louise, “A New Cryptography Scheme for Selective Broadcasting.” In proceedings of IEEE International Conference on Information and Computer Networks, ICICN 2011, Guiyang, China, 2011.
- [15] D.-Q. Nguyen, L. Lamont, and P.-C. Mason, “On Trust Evaluation in Mobile Ad-hoc Networks.” In proceedings of Proceedings of the 1st International ICST Conference on Security and Privacy in Mobile Information and Communication Systems, MobiSec '09, Turin, Italy, 2009.
- [16] A. Perrig, R. Szewczyk, J.-D. Tygar, V. Wen, and D.-E. Culler, “Spins: Security Protocols for Sensor Networks.” *Wireless Networks*, pp. 521–534, 2002.
- [17] B. Schneier, “Applied Cryptography (2nd Ed.): Protocols, Algorithms, and Source Code in C.” John Wiley & Sons, Inc., New York, NY, USA, 1995.
- [18] N. Stakhanova, S. Basu, J. Wong, and O. Stakhanov, “Trust Framework for P2P Networks Using Peer-Profile Based Anomaly Technique.” In proceedings of the 2nd International Workshop on Security in Distributed Computing Systems (SDCS), ICDCSW '05, pp. 203–209, Washington, DC, USA. IEEE Computer Society, 2005.
- [19] Y. Sun, W. Yu, Z. Han, and K.-J. Liu, “Trust Modeling and Evaluation in Ad-hoc Networks.” In proceedings of IEEE Global Telecommunications Conference, GLOBECOM '05, St. Louis, MO, USA, 2005.
- [20] G. Theodoropoulos and J.-S. Baras, “Trust Evaluation in Ad-hoc Networks.” In proceedings of the 3rd ACM workshop on Wireless security, WiSe '04, New York, NY, USA. ACM, 2004.
- [21] C.-N. Ziegler and G. Lausen, “Propagation Models for Trust and Distrust in Social Networks.” *Information Systems Frontiers*, pp. 337–358, December 2005.

Dang-Quan Nguyen is a research scientist at the Communications Research Centre, Ottawa, Canada. He has obtained his Master degree in 2003 and his PhD in 2006, both in Computer Science at University of Paris VI (Pierre and Marie Curie), France. He has undergone various research projects at INRIA (France), Orange's Labs (France) and CRC (Canada) under major government grants and contracts. His research interest includes quality of service in MANETs, trust-based security and cryptography.

Louise Lamont is the Research Manager for the Mobile ad hoc and Sensor Networking Group at the Communications Research centre. In this position Louise is responsible for identifying novel research areas for study at CRC and for proposing, implementing and securing funds for new projects in support of major client requirements such as DND. She manages several state-of-the-art laboratories for the conduct of research as well as technical demonstration in the area of mobile ad hoc and sensor networks. She is also responsible for establishing and maintaining liaison, collaboration and partnership with R&D groups at CRC and with external national and international organizations.

Modelling and simulation of complex systems: an approach based on multi-level agents

Alain-Jérôme Fougères¹

¹University of Franche-Comté / Computer Science Laboratory (LIFC)
Montbéliard, France

Abstract

A complex system is made up of many components with many interactions. So the design of systems such as simulation systems, cooperative systems or assistance systems includes a very accurate modelling of interactional and communicational levels. The agent-based approach provides an adapted abstraction level for this problem. After having studied the organizational context and communicative capacities of agent-based systems, to simulate the reorganization of a flexible manufacturing, to regulate an urban transport system, and to simulate an epidemic detection system, our thoughts on the interactional level were inspired by human-machine interface models, especially those in "cognitive engineering". To provide a general framework for agent-based complex systems modelling, we then proposed a scale of four behaviours that agents may adopt in their complex systems (reactive, routine, cognitive, and collective). To complete the description of multi-level agent models, which is the focus of this paper, we illustrate our modelling and discuss our ongoing work on each level.

Keywords: Agent-Based System, Agent Modelling, Agent Behaviour, Complex System, Multi-Level Agent

1. Introduction

The main objective of our research over the last decade has been agent-based simulation and complex system (CS) modelling. After studying the agent-based systems (ABS) organizational context, to simulate the reorganization of flexible manufacturing and regulate an urban transport system, we focused on modelling agents with strong communication skills, which may be used as building elements for the design of assistance systems to CS users. Our thoughts on the interactional level were inspired by models from human-machine interfaces field, especially those of the "cognitive engineering" approach [6]. Then we suggested a first agent model and defined the communication model of these agents [10,11,29]. Next, we realized that the agent architecture could vary (multi-level architecture), so as to support the more or less cognitive tasks that agents perform. We then defined the different granularities of these agents by placing them on a scale of behaviours inspired by Ramussen's three-level scale [31] to describe human operator behaviour.

CS are "made of many components with many interactions" [34]. The CS design (cooperative systems, assistance systems, etc.) then includes a very specific modelling of interactional and communicational levels.

Moreover, according to Morin [22], CS designer "must have a method that allows to design the multiple points of view, and to move from one point of view to another."

In this sense, Wooldridge [37] and Jennings [19] have argued that agents are a new paradigm for CS engineering; they suggest a satisfactory response to three common techniques for reducing the software complexity: decomposition, abstraction and organization. Jennings [19] raises two hypotheses of adequacy and formation: (1) agent approach can significantly improve our ability to model, design and build complex and distributed software systems, (2) in addition to being able to design and build CS, agent approach is destined to become a major paradigm of software engineering.

From the IAD field [36], ABS, and before that the actor model [17], all withhold the basic principle of the knowledge and information distribution necessary to solve a problem on a set of interacting agents, capable of pursuing and achieving a common goal. An agent is an active, interactive and proactive entity of a system. It is often differentiated according to its cognitive (social metaphor) or reactive (biological metaphor) nature; we frequently define the granularity of agents according to their degree of knowledge and functional complexity.

A software agent, according to the Newell and Simon model [24], is an autonomous information processing system that means it is composed of reception and transmission devices, a processor and a memory (knowledge base). An agent-based system is a society of autonomous agents working together to reach a common goal from interaction, communication or transaction. For a first understanding of the agent paradigm, we can consider that an agent is a computer system located in an environment, in which it can act, possibly in interaction with others agents, and this in complete autonomy [19]. Autonomy is for us the main characteristic of an agent, relating to the object paradigm. It is realized by: (1) an independent computer process, (2) an individual memory (knowledge/data), and (3), an ability to interact (perception/reception, communication/action).

The paper is organized as follows. In the second section, the adequacy of the agent paradigm to model and simulate complex systems, and then a model of multi-level agents is suggested. In the third section, an

illustration and a discussion of each of the four agent behaviour levels was drawn. Finally, in the last section, the conclusions of this research are presented.

2. Agents for Complex Systems Simulation

2.1. Agent Paradigm

There are many definitions of the agent paradigm [19, 37], supported by typology proposals, but new types of agents continue to emerge [35]. It is therefore difficult to establish a consensus. However, through these definitions we observe that three functions characterize agent activity: to perceive, decide, and act. An agent has its own knowledge. It acts in autonomy to reason and decides according to its objectives, its interactions with other agents in the system, and its environment perception. By extension, considering cognitive agents, experts of this domain generally agree on the following characteristics: intentionality, rationality, commitment, adaptability, and "intelligence". ABS are systems that allow distributing agents, communicating, autonomous, reactive, skilful, and finalized entities. They form intelligent solver networks, weakly bound, working together to solve problems beyond their individual capabilities and knowledge [18].

Among the suggested agent architectures with a cognitive orientation, the *BDI* model (Belief, Desire, and Intention) is best known [32]. It is built around three concepts inspired by human behaviour models: (1) beliefs, based on agent knowledge, (2) desires, corresponding to the knowledge that agent would express, and (3) intentions, or actions, that agents decide to do. Thus, a software agent, which has a desire, can transform it into intention when it knows it can achieve this; it is then left to act!

2.2. Agent Organisation and Interaction

Work in sociology or science organizations have always interested the ABS community, as a source of modelling [38]; the organizational configuration type scale by Mintzberg [21], from hierarchical structure (centralization) to the professional bureaucracy (decentralization), via the "Adhocracy" (mutual adjustment between groups), is one example.

The inherent problems to the partial agent knowledge in the pursuit of local goals or interlaced agents require the development of advanced coordination mechanisms [9]. An organization must allow ABS to behave as a coherent whole, to solve a problem uniquely. It controls and coordinates the interactions between agents of the system, and structures their activities.

Communication and interaction are interrelated. Following the definitions proposed for man-machine dialogue [5], it is possible to distinguish these two concepts as follows: *interaction* is an exchange between agents and their environment, this exchange depends on the intrinsic properties of the environment in which agents are active; agent *perception* may be passive by

receiving messages/signals, or active, when it is the result of voluntary action; *communication* is an exchange between agents themselves, using a language. In addition, the concept of interaction is associated with the protocol and interaction pattern concepts, allowing us to specify how and with whom an agent can interact - engineering has been suggested [9]. Interaction specifications differentiate levels of interaction, such as micro or macro levels. As in the case of human communication, communicative interactions among agents can obviously be linked together and produce information exchanges or dialogues [2].

In most of ABS, the agent behaviour in interaction consists of three phases: (1) information reception from another agent, or the perception of a change in its environment, (2) interpretation of this event taking into account other agents, (3) sending a message or taking action in an environment modification. If interactions between agents are frequently communicative, they also involve action coordination, cooperation and negotiation [39].

2.3. Agent-Based Systems Design

The ABS design, often questioned from process or methodological perspectives [4], presumes that the designer proceeds with a local vision to respect the fact that each agent manages its own knowledge and actions (autonomy). The support languages of ABS design are numerous [3].

In [11] we proposed an ABS design method in four stages making extensive reference to *AUML* (Agent Unified Modelling Language) [1]: (1) making the use case diagrams (services provided by ABS); (2) for each case, sequence diagrams carry out the interactions (message exchanges and scheduling) between the agents involved in this reference case; (3) from the sequence diagrams, which identified agents, objects and their interactions, making the class diagram: objects are associated with classes, exchanged messages (service requests between objects) are translated into operations on classes, parameters associated with operations are translated into class attributes - it may be possible to complete this diagram by a collaboration diagram; (4) from the class diagram, defining the behaviour of each agent (class agent) with a state diagram or an activity diagram. The description of roles played by various cooperative agents is insufficient since the notion of role, is non-existent from *UML*, and isn't the subject of a few short extensions in *AUML* [1]. It focuses on collaboration diagrams and sequence diagrams.

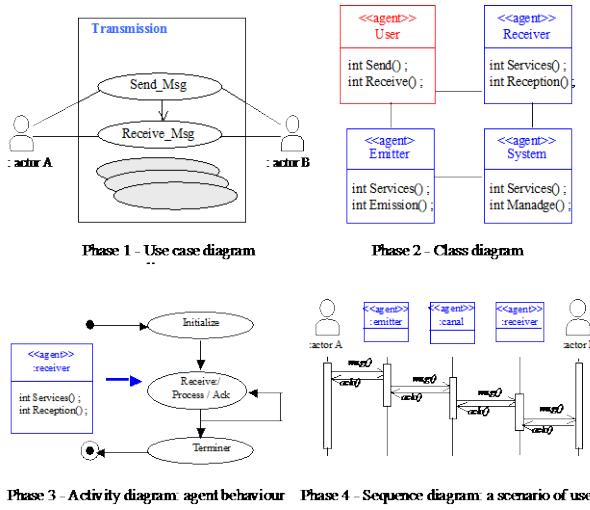


Fig. 1. *Agent design methodology, illustrated by a specification of "Message Transmission"*

2.4. Proposition of Agent-Based Systems Modelling

After studying the ABS organizational context, to simulate the reorganization of a flexible manufacturing system and the regulation of an urban transport system, we focused on the modelling of agents with strong communication skills, which may be used as building the foundation for the design of complex systems such as cooperative systems and user assistance systems. Our thoughts about this interactional level were fed by models from the man-machine interface field, especially those of the "cognitive engineering" approach [6]. We then proposed a variable granularity agent model based on Rasmussen's three level scale [10, 11, 28].

2.4.1 Notations

Following notations and definitions are used in this modelling section:

$A = \{\alpha_i\}$ is the agents finite set, $i \in I_A$, $I_A = \{1, 2, \dots, q_A\}$;

$I = \{t_i\}$ is the interactions finite set defined for all agents, $i \in I_I$, $I_I = \{1, 2, \dots, q_I\}$;

$P = \{\rho_i\}$ is the roles to be played by all agents finite set, $i \in I_P$, $I_P = \{1, 2, \dots, q_P\}$;

$O = \{o_i\}$ is the agent organizations into communities finite set, $i \in I_O$, $I_O = \{1, 2, \dots, q_O\}$;

$\Sigma = \{\sigma_i\}$ is the states finite set of agent-based system, $i \in I_\Sigma$, $I_\Sigma = \{1, 2, \dots, q_\Sigma\}$;

$\Sigma_{\alpha_i} \subseteq \Sigma$ is the states finite set of agent α_i ;

$\Sigma_{M_{\alpha_i}} \subseteq \Sigma$ is the states finite set of agent-based system that agent α_i knows;

$\Pi = \{\pi_i\}$ is the observations finite set, $i \in I_\Pi$, $I_\Pi = \{1, 2, \dots, q_\Pi\}$;

$\Pi_{\alpha_i} \subseteq \Pi$ is the observations finite set that agent α_i can do;

$\Delta = \{\delta_i\}$ is the decision rules finite set, $i \in I_\Delta$, $I_\Delta = \{1, 2, \dots, q_\Delta\}$;

$\Delta_{\alpha_i} \subseteq \Delta$ is the decision rules finite set that agent α_i can trigger;

$\Gamma = \{\gamma_i\}$ is the actions/reactions finite set of all agents, $i \in I_\Gamma$, $I_\Gamma = \{1, 2, \dots, q_\Gamma\}$;

$\Gamma_{\alpha_i} \subseteq \Gamma$ is the actions finite set that agent α_i can process;

$\Gamma_{\Lambda_{\alpha_i}} \subseteq \Gamma$ is the specific communication acts finite set that agent α_i can process;

$\Omega = \{\omega_i\}$ is the interpretations finite set of all agents, $i \in I_\Omega$, $I_\Omega = \{1, 2, \dots, q_\Omega\}$;

$\Omega_{\alpha_i} \subseteq \Omega$ is the finite set of interpretations of observations made by agent α_i ;

$K = \{\kappa_i\}$ is the knowledge finite set of, $i \in I_K$, $I_K = \{1, 2, \dots, q_K\}$;

$K_{\alpha_i} \subseteq K$ is the knowledge finite set of agent α_i , with $K_{\alpha_i} = P_{\alpha_i} \cup \Sigma_{\alpha_i} \cup \Sigma_{M_{\alpha_i}}$;

$E = \{e_i\}$ is the events finite set, $i \in I_E$, $I_E = \{1, 2, \dots, q_E\}$;

$E_{\alpha_i} \subseteq E$ is the events finite set that agent α_i can observe;

$X = \{\chi_i\}$ is the conditions finite set, $i \in I_X$, $I_X = \{1, 2, \dots, q_X\}$;

$X_{\alpha_i} \in X$ is the conditions finite set associated to the internal states of agent α_i ;

$N = \{\nu_i\}$ is the configuration networks finite set of, $i \in I_N$, $I_N = \{1, 2, \dots, q_N\}$;

$\Upsilon = \{\psi_i\}$ is the connexions between agents finite set, $i \in I_\Upsilon$, $I_\Upsilon = \{1, 2, \dots, q_\Upsilon\}$;

$\Lambda = \{\lambda_i\}$ is the speech acts finite set, $i \in I_\Lambda$, $I_\Lambda = \{1, 2, \dots, q_\Lambda\}$;

$H = \{\eta_i\}$ is the messages finite set, $i \in I_H$, $I_H = \{1, 2, \dots, q_H\}$;

$T = \{\tau_i\}$ is the type of messages finite set, $i \in I_T$, $I_T = \{1, 2, \dots, q_T\}$;

$\Phi_{\Pi(\alpha_i)} : \Sigma \times \Sigma_{M_{\alpha_i}} \rightarrow \Pi_{\alpha_i}$ is the function of observations of agent α_i ;

$\Phi_{\Delta(\alpha_i)} : \Pi_{\alpha_i} \times \Sigma_{\alpha_i} \rightarrow \Delta_{\alpha_i}$ is the function of decisions of agent α_i ;

$\Phi_{\Gamma(\alpha_i)} : \Delta_{\alpha_i} \times \Sigma \rightarrow \Gamma_{\alpha_i}$ is the function of actions of agent α_i .

2.4.2 Modelling

The formal approach we follow to model and design CS is to define the modular architecture of agents, to define their model of interaction, communication and knowledge and to respect a rigorous methodology for acquiring expertise. Then, an ABS M is described by a 4-tuple (1):

$$M = \langle A, I, P, O \rangle \quad (1)$$

where A is a set of agents, I is the set of interactions defined for agents of A , P is the set of roles to be played by agents of A , and O is the set of organizations of agents into communities.

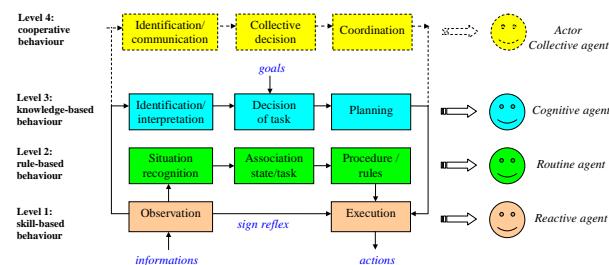


Fig. 2. *Variable agent behaviour, based on Rasmussen's model.*

Many agent structures known as “cognitive” are inspired by the cycle \langle perceive, decide, act \rangle [23]. However, our agent model [10, 11] is rather inspired by Rasmussen’s three-level operator [31]: 1) reflex-based behaviour, 2) rule-based behaviour, and 3) knowledge-based behaviour with interpretation, decision and plan (Figure 2). We interpreted this model as a model of process for agents. The latter are both cognitive and reactive. Moreover, they have behaviours adapted to the tasks they perform. We added one level at this scale to include behaviour based on cooperation. We call *actor* (or collective agent) a system of cooperative agents in which the behaviour is defined by collective decision tasks and collective coordination tasks [13].

Agents developed in our various projects, whose behaviour is illustrated in the following, can perform reflex actions (automatic), routine actions, and actions in new situations (creative or cooperative situations). These models are based on the sequential character of the human cognitive system (serialization in the symbolic level), not excluding a certain degree of parallelism in the processing of sensorimotor signals [34]. Thus, an

agent $\alpha_i \in A$ can evolve on one of the first three levels of our scale (Figures 2 and 3); it is described by one of the following tuples (2, 3, 4):

$$\alpha_i = \langle \Phi_{\Pi(\alpha_i)}, \Phi_{\Gamma(\alpha_i)} \rangle \quad (2)$$

$$\alpha_i = \langle \Phi_{\Pi(\alpha_i)}, \Phi_{\Delta(\alpha_i)}, \Phi_{\Gamma(\alpha_i)}, K_{\alpha_i} \rangle \quad (3)$$

$$\alpha_i = \langle \Phi_{\Pi(\alpha_i)}, \Phi_{\Omega(\alpha_i)}, \Phi_{\Delta(\alpha_i)}, \Phi_{\Gamma(\alpha_i)}, K_{\alpha_i} \rangle \quad (4)$$

where $\Phi_{\Pi(\alpha_i)}$ is the function of observations of agent α_i ; $\Phi_{\Omega(\alpha_i)}$ is the function of interpretations of agent α_i ; $\Phi_{\Delta(\alpha_i)}$ is the function of decisions of agent α_i ; $\Phi_{\Gamma(\alpha_i)}$ is the function of actions of agent α_i ; K_{α_i} is the finite set of knowledge of agent α_i - the knowledge contained in its memory, among which are the decision rules, the values of the domain, and the acquaintances and/or networks of affinities between agents, along with dynamic knowledge (observed events, internal states, etc.). The resource management associated with these various functions is provided by the set M_G of managers: $M_G = \{M_H, M_\Gamma, M_K\}$, where M_H is the messages manager, M_Γ is the actions manager and M_K is the knowledge-base manager (Figure 3).

The decision rules Δ_{α_i} of the agent α_i , gathered in its knowledge base, are described by a 3-tuple (5):

$$\Delta_{\alpha_i} = \langle E_{\alpha_i}, X_{\alpha_i}, \Gamma_{\alpha_i} \rangle \quad (5)$$

where E_{α_i} is the set of events that agent α_i can observe, X_{α_i} is the set of conditions associated to the internal states of agent α_i , and Γ_{α_i} is the set of actions that agent α_i can perform. For instance, let us consider the decision rule δ_1 with: (1) $\varepsilon_1 := \langle \text{inform}, \alpha_{f_1}, \alpha_{r_k}, t = 2, V \rangle$; (2) $\chi_1 := \langle V = \text{sup}(0.4) \rangle$; (3) $\gamma_1 := \langle \text{diffuse}, \alpha_{f_1}, F, t=2, V \rangle$.

This rule means that: (1) depending on following event ε_1 : function agent α_{f_i} ($\alpha_{f_i} \in F, F \subseteq A$) receives a message of type t whose value is equal to 2 (corresponding to the transmission of a value) by which a requirement agent α_{r_k} ($\alpha_{r_k} \in R, R \subseteq A$) informs α_{f_i} of its value V ; (2) under condition χ_1 “ V must be greater than the threshold value 0.4”; (3) action γ_1 will then be triggered: agent α_{f_i} will communicate this information to all function agents of the set F . Actions of agent α_i are controlled and memorized by a manager M_{Γ_α} .

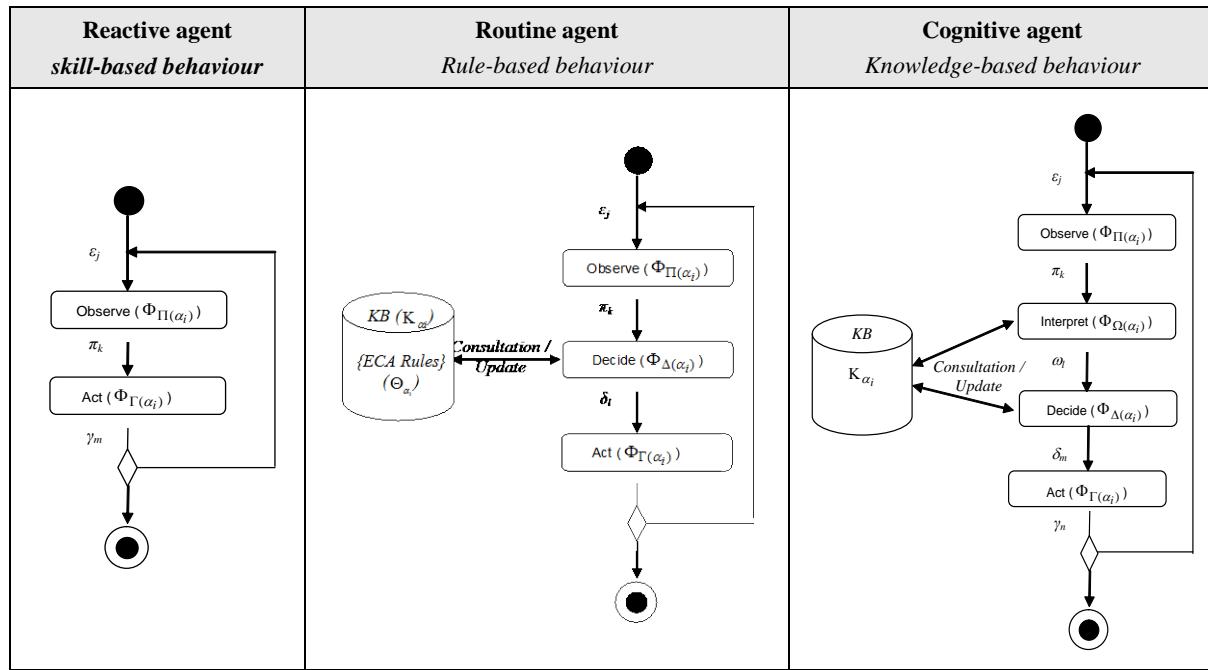


Fig. 3. Activity diagrams for the behaviour of agents on the first three level of our scale

Agent communicational interactions. Communication is the primary mechanism of interaction of an agent with its agent community. To communicate with other agents, an agent can exchange messages in syntax of an interaction language based on the concept of speech acts [33]. These information exchanges are controlled by a communication protocol in which a response is required for some speech acts (ask/[accept, refuse], inform/confirm, propose/[confirm, refuse], evaluate/[agree, disagree], etc.). We have compiled a lexicon of performative verbs we use in our applications ($\Lambda = \{inform, diffuse, ask, reply, confirm...\}$). The basic elements of this language (variables and primitives) are presented in Table 1. These speech acts are sufficient to enable agents to perceive the intention associated with the proposition content in a message. In the course of an interaction, an agent chooses its destination agent according to its intentions, the activity context and the state of its acquaintances.

A communication act $\lambda_{s,r}$ exchanged between two agents ($\lambda_{s,r} \in \Gamma_{\Lambda_{\alpha_i}}$) is defined by a 5-tuple (6):

$$\lambda_{s,r} = <\lambda, \alpha_s, \alpha_r, \tau, \eta> \quad (6)$$

where $\lambda \in \Lambda$ is a speech act denoted by a performative verb, α_s is the source agent of communication, α_r is the receiver agent, $\tau \in T$ is the type of message and $\eta \in H$ is the message itself, which can be an assertion, a question, a response, etc.

Some elements of language	Significance
$\alpha, \varepsilon, \gamma, \eta, \beta, \tau$	respectively are agent, event, action, message, speech act and type of message
$inform(\alpha_s, \alpha_r, \tau, \eta)$	α_s sends to α_r the fuzzy message η of type τ
$diffuse(\alpha_s, [\alpha_l], \tau, \eta)$	α_s sends to the list $[\alpha_l]$ the fuzzy message η of type τ
$ask(\alpha_s, \alpha_r, \tau, \eta)$	α_s asks to α_r the fuzzy request η of type τ
$answer(\alpha_s, \alpha_r, \tau, \eta)$	α_s answers α_r the fuzzy message η of type τ
$confirm(\alpha_s, \alpha_r, \tau, \eta)$	α_s confirms to α_r that it is agree with fuzzy message η of type τ
$propose(\alpha_s, \alpha_r, \tau, \eta)$	α_s proposes to α_r the proposition contained in the fuzzy message η of type τ
$against-propose(\alpha_s, \alpha_r, \tau, \eta)$	α_s against-proposes to α_r the proposition contained in the fuzzy message η of type τ
$refuse(\alpha_s, \alpha_r, \tau, \eta)$	α_s refuses the proposition proposed by α_r , contained in the fuzzy message η of type τ
$accept(\alpha_s, \alpha_r, \tau, \eta)$	α_s accepts the proposition proposed by α_r , contained in the fuzzy message η of type τ
$order(\alpha_s, \alpha_r, \tau, \eta)$	α_s orders α_r to do the task contained in the fuzzy message η of type τ

Table 1. Interaction language for cooperative agents

Coordination and organization of agents. Interactions between agents of complex systems are not just communicational, they also involve cooperation and action coordination required to achieve the common goals of the agent system. Agent-oriented coordination models focus on the behaviour of agents in order to achieve a coordinated system. Initially organized in communities, agents are all involved in the same activity of production or problem solving. During activities, inter and intra-community networks of affinities between agents can emerge. Therefore, the coordination of agents and the self-organization of the communities are carried out by message exchange (mutual adjustment or emergence of networks of affinities, for examples).

Agent knowledge. An agent of a complex system has four kinds of knowledge: (1) domain knowledge, (2) functional skills to operate on states and values according to the processes defined in the complex system, (3) knowledge to control the agent's activity (decision rules), and (4) knowledge to interact with other agents (language and protocol of communication, acquaintances, networks of affinities, etc.). This knowledge is constantly evolving during tasks processed in the complex system, following the interactions between agents and the interventions of human users.

3. Illustration and Discussion

During the past decade, to validate the use of the proposed scale of agent behaviour (see Figure 2), we designed a set of agent-based systems and we also made numerous modelling. In the following we illustrate the agent modelling for each of our four-level scale with a representative application.

3.1. Level 1 (reactive agent): Agents to Design an Epidemic Simulation-Detection System

The main application for this level focused on the modelling and designs of simulation agents for an ABS called *SIMBADE* [10]. *SIMBADE* is an agent platform that aims to simulate and detect epidemics. This platform, designed in accordance with the French public health organization, allows simulating cases of disease (local or scattered) and regularly reporting diagnostic epidemic, from the messages exchanged by hierarchical agents of the detection system. *SIMBADE* combines complexity and clarity of presentation. Indeed, *SIMBAD* is composed of three subsystems (Figure 5.a): (1) an ABS for the simulation of epidemics, including agents from level 1 which interest us especially in this section; (2) an ABS for the detection of epidemics, including agents of Level 3; and (3) a system of decision support manager of medical knowledge to diagnose diseases and epidemics. Each agent of detection system has its own knowledge in relation to its role within the organization. Decision making is thus distributed. The simulation system is defined to control the two other systems and to understand what would be an autonomous system deployed on the public health system.

The ABS of epidemics simulation, related to artificial life systems, is composed of two types of agents at level 1: (1) *<contaminant>* agents, carriers of diseases known by the system, including reportable diseases, and (2) *<individual>* agents who consult their doctors if they are contaminated. An agent *<contaminant>* is introduced locally by an agent of simulation activating disease (influenza, meningitis, pertussis, listeriosis, etc.). This agent may, because of its proximity, infect healthy agent *<individual>*. It becomes bearer of the disease and may, in turn, move, contaminate other agents of its environment. Once infected, a *<Patient>* agent consults a *<Doctor>* agent from the doctors network (or medical community) (see scenario of Figure 5).

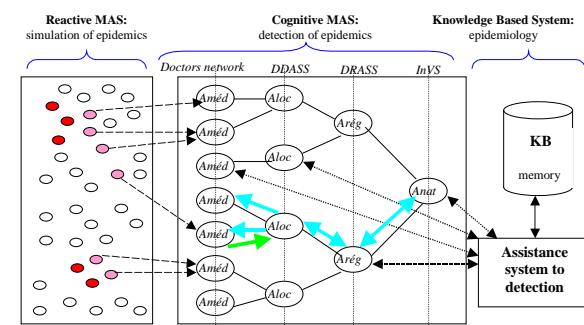


Fig. 4. The three-level architecture of *SIMBADE*

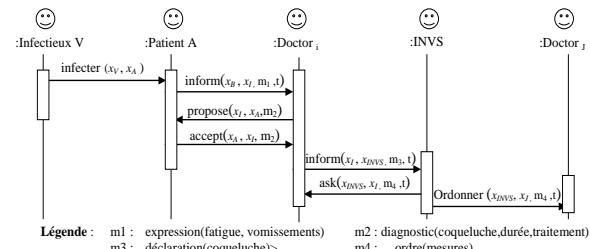


Fig. 5. Scenario describing the contamination of a *<Patient>* agent, leading to mandatory reporting.

The agent behaviour, defined by the first level of our scale, is entirely satisfactory for this simulation activity. The reflex behaviour is inherently simple to model; we have not seen fit to reproduce it to validate the model at the first level.

3.2. Level 2 (routine agent): Agents to Design a Product Configuration System

The main application for this level focused on the modelling and designs of configuration agents for an ABS is called *APIC* [28]. *APIC* is an agent platform of collaborative design for an optimization of product integrated configuration. The configuration of product families is a collaborative and distributed process. Actors in the configuration access the *APIC* platform, using a specific μ-tool to their domain [20]. These μ-tools communicate with the platform through an interface agent. In *APIC*, any configuration item is an

agent. Then, each specification (characteristic of product), function, solution (component of product) and constraint, is represented by an agent. Agents are organized in four communities that communicate heavily with each other and the interface agents (Figure 6): (1) community of *<requirement>* agents, (2) community of *<function>* agents, (3) community of *<solution>* agents, and (4) community of *<constraint>* agents. Each of these communities provides constraints on the process of configuration.

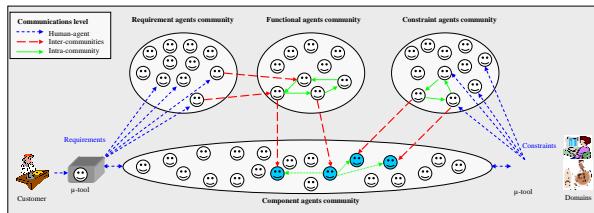


Fig. 6. Agent-based architecture of APIC platform: configuration/communication/cooperation levels

Note that APIC agents, defined by level 2 on our scale (using ECA rules to decide on their actions), were designed as fuzzy [16], in the sense that, not only the data they process are fuzzy (fuzzy knowledge), but their behaviours themselves are fuzzy. Indeed, the inter/intra-communities interactions are weighted by fuzzy values, this allows to weaken them, strengthen them, or even to inhibit them [29]. They operate in an environment defined by fuzzy variables and constraints of configuration.

Figure 7 shows a typical scenario of configuration highlighting the communicative context between level 2 agents of APIC. Consider the first exchanges. An agent *<Ri>* sends a message *M1* of type *T* (fuzzy value that characterizes it in the Requirements network, for example) to an agent *<Fj>*, then, according to the defined communication protocol, awaits an acknowledgment message. This acknowledgment message will allow agent *<Ri>* to know that agent *<Fj>* has received the message *M1* and was able to handle it. For its part, agent *<Fj>* consults its ECA rules. One of them includes: *Event* [receiving a message of type *T*], *Conditions* (value > threshold, for example), and *Action* [diffusion to entire functions community]. Then, agent *<Fj>* broadcasts the received message, and when it received all acknowledgments of his own message, it can send the acknowledgment expected by agent *<Ri>*. And so on.

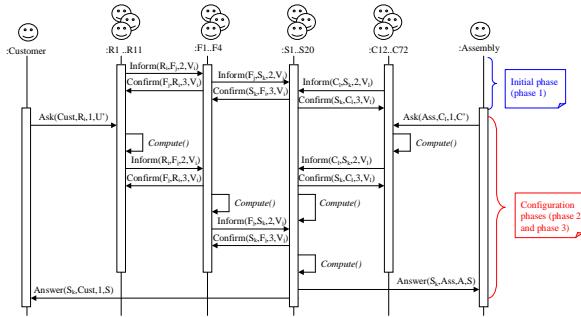


Fig. 7. Typical scenario for product configuration

The behaviour and knowledge (decision rules) of defined agents is entirely satisfactory for this configuration activity [29]. To further validate the model in this second level, we have recently reproduced for modelling simulation agents of a push service (information on mobility) for an on-demand transport system. Other cases of modelling in level 2 have been proposed [12, 13].

3.3. Level 3 (cognitive agent): Agents to Design an Assistance System for Student Projects Management

The main application for this level focused on the modelling and design of assistant agents for an ABS called *iPédagogique* [25]. *iPédagogique* is an agent-based platform that aims to track, manage, and evaluate students' projects. Management of students' project is a complex activity, cooperative and low-instrumented. After designing functionalities for the learning environment *iPédagogique*, it is natural that we were interested in the right level of assistance to provide for facilitating its use. Then, *iPédagogique* proposed an ideal testing environment for the design of an agent-based system in level 3 (high cognitive skills and specific knowledge to each agent). Modelling work has resulted in developing five assistant agents: (*<aCourse>*, *<aPM>*, *<aUser>*, *<aForm>*, *<aTutorial>*).

During the semester, the assistant of project management *<aPM>* simplifies communication, project management and monitoring: for the concerned teacher, assisting in the information broadcasting task, project registration, schedules, and group monitoring, etc.; for student projects groups, assisting in the project registration tasks, compliance schedules, sharing roles, delivery of documents, making appointments, phases constraint reminders, etc.; for tutor projects, assisting in the tasks of monitoring schedules for project groups, receiving documents, making appointments, etc.

The cooperative behaviour of agent *<aPM>* to a group of students comes in two modes: an intervention mode and an informational mode (a reminder during the connection of students, for example). As an illustration, Figure 8 shows the reaction of *<aPM>* following the delay of a project group during the analysis phase - this should normally result in the delivery of a requirements

document to the tutor of the group. An interaction will be maintained with the student responsible for the phase in question. Then $\langle aPM \rangle$ observes proposals from students A and B and tutor, and decides to inform according to these defined ECA rules.

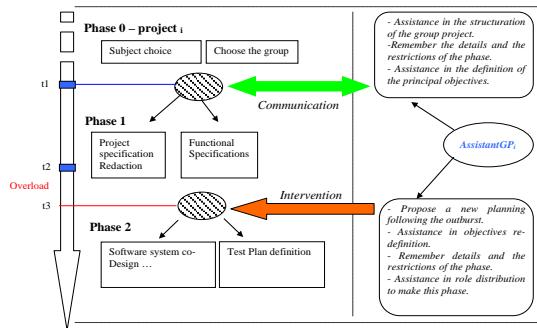


Fig. 8. $\langle aPM \rangle$ intervention following an overflow phase

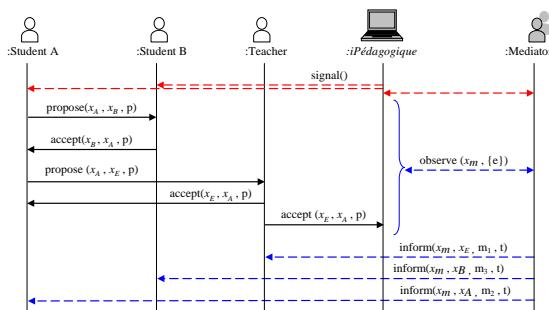


Fig. 9. Interactions between actors and Mediator during $\langle aPM \rangle$ intervention

The behaviour of these assistant agents is entirely satisfactory for this activity of assistance to project management. To further validate the model in this third level, we have recently reproduced for modelling five design roles with cognitive agents ($\langle Prescriber \rangle$, $\langle Legislator \rangle$, $\langle Designer \rangle$, $\langle Evaluator \rangle$, and $\langle Observer \rangle$), to realize a demonstrator with the aim of validating an advanced model of mechanical design [8]. Other cases of modelling have been proposed [10, 27].

3.3. Level 4 (collective agents): an Actor of Mediation for a Functional Analysis System

The main application for this level is focused on the modelling and designs of an agent-based system of cooperative mediation called *Mediator* [26, 27]. *Mediator* is an artificial actor integrated in collaborative systems. Its role is to assist users in their cooperative activities of design, project management, functional analysis, or software development. The *Mediator* is designed as a group of agents; each agent has specific skills of cooperation (communication, co-memorization, co-production, coordination, and control_process, that we called the 5-Co [20]).

The *Mediator* is part of a group of human actors during a cooperative functional analysis activity, which means it

is able to interact with them. For that it must perform cognitive tasks of observation, interpretation, decision and action.

The projection of the agents' skills in the *Mediator* activity pattern, allows to define that: the task "observe" will be performed by agents $\langle Observer \rangle$ specialized in the acquisition of cooperation information mediated by the cooperative system; the task "interpret" and "decide" will be performed by agents $\langle Knowledge \rangle$ and $\langle Control \rangle$; monitoring of cooperative activity will be performed by the agent $\langle Monitoring \rangle$; memorization by the agent $\langle Memorization \rangle$, and the task "act" relating to the cooperation communication information (communication to act), will be performed by the agent $\langle Communication \rangle$. Coordination is centralized at the agent $\langle Control \rangle$. Figure 10 shows the agent architecture of *Mediator*. The cooperative system is artificially divided into two parts to facilitate understanding of the scheme. $\langle Observer \rangle$ agents are specialized by modes of cooperation, while $\langle Communication \rangle$ agent is unique (ie, performing tasks of similar nature).

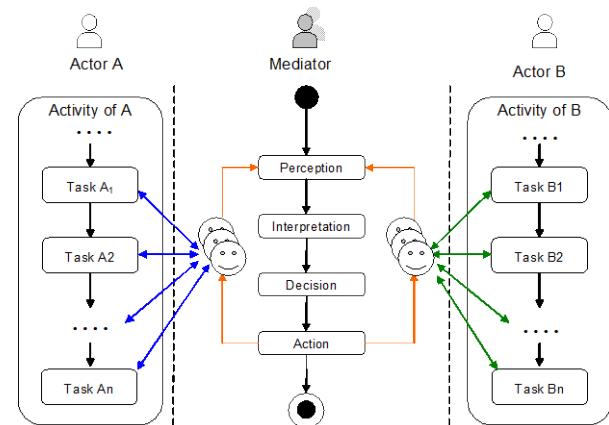


Fig. 10. Functional architecture of the Mediator

To make explicit communication between *Mediator* and designers, we have made a strong case for cooperation: there can be interaction without a minimum level of cooperation, especially in the case of a mix of human and artificial actors. Communicative interactions are of two types: macro-interactions between the designers and the *Mediator*, and micro-interactions between the agents of *Mediator*. Figure 11 shows micro-interactions between the five agents comprising *Mediator*, resulting from the observation of a designer's proposal.

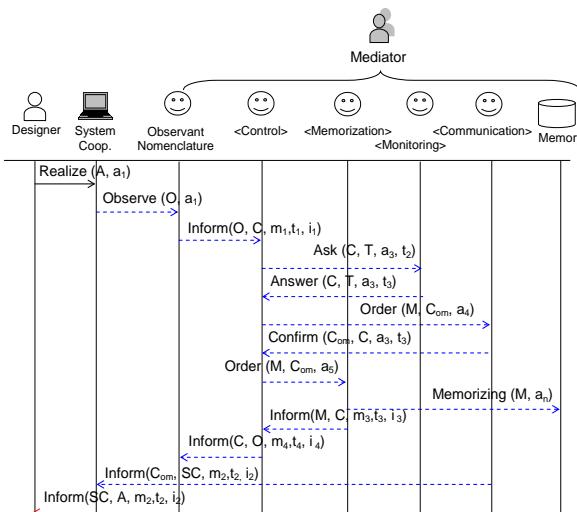


Fig. 11. Interactions between agents of the Mediator, cooperative system and memory

The cooperative behaviour of these agents, defined by the fourth level of our scale, is entirely satisfactory for this activity of cooperative mediation. To further validate the model in this fourth level, we have recently reproduced agents for modelling simulation for modelling design communities of APIC. These communities are seen as actors composed by the fuzzy cooperating agents previously described in the second level [28].

5. Conclusion and Perspectives

In this paper, we presented a generic framework for modelling agents defined for the modelling and design of complex systems (cooperative systems, assistance systems...). The correlated formal approach is to define a modular architecture for designing the various cognitive processes of agents, to respect a rigorous methodology to acquire expertise of each agent, to define their model of knowledge, and their interaction and communication patterns.

Agent activity in a complex system is very variable (from simple reflex reaction to complex cognitive decisions). So we have adapted the behaviour of agents modelled in our various applications. We then proposed a scale of behaviours which completes the scale proposed by Rasmussen. This scale consists of four levels we have presented in this article.

Recently, from the perspective of modelling communication and cooperation in CS, we have identified two levels of interaction situation: (1) micro-interaction situations, such as intra-Mediator cooperation (between agents of Mediator) or intra-community communication in APIC platform, for example; (2) macro-interaction situations, such as cooperation between the designers and the Mediator, or the communication between designers and requirement agents in APIC platform, for example. This extends our work on modelling of the interactions in complex

systems with mixed prospects for communication between software agents and humans.

We are now working on a better understanding and implementation of level changes of agent behaviour during their activities [14]. The current extension of the APIC platform to other design tasks then offers an experimental context to test the evolution of our model.

Références

- [1] B. Bauer, and J. Odell, "UML 2.0 and agents: how to build agent-based systems with the new UML standard", *Engineering Applications of Artificial Intelligence*, Vol. 18, 2005, pp. 141-157.
- [2] J. Bentahar, K. Bouzoubaa, B. Moulin, "A Computational Framework for Human/Agent Communication Using Argumentation, Implicit Information, and Social Influence", in *Proc. of the 2006 IEEE/WIC/ACM int. conf. on Web Intelligence and Intelligent Agent Technology*, 2006, p. 372-377.
- [3] F. Bergenti, M.-P. Gleizes, F. Zambonelli, "Methodologies and Software Engineering for Agent Systems", Kluwer, 2004.
- [4] P.K. Biswas, "Towards an agent-oriented approach to conceptualization", *Applied Soft Computing*, Vol. 8, No. 1, 2008, pp. 127-139.
- [5] J. Caelen, A. Xuereb, "Interaction et pragmatique. Jeux de dialogue et de langage", Hermès, 2007.
- [6] S. Card, T. Moran, A. Newell, "The Psychology of Human-Computer Interaction", Lawrence Erlbaum, New Jersey, 1983.
- [7] D. Choulier, "Towards a new theory for design activity reasoning", *1st Int. Conf. On Design Creativity, (ICDC2010)*, Kobe, Japan, November 29 – December 1, 2010.
- [8] D. Choulier, A.-J. Fougères, E. Ostrosi, "Analysis of a theory for design activity reasoning through a multiagent approach", *Proceedings of TMCE 2012*, May 7-11, 2012, Karlsruhe, Germany, to appear.
- [9] A. El Fallah-Seghrouchni, "Modèles de coordination d'agents cognitifs", in *Principes et architecture des systèmes multi-agent*, J.-P. Briot et Y. Demazeau (Dir.), p. 139-176, Hermès, 2001.
- [10] A.-J. Fougères, "Model of cognitive agents to simulate complex information systems", in *Proc. of IEEE Int. Conf. on Systems, Man and Cybernetics, (SMC'02)*, Hammamet, Tunisia, October 6-9, 2002.
- [11] A.-J. Fougères, "Agents to cooperate in distributed design process", in *Proc. of IEEE Int. Conf. on Systems, Man and Cybernetics, (SMC'04)*, pp. 2629-2634, The Hague, Netherlands, Oct. 10-13 2004.
- [12] A.-J. Fougères, "Agent-based micro-tools development for a co-operative design platform", *ITI 3rd Int. Conf. on Information and Communication Technology, (ICICT'05)*, Cairo, Egypt, December 5-6, 2005.
- [13] A.-J. Fougères, "Agent-Based μ-Tools Integrated into a Co-Design Platform", *Int. J. of Computer Science Issues*, Vol. 7, No. 3-8, 2010, pp. 1-10.
- [14] A.-J. Fougères, and E. Ostrosi, "Fuzzy agents communities for product integrated configuration", in *Proc. of the 11th Int. Conf. on Intelligent Systems Design and Applications (ISDA'11)*, November 22–24 2011, Cordoba, Spain.
- [15] A.-J. Fougères, and V.E. Ospina, "A mediation system to facilitate cooperation in a Co-design platform", in

- Proc. of the 10th Int. Conf. on Intelligent Systems Design and Applications (ISDA'10)*, pp. 664-669, November 29 – December 1 2010, Cairo, Egypt.
- [16] N. Ghasem-Aghaee, and T.I. Ören, "Towards Fuzzy Agents with Dynamic Personality for Human Behavior Simulation", in *Proc. of SCSC 2003*, Montreal, Canada, July 20-24 2003, p. 3-10.
- [17] C. Hewitt, P. Bishop, R. Steiger, "A universal modular actor formalism for artificial intelligence", in *Proceedings of the 3rd IJCAI*, 1973, p. 235-245, Standford, CA.
- [18] N.R. Jennings, K. Sycara, M. Wooldridge, "A Roadmap of Agent Research and Development", *Autonomous Agents and Multi-Agent Systems*, Vol. 1, No. 1, 1998, pp. 7-38.
- [19] N.R. Jennings, "On agent-based software engineering", *AI*, Vol. 117, No. 2, 2000, pp. 277-296.
- [20] J.-P. Micaëlli, and A.-J. Fougères, "L'Évaluation créative", Presses de l'UTBM, Belfort, 2007.
- [21] H. Mintzberg, "The structuring of Organization", Englewood Cliffs, NY, Prentice Hall, 1979.
- [22] E. Morin, "La Méthode. 1. La nature de la nature", tome I, Editions du Seuil, Paris, 1977.
- [23] U. Neisser, "Cognition and reality: principles and implications of cognitive psychology", Freeman, 1976.
- [24] A. Newell, H.A. Simon, Human problem solving, Englewood Cliffs, Prentice-Hall, 1972.
- [25] V.E. Ospina, and A.-J. Fougères, "Knowledge modelling for Mediation system based on co-operative task: an e-learning application", *ITI 3rd International Conference on Information and Communication Technology*, Cairo, Egypt, December 5-6 2005.
- [26] V.E. Ospina, and A.-J. Fougères, "The mediator: an artificial actor integrated in a cooperative design system", in *Proc. of the 5th IEEE Int. Conf. on Computational Cybernetics, (ICCC'2007)*, pp. 151-159, Gammarrath, Tunisia, October 19-21, 2007.
- [27] V.E. Ospina, and A.-J. Fougères, "Agent-based Mediation System to Facilitate Cooperation in Distributed Design", *WSEAS Transactions on Computers*, Vol. 6, No. 8, 2009, pp. 937-948.
- [28] E. Ostrosi, and A.-J. Fougères, "Optimization of Product Configuration Assisted by Fuzzy Agents", *Int. J. on Interactive Design and Manufacturing*, Vol. 15, No. 1, 2011, pp. 29-44.
- [29] E. Ostrosi, A.-J. Fougères, M. Ferney, D. Klein D., "A fuzzy configuration multi-agent approach for product family modelling in conceptual design", *Journal of Intelligent Manufacturing*, Published online: 25 May 2011, DOI: 10.1007/s10845-011-0541-5.
- [30] J. Peng, A.-J. Fougères, S. Deniaud, M. Ferney, "Dynamic Shared Context Processing in an E-Collaborative Learning Environment", *Int. J. of Computer Science Issues*, Vol. 7, No. 5, 2010, pp. 1-10.
- [31] J. Rasmussen, "Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 13, 1983, pp. 257-266.
- [32] A.S. Rao, and M.P. Georgeff M.P., "BDI agents: from theory to practice", in *Proceedings of ICMAS 95*, 1995, San Fransisco.
- [33] J.R. Searle J.R., "Speech acts", Cambridge University Press, 1969.
- [34] H.A. Simon, "The sciences of the artificial", Cambridge (MA): MIT Press, 1969.
- [35] J. Tweedale, N. Ichalkaranje et al., "Innovations in multi-agent systems", *Journal of Network and Computer Applications*, Vol. 30, No. 3, 2007, pp. 1089-1115.
- [36] G. Weiss G., "Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence", MIT Press, 1999.
- [37] M. Wooldridge, "Agent-based Software Engineering", in *IEE Proceedings on Software Engineering*, Vol. 144, No. 1, 1997, pp. 26-37.
- [38] F. Zambonelli, N.R. Jennings et al., "Organizational abstractions in the analysis and design of multi-agent systems", in *Proc. of Agent-Oriented Software Engineering*, P. Ciancarini and M. Wooldridge (Eds.), Berlin, Springer, Vol. 1957, No. 1, 2001, pp. 235-251.
- [39] X. Zhang, V. Lesser, T. Wagner, "Integrative negotiation in complex organizational agent systems ", in *Proc. of the IEEE/WIC Int. Conf. on Intelligent Agent Technology (IAT'03)*, Halifax, Canada. October 13-17 2003.

Alain-Jérôme Fougères is Computer Engineer and PhD in Artificial Intelligence from University of Technology of Compiègne (UTC). He is currently member 1) of the Laboratory of Computer Science (LIFC) of the University of Franche-Comté, where he conducts research on mobility mediation for on demand transport systems; and 2) of the Laboratory of Mecatronics3M (M3M) at University of Technology of Belfort – Montbéliard (UTBM), where he conducts research on cooperation in design. His areas of interests and scientific contributions concern: the natural language processing, the knowledge representation, the agent-based system design (architecture, interactions, communication and co-operation problems). In recent years, his research has focused on the context of co-operative work (mediation of cooperation and context sharing).

SMARTNotes : Semantic annotation system for a collaborative learning

Driss BOUZIDI¹, Rachid ELOUAHABI², Noreddine ABGOUR³

¹ Department of mathematics and computer Science, Hassan II Ain Chock University,
Faculty of Sciences, Casablanca, Morocco

² Department of mathematics and computer Science, Moulay Ismail University,
Faculty of Sciences, Meknes, Morocco

³ Department of mathematics and computer Science, Hassan II Ain Chock University,
Faculty of Sciences, Casablanca, Morocco

Abstract

Creating situations that promote collaboration between learners and/or their tutor is generally based on traditional communication tools. These are an important means to exchange ideas among learners, validate and consolidate their learning. However, we note that the large volume of messages exchanged between students and the tutor generates unwanted noise that can cause problems of disorientation for the majority of learners and cognitive overload.

Through a solution based on the use of annotations as a support for collaboration, we sought, first, to stimulate interaction and facilitate collaborative activities between learners and their tutor in an activity of understanding of a distance-learning course. Secondly, we have proposed to create connections, through automatic annotations, linking parts of the course to the most relevant messages posted in a discussion forum. The degree of relevance of these messages is based on a customized classification according to the profile of the learner and the objective of the course, as well as the integration of a semantic search done by applying a thesaurus to the LSA method.

Keywords: Annotation systems, collaborative learning, discussion forums, messages classification, Semantic Web, Latent Semantic Analysis, LSA.

1. Work context and problematic

On the basis of the logic of classical training, where the learner who wishes to understand the concepts of a course, taking full advantage of the discussions that take place in the classroom to dispel ambiguities and/or more detail on these concepts, the e-learning solutions have encased the

training content of technological tools of communication. These systems gather in one place all the necessary tools to learners and tutors to follow learning activities. These tools enable communication (e-mail, forums, chat, etc...), share resources and files (shared bookmarks, virtual libraries, etc.), and even offer distance courses (the case of videoconference session, etc.)[1].

These tools are an important means to conceal the feeling of loneliness within the learner, and encourage interaction with the tutor and peers, this allows him to have/provide support of/to other learners [2].

However, this multiplicity of tools for communication and information sharing, was not without its drawbacks, the fact that distance training often takes place in asynchronous session (the learners are not obliged to follow courses at the same time), usually when the student logs in to the online learning platform, he is faced with a large number of messages that emerged in an exponential and uncontrolled way in the communication spaces, during its disconnection [3].

Generally we see that the important number of messages exchanged between students and tutor to generate unwanted noise that is proportional to the number of participants [4]. Read all that was exchanged in these different sources of information becomes a difficult and harmful task. It can, therefore, lead to problems of disorientation for the majority of learners and induce cognitive overload.

Also searching for information in these environments is a difficult task that requires more patience, the learner finds it difficult to distinguish the types of messages and know their level of educational importance (personal exchanges, administrative exchanges, discussions on the content, group exchanges, etc.).

In addition, a need for support is often felt during the progress of the learner in understanding the course[5], the fact that the messages exchanged between learners relating to parts of the course are scattered in various communication spaces (mail, forum discussion, chat, etc.) and so set outside course document, establishing a link between the messages exchanged in these different knowledge bases and content of the source course of these exchanges, generally goes through external references to the course document. The learner finds it difficult to establish the connections between the messages produced mainly in the discussion forums and electronic mail, and the parts of its course. This tends to make the learning process much more cumbersome, inducing cognitive overload for the learner.

The learner finds himself lost in this rich mine of information but unfortunately not easy to handle, pushing him not to use it properly as a source to supplement and enrich his learning process.

In this article we present the specifications of our annotation tool called SMARTNotes to support the learner in learning activities, and we look at the problem of linking the digital course materials and exchanges produced through communication tools, particularly discussion forums. We propose an approach to automatically generate annotations on the course document leading to messages produced through the communication tools. This will allow on the one hand to remain open on the choice of communication tool within the constraints of the learning activities and also to avoid the learner the effort to link between the course document and exchanges of posted messages in the bases of these communication tools.

2. Specification of SMARTNotes annotation system

Because communication tools, find their relevance only if they are used in a context providing the best possible interactions with respect to exchanges around the course content, we propose a solution that incorporates the messages produced by learners and the tutor in this latter.

It was inspired by the practice of annotation of paper documents, which is a practice frequently used by readers to write personal comments on its margins [6] [7]. This activity allows annotators to express their views, to build a stable and written memory of the passages that they annotate [8].

The annotation on the Web offers more to the reader the opportunity to share his notes. Thus, the person consulting

an annotated document can acquire annotations attached to this document. This allows asking questions, giving advice and discussing problem solving in group, by using the annotations. The learner can have a complete and an enriched overview of his course document containing its messages (advice, questions, answers, references, etc...), and those of others attached to their corresponding parts in the document. This will save the effort to the learner to make associations between his notes (and those of his colleagues), and parts of the course sources of these interactions.

It is in this direction that we have proposed a collaborative annotation system, we have called SMARTNotes. This system aims to provide learners (and tutor) with support tools for fostering collaboration through the course document. The learner can interact, validate and enrich his course by the notes generated during the collaboration with his tutor and his peers.

2.1 Definition and objective of the annotations

The annotation is an action performed by setting a mark on an object. It is a sign of mental state that the reader gets about the annotated item [7]. In our context, the latter representing the target to which the annotation is linked, can be a collection of documents, a document, or any part thereof (paragraph, sentence, word, image ...), or even another annotation.

Studies on annotations [5][9][10], have shown that annotations made in a shared document can be used either by the annotator himself for personal use (support for active reading, customization, argumentation, etc.), or a consultant of the annotated object, be it a human (ownership, guidance and counseling, discussion and collaboration on the document, etc.) or a machine (automatic generation of abstract indexing for extensive research, tracking interactions, etc.).

2.2 Collaborative annotation systems

Several annotation systems have been developed, some addressing common issues and others specific to them. There are two types of field of use of such systems, either to index web resources to facilitate research, or to facilitate communication in an activity of understanding a document or the completion of a joint work involving several actors. Among these systems, we can mention, Annotea which is an experimental project of consortium W3C aiming to develop an environment of shared and collaborative annotation [11]. By using open standards of W3C, in particular RDF (Resource Description Framework), it arises as an objective to promote interoperability between applications that exchange annotations in the form of metadata. CoNote is also a collaborative annotation system developed at Cornell University [12]; it focuses on rights of access to annotations for a group of people who share a document. The annotations in the Yawas system are dedicated to the automatic and customized classification of annotated documentation. It shows that the classifications obtained

through annotations are more accurate than those obtained using the full document [13].

2.3 SMARTNotes collaborative annotation tool

In our SMARTNotes system, the annotator (especially a learner) interacts with the document he consults ; he is actively involved in enrichment. He is no longer seen as a simple passive reader, but he becomes more of a reader / a writer of the document; he completes it by his own understanding by annotating it as and when he progresses in reading. These notes serve as a means of locating relevant information and clarification in the form of comments.

2.3.1 The annotation representation model

One of the basic elements in the establishment of an annotation system is how to organize information about the object in the annotated document to be manipulated and implemented without any difficulty.

We were inspired by the model proposed by the W3C consortium which is an annotation as a set of metadata (attributes of annotation) and the body of the annotation (the content of the annotation). Properties induced by this model are, firstly, the opening notes to share with other systems that meet this standard and also the possibility of the scalability of this model to support new annotation types.

Metadata are defined as an RDF schema. This latter represents a set of specifications aiming to standardize the modeling of annotations to ensure interoperability between applications exchanging metadata.

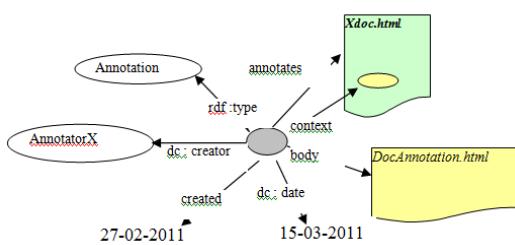


Fig. 1 RDF model of an annotation proposed by W3C

The annotation shown in the figure above was created on 27-02-2011 by the AnnotatorX, and subsequently amended on 15-03-2011. It is associated to the context located in the Xdoc.html document and its content is put into the document DocAnnotation.html.

2.3.2 Definition of semantic annotation

The formats used to annotate a document differ from one annotator to another; each makes his annotations according to his own semantics. This can lead to an overload in reading annotations; a learner cannot know the subject of an annotation (question, answer reference, etc.) unless he reads its contents. Also, the tutor has to read the

content of any posted messages if he wants to recognize the learners with learning difficulties (ask more questions) from those who are not (offer more responses).

We feel it is extremely interesting to type annotations in SMARTNotes we defined taxonomy of acts of dialogues adapted to the type of interactions that the annotator can do. It is proposed to extend the RDF model to also support the type of the annotation. We categorized all the annotations according to their purpose and their semantics. Each category is associated with specific types expressing the subject of the action of the annotator.

When an annotator (learner/tutor) decides to create an annotation, the system asks him via a semi-structured type of annotation to be set. This technique did seem to us a little heavy for the annotator, but it is most relevant in the field of communication. On the one hand, the fact of assigning a type to annotation leads the annotator to ask about the specific purpose of this [14]. On the other hand, this semantization will enormously facilitate the automatic analysis of the behavior of the annotator in his group [15].

Category	TYPE	DESCRIPTION
Highlighting	Important	Allows emphasizing the selected part to give more value compared to the other parts of the document. The annotator can explain, with a comment, the cause for which he judges that a passage Highlighting is important.
	Useless	Allows to cross out the selected part to indicate that it is useless. The annotator can explain, with a comment, the cause for which he judges that a crossed out passage is useless.
Consign	Advice	The annotator can propose advice in the form of annotation to direct another annotator. For example, a tutor (even a learner) can give advice to another learner who is facing some learning difficulties.
	Explanation	Allows to add an explanation
Conversation	Question	Allows to ask question
	Answer	Allows to answer a question
	Discussion	Allows to open a discussion forum which can be carried out between two or several annotators on an annotated object. An annotation of the discussion type is a made up

		of annotation which can include an annotation of the question type, answer, advice, explanation.
Indication	Reference	Represents a reference to another resource that can be an internal passage in the document, an annotation on the document or to an external document. This allows the reader to refer to a reading in relation to the passage annotated.
	Alert	allows the annotator to plan future temporal actions (annotation performed by the system according to a given date) and non-temporal (type of annotations to read).

Table 1. The categories and types of annotations in SMARTNotes

2.3.3 Annotation based discussion forums

We have proposed in the first version of SMARTNotes collaboration support between users through chat rooms based annotation. The annotator can post a message as a comment, question and / or response by annotating annotation. The conversational type of annotation is then presented as a discussion thread. The annotations are attached together and presented as a tree structure similar to the classic discussion forums. The advantage of this approach is that the context of the discussion-forum annotation based is defined by default by the content of the annotated part. This allows for even easier and direct exchange on the parts of course document.

3. Automatic semantic annotation of course documents

As we have previously reported, the communication tools have an important place in an e-learning platform; they have a very important role in stimulating interaction between the learners and their tutor. Each tool has its place in the learning activities according to many factors: objectives sought, characteristics of the target audience, time and technical constraints. For these reasons, we found it beneficial to create links between our SMARTNotes system and other communication tools. We thought it better to integrate on our system the ability to generate automatic annotations filed on the course document leading to the messages produced through these tools. This will allow on the one hand to remain open on the choice of communication tool according to the constraints of the learning activities and also to enable the learner avoid efforts to make the connections between the course document and exchanges of posted messages in the bases of these communication tools.

To do this, we proposed an approach based on two steps: the first allows a semantic classification of posted messages through the communication tools. This classification is based on: (1) the creation of a thesaurus based on the interests of the learner and the objectives of the course, (2) adapting the LSA method (Latent Semantic Analysis) to group the posted messages with theatics that are semantically close.

The second step of our approach is to automatically generate annotations linking parts of the course document to the messages exchanged in discussion forums in connection with these parties.

4. Semantic classification of posted messages in a discussion form

To facilitate the search of exchanged messages, the majority of communication tools use a classification based on keywords chosen by the learner; the results returned are generally independent of the conceptual intentions and areas of interests of the latter; this makes them in most cases ineffective and unintelligible. These problems are increasing on the one hand with the volume and variety of messages exchanged in discussion forums and also because of the synonymy and/or polysemy problems.

We propose to include in SMARTNotes a semantic search process delivering messages according to the most appropriate interests of the learner and the objectives of the training undertaken. This research does not require direct involvement of the learner. All relevant information to the search query are implicitly acquired from the profile and training objectives set by the author of the course.

As mentioned earlier, the important volume of messages exchanged through the communication tools often generate unwanted noise, making their reading a difficult and non practical operation. The purpose of our work is to better exploit its messages for an instant support to the learner in his activities of construction of knowledge through the course document.

4.1 Support for the learner profile in the classification of messages

The user profile is the subject of attention in several areas in particular that of education. It is represented by a set of educational and general information on the learner that are useful to establish an adaptive learning. The specifications of standard models have proposed structuring the information into categories: identification, access, relationships with others, etc.. In this article, we are interested in the IMS-LIP (Instructional Management Systems Global Learning Consortium for Learner Information package) standard, which offers a rich profile model used in most current systems of learning, and having the category "GOAL" describing personal learning objectives and aspirations of the learner.

We then propose to exploit this category to achieve an automatic classification of messages exchanged between learners, which will be adapted to the learner profile.

For a better classification of messages, we also add to the information in this category keywords associated with learning objects. These keywords are defined by the author of the course during the creation of the latter. We have adopted the model proposed by LOM (Learning Object Metadata) as a de facto standard and widely used. LOM provides the element of "Keyword" in the category "GENERAL" dedicated to store the keywords related to a learning object.

All keywords deducted of "GOAL" of the learner profile and the "Keyword" element of learning objects (the course document), is the query we will use to search for posted messages in the discussion forum.

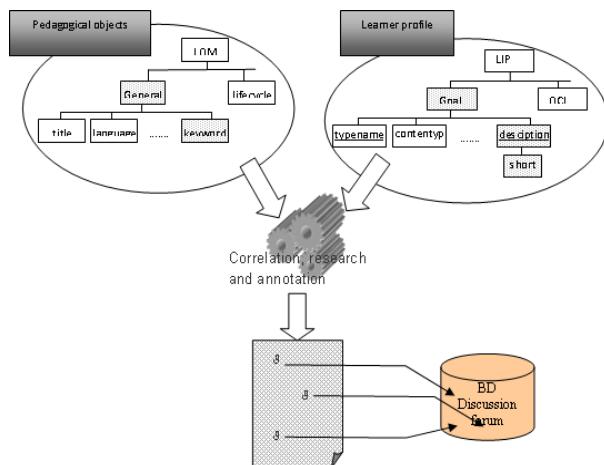


Fig. 2 Process of automatic generation of annotations

4.2 Semantic classification of messages based on LSA

Instead of searching the messages based solely on the terms set by the learner profile and keywords associated with learning objects, resulting in most cases in restrictive findings because of synonymy and polysemy problems, we propose (a) to extend this research to the terms semantically related to them by referring to a thesaurus. (b) classify messages by using the LSA method that will be completed by the measure of similarity of messages based on the application message [16].

4.2.1 Process of creating the thesaurus

The thesaurus is as an instrument of control and structuring of the vocabulary; it contributes to the consistency of indexing and facilitates the search for information to modulate the rate of recall and precision in the identification [17].

We then created a thesaurus in the area of information technology from the whole corpus of messages in different topics (e.g., system, language, technology, etc.).

To build our corpus, we did a search on each message to select the terms which include more information. As a result, we have created the semantic relationships (hierarchical, equivalence and association) between these terms.

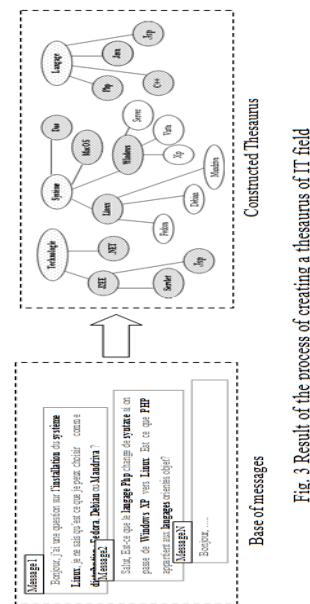


Fig. 3 Result of the process of creating a thesaurus of IT field

At the end of this process, we obtain a global organization of all the terms of our corpus according to semantic relationships that will generate the basic thesaurus for our research.

4.2.2 Classification of messages by using LSA method

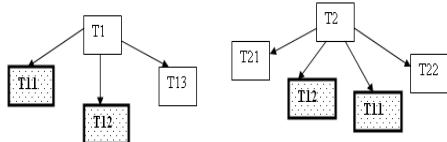
The Thesaurus, designed in the first stage, will be applied in the construction phase of the lexical table (Terms / messages) that will precede the application of the LSA method (Latent Semantic Analysis).

The LSA is a method to determine similarities between documents; a document can be a text, a paragraph or even a sentence or a word [18]. For this, each message is represented by a vector $m = (d_1, d_2, d_3, \dots, d_n)$ called lexical profile in which the j th component d_j represents the weight (or importance), the message m , the indexing term t_j associated to the i th dimension of the vector space. Then we proceed to the normalization of this matrix [19].

For successful application of the thesaurus during the construction phase of the lexical table we have adopted an approach which is to include more keywords mentioned by the user, the specific terms associated with them via the thesaurus and common to them while avoiding repetition.

T1 and T2 two terms mentioned by the user. T11, T12, T13, are terms specific to T1 and T21, T12, T11, T22-specific to T2. We note that T11, T12 are common to T1 and T2.

introduced by the user without being given in the latter [19].



The generated lexical table as follows :

	Messages
T1	
T11	Occurrences of terms of each message
T12	
T2	

Table. 2 Lexical table includes all the terms in semantic relationships with keywords

For all the messages, we take into consideration all those coming from our basic message and the query keywords that will be considered also as a message. To classify the messages we use the LSA method that will be completed by the similarity measure of basic messages with the message request. To do this, we chose the cosine which is a simple measure in terms of computation and accurate in terms of results compared, for example, to the Euclidean distance and the scalar product [20].

Following the first approach, the overall architecture of our system can be summarized as per the following diagram:

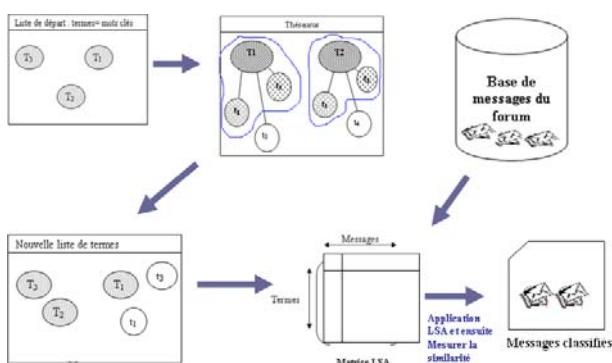


Fig. 4 General architecture of semantic classifier module

The application of this approach shows that the research carried out by applying a thesaurus to the LSA gives a more relevant result than that obtained from the application of the LSA only. The results obtained return messages whose context is one of the themes of the terms

5. The SMARTNotes architecture system

5.1 The annotation system architecture

The standard architecture of a system of annotation is primarily based on the notion of intermediary that provides the interface between the client and the Web server [13], [23]. It is responsible for processing any transaction on the annotations between client/server and consists of four complementary modules to ensure the execution of the annotation:

- The **Interceptor** is solicited at each request sent by the client browser. If the application requests the loading of a page, the interceptor sends it to the Web server, get the result, then needles is to the Composer module.
- The **composer** is responsible for include the annotations extracted from the annotations base (if any) in the requested page and returns the result to the Interceptor module. The latter returns the annotated page to the client so that it can finally be loaded by the browser.
- The **Annotation Management module** is called by the interceptor to update the database of annotations on requests from the client browser to do so.
- The **User Management module** provides the management of users and access rights to the annotations.

We have distinguished three types of implementation of the architecture of an annotation system according to the position where the intermediary is installed:

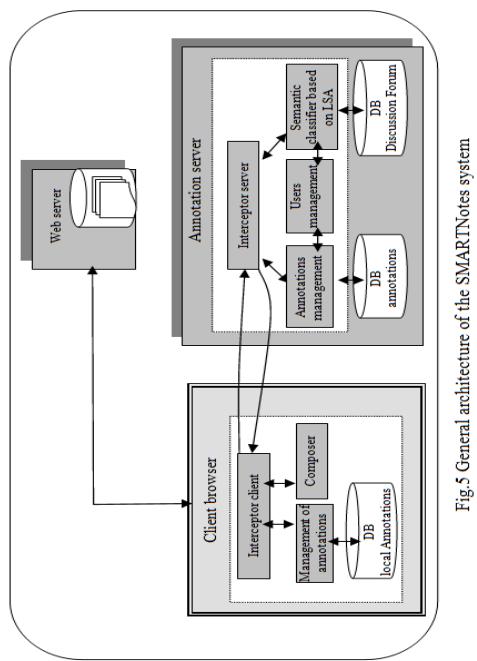
- The first architecture proposes the intermediary to the Web server. The proposed annotation features are limited to web pages published on it
- The second architecture places the intermediary on a particular server set independently of the client and Web servers. This proxy server follows the standard pattern of an annotation system; it acts as an interface between the client and Web servers and manages pages with annotations on its base [24]. Among the weaknesses of this architecture is that we mainly found the response time slow. All tasks are performed by the proxy server, the search for the requested page, the extraction of annotations associated with it from the base, the integration of these annotations on the page, and return the response to the browser applicant, make of this server a major choke point.
- In the third architecture the intermediary is put near the client, as a plug-in or external application (eg an applet). The aim is to enrich the browser functions to manipulate the annotations of a web page. This architecture has more advantages than the previous two: ability to annotate web documents stored locally, more flexibility and control of interactions of the annotator and avoiding the bottleneck at the server. However, opposed to this, it remains dependent on the

type of browser used and the possibilities of sharing annotations and collaboration on web documents are painfully managed.

5.2 Main components of SMARTNotes system

In SMARTNotes, we propose an architecture based on an intermediary divided between the client and server. The first part is set at the browser as an extension, allowing the user to annotate a document, when loaded by the browser, without using the annotation server. The interceptor module provides communication between the client and the server; you can either retrieve the annotations associated with the document loaded from the server or to transmit the updates made. The composer module is also placed on the browser; it is based particularly on the DOM and XPointer technologies to include annotations in the web page being loaded.

The second part of the intermediary server is placed on the annotation to intercept client requests and manage the annotations present in the base. The semantic classification module of the messages posted in the discussion forum enables to deliver the most appropriate messages according to the interests of the learner and the objectives of the training undertaken.



This architecture has several advantages, namely:

- Respect of confidentiality: providing the ability to save annotations locally with the client.
- The off-line annotation of documents: the integration of the intermediate level clients can annotate a document (local or remote after the download on the browser) without having to connect to the server annotation. The annotator can make a backup of annotations locally or export batch annotation on the

server; this will also avoid the problem of the bottleneck at the server level annotation.

- Sharing of annotations and collaboration on the document in a simple way: they are based on export operations and synchronization between the server and annotation client. Each time the server intercepts an update of a page, it sends a message to the client connected to the server and logging on the same page. The automatic refresh of the page is done on the client terminal only after confirmation of the user warned.
- Response time very small: the set up of the composer and the manager of annotations at the level of client allows on the one hand to manipulate the annotations of pages with very low response time. Secondly, reduce the frequency of use of the annotation server as backup and search annotation on its base. And therefore also have long response times on the server annotation.

5.3 Processus of Integration of annotations in the course document

In SMARTNotes, communication between the client and the server is based on an exchange annotation XML. We have defined a database link associating the content of the annotation, which is stored in the annotations or discussion forum, and the corresponding part of the course. This solution enables the reader to view the document associated with annotated annotations in a transparent manner and unchanged the source document.

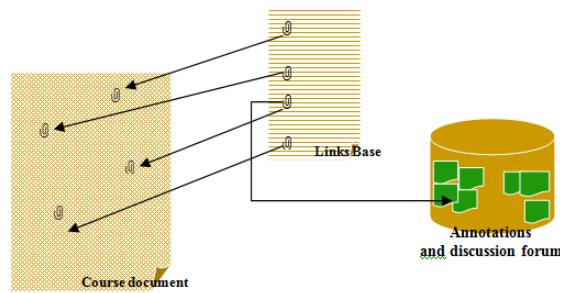


Fig.6 Base of links joining the basis of exchange information to the course parts

The connecting links between the base of exchange information and to the document to be annotated are put into a database of links regardless of the course document which solves the problem of copyright of the course. Thanks to the XLink technology which offers new mechanisms making hypermedia documents more flexible, we defined this XML document called "basic links" without text links that are totally separated from the source document.

6. Conclusion and perspectives

In this work we have presented barriers that a learner may face in the misuse of traditional communication tools in an activity of understanding of a distance-course. In particular, the large volume of messages posted in these

areas can lead to disorientation for the majority of learners and cognitive overload. We have proposed to put the hypermedia course documents in a broader dimension to centralize much more interaction and exchange of ideas for successful learning. Through a solution based on the use of annotations as a medium for collaboration, we sought, first, to stimulate interaction and facilitate collaborative activities between learners and their tutor in an activity of understanding of a course. Secondly, we have proposed to create connections, through automatic annotations linking parts of the course to the most relevant messages posted in a discussion forum. The level of relevance of these messages is based on a customized classification according to the profile of the learner and the objective of the course as well as the integration of a semantic search done by applying a thesaurus to the LSA method.

The design of the functional architecture of our annotation tool has been defined; it meets a number of requirements such as independence from the platform hardware / software, interoperability, scalability to support other types of annotations. Also, in order to verify the feasibility of our proposal based mainly on XML technology, unit tests validation were successfully completed. The application of our approach to semantic classification on a corpus of messages posted through a discussion forum of the MOODLE platform has shown relevant results with the LSA method.

We are now in a period of experimentation with our system and we plan to include a recommendation system that draws in the documents annotated by learners and tutors, and automatically offers educational resources for active learners without having to explicitly request their feedback.

References

- [1] N. ABGOUR, "Schéma d'autorisation pour les applications réparties sur Internet", Thèse de l'Institut National Polytechnique, Toulouse, Rapport LAAS, 2004
- [2] Henri, F. & Lundgren-Cayrol, K « Apprentissage collaboratif à distance, téléconférence et télédiscussion.» Rapport interne n°3.Montréal, Canada; Centre de recherche LICEF, 1996.
- [3] D.Bouzidi , R.Ajhoun, A.Benkiran, "The use of the annotations in a collaborative activity of reading an online hypermedia course document". IEEE 1-4244-0406-1, 2006
- [4] Marc Walckiers and Thomas De Praetere « L'apprentissage collaboratif en ligne, huit avantages qui en font un must ». Vol.2 Distances et savoirs 2004, ISSN 1765-0887.
- [5] R.Elouahbi A.Nassir and S.Benhlima, "Design and implementation of adaptive sequencing based on graphs", 11th International Educational Technology Conference, Turkey, 2011, p 233-238.
- [6] C. Marshall, "Annotation: from paper books to the digital library ", In Proceedings of DL '97, 1997, pp.131-140.
- [7] M.Veron « Modélisation de la composante annotative dans les documents électroniques » Rapport de stage du DEA Représentation des Connaissances et Formalisation du Raisonnement, UPS-IRIT, Toulouse, Juin 1997.
- [8] D.Bouzidi, « Collaboration dans les cours hypermédiis du système de télé-enseignement SMART-learning », Thèse de Doctorat, Spécialité Informatique, Ecole Mohammadia d'Ingénieurs (EMI), Université Mohammed V-Agdal, Rabat, 2004.
- [9] JJ Cadiz, A.Gupta, J. Grudin "Using Web Annotations for Asynchronous Collaboration Around Documents", Conference on Computer Supported Cooperative Work (CSCW2000), Philadelphie, Pennsylvanie, USA, 2000.
- [10] L. Denoue, « De la création à la capitalisation des annotations dans un espace personnel d'informations ». Thèse de doctorat (PhD), Université de Savoie, Octobre 2000.
- [11] M.R Koivunen, "Annotea : Applying Semantic Web Technologies to Annotations", in "Semantic Web Kick-Off en Finlande, Vision, Technology, Research and Applications", HIT Publication, Hyvonen Eero (ed), p.213,226, 2002
- [12] J.Davis and D.Huttenlocher, "Shared Annotation for Cooperative Learning". Proceedings of the Conference on Computer Supported Cooperative Learning, 1995
- [13] L.Denoue & L.Vignollet, "Yawas un outil d'annotation pour les navigateurs du Web", IHM'99Montpellier, France, 1999.
- [14] G.Sébastien, "Apprentissage collectif à distance. SPLACH : un environnement informatique support d'une pédagogie de projet." Thèse de Doctorat, Spécialité Informatique, Université du Maine, 2001
- [15] D.BOUZIDI, R.AJHOUN, A.BENKIRAN, "Les annotations pour un apprentissage collaboratif", JOCAIR'2006, colloque International, France 2006 pp.372-382.
- [16] S.LGARCH, D.BOUZIDI, S.BENNANI, "Une approche sémantique de classification de messages d'un forum de discussion basée sur la méthode LSA", WNGN Workshop on Next Generation Network, pages : 3-8, Maroc 2008, pages 3-8.
- [17] L.Saadani et S.Bertrand-Gastaldy, "Cartes conceptuelles et thésaurus : essai de comparaison entre deux modèles de représentation issus de différentes traditions disciplinaires ",2000
- [18] Y.Bestgen, « Analyse sémantique latente et segmentation automatique des textes », 2004
- [19] G. Salton, « Autoamtic Text Processing – The Transformation Analysis and Retrieval in Full Text Information Systems», 1989
- [20] H.Zargayouma « Indexation sémantique de documents XML » Thèse de doctorat, 2005
- [21] S. LGARCH, "Une approche sémantique de classification de messages d'un forum de discussion basée sur la méthode LSA", Rapport de DESA, Mohammadia Engineering School, Mohamed V-Agdal University, Rabat, Morocco, 2008.
- [22] V.Vasudevan and M. Palmer, « On web annotations : Promises and pitfalls of current web infrastructure». Actes de 32nd Hawaii International Conference on System Sciences, Maui, Hawaii, 1999
- [23] I.Ovsianikov, M. Arbib, et T. McNeill, « Annotation technology». Int. J. Human-Computer Studies, 50 :329–362
- [24] E.Desmontils, C.Jacquin & L.Simon, « Vers un système d'annotation distribué », Rapport de recherche n°03-01, Nantes, Institut de Recherche en Informatique Université de Nantes, 2003, 16 p.

Driss BOUZIDI is an Assistance Professor in Computer Science at Hassan II Ain Chock University, Faculty of Sciences Casablanca, Morocco. He received a Ph.D. degree in computer engineering from Mohammed V University in 2004 on "Collaboration in hypermedia courses of e-learning system SMART-Learning". Dr. Bouzidi's research interests are in Distributed multimedia Applications, Collaboration systems, and e-learning. He was vice-chair of the international conference NGNS'09 (Next Generation Network and Services) and the treasurer of the two research associations e-NGN and APRIMT.

Rachid ELOUAHBI received his doctorate degree in Computer Science from Mohammadia School of Engineering Morocco in 2005. He is currently an Assistant Professor at the University of Moulay Ismail, Meknès. His research focused on the area of adaptive learning systems, course sequencing and learning technologies. He works on the project Graduate's Insertion and Assessment as tools for Moroccan Higher Education Governance and Management in partnership with the European Union.

Noreddine ABGOUR is currently an Assistant Professor at the Faculty of Sciences Ain Chok, Casablanca, Department of Mathematics and Computer Science. He is also a PhD in Computer Science, area of distributed applications, from The Polytechnic Institute of Toulouse. In addition to teaching, his research interests are in the field of security and collaborative systems.

An Ensemble Method for Validation of Cluster Analysis

Sunghae Jun

Department of Bioinformatics and Statistics, Cheongju University
 Cheongju, Chungbuk 360-764, Korea

Abstract

Clustering is more subjective work than classification and regression. Though classification and regression have many general validation measures, clustering has few validation measures. Also, it is difficult to develop general measure of cluster validation. So, many evaluation measures have been published for cluster validation. In this paper, we propose an ensemble method of validation for cluster analysis. We use voting approach to some validation measures of cluster analysis. To verify our improved performance, we make experiments by some objective data sets from UCI machine learning repository.

Keywords: Cluster Analysis, Cluster Validation, Ensemble Method, Voting, Internal measures, Stability measures.

1. Introduction

Clustering is a rudimentary and exploratory approach to start of multivariate data analysis [1]. Also, clustering is grouping data points into clusters and then the points within a cluster are highly similar to each other [2]. That is, the variance of the points within same cluster is small and the variance of the points in other clusters is large. Recently, clustering was used in diverse fields of bio data analysis such as bioinformatics [3-10]. But, unlike classification and regression, there is no general validation measure in the clustering. Many validation measures of the clustering results have been published in diverse fields [3],[11-15]. They gave improved performance on case by case. That is, we should select a cluster validation measure for given data set. So, in this paper, we propose an ensemble method as a general evaluation for cluster validation. We use many measures such as internal and stability validations in our proposed method. All measures used in our method will be voted to validate the clustering result. We will make experiment to verify our improved performance. Three data sets from UCI machine learning repository will be used in our experiment.

2. Cluster Validation using Ensemble Method

We found various measures for cluster validation [16-18]. These measures were based on internal and external properties of clustering results [19-22]. In this paper, we divide validation measures to two types, internal and stable.

First, internal validation measures reflect the compactness, connectedness, and separation of the cluster partitions [3]. From previous researches, we knew that it is difficult to evaluate clustering result by a validation measure. In this paper, we propose an ensemble method of validation for cluster analysis. We vote the results of some validation measures for efficient cluster validation. In our study, we use three internal validation measures which are connectivity, Silhouette width, and Dunn index. Let N and P denote the numbers of observations and variables respectively. We can define the connectivity as follow [3].

$$Connect(C) = \sum_{i=1}^N \sum_{j=1}^R x_{i,n_{ij}} \quad (1)$$

Where n_{ij} is defined as the j th nearest neighbor of observation i . C has k disjoint clusters, (C_1, \dots, C_k) . R is used as the number of nearest neighbor. Also, we can validate the clustering result as the following.

$$x_{i,n_{ij}} = \begin{cases} 0 & \text{if } i \text{ and } j \text{ are in same cluster} \\ \frac{1}{j} & \text{otherwise} \end{cases} \quad (2)$$

This value is ranges between 0 and ∞ . The cluster validation is better by minimizing the connectivity value. Next, Silhouette width is defined as the average value of all observations' Silhouette values [3],[11]. This has the value between -1 and 1. The cluster validation is good when the value is close to 1 and is bad when the value is close to -1. The Silhouette width of observation i is defined as follow [11].

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (3)$$

Where $a(i)$ is the distance measure as follow.

$$a(i) = d(i, A) \quad (4)$$

This is defined as the average distance between i and all other observations in cluster A . Also, $b(i)$ is the average

distance of i to the observations in the nearest neighbor cluster. In this paper, we use Dunn index as third internal measure. This index is a ratio as follow [3].

$$Dunn_{index} = \frac{\min(O)}{\max(I)} \quad (5)$$

Where O is the distance between observations not in the same cluster and I is the intra-cluster distance.

Next stability measures validate the clustering results using all instances of given data columns [3]. These measures are average proportion of non-overlap (APN), average distance (AD), average distance between means (ADM), and figure of merit (FOM) [3],[12-13]. So, we validate the results from clustering using internal and stability measures. Next table shows the criteria of all validation measures in this paper.

Table 1: Criteria of validation measures

<i>Validation measures</i>		<i>Range of value</i>	<i>Criteria of well clustering</i>
Internal	Connectivity	(0, ∞)	Minimized
	Silhouette	(-1,1)	Near 1
	Dunn index	(0, ∞)	Maximized
Stability	APN	(0, 1)	Close to 0
	AD	(0, ∞)	Minimized
	ADM	(0, ∞)	Minimized
	FOM	(0, ∞)	Minimized

Using these criteria for the clustering results, we can evaluate the clustering results as following figure.

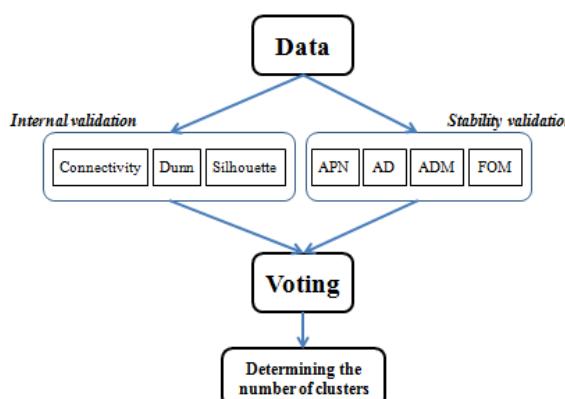


Fig. 1 Proposed ensemble method of cluster validation.

In this paper, we vote the clustering results from internal and stability validations.

3. Experimental Result

To verify improved performance of our research, we made experiments using objective data sets from UCI machine learning repository [23]. We used ‘Abalone’, ‘Glass identification’, and ‘Yeast’ in our experiment. Next table shows the numbers of instances and variables of the data sets.

Table 2: Data sets from UCI machine learning repository

<i>Data set</i>	<i>Number of instances</i>	<i>Number of variables</i>
Abalone	4177	8
Glass	214	9
Yeast	1484	8

These were used for evaluating performance of the clustering results by internal and stability validations. Next table shows the internal validation result of Abalone data set.

Table 3: Internal validation: Abalone

<i>Clustering algorithm</i>	<i>Internal validation</i>	<i>Number of clusters</i>	<i>Validation value</i>
hierarchical	Connectivity	2	3.8286
	Dunn	9	0.2233
	Silhouette	2	0.6268
K-means	Connectivity	2	1.0552
	Dunn	9	0.3994
	Silhouette	2	0.5537
PAM	Connectivity	2	0.0000
	Dunn	9	0.1648
	Silhouette	3	0.5288
CLARA	Connectivity	2	0.0000
	Dunn	9	0.0203
	Silhouette	10	0.5179

The occurred ratio of the number of clusters was 2 with 50%. So, using internal validation, we decided the number of clusters to 2 in Abalone data set. Another clustering result of Abalone data set is shown in the following table.

Table 4: Stability validation: Abalone

<i>Clustering algorithm</i>	<i>Stability validation</i>	<i>Number of clusters</i>	<i>Validation value</i>
hierarchical	APN	2	0.0054
	AD	10	1.0658

	ADM	10	0.2531
	FOM	10	0.4465
K-means	APN	2	0.0707
	AD	10	1.0526
	ADM	4	0.2529
	FOM	10	0.4402
PAM	APN	2	0.0486
	AD	10	0.8956
	ADM	2	0.2176
	FOM	7	0.4353
CLARA	APN	2	0.0488
	AD	10	0.9560
	ADM	2	0.2220
	FOM	10	0.4331

This result was based on stability validation. We selected 10 as a proper number of clusters of Abalone data set. This result was different from the result of internal validation. So, we voted two clustering results as follow.

Table 5: Ensemble result of Abalone data

# of clusters	Frequency	Ratio (%)
2	12	42.86
3	1	3.57
4	1	3.57
7	1	3.57
9	4	14.29
10	9	32.14

From the ensemble result of internal and stability validation, we found that 2 was the number of clusters for Abalone data set. Next we made another experiment using Glass identification data for verifying our research.

Table 6: Internal validation: Glass identification

Clustering algorithm	Internal validation	Number of clusters	Validation value
hierarchical	Connectivity	2	3.8579
	Dunn	2	0.4934
	Silhouette	2	0.6343
K-means	Connectivity	2	8.8643
	Dunn	2	0.1904
	Silhouette	4	0.5879
PAM	Connectivity	2	21.2698

	Dunn	3	0.1244
	Silhouette	3	0.5822
CLARA	Connectivity	2	17.8778
	Dunn	4	0.1633
	Silhouette	4	0.5883

We decided 2 as the number of clusters for Glass identification using internal validation. The frequency ratio of 2 was 58.33%. Next table shows the clustering result of stability validation for Glass identification.

Table 7: Stability validation: Glass identification

Clustering algorithm	Stability validation	Number of clusters	Validation value
hierarchical	APN	2	0.0010
	AD	10	1.6310
	ADM	2	0.0186
	FOM	10	0.4963
K-means	APN	4	0.0368
	AD	10	1.3887
	ADM	4	0.2868
	FOM	10	0.4763
PAM	APN	2	0.0523
	AD	10	1.2717
	ADM	4	0.1987
	FOM	10	0.4187
CLARA	APN	3	0.0351
	AD	10	1.3614
	ADM	3	0.1555
	FOM	10	0.4284

In the result of stability validation, we selected 10 as the proper number of clusters. We got the ensemble result from the internal and stability validations in the following table.

Table 8: Ensemble result of Glass identification data

# of clusters	Frequency	Ratio (%)
2	10	35.71
3	4	14.29
4	6	21.43
10	8	28.57

So, we found that the optimal number of clusters was 2 with the frequency ratio of 35.71%. Last experimental data

set was Yeast in the paper. Next table shows the clustering result based on the internal validation measures.

Table 9: Internal validation: Yeast

Clustering algorithm	Internal validation	Number of clusters	Validation value
hierarchical	Connectivity	2	2.9290
	Dunn	2	0.5038
	Silhouette	2	0.6229
K-means	Connectivity	2	82.6579
	Dunn	7	0.0782
	Silhouette	4	0.2864
PAM	Connectivity	2	60.1119
	Dunn	6	0.0465
	Silhouette	2	0.3068
CLARA	Connectivity	2	60.3103
	Dunn	3	0.0560
	Silhouette	2	0.2971

We decided the number of clusters to 2 with 66.67% using internal measures. Next table shows the clustering result of Yeast data set by stability validation measures.

Table 10: Stability validation: Yeast

Clustering algorithm	Stability validation	Number of clusters	Validation value
hierarchical	APN	2	0.0000
	AD	10	0.3101
	ADM	2	0.0000
	FOM	10	0.0865
K-means	APN	6	0.0723
	AD	10	0.2495
	ADM	2	0.0320
	FOM	10	0.0850
PAM	APN	2	0.1443
	AD	10	0.2466
	ADM	3	0.0416
	FOM	10	0.0858
CLARA	APN	2	0.1256
	AD	8	0.2598
	ADM	2	0.0427
	FOM	10	0.0862

The number of clusters was decided to 10 with the frequency ratio of 43.75% using stability validation

measures. So, we voted the results for our ensemble method. Next table shows the ensemble result for determining the number of clusters for Yeast data set.

Table 11: Ensemble result of Yeast data

# of clusters	Frequency	Ratio (%)
2	14	50.00
3	2	7.14
4	1	3.57
6	2	7.14
7	1	3.57
8	1	3.57
10	7	25.00

From the ensemble result, we decided that the number of clusters for Yeast data set was 2 with the frequency ratio of 50.00%.

4. Conclusions

We proposed an ensemble method for cluster validation. In this paper, we used the internal and stability measures for validating the cluster result. Also, we voted the clustering results from internal and stability validations. To verify the performance of proposed method, we made experiment using some objective data sets from UCI machine learning repository. This paper tried to find the number of clusters objectively using the ensemble approach by voting the clustering results. But, we need more general ensemble method for cluster validation. This is our future work.

References

- [1] R. A. Johnson, and D. W. Wichern, Applied Multivariate Statistical Analysis, Prentice Hall, 1992.
- [2] J. Han, and M. Kamber, Data Mining Concept and Techniques, Morgan Kaufmann, 2001.
- [3] G. Brock, V. Pihur, S. Datta, and S. Datta, “clValid: An R Package for Cluster Validation”, Journal of Statistical Software, Vol. 25, Iss. 4, 2008, pp. 1-20.
- [4] J. L. DeRisi, V. R. Iyer, and P. O. Brown, “Exploring the metabolic and genetic control of gene expression on a genomic scale”, Science, Vol. 278, No., 5338, 1997, pp. 680-686.
- [5] M. B. Eisen, P. T. Spellman, P. O. Brown, and D. Botstein, “Cluster analysis and display of genome-wide expression patterns”, Proc. Natl. Acad. Sci. USA, Vol. 95, No. 25, 1998, pp. 14863-14868.
- [6] V. Bhattacherjee, P. Mukhopadhyay, S. Singh, C. Johnson, J. T. Philipose, C. P. Warner, R. M. Greene, and M. M. Pisano, “Neural crest and mesoderm lineage-dependent gene expression in orofacial development”, Differentiation, 2007.

- [7] G. J. McLachlan, R. W. Bean, and D. Peel, "A mixture model-based approach to the clustering of microarray expression data", *Bioinformatics*, Vol. 18, No. 3, 2002, pp. 413-422.
- [8] D. Dembele and P. Kastner, "Fuzzy C-means method for clustering microarray data", *Bioinformatics*, Vol. 19, No. 8, 2003, pp. 973-980.
- [9] L. Fu and E. Medico. FLAME, "A novel fuzzy clustering method for the analysis of DNA microarray data", *BMC Bioinformatics*, Vol. 8, No. 3, 2007, pp. 1-15.
- [10] S. Sarmah, and D. K. Bhattacharyya, "An Effective Technique for Clustering Incremental Gene Expression data", *International Journal of Computer Science Issues*, Vol. 7, Iss. 3, No. 3, 2010, pp. 31-41.
- [11] P. J. Rousseeuw, "Silhouette: a graphical aid to the interpretation and validation of cluster analysis", *Journal of Computational and Applied Mathematics*, Vol. 20, 1987, pp. 53-65.
- [12] S. Datta and S. Datta, "Comparisons and validation of statistical clustering techniques for microarray gene expression data", *Bioinformatics*, Vol. 19, No. 4, 2003, pp. 459-466.
- [13] K. Y. Yeung, D. R. Haynor, and W. L. Ruzzo, "Validating clustering for gene expression data", *Bioinformatics*, Vol. 17, No. 4, 2001, pp. 309-318.
- [14] Y. Abofathy, and B. Zarei, "A New Approach for Optimal Clustering of Distributed Programs Call Flow Graph", *International Journal of Computer Science Issues*, Vol. 7, Iss. 4, No. 3, 2010, pp. 37-43.
- [15] P. Agarwal, M. A. Alam, and R. Biswas, "Issues, Challenges and Tools of Clustering Algorithms", *International Journal of Computer Science Issues*, Vol. 8, Iss. 3, No. 2, 2011, pp. 523-528.
- [16] M. K. Kerr and G. A. Churchill, "Bootstrapping cluster analysis: assessing the reliability of conclusions from microarray experiments", *Proc. Natl. Acad. Sci. USA*, Vol. 98, No. 16, 2001, pp. 8961-8965.
- [17] K. Y. Yeung, D. R. Haynor, and W. L. Ruzzo, "Validating clustering for gene expression data", *Bioinformatics*, Vol. 17, No. 4, 2001, pp. 309-318.
- [18] S. Datta and S. Datta, "Comparisons and validation of statistical clustering techniques for microarray gene expression data", *Bioinformatics*, Vol. 19, No. 4, 2003, pp. 459-466.
- [19] F. D. Gibbons and F. P. Roth, "Judging the quality of gene expression-based clustering methods using gene annotation", *Genome Res.*, Vol. 12, No. 10, 2002, pp. 1574-1581.
- [20] I. Gat-Viks, R. Sharan, and R. Shamir, "Scoring clustering solutions by their biological relevance", *Bioinformatics*, Vol. 19, No. 18, 2003, pp. 2381-2389.
- [21] N. Bolshakova, F. Azuaje, and P. Cunningham, "A knowledge-driven approach to cluster validity assessment", *Bioinformatics*, Vol. 21, No. 10, 2005, pp. 2546-2547.
- [22] S. Datta and S. Datta, "Methods for evaluating clustering algorithms for gene expression data using a reference set of functional classes", *BMC Bioinformatics*, Vol. 7, 2006, pp. 397.
- [23] UCI machine learning repository, archive.ics.uci.edu/ml.

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 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.

An Efficient Method for Mining Event-Related Potential Patterns

Seyed Aliakbar Mousavi¹, Muhammad Rafie Hj Arshad², Hasimah Hj Mohamed³ and Saleh Ali Alomari⁴

^{1, 2,3,4} School of Computer Sciences, Universiti Sains Malaysia
Pulau Penang 11800, Malaysia

Abstract

In the present paper, we propose a Neuroelectromagnetic Ontology Framework (NOF) for mining Event-related Potentials (ERP) patterns as well as the process. The aim for this research is to develop an infrastructure for mining, analysis and sharing the ERP domain ontologies. The outcome of this research is a Neuroelectromagnetic knowledge-based system. The framework has 5 stages: 1) Data pre-processing and preparation; 2) Data mining application; 3) Rule Comparison and Evaluation; 4) Association rules Post-processing 5) Domain Ontologies. In 5th stage a new set of hidden rules can be discovered base on comparing association rules by domain ontologies and expert rules.

Keywords: EEG, MEG, Event-related Potentials (ERP), Data Mining, ontology.

1. Introduction

The Electroencephalographies (EEG) [1] are electric signals and Magnetoencephalographys (MEG) [2] are magnetic fields produced by electrical currents in cortex. MEG and EEG are used to record brain activities. There are other methods of imaging brain signals such as functional Magnetic Resonance (fMRI) and Imaging and Position Emission Tomography (PET). In comparison EEG/MEG method have two main advantages; direct measure of neuronal activity and best temporal resolution by milliseconds. Unlike fMRI and PET, all sensing and processing will take place within few hundred milliseconds so it gives a very good representation of time course of brain activity to control EEG signals in time. Fig.1 is an example of EEG recorded data that channels are shown vertically and time courses horizontally. EEG patterns are recorded using a set of sensors called Net which is put all around head and well connected to scalp. Therefore the electrical activities close to sensors (channels) are captured even though they are very small. Fig.2 is an example of MEG data.

Fig.1. EEG recorded data that shows the channels vertically, time courses horizontally and events on top

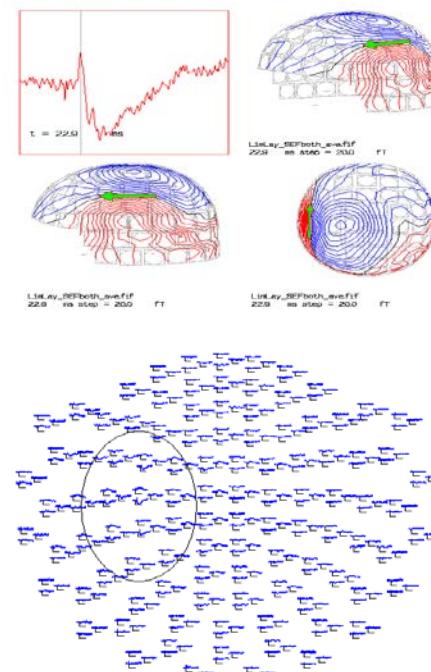
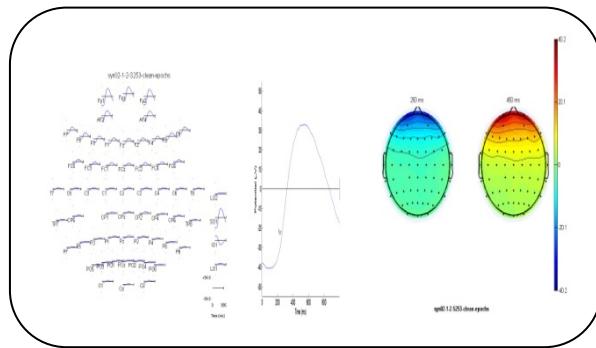


Fig.2 is MEG data that is mapped on scalp.

As shown in Figure 3 the event-related potentials (ERP) [3] are physiological patterns created by averaging segments of EEG/MEG, time-locking to event of interest (e.g., onset of a visual or auditory stimulus) such as internal and external stimulus or response onset. This notion has been generalized under the term event-

related responses to include triggered events. In the latter case the recorded signal portrays the neural activity related with cognitive processes like performing a given memory task or during decision making. In both cases the target activity is accompanied by (and occasionally modulated by or even interrelated with) the ubiquitous spontaneous activity of the brain [12]. Signals that are not event-related tend towards zero as the number of averaged trials in experiment increase. In this way, averaging increases the signal-to-noise ratio (SNR) and provides measures of electrical activities that are specifically linked to stimulus processing [11]. ERP is spatial, overlapping, and temporal, so it's difficult to extract the features by normal tools and methods. In the other hand, the data contains a lot of noise (e.g. Artifacts) and other brain activities which superpose the target signals. ERP provides the measures of brain activities that are related to stimulus processing. The ERP is characterized by the time course, polarity and scalp's topography. It might be early or late in the time respond, either positive or negative and distributed over scalp topography.



In Fig.3 there is an example of ERP with its properties from EEGLab. [7] The "P300" component (Signal) which was extracted from the superposed (Overlapped) data using Independent Components Analysis (ICA) has a peak latency of approximately 300 to 500ms and is positive over frontal areas of the scalp.

2.0 Related Research

Human Brain is continuously dispersing huge amount of electromagnetic waves. Brain's electromagnetic fields are actual carriers of conscious experience [17]. Brain waves are unprocessed data which contains knowledge about brain functions. They act as input in knowledge discovery process and data mining. The brain wave patterns are important in computational neuroscience and neuro-Informatic research. The reason those patterns are retrieved is because of usefulness in recognizing brain disorders; such as by identifying the type of seizure occurring in the brain, and in diagnosing epilepsy. We can also examine the

problem of dementia which is difficult to diagnose by ordinary clinical diagnosis. Furthermore, those patterns provide a practical method of identifying patients who suffer from mental health problems or are in a comatose condition. In the clinical field, EEG is applied in the neurophysiological area and also being studied in an attempt to provide a new way of communication between the human brain and computers by reading the signal extracted from the brain via the computer interface. The extracted knowledge from EEG and MEG patterns leads neuroscientist and neurosurgeons to find the patterns associated with Cortex in-functionality, Neuron illness, brain capability analysis, [4] and responses to outside stimulus in different situations. The knowledge about brain patterns can be used in commercial marketing and decision making known as Neural-Economy to increase sale and find the profitable customers.

Researches in Computational Neuroscience are depended on data and in-order to process data, there is an increasing need for advance tools and concepts. The concept of building a framework to do knowledge extraction and representation has been proposed by Dou, D., Frishkoff, G., Rong, J., Frank, R., Malony, A., & Tucker, D. in 2007 as a project called Neuro Electro Magnetic Ontologies (NEMO). In NEMO project the EEG brain waves are fed to data mining process to extracted rules and domain ontology. NEMO project along with previous researches have built frameworks and determined methods. However most of those researches have same problem which is wrong classification of ERP patterns. Those patterns split across wrong classes. Dou, D suggested a further refinement of ERP during decomposition. The reason is that ERP is spatial, temporal and with high dimension at each time point. Many parts of the brain are simultaneously active and that would overlap the other patterns to target signals which would cause uncertainty in data mining progress. He also advised to employ spatial and dense merits that capture temporal and spatial attributes more accurately, and it may reduce the misallocation of patterns variance. He suggested evaluating the machine generated rules with "Gold Standards" which has been established before by experts in ERP domain.

The main challenge which is mentioned in most of ERP researches is the establishment of robust computational analysis methods such as data decomposition (Pre-processing) and data mining to correctly interpret (classifying) ERP patterns and relate them to specific brain functions [5]. M. Sabeti, S.D. Katebi and R. Boostani [19] suggested a new approach for classification of EEG signals in schizophrenic

patients. They used autoregressive model parameters, band power and fractal dimension features for classification. They employ both linear discriminant analysis (LDA) and adaptive boosting (Adaboost) [19] to classify the reduced feature set. They attained a classification with high accuracy by LDA and Adaboost, respectively. M. Sabeti and R. Boostani [19] have also tried to determine the more informative channels in EEG experiments of 20 schizophrenic patients. Gwen A. Frishkoff, RobertM. Frank, Jiawei Rong [9], suggested an automatic classification and labeling of brain electromagnetic patterns. They designed an ontology-base system to classify EEG ERP patterns. They focus on specification of rules and concepts that capture expert knowledge of ERP patterns in word recognition. Their contribution is implementation of rules in an automated data processing and labeling stream along with data mining techniques that combine top-down (knowledge-driven) with bottom-up (data-driven) method to refine the ERP machine generated rules. C.Davatzikos, K. Ruparel, Y. Fan and D.G. Shen [22] used data mining machine learning to classify spatial fMRI patterns in application of lie detection. Classified, high-dimensional and non-linear pattern in functional magnetic resonance (fMRI) images is used to discover the spatial patterns of brain activities associated with lie and truth.

The methodology in most of related publications consisted of 4 phases. First is data collection and acquisition; in this part the signals are acquired from brain and collected by tools/softwares such as Neuroscan and WinEEG [22] and converted to input data for next phase. WinEEG allows the recording, editing and analysis of continuously recorded EEG and Evented Related Potential data (with the use of PsyTask), and performs database comparisons. Neuroscan provides a set of systems for the acquisition and analysis of EEG and ERP data. The systems are integrated platforms, designed to allow uncompromising solutions for seamless recording and analysis of EEG data across a variety of domains. Neuroscan has developed multiple hardware and software systems that combine to build the ideal platform for a particular area of research.

The second phase is data analysis and decomposition; ERP data contains mixture of brain signals (patterns) and noises. The aim of this phase is to separate signals from noises and disentangle overlapping patterns. [5] There are variety of known tools for data preprocessing and decomposition. For e.g. Net Station [8] is a suite of tools which includes data cleaning, statistical extraction and visualization techniques. EEGLAB [7] is a Matlab toolbox that provides advanced statistical

methods for EEG/MEG and ERP processing, with Independent component analysis ICA and Joint time-frequency analysis TFA. Dien PCA toolbox [6] is Principal component analysis PCA tolls that are optimized for ERP data decomposition.

Jia-Cai Zhang and Xiao-Jie Zhao [21] have suggested ICA for decomposition of ERP EEG patterns. "ICA can blindly separate the input ERP data into a sum of temporally independent and spatially fixed components arising from distinct or overlapping brain regions." They used ICA to illustrate "the P300 components in two ERPs recorded under various conditions or tasks. Both are contributed from a few independent sources. ICA decomposition also indicates a new method to compare P300 components between two ERPs induced by two related tasks". After decomposition in this phase there is summary extraction from PCA/ICA components.

The third phase is data mining. Jiawei Rong and Dejing Dou [5] have suggested to split data mining process into two steps; unsupervised and supervised learning. In unsupervised learning; the clustering is applied by using Expectation Maximization EM, it is trying to form the clusters base on summary metrics extracted from e.g PCA, and then the observations that belong to same pattern are expected to map to the same cluster. Later in supervised classification, decision tree is used to generate rules for each cluster and classify them. After classified rules for each cluster is generated, the fourth phase is rule comparison with expert rules. This phase is called evaluation part. Dejing Dou and Gwen Frishkoff [5] have suggested an ERP evaluation phase to refine the rules by ERP expert standards. In this phase the machine generated rules are compared with expert rules and the results would be the ERP domain rules.

ERP sharing is another important issue that concerns experts in this domain. Paea LePendu and Dejing Dou [20] have proposed an automatic method for modeling a relational database that uses SQL triggers and foreign-keys to efficiently answer positive semantic queries about ground instances for Semantic Web ontology. They applied this methodology to the study of brain EEG and ERP data. Dejing Dou [5] has suggested Web ontology Language (WOL) to publish and share ERP domain ontologism on the Web to be accessed by ERP experts.

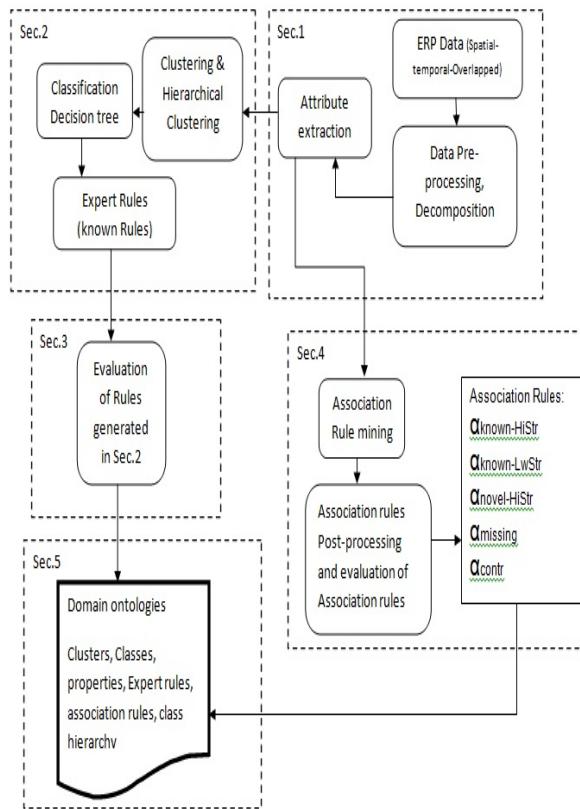


Fig.4. Neuroelectromagnetic Ontology Framework (NOF)

3. Neuroelectromagnetic Ontology Framework (NOF)

In Fig.4 there is an illustration of Neuroelectromagnetic Ontology Framework (NOF). first step is data pre-processing following by second step which is data mining. The third step is to evaluate machine generated rule. And finally the last step is association rule mining. The generated rules later will be evaluated with domain ontologies and refined to create the domain hidden rules.

4. Data preprocessing

4.1 Data Decomposition

ERP data is a mixture of noise which is a combination of artifacts and brain activities that is not related to events, and signals which is the target and the brain pattern. Decomposition methods help detecting noises and separating signals from overlapping patterns. There are many decomposition methods available such as PCA (Principle Component Analysis), Temporal PCA, and ICA (Independent Component Analysis) and Wavelet. In the present paper, ICA is proposed to

decompose the data. ICA is a method for separating a multivariate signal into additive subcomponents supposing the mutual statistical independence of the non-Gaussian source signals. The ICA toolbox in Matlab [6] is a useful tool in order to decompose the Matlab-friendly data. The datasets are used as input to the ICA is formed by variables that corresponding to time points as latency in ERP.

4.2 ERP Attributes extraction

After Decomposition, a summary of attributes of processed data is extracted. It represents spatial, temporal and functional dimensions of patterns of interest. In this study 13 attributes has been used by adding more spatiotemporal ones. In table 1, there is a list of attributes that are used.

Table 1: shows 13 ERP extracted attributes

Attribute	Description
SP-min(ROI)	channel grouping(ROI) to which the min channel belongs
SP-max	channel with max weighting for factor FA
SP-max (ROI)	channel grouping(ROI) to which the max channel belongs
SP-min	channel with min weighting for factor FA
IN-min	min amplitude
IN-max	max amplitude
IN-mean	mean amplitude for a specified channel set
ROI	region of interest
SP-cox	cross-correlation between Factor(FA) topography and topography of target pattern
TI-max	max latency(time of max amplitude)
EVEN	Event type (stimon, respon, EKG-R, etc.)
STIM	stimulus
MOD	modality of stimulus

5. Data mining processes

5.1 Unsupervised learning: ERP Patterns Clustering

Usually, ERP patterns are define through a “manual” process that involves visual assessment of peaked-major-averaged ERP data. Nevertheless, the appropriate definition of a target pattern, its operationalization, and measurement across individual subjects, can vary substantially across research groups. While this method can lead to agreement on the high-level rules and concepts that characterize ERP patterns

in a given domain, operationalization of these rules and concepts is highly variable across research labs.

The clustering method that is used in this paper is the PCA Clustering method which is formally explained and used in EEGLAB [7]. Before clustering, we need to prepare the data. This requires, first, identifying the components from each dataset to be entered into the clustering, then computing component activity measures for each study dataset. For this purpose, for each dataset component, the pre-clustering function first computes desired condition-mean measures used to determine the cluster 'distance' of components from each other. The condition means used to construct this overall cluster 'distance' measure may be selected from a palette of standard EEGLAB measures: ERP, power spectrum, ERSP, and/or ITC, as well as the component scalp maps (interpolated to a standard scalp grid) and their equivalent dipole model locations (if any). [7] After clusters are formed up and based on the clustering result, we can generate ontology classes.

5.2 Hierarchical Clustering

In this part it is expected to automatically discover the patterns those are found during clustering and arranged hierarchically, then generate taxonomy of the patterns. We may put whole of data in one cluster to sub-divide this cluster into 2 clusters then after that each cluster will be sub-divided repeatedly [9], in each step we take out the majority of data instances of one pattern into one cluster and label it. In another method we may determine the number of clusters based on the previous clustering step and put the data into clusters then try to merge each 2 clusters that are close to each other to form up a higher cluster. This process is done repeatedly until all data instances are put into one cluster. The discovered hierarchy (class taxonomy) can be represented in form of ontology classes and added into the ERP ontology.

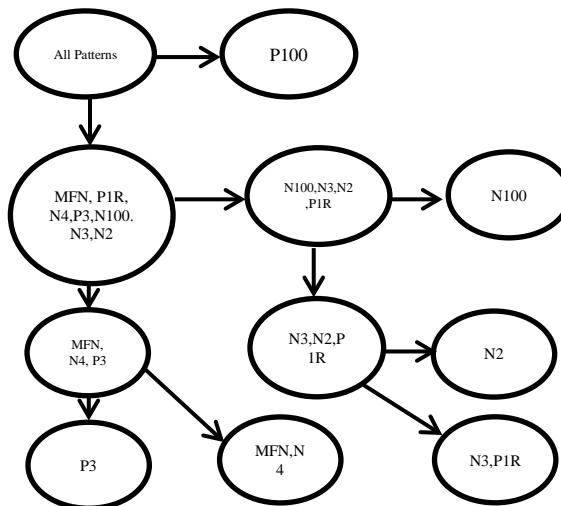


Fig.5 hierarchical clustering [5]

5.3 Supervised learning: Cluster-based classification

In clustering, data is automatically partitioned into clusters. The main purpose of classification in this section is to develop rules that accurately assign designated data to clusters. Therefore, we use classification methods such as decision tree to develop mentioned rules. We use C5.0 classification algorithm [16] to classify factors in each cluster. C5.0 is a new decision tree learning algorithm. In other researches C4.5 is mainly used for classification but C5.0 have some significant advantages over C4.5 such as accuracy, speed and memory usage. C5.0 works well for continuous values. Some ERP attributes have continuous values. On the other hand, the classification rules derived from the decision tree are meaningful to human experts. Although data in each cluster can be labeled with cluster names not respective to expert labels. When the clustering process reach to optimized point, then there is no need for the need for expert labeling of patterns, this will be providing a considerable savings in time and a sizable gain in information processing for ERP analysts. The rules will be represented in form of ontology rules to be added into ERP ontology database.

6. High-level Knowledge Generation

6.1 Rule comparison with domain experts

In decision tree we can generate ERP domain rule that would give a chance to compare them with the domain experts' rules which are manually generated. It is helpful to label the clusters in case their name is

arbitrarily assigned. These rules could be a good reference for domain experts.

6.2 Association Rule Mining

In clustering and Classification sections the main goal is to distinguish the classes and its properties. Along these classes, the relationship between properties and patterns are important to domain experts and is an input in ERP ontology. Association rule mining finds frequencies among patterns in certain data sets. In this paper, association rule mining is used to seek the properties that frequently co-occur for the specific ERP pattern factors. Values of each attribute are quantified by using their splitting point value in the decision tree. After converting the numeric values of each attribute to categorical values. Then, Apriori algorithm [14] in Weka [13] will be used to find association rules of these attributes.

7. Association rules Post-processing

Gunjan Mansingh, Kweku-Muata Osei-Bryson, Han Reichgelt [15] have proposed a generic method for pruning and partitioning the output of association rule induction. This approach can be used to combine the knowledge represented in ERP ontology with an objective measure Reliability (r) to create meaningful partitions in a set of extracted association rules. The partitions that are interesting to the domain experts are; 1. Novel rules with high strengths $\alpha_{\text{novel-HiStr}}$ – Extracted association rules those are not part of the ontology and have a Reliability value greater than a user-specified threshold. Rules in this partition may lead to the formation of new beliefs and additions to the ontology. 2. Known rules with high strengths $\alpha_{\text{known-HiStr}}$ – Extracted association rules that are already part of the ontology and have a Reliability value greater than a threshold. These rules strengthen previous beliefs of ERP patterns which have been induced by the domain expert(s). 3. Known rules with low strengths $\alpha_{\text{known-LwStr}}$ – Extracted association rules that are part of the ontology and have a Reliability value lower than a user-specified threshold. These rules represent aspects of an expert's knowledge that are not supported by data which suggests that the corresponding beliefs in the ERP domain that may no longer be true. When such rules are extracted the domain experts may investigate the reasons behind such an occurrence. 4. Missing rules α_{missing} – Expert Rules in ERP domain were not extracted by association rule induction. Presence of these rules implies that either the threshold values are incorrect or that the domain knowledge is not supported by data. 5. Contradictory rules α_{contr} – Rules

extracted by association rule induction with high strengths that contradict Expert rules. In order to detect the presence of such rules, the representation language in which the rules and the original ontologies were expressed must allow for the expression of negations.

In order to prune and create partitions in the output of the association rule induction process we use the domain ontology and reliability measures. the thresholds should be determined by experts; β_{sup} (minimum support), β_{conf} (minimum confidence) and π_{min} (minimum reliability). We extract the previously known rules $\alpha_{\text{pre-known}}$ (or expert rules) from domain ontology. Then we constrain the association rules by Support and Confidence thresholds β_{sup} β_{conf} therefore a new set of association rules based on expert constraints α_{AREC} are represented then for each α_{AREC} , the π_{rel} (reliability) need to be calculated.

Fourth the definitions of rules that are found in the process are below.

$\alpha_{\text{known-HiStr}} = \alpha_{\text{AREC}} \cap \alpha_{\text{pre-known}} \{ \pi_{\text{rel}}, \alpha_{\text{AREC}} \geq \pi_{\text{min}} \}$
$\alpha_{\text{known-LwStr}} = \alpha_{\text{AREC}} \cap \alpha_{\text{pre-known}} \{ \pi_{\text{rel}}, \alpha_{\text{AREC}} < \pi_{\text{min}} \}$
$\alpha_{\text{novel-HiStr}} = \alpha_{\text{AREC}} - \alpha_{\text{pre-known}} \{ \pi_{\text{rel}}, \alpha_{\text{AREC}} \geq \pi_{\text{min}} \}$
$\alpha_{\text{missing}} = \alpha_{\text{pre-known}} - \alpha_{\text{AREC}}$
$\alpha_{\text{contr}} = \{ A \in \alpha_{\text{AREC}} \mid \exists B \in \alpha_{\text{pre-known}} : A = -B \}$

8. Conclusion

We gave a brief overview of ERP domain knowledge in MEG EEG brain signals, and introduce Neuroelectromagnetic Ontology Framework (NOF) for mining event-related potentials as well as, data-mining process and results. The aim for this paper is to develop an infrastructure for mining, analysis and sharing the ERP domain ontologies in EEG, MEG data. The proposed Neuroelectromagnetic Ontology Framework (NOF) for mining ERP patterns contains 5 stages: 1) Data pre-processing and preparation: in this phase we collect brain signals by Netstation software and filter the noises along with decomposition of overlapping signals ; 2) Data mining application: in this phase we apply data mining which include unsupervised and supervised methods; 3) Rule Comparison and Evaluation: in this phase we refine the expert rules and evaluate the machine generated rules; 4) Association rules Post-processing: in this phase we

do association rule mining and refine the association rules by domain ontologies to produce new set of hidden rules. 5) Domain Ontologies: in this phase we develop the ERP ontologies from the results of previous phases. The domain ontologies are ERP inside knowledge and will be stored in a knowledge base. The results shows machine generated rules are good reference for expert rules and these rules can be evaluated by rules from Decision tree and Association rule mining. On the other hand, in Association rules Post-processing a new set of hidden rules can be discovered based on association rules and domain ontologies to improve and refine expert rules. The outcome of this research is a Neuroelectromagnetic knowledge-based system which contains the ERP domain ontologies. This System can refine the existing expert knowledge and produce ERP inside knowledge which will be reference for experts in the field of neuroscience. This knowledge extends the expert understanding of brain functions and will be used in finding the causes of brain disorders Cortex infunctionality, Neuron illness, brain capability analysis, and responses to outside stimulus in different situations. Although it can be used in commercial marketing known as Neural-Economy to increase sale and find the profitable customers.

Acknowledgments

This work was supported by a grant from the University Sains Malaysia. We thank Professor Jafri Malin Abdullah Department of Neuroscience, School of Medical Sciences and Prof. Rahmat Budiarto for helping us in this paper.

References

- [1] Niedermeyer E. and da Silva F.L. (2004). *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Lippincot Williams & Wilkins.
- [2] Cohen D. "Magnetoencephalography: evidence of magnetic fields produced by alpha rhythm currents." *Science* 1968; Vol.5, No.6, pp.161-784
- [3] Coles, Michael G.H.; Michael D. Rugg (1996). "Event-related brain potentials: an introduction". *Electrophysiology of Mind*. Oxford Scholarship Online Monographs. pp. 1-27.
- [4] Kam Swee Nga, , Hyung-Jeong YangCorresponding , Sun-Hee Kima, "Hidden pattern discovery on event related potential EEG signals". Department of Computer Science, Chonnam National University, South Korea, Vol. 2, No.1, PP. 123-129.
- [5] Rong, J., Dou, D., Frishkoff, G., Frank, R., Malony, A., & Tucker, D. (2007). A semi-automatic framework for mining ERP patterns. *Proceedings of the IEEE International Symposium on Data Mining and Information Retrieval (DMIR-07)*, Ontario, Canada. pp. 329-334.
- [6] Dien, J. PCA Toolbox (version 1.7).
- [7] EEG lab. <http://www.sccn.ucsd.edu/eeglab/>.
- [8] NetStation Technical Manual. <http://www.egi.com>.
- [9] Frishkoff, G., Frank, R., Rong, J., Dou, D., Dien, J., & Halderman, L. (2007). A framework to support automated ERP pattern classification and labeling. *Computational Intelligence and Neuroscience*, Vol.1, No.3, 2007, Article ID 14567.
- [10] LePendu, P., Dou, D., Frishkoff, G. , & Rong, J. (2008). Semantic data modeling: Methods for ontology-based queries and an application to brainwave data., *Proceedings of the 20th International Conference on Scientific and Statistical Database Management (SSDBM-08)*, Hong Kong, China.Vol.20, No.2, pp.220-228.
- [11] Dou, D., Frishkoff, G., Rong, J., Frank, R., Malony, A., & Tucker, D. (2007). Development of NeuroElectroMagnetic Ontologies (NEMO): A framework for mining brain wave ontologies, *Proceedings of the Thirteenth International Conference on Knowledge Discovery and Data Mining (KDD2007)*, pp. 270-279, San Jose, CA.
- [12] Christos N. Zigkolis, Nikolaos A. Laskaris, "Using conditional FCMtomine event-related brain dynamics" *Journal of Computers in Biology and Medicine* 39 (2009) 346 – 354, Artificial Intelligence & Information Analysis Laboratory, Department of Informatics, Aristotle University, Thessaloniki, Greece
- [13] Weka 3: Data Mining Software in Java. <http://www.cs.waikato.ac.nz/ml/weka/>.
- [14] R. Agrawal and R. Srikant. Fast Algorithms for Mining Association Rules. In *Very Large Data Bases (VLDB) Conference*, pages 487–499. Morgan Kaufmann, 1994.
- [15] Gunjan Mansingh, Kweku-Muata Osei-Bryson, Han Reichgelt. "Using ontologies to facilitate post-processing of association rules by domain experts" *International Journal of Information Sciences* 181 (2011) 419–434.
- [16] Ron Kohavi, Ross Quinlan. "Decision Tree Discovery" robotics.stanford.edu
- [17] Johnjoe McFadden (2002). "The Conscious Electromagnetic Information (Cemi) Field Theory: The Hard Problem Made Easy?". *Journal of Consciousness Studies*. Vol. 9, No.8, pp 45–60.

- [18]http://sccn.ucsd.edu/~arno/fam2data/publicly_available_EEG_data.html.
- [18]http://sccn.ucsd.edu/~arno/fam2data/publicly_available_EEG_data.html.
- [19] M. Sabeti, S.D. Katebi, R. Boostani, G.W. Price, "A New Approach for EEG Signal Classification of Schizophrenic and Control Participants" Elsevier Journal of Expert Systems with Applications, Volume 38, Issue 3, March 2011, Pages 2063-2071 (ISI)
- [20] Paea LePendu, Dejing Dou, Gwen A. Frishkoff, Jiawei Rong: Ontology Database: A New Method for Semantic Modeling and an Application to Brainwave Data. SSDBM 2008: 313-330 [22] C. Davatzikos, M. Acharyya, K. Ruparel, D.G. Shen, J. Loughhead, R.C. Gur, and D. Langleben , "Classifying spatial patterns of brain activity for lie-detection", NeuroImage, 28 (3), 663-668, 2005.
- [21] Jia-Cai Zhang, Xiao-jie Zhao, Yi-Jun Liu, Li Yao: Mining the Independent Source of ERP Components with ICA Decomposition. ISNN (2) 2006: 592-599
- [22] <http://www.novatecheeg.com/wineeg.html>

First Author Seyed Aliakbar Mousavi Aslarzanagh is a researcher in University Science Malaysia, School of Computer Science. He obtained E-commerce Bachelor degree from Multimedia University in 2010 and continued his study to a research degree in Computer Science. He is currently cooperating with Department of Neuroscience in USM Kubang Kerian and multimedia research group in an interdisciplinary fields of Neuroinformatic and Data Mining Knowledge discovery.

Second Author Assoc Prof Muhammad Rafie Hj. Mohd. Arshad achieved his Bachelor degree from Macalester College, Minnesota, U.S.A. and his MBA-MIS from Dallas, Texas, U.S.A.

Third Author Hasimah Hj Mohamed has obtained her bachelor degree from Universiti Teknologi Malaysia and her MSc from Universiti Sains Malaysia. Currently she is a senior lecturer at Universiti Sains Malaysia, and also a PhD candidate at National University of Malaysia.

Fourth Author Saleh Ali K. Al-Omari has obtained his Bachelor degree in Computer Science from Jerash University, Jordan in 2004-2005 and Master degree in Computer Science from Universiti Sains Malaysia, Penang, Malaysia in 2007. Currently, He is a PhD candidate at the School of Computer Science, Universiti Sains Malaysia. He is the candidate of the Multimedia Computing Research Group, School of Computer Science, USM. He is a member and reviewer of several international journals and conferences (IEICE, JDCTA, IEEE, IACSIT, AIRCC, etc). His main research area interest are in includes Peer to Peer Media Streaming, Video on Demand over Ad Hoc Networks, MANETs, and Multimedia Networking, Mobility.

LR-WPAN Formation Topologies using IEEE 802.15.4

Salman Naseer¹ and S R Chaudhry²

¹ Department of Information Technology, University of the Punjab
Gujranwala Campus, Punjab 52250, Pakistan

² Communication Network Research Center (CNRC), Department of Computer Science
COMSATS Institute of Information Technology, Lahore, Pakistan

Abstract

IEEE 802.15.4 protocols are gaining interests in both industrial and research fields as candidate technologies for (WPAN) Wireless Personal Area Networks, (WSN) Wireless Sensor Network and control Wireless Networks applications. This paper analyzes multiple topologies such as Cluster-Tree, Mesh and Star in various scenarios to compare different performance metrics (Throughput, traffic sent, traffic received, Load, End-to-end Delay and etc). In this analysis it is found that Cluster-Tree topology is best as compared to Mesh and Star topology because it take 20% and 45% load greater as compared to Mesh and Star Topology respectively. Similarly its throughput, Traffic Sent, Traffic Received and Delay is better than the other two topologies.

Keywords: LR- WPAN (Low Rate- Wireless Personal Area Network), ZC (ZigBee Coordinator), ZR (ZigBee Router), ZED (ZigBee End Device), Cluster-Tree, Mesh, Star and IEEE 802.15.4

1 - Introduction

In the modern era we are getting benefit from different electronic appliances, it is compelling need to organize them in this way so that devices can communicate wirelessly. It is obviously preferable to establish wireless network. Wireless networks have changed our lives as the internet has revolutionized this universe. The future is of wireless network and WPAN, will be used in different embedded application like home appliances, military control system, medical, industry etc. These devices are battery operated and communicate in a specific range using wireless radio waves.

Latest standard of WPAN with low data rate and energy efficient protocol is IEEE 802.15.4 also called ZigBee [1]. ZigBee gets its name from the honeybee “zig-zag dancing of honeybees to give information to other bees for new food”[2]. It is an open standard for WPAN in monitoring and control

fields and introduced by ZigBee Alliance (An organization of more than 150 companies [3].

A Network layout of proposed IEEE 802.15.4 WPAN based topologies is shown in Figure 1(a).

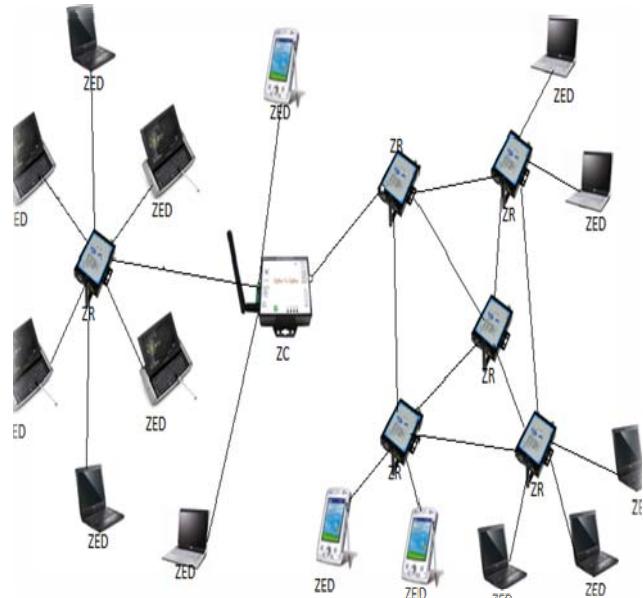


Figure 1(a): IEEE 802.15.4 Network Layout

In Europe and US 802.15.4 operate in the 2.4GHz band or the 868MHz and 915MHz ISM (industrial, scientific, and medical) bands @ 20,40 and 250 kbps by using CSMA/CA and slotted CSMA/CA.[4].

The IEEE 802.15.4 specifies the Physical Layer of (LR-WSNs) Low-Rate Wireless Sensor Networks and the (MAC) Medium Access control sub-layer [5]. ZigBee standard protocol introduces the cross platform communication independent of hardware and software. The protocol stack of 802.15.4 is shown here in diagram.[3]

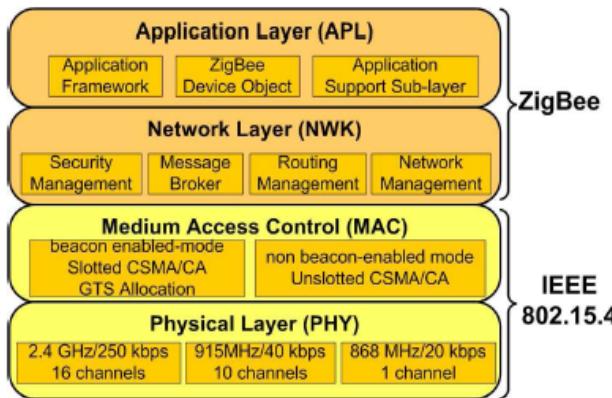


Figure 1(b): Protocol Stack of ZigBee

According to 802.15.4 standard there are three types of devices. 1) Coordinator, 2) Router, and 3) end device. Routers maintain the path to destinations and send data towards the desired device. The coordinator also performs the functionality of the router plus it creates, maintains and manages the network. Both router and coordinator are called (FFDs) *Fully Functional Devices*. Because they can implement all the functions of ZigBee standard. End devices also called (RFDs) *Reduced Functions Devices*. They receive data from devices called sensors arrange this data into packets and send to destination devices [6]. These devices operate in a short range of distance 10 to 20 meters.

In this paper we are analyzing three different topologies star mesh and cluster. The novelty of the work is in the performance of the parameters can be measured by different simulations. These results will be helpful to configure the ZigBee and to select a suitable topology according to situation.

This paper is organized in five parts. Part one describe the brief introduction of WPAN and 802.15.4 standard, in part two we describe the topologies w.r.t WPAN briefly, in part three there are proposed scenarios with multiple hybrid topologies, part four has detailed analysis and discussion of WPAN purposed topologies and part five has final conclusions based on our discussion.

2 - WPAN TOPOLOGIES

Topologies are the physical arrangements of the devices in the network, we discuss three different WPAN topologies in this paper. We use three type of devices in these topologies

- (1) ZC ZigBee Coordinator, it is a FFD fully functional device, act as a PAN coordinator which configure and maintain the network[7]

- (2) ZR ZigBee Router, it is also a FFD fully functional device, coordinate with ZC, and manages the multi-hop routing.[7]
- (3) ZED ZigBee End Device, it is a RFD reduced functional device, it is an end device of ZigBee Network not perform routing, other devices cannot communicate through ZED[7]

2.1 - Star Topology

In case of star topology ZC is central device called ZigBee Coordinator ZC chooses the PAN ID which will not be used by other ZC in the range. It is a centralized network i.e all devices ZED in it cannot communicate directly but can communicate via ZC[7]. Star topology is shown in Figure 2(a)

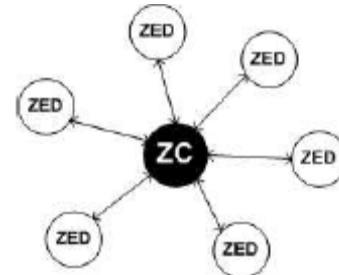


Figure 2(a): Star Topology

2.2 - Mesh Topology

It is decentralized network all devices can communicate directly with each other in their radio range[7]. It is a robust and flexible topology. Figure 2(b) shows the mesh topology with ZC ZigBee Coordinator, ZR ZigBee Routers, and ZED ZigBee end devices.

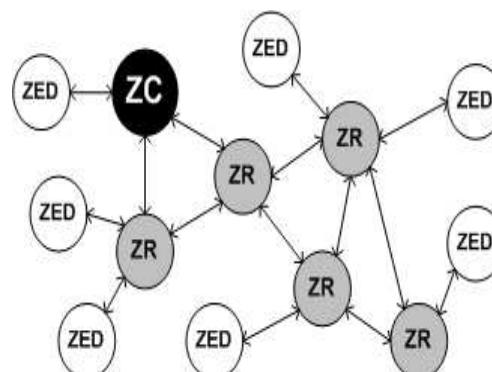


Figure 2(b): Mesh Topology

2.3 - Cluster-Tree Topology

Figure 2(c) shows the cluster topology which is just like the mesh network work with beacon-enabled mode having only one path between pair nodes[7]. In figure ZC

manages the whole network and coordinate with three clusters, these clusters are managed by ZRs.

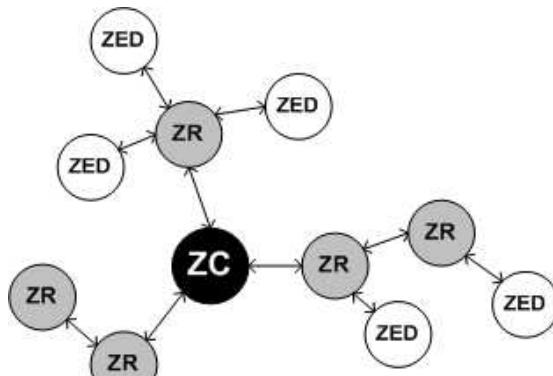


Figure 2(c): Cluster-Tree Topology

3 - Methodology Overview, Network Design and Simulation.

Simulation modeling is the best approach to develop the system w.r.t time as well as cost[8]. The module ZigBee 80215.4 of OPNET Modeler 14.5 is used to develop the simulations. This version supports ring, star , cluster and mesh topologies[8]. Equal number of ZC,ZR and ZED are used in all topologies as briefly discussed here.

Star topology is shown in figure 3(a). A single PAN coordinator is the central device in the star and all end devices arranged around this PAN coordinator. No two end devices can communicate directly with each other without the help of PAN coordinator. End device first send data to PAN coordinator and then PAN coordinator send data to other particular end device.



Figure 3(a) Purposed Star Topology

Table 3.1: Purposed Simulation Parameters of Physical Layer

Structure of Mesh topology is shown in figure 3(b). One PAN coordinator to manage the PAN but any end device can send data to any other end device in its range.



Figure 3(b) Purposed Mesh Topology

Cluster-Tree Topology is shown in the figure 3(c), having 3 Routers. Which manage each network locally and can communicate through PAN Coordinator. It is most popular topology due to its scalable nature w.r.t geographical area.



Figure 3(c) Purposed Cluster-Tree Topology

These are the three purposed topologies which are configured in OPNET according to different simulation parameters of Physical layer, media access control, carrier sense multiple access and Application traffic as shown in the following tables:

Physical Layer	
Data rate	Data rate
Receiver Sensitivity	-85 dB
Transmission Band	2.4 GHz
Transmission Power	0.05 W

Table 3.2: Purposed Simulation Parameters of MAC and CSMA

MAC	
ACK wait time	0.05
Total Retransmissions	5
CSMA	
Exponent of minimum back off	3
Exponent of maximum back off	4
Carrier sense duration	0.1

Table 3.3: Purposed Application Traffic

Parameters	Application Traffic					
	Device Type	Inter-arrival time of Packet	Size of Packet	Start Time	Stop Time	Destination
Star Topology	ZC	Constant (1.0)	Constant (1024)	Uniform (20,21)	Infinity	All ZCs and ZRs
	ZR	Constant (1.0)	Constant (1024)	Uniform (20,21)	Infinity	ZC
	ZED	Exponential (1.0)	Exponential 1024	Exponential (1.0)	Infinity	ZC
Mesh topology	ZC	Constant (1.0)	Constant (1024)	Uniform (20,21)	Infinity	All ZCs and ZRs
	ZR	Constant (1.0)	Constant (1024)	Uniform (20,21)	Infinity	All ZCs and ZRs
	ZED	Exponential (1.0)	Exponential 1024	Exponential (1.0)	Infinity	Parents
Cluster-Topology	ZC	Constant (1.0)	Constant (1024)	Uniform (20,21)	Infinity	ZC and ZRs
	ZR	Constant (1.0)	Constant (1024)	Uniform (20,21)	Infinity	All ZCs and ZRs
	ZED	Exponential (1.0)	Exponential 1024	Exponential (1.0)	Infinity	ZRs

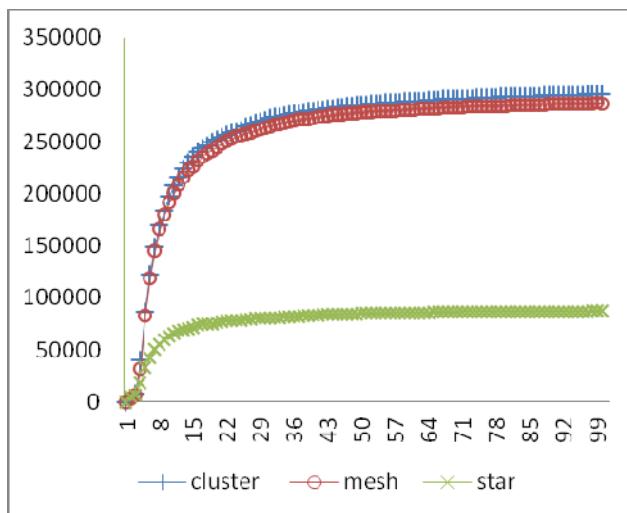
4 - Measurement Results

In this section we describe the results obtained by the simulations. As already it is stated that we have taken three topologies star, mesh and cluster tree with equal number of ZCs, ZRs and ZEDs. Different parametric results (Throughput, Traffic sent, Traffic received, Load, end-to-end Delay) have been explained here that show the impact of performance on different topologies.

4.1 – Throughput

Represents the total number of bits (in bits/sec) forwarded from 802.15.4 MAC to higher layers in all WPAN nodes of the network.

Graph 4(a) shows throughput of Cluster-Tree, Mesh and Star topology respectively.

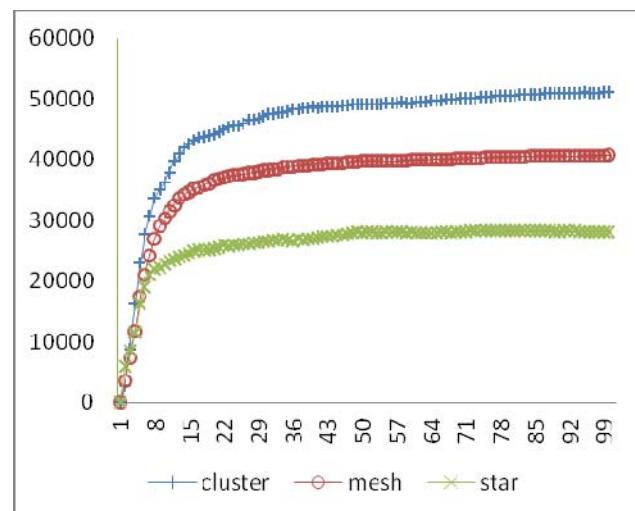


Graph 4(a): Throughput of Cluster-Tree, Mesh and Star Topologies (bits/sec)

Graph 4(a) shows that throughput of Cluster is 295.632 Kbps, 287.046 for mesh and 87.046 for star. This observation shows that throughput is maximum for cluster topology. Because Cluster has four fully functional devices and it is a beacon enabled topology where each cluster is managed separately by PAN routers and then joined with PAN coordinator which reduces the number of collisions and retransmissions. This graph also shows that throughput is minimum in case of star topology because it has only one PAN coordinator ZC and all other devices act as end devices ZEDs. All these ZEDs can communicate through a single ZC which leads to lower throughput. Moreover all nodes in Mesh communicate with each other, so the communication between ZEDs to ZEDs are not efficient than communication between ZEDs to ZC or ZR. Hence Mesh has fewer throughput than Cluster-Tree.

4.2 - Traffic Sent

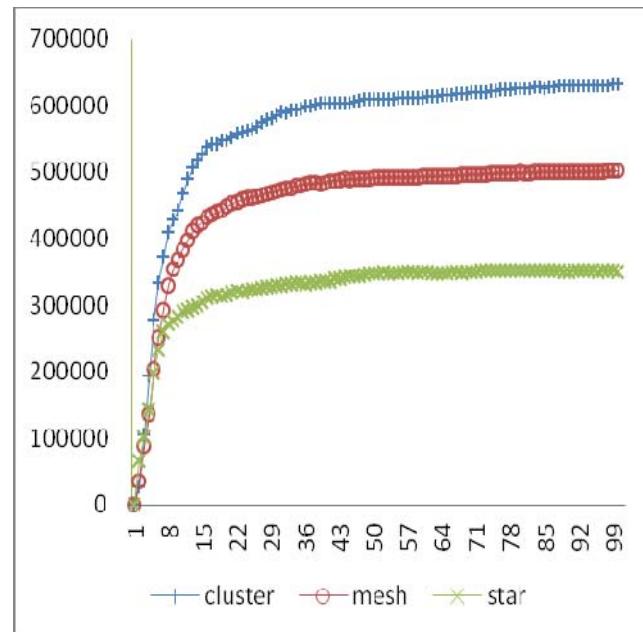
Traffic transmitted by all the 802.15.4 MACs in the network in bits/sec. While computing the size of the transmitted packets for this statistic, the physical layer and MAC headers of the packet are also included. This statistics include all the traffic that is sent by the MAC via CSMA-CA. It does not include any of the management or the control traffic, nor does it include ACKs. Graph 4(b) shows the traffic sent for all topologies. Traffic sent for cluster topology is 51.083 Kbps, 40.732Kbps for Mesh and 28.186 Kbps for Star. This observation shows that Traffic sent is maximum for Cluster-Tree topology. Because it uses ZC and ZRs which manages their own routing tables which are used in traffic generations. Lower collision and packet drop rate leads to high traffic sent for Cluster-Tree topology. This graph also shows that traffic sent for Star is minimum due to one ZC there are more collision and retransmissions



Graph 4(b): Traffic Sent for Cluster-Tree, Mesh and Star Topologies (bits/sec)

4.3 - Traffic Received

Represents the total traffic successfully received by the MAC from the physical layer in bits/sec. This includes retransmissions. Graph 4(c) shows the traffic received by all the topologies. Traffic received for Cluster-Tree Topology is 631.428 Kbps, 501.736 Kbps for Mesh and 351.460 Kbps for Star. This result shows that the traffic received is maximum for Cluster-Tree topology because ZEDs communicate through ZCs and ZRs which leads to less collision and less packet drop and results to high traffic received

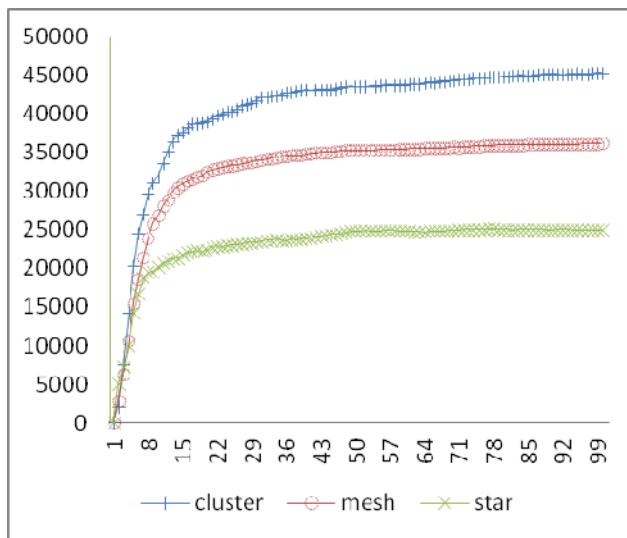


Graph 4(c): Traffic Received for Cluster-Tree, Mesh and Star (bits/sec)

This result also shows that the traffic received is minimum for Star Topology because there all devices communicate through a single ZC which causes high collision rate and high packet drop rate results to lower traffic rate. Same is the case with the Mesh topology but only a few ZEDs are in direct communication of ZCs

4.4 – Load

Represents the total load (in bits/sec) submitted to 802.15.4 MAC by all higher layers in all WPAN nodes of the network. Graph 4(d) shows the load of all topologies. Load for Cluster-Tree topology is 45.351 Kbps 36.218 Kbps for Mesh and 25.006 Kbps for Star. This results also shows that the load is maximum in case of Cluster-Tree topology.

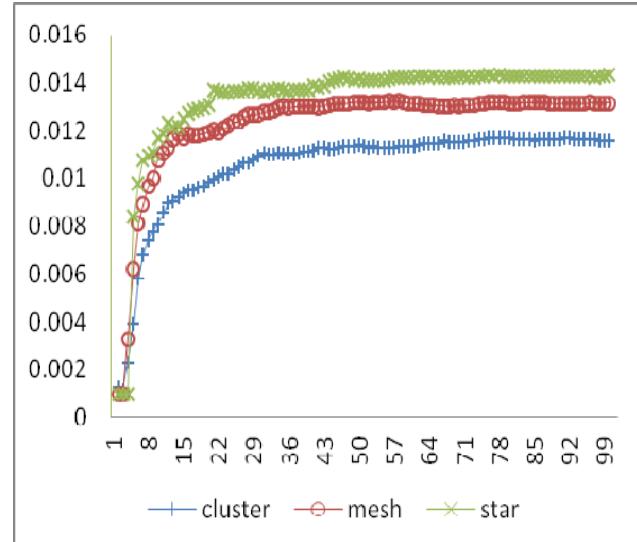


Graph 4(d): Load for Cluster-Tree, Mesh and Star Topologies (bits/sec).

4.5 - Media Access Delay

The total of queuing and contention delays of the data frames transmitted by the 802.15.4 MAC. For each frame, this delay is calculated as the duration from the time when it is inserted into the transmission queue, which is arrival time for higher layer data packets and creation time for all other frames types, until the time when the frame is sent to the physical layer for the first time.

Graph 4(e) shows the media access delay of all topologies. Media access Delay for Cluster-Tree topology is 0.01159 sec, for Mesh 0.01313 Sec and for Star is 0.01433 Sec. This graph result shows that the minimum media access delay is for Cluster-Tree Topology and maximum for Star Topology.



Graph 4(e): Media Access Delay for Cluster-Tree, Mesh and Star (sec)

6 – Conclusions

Performance of WPAN ZigBee IEEE 802.15.4 has been analyzed in detail with the help of three different topologies Cluster-Tree, Mesh and Star. Performance of these WPAN Topologies has been analyzed with the help of different parameters like Throughput, Traffic Sent, Traffic Received, Load and Media access delay. This type of analysis is missing in literature which is helpful to understand and to implement 802.15.4 Network. The summarized results are given in the following table 6(a).

Table 6(a): Comparisons of Cluster-Tree, Mesh and Star Topologies

Comparisons	Cluster-Tree	Mesh	Star
Throughput (Kbps)	295.632	287.046	87.046
Traffic Sent (Kbps)	51.083	40.732	28.186
Traffic Received (Kbps)	631.428	501.736	351.460
Load (Kbps)	45.315	36.218	25.006
Media Access Delay (Sec)	0.01159	0.01313	0.01433

From the discussion in Section – 4 and summarized results as shown in Table 6(a) it is concluded that for WPAN ZigBee 802.15.4 the best and efficient topology is Cluster-Tree topology as compared Star and Mesh topology.

Acknowledgements

We would like to pay special thanks to our respected institutions for providing us the facilities to conduct the research on the topic. We acknowledge the people at research center CNRC for great help and support.

References

- [1]. Sukhvinder S. Bamber; Ajay K. Sharma; “ Comparative Performance Investigations of different scenarios for 802.15.4 WPAN ” in IJCSI International Journal of Computer Science Issues. Vol. 7, Issue 2, No 4, March 2010.
- [2]. P. Ramanathan; Pradip Manjrekar; “Wireless sensor network for monitoring a patient’s physical conditions continuously using ZigBee” in Indian Journal of Science and Technology. Vol. 4 No. 8, Aug 2011: ISSN: 0974- 6846.
- [3]. Andre Cunha; Anis Koubaa; Ricardo Severino, Mario Alves; “Open-ZB: an open source implementation of the IEEE 802.15.4/ ZigBee Protocol Stack on Tiny OS” in 4th IEEE international Conference 2007.
- [4]. C. Evans-Pughe, “Is the ZigBee Wireless Standard, promoted by an alliance of 25 firms, a big threat to Bluetooth?.” In IEEE Review, Vol. 49, pp.28-31, March 2003.
- [5]. IEEE-TG15.4, "Part 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (LR-WPANS)," in IEEE standard for Information Technology, 2003.
- [6]. Jin-Shyan Lee; “An Experiment on Performance Study of IEEE 802.15.4 Wireless Networks” Published in Proc. IEEE International Conference on Emerging Technologies and Factory Automation, Catania, Italy, Sept. 2005, vol. 2, pp. 451-458.
- [7]. Anis Koubaâa ; André Cunha; Mário Alves; Eduardo Tovar; “TDBS: a time division beacon scheduling mechanism for ZigBee cluster-tree wireless sensor networks” in Real-Time Syst. DOI -10.1007/s11241-008-9063-4, Springer Science + Business Media, LLC 2008
- [8]. Abhiram Devineni; “Performance Evaluation of Body Area Network Using ZigBee Protocol” published in http://sdsuspace.calstate.edu/xmlui/bitstream/handle/10211.10/1066/Devineni_Abhiram.pdf?sequence=1
- [9]. IEEE 802.15.4 OPNET Simulation Model, <http://www.open-zb.net/>
- [10]. OPNETs ZigBee Model, <http://www.opnet.com/>
- [11]. IEEE. (2010). *IEEE 802.15.4 Task Group Web Site* [Online]. Available: <http://www.ieee802.org/15/pub/TG4.html>

Salman Naseer Completed his M.Sc. Computer Science from University of the Punjab Lahore Pakistan in 2004. And now MS leading to PhD in Computer Science (Specialization in computer Networks) in progress from Comsats Institute of Information Technology Lahore Pakistan. Currently he is working as a lecturer in Department of Information Technology University of the Punjab Gujranwala Campus.

Dr. Saqib Rasool Chaudhry is affiliated as visiting research fellow with the Wireless Networks and Communications Centre (WNCC) at Brunel University. He also held visiting appointments at King's College London (CTR). He is a permanent member of the Institute of Electrical and Electronics Engineers and Institute of Engineering and Technology. He completed his doctorate in Network and Communication engineering from the Brunel University in 2010. He was awarded EPSRC scholarship for his PhD at Brunel University, UK. From 2005 to 2009, he served as Senior Research Associate and Team leader at the Brunel University,

Dewy index based Arabic Document classification with Synonyms Merge Feature Reduction

E. M Saad, M H Awadalla and A F Alajmi

Communication & Electronics Dept., Faculty of Engineering, Helwan University
Egypt

Abstract

Feature reduction is an important process before documents classification. The classification performance is impact by the quality of the selected. A new semantic approach is presented using synonym merge to preserve features semantic and prevent important terms from being excluded. The resulting feature space were then processed with five feature selection methods, ID, TFIDF, CHI, IG and MI. experiment show that classification performance is increased after merging terms and yielding best performance for CHI and IG selection method. A promising classification technique is presented based on Dewey decimal classification system, which uses filtered indexes and three levels of classes from Dewey system to classify and label Arabic documents. The technique shows along with synonyms merge a promising result.

Keywords: Dimension reduction, Arabic text Classification, synonyms.

1. Introduction

Over the past decade there was a rapid growth of data, which raised the question of the ability to process accurate information from text. Text mining and, natural language processing aim at information extraction and organizing. The tasks of text mining such as classification assigns labels to texts (documents) based on their similarities to the assigned classes, Whereas, clustering groups texts which have similar conceptual contents. Text summarizing, summarizes the text according to its base idea, and question answering searches a precise answer for a given question from a collection of texts. Whichever the task in question, the collection of text (Data set) must be processed first in order to extract useful information. One of the main preliminary requirements of text processing is feature extraction. In the context of text, features are words that appear in the text, but words at it appears in text are not a very good representation. Further processing is needed to produce meaningful and unified form of features, such as stemming, and stop word removing, which are grammatical words that by removing the efficiency of classification increases. Features are then represented. vector space model (VSM) [1] is a is an algebraic model for representing text documents, which uses category as the class label and the word or phrase as features of vector space. There are different methods to generate the values of features such as statistics based, grammar based, semantic based, etc.

There are two problems associated with features, the large scale of features, and the importance of a particular feature for understanding the text in question. For many applications dimension reduction is needed before further processing. Dimension reduction (DR) is done to reduce the large feature space extracted from the text. Another reason for reducing feature space is to eliminate noisy components to better represent [2], because it could handle other common words that were not eliminated by the stop word removal process. Dimension reduction (DR) also have the advantage of reducing the computing time, and cost, thus makes it more suitable requirement wise [3] for further processing to improves the text classification (task) efficiency and performance and, provides better data understanding that improves the clustering and classification result [4] by removing redundant and irrelevant terms from the corpus.

There are two classes of dimension reduction techniques, feature selection (FS), and feature extraction (FE) [3][5][6][7][8][9]. FE reduce data dimension by projecting the high dimensional data into lower space through transformations. FS searches a subset of the most representative features according to some criterion.

The goal of this paper is to present a semantic approach for dimension reduction. The approach is based on the fact that different words may have similar meaning (synonyms), and by looking at the count of the words in a text (TF, DF, etc.) alone we may exclude important features from the list of feature which might affect the quality of the classification process. Thus a synonym merge approach is presented to reduce the dimension of the feature space and maintain the semantic relation between the features. Furthermore a Dewey based categorization technique is presented, which uses a derived list of indexes from Dewey Decimal indexing system and three level of classification labels. Where a document may belong to more than one class and the classes are labeled using Dewey main category and two subcategories.

In the next section a review of related work to dimension reduction is presented. Section 3 shows different reduction techniques. Section 4 discusses the synonym merge reduction technique. In section 5 a Dewey index based classification method is explained. Experiment

result is shown in section 6 and, work is concluded in section 7.

2. Related Work

There are two classes of dimension reduction techniques, feature selection (FS), and feature extraction (FE). Feature selection selects a representative subset of the input feature set, based on some criterion. An estimate function is used to rank original features according to the calculated score value for each feature. This value represents the quality or importance of a word in the collection. The features then ordered in descending or ascending order to the values, and then select a suitable number of words of highest orders. Feature selection methods are classified into four categories [11]: filter, wrapper, hybrid and embedded, the hybrid method has better efficiency in time complexity and granularity.

The two main feature selection methods are wrapper and filtering [2], [10]:

- 1- Wrapper (unsupervised): select a subset of features by evaluation function based on learning algorithms that will take these selected features [2]. Also called subset selection [4], it searches the set of possible features for the optimal subset. Wrapped method applied to high dimension feature space is NP hard optimization problem, which can hardly run effectively due to the time complexity
- 2- Filtering (supervised): The base in a filtering method is to set up a criterion (evaluation function, score function, feature evaluation index, filtering function) for measuring the feature so that whether it should be remained can be then decided. The most important features, whose evaluation function values are the largest or the smallest, are kept, and the others are filtered. [2]. Feature filtering ranks the features by a metric and eliminates all features that do not achieve an adequate score [4] defined by the threshold. The speed of filter method is fast because it does not consider the combination of different feature, but it generally catches coarse results only and that is why it is simple but efficient [7].

Examples of score function used are document frequency (DF), Chi-square (CHI) (X² statistics), information gain (IG), mutual information gain (MI), term strength, term contribution (TC)[2]. IG is one [12] of the most effective techniques. term frequency (TF), document frequency (DF), information gain (IG), mutual information (MI), odd-ratio, Chi-square [13][11].

On the other hand, feature extraction methods transforms the original features into new, lower dimension space and creates new features, by computing new features as some function of the old ones. They are categorized into linear and non-linear algorithms [6]. The new space sometimes called "concept Space" [3] or latent sub-space [7]. Feature Extraction is helpful in

solving the problems related to synonymy and polysemy [9], were the loss of great part of [8] useful information of original feature set are avoided. Transformed features generated by feature extraction may provide a better discriminative ability than the best subset of given features, but these new features may not have a clear physical meaning [8]. The complexity of feature extraction algorithms are often too high to be applied on large scale text processing tasks [6]. Furthermore, it is difficult to provide a direct semantic interpretation to the new features. Examples of linear feature reduction are fisher discriminant method, principal component analysis PCA, Latent semantic analysis LSA, linear discriminant analysis, maximum margin criterion (MMC) and orthogonal centroid algorithm (OCA). Non-linear feature extraction transformation algorithms are Locally Linear Embedded (LLE), ISOMAP and Laplacian Eigenmaps.

In the next two section a review of previous work done on feature reduction.

2.1 Feature Selection Methods:

Feature selection algorithms are widely used in the area of text processing due to their efficiency [6].

Fouzi et al [20] compares five reduction techniques, root-based and, light stemming, document frequency DF, tfidf, and latent semantic indexing LSI. Then it shows that df , tfidf, and lsi methods were superior to the other techniques in term of classification problem. Furthermore, Savio [21] discusses four reduction methods DF, category frequency – Document frequency CF-DF, TFIDF, and principal component analysis PCA. The experiment shows that reduction of DF=15.2%, CF-DF=36.4%, TFIDF=78.8%, and PCA=98.9%. Which conclude that PCA is the most effective method in term of reduction with some decrease in classification efficiency.

Yang et al. [12] experimented with the first five score functions on Reuters21578 collection, and concluded that DF, CHI, and IG [14] are more effective than the others. The functions were Document frequency DF, Information gain IG, mutual information MI, X₂-test CHI, and term strength.

Wang [2] Applies feature reduction method called variance-mean based feature filtering that aims at keeping the best features and at the same time improves performance. Features are represented as terms-documents matrix, and the values are the probability that term t will occur in document i. Then two vectors mean E and variance D are computed. The variance of E is computed to show degree of dispersion among classes and, the mean of D are computed, which shows average level of the degree of variability within every class term t can show. The bigger D(E) the more distinguishable among classes using that term w. and the smaller E(D) the more cohesive within each single class averagely using that term w is. So the more distinguishable among classes and cohesive within each single class using that

term w is the more possible the term t should remain. D(E) and E(D) based criterion can be used to evaluate the importance of the candidate term t , with the evaluation function $F = \beta * D(E)/E(D)$ where β is a tuning parameter. If $F > f$ (threshold) then the term is selected. Comparison with DF and CHI results in similar performance. Performance reaches 0.92 of macro-f1 value with order 400.

Wangi [16] investigates the use of Hill Climbing (HC), Simulated annealing (SA), and threshold accepting (TA) optimization techniques as feature selection techniques to reduce dimension of an e-mail, and improve the classification filter performance. It was found that simulated annealing has the best performance. The approach starts with transferring the email to vectors of TF-IDF. Then apply the feature selection techniques to choose best discriminating feature sets. Performances were compared with Linear Discriminant Analysis and the accuracy was 90% for LDA, 93.6 for HC, 94.6 for TA and 95.5 for SA.

Zifeng [5] propose an approach of feature selection based on Constrained LDA (CLDA) to overcome the problem of un-interpretable features resulting from LDA transformation. Like LDA, the proposed approach will find a subset of features which maximize the discriminant capability between classes with a linear. The work is based on selecting but not transforming features by LDA to preserve structure information between-class and within-class for text categorization. Constrained LDA (CLDA) models feature selection as a search problem in subspace and finds optimal solution subject to some restrictions. The CLDA is transformed into a process of scoring and sorting of features. Experiments on 20 Newsgroups and Reuters-21578 show that CLDA is consistently better than information gain IG and CHI with lower computational complexity. Ren [22] propose an improved LAM feature selection algorithm (ILAMFS). The algorithm is based on combining the gold segmentation and the LAM algorithm based on the characteristics and the category of the correlation analysis, filtering the original feature set, and retaining the feature selection with strong correlation and weak category. Then, weighted average and Jaccard coefficient of feature subsets make redundancy filtering. Finally, obtains an approximate optimal feature subset. LAM algorithm has been improved, an improved LAM proposed feature selection algorithm (ILAMFS). the accuracy of selection in the threshold, feature selection and, efficiency of the running time are improved.

Xi [9] propose a two-stage feature selection method based on the Regularized Least Squares-Multi Angle Regression and Shrinkage (RLS-MARS) model. First, measure the features, and select the important features, by applying a new weighting method, the Term Frequency Inverse Document and Category Frequency Collection normalization (TF-IDCFC), and using the category information as a factor. Next, the RLS-MARS model is used to select the relevant information, while the Regularized Least Squares (RLS) with the Least

Angle Regression and Shrinkage (LARS) can be viewed as an efficient approach. The experiments demonstrate the effectiveness of the new feature selection method for text classification in several classical algorithms: KNN and SVMLight. The performance of the new algorithm is similar to the feature selection by χ^2 CHI statistics less number of features was chosen and, outperforms χ^2 methods when the dimensionality grows higher. Furthermore, features are selected using hybrid FS method based on improved particle swarm optimization PSO and, support vector machine SVM in JIN.[11], named FS_PSO_TC. It integrates the advantages of term frequency-inverse document frequency DF-IDF as inner-class measure and Chi-square as inter-class, and introduce a feature selection method based on swarm intelligence. The improved particle swarm optimization used to select fine features on the results of coarse grain filtering, and utilizing support vector machine to evaluate feature subsets and taking the evaluations as the fitness of particles. Experiments show the method reducing effectively the high dimension while catching better categorization efficiency.

A collaborative filtering method in [15] is used to reduce feature space, which utilizes traditional feature reduction techniques along with a collaborative filtering method for to predict the value of missing features for each class. Information Gain (IG) used to identify non-trivial noun phrases with semantic meanings in the documents, noun phrases (NP) chunking is adopted for this purpose. Chunking groups together semantically related words into constituents. Experiment indicates an improvement in classification accuracy over the traditional methods for both Support Vector Machines and AdaBoost classifiers.

2.2 Feature Extraction Methods

Feature extraction methods project the high dimension feature space in to a lower one. Furthermore, rough set theory has been applied to feature reduction area. [14] Rough set theory can discover hidden patterns and dependency relationships among large number of different feature terms in text datasets. No additional information about the data is required such as thresholds or domain knowledge. Essential part of information can be identified through generating reducts (i.e., the minimal sets of attributes which has the same distinguish capability as the original set of attributes), thereby reducing the irrelevant/redundant attributes as well as maintain the discernibility power with respect to the task of pattern recognition or classification [14].

Jensen [13] reviews techniques that preserve the underlying semantics of the data, using crisp and fuzzy rough set-based methodologies. This paper reviews techniques, which employs rough set based methodology which belongs to rough sets, fuzzy rough sets, and rough set-based feature grouping. It was shown how fuzzifying a particular evaluation function, the rough set dependency degree, can lead to group and individual

selection based on linguistic labels—more closely resembling human reasoning.

A set-based case-based reasoning (CBR) approach is proposed [14] to tackle the task of text categorization (TC). The initial work of integrating both feature and document reduction/selection in TC using rough sets and CBR properties is presented. Rough set theory is incorporated to reduce the number of feature terms through generating reducts. Two concepts of case coverage and case reachability in CBR are used in selecting representative documents. The main idea is that both the number of features and the documents are reduced with minimal loss of useful information. Experiments on the text datasets of Reuters21578 show that, although the number of feature terms and documents are reduced greatly, the problem-solving quality in term of classification accuracy is still preserved. In average, 43.6% documents can be reduced and the classification accuracy is still preserved.

Chouchoulas [23] proposed a Rough Set-based approach (RSAR) and test it using Email messages; based on their work, Bao developed a rough set-based hybrid method using Latent Semantic Indexing (LSI) and Rough Set theory to TC

Cheng and Zhang [18] propose a method TFERS based on rough set theory and correlation analysis, in which a new formulation for attribute importance is proposed based on the classification capability of attributes. This formulation also avoids the recalculation of attribute importance. In text preprocessing phase, the term vector space representation of text is extended to concept ('synset') level based on Wordnet. As a result, dimension of the feature vector is reduced. A complete text feature extraction method TFERS is proposed, which includes text preprocessing, construction of the text feature vector, calculation of attribute significance, and attributes reduction. In the process of attributes reduction, correlation analysis is incorporated in order to get satisfactory feature reduction. The results of the simulation experiment and text classification show the validity of TFERS.

Shaiei and wang [17] conduct a study on three different document representation methods for text used together with three Dimension Reduction Techniques (DRT). The three Document representation methods considered are based on the vector space model, and they include word, multi-word term, and character N-gram representations. The dimension reduction methods are independent component analysis (ICA), latent semantic indexing (LSI), and a feature selection technique based on Document Frequency (DF). Results are compared in terms of clustering performance, using the k-means clustering algorithm. Experiments show that ICA and LSI are clearly better than DF on all datasets. For word and N-gram representation, ICA generally gives better results compared with LSI. Experiments also show that the word representation gives better clustering results compared to term and N-gram representation. The results show that for all datasets, clustering quality using ICA is better than using LSI in the whole range of

dimensionalities investigated. For low dimensionalities, especially lower than 50, for all datasets, the DF based method has the worst performance among the dimension reduction methods used.

Ren[19] performs dimension reduction using Linear discriminant analysis (LDA) to maximize class separability in the reduced dimensional space.

Haifeng [8] present a weighted method based on the sample distribution, which will make the between-class and within-class scatter matrixes with poor scatter be weighted, to enhance the categorization ability after dimensional reduction and to improve the dimensional reduction effect of linear feature extraction method based on scatter difference. experiment show this method is superior to the original maximum scatter difference method in precision rate and recall rate

Thang and, cheng [7] propose a feature reduction method based on probabilistic mixture model of directional distributions. The main idea states that if documents can be viewed as directional data, so can words in the current context. Attributes of a word data point will be its frequencies of appearance in the documents. A mixture model of von Mises-Fisher distributions is applied for clustering the word space, Which results in a set of mean vectors, each of which potentially represents a group of words of the same topic. A projection matrix is then created based on word-to-mean cosine measure, by calculating the cosine distance of each word-mean vector pair. Hence, after the linear transformation, documents in the reduced-dimension space have the number of attributes equal to the number of potential topics in the document corpus [7]. A mixture of distributions is utilized to decompose the word space into a set of sub-topics, which are represented by their mean vectors. Through this matrix, the document corpus is transformed into a new feature space of much lower dimension. Experiments on various benchmark datasets shows that proposed method performs comparably with Latent Semantic Analysis (LSA), and much better than standard methods such as Document Frequency (DF) and Term Contribution (TC).

There was some work done to produce a universal dimension reduction methods [3].were both feature selection and feature extraction techniques are applied.

Zhu [3] applied feature selection and feature extraction to SVMs. In the feature selection case, experimental results show that when the linear kernel is used for SVMs, the performance is close to the baseline system, and when nonlinear kernel is employed, feature selection methods get the performance decrease sharply. On the contrary, principal component analysis (PCA), one of feature extraction methods, gets excellent performance with both linear and nonlinear kernel functions. It examines the ability of feature selection methods to remove irrelevant features, and combine PCA, a feature extraction method, with SVMs as a solution to the problem of synonym, and study the influence of different kernel functions on the performance of dimension reduction methods. Experimental results over

two different datasets show that when the linear kernel is employed for SVMs, feature selection methods achieve better performance compared to the baseline system. However, when the polynomial kernel is combined with feature selection methods, the performance decreases dramatically, and is much worse than the baseline system. On the contrary, PCA perform well no matter which kernel is employed. Employ dimension reduction methods, feature selection and feature extraction, for SVMs as the preprocessing of text categorization.

Ning and Yang [6], propose feature selection algorithm called Trace Oriented Feature Analysis (TOFA). The main function of TOFA is a unified framework that integrates feature extraction algorithms such as unsupervised Principal Component Analysis and supervised Maximum Margin Criterion. TOFA can process supervised problem and, unsupervised and semi-supervised problems. Experimental results on real text datasets demonstrate the effectiveness and efficiency of TOFA. The main contributions of this paper are: (1) by formulating the feature extraction algorithms as optimization problem in continuous solution space a unified feature extraction objective function, where many commonly used previous work are special cases of this unified objective function; (2) by formulating the feature selection algorithms as the optimization problem in a discrete solution space, we propose a novel feature selection algorithm by optimizing our proposed unified objective function in the discrete solution space; and (3) through integrating the objective function of feature extraction and solution space of feature selection, our proposed feature selection algorithm can find the optimal solution according to the unified objective function. Experimental results on real text datasets show the effectiveness and efficiency of TOFA for text categorization.

2.3. Classification

Text classification is an important task of text processing. A typical text classification process consists of the following steps: preprocessing, indexing, dimensionality reduction, and classification [24]. A number of statistical classification and machine learning techniques has been applied to text classification, including regression models like linear least square fit mapping LLSF [12], K-Nearest Neighbor classifiers [5][9][12], Decision Tress, Bayesian classifiers, Support Vector Machines [2][3][15][19][9][11][6-6], Neural Networks [20][21] and, AdaBoost [15]. SVM has been applied to text classification [11] and achieved remarkable success, [17] proposed a hybrid method used transductive support vector machine (TSVM) and simulated annealing (SA), which selected top 2 thousands high CHI-square value features to form dataset and gained better classification results compared to standard SVM and TSVM.

3. Feature Selection techniques:

3.1 Stemming:

Root based stemming and light stemming results in a considerable reduction in feature dimension. Root-based depends on pattern matching to extract the root of the word after prefix, suffix, and infix removal. This technique reduces over 40% of the feature but it does not preserve the semantic of the features, because the root could generate many words with different meaning. on the other hand, light stemming which strips off prefixes and suffixes without removing the infixes has more ability of preserving the meaning and reduces the feature space 30%, but it still have the problem of dealing with two similar words (in-term of semantic) as different feature because of their difference in infixes.

We developed a technique based on morphological word weights using HMM [25]. A hidden markov model is used to match a word with a pattern and thus remove prefixes and suffixes. The pattern then transformed to a unified pattern called Masdar (مصدر). This technique reduces the feature space by 40% and at the same time preserves the semantic of the features.

3.2 Document frequency (DF)

Document frequency refers to the number of documents that a feature appears in. The selection of features is based on the high value of DF. By experiment 60% reduction achieved by removing terms which occurs only in one document. DF can be used as a criterion for selecting good terms [17]. The main idea behind using document frequency is that rare terms either do not capture much information about one category, or they do not affect global performance. DF is simple and, as effective as more advanced feature selection methods [10-24].

3.3 Term Frequency- Inverse Term Frequency (TFIDF):

Term-Frequency-Inverse-Term-Frequency is formulated as:

$$\text{tf-idf}(t, d) = \text{tf}(t, d) \times \text{idf}(t)$$

Where the term frequency refers to the number of occurrences of term t in document d , and, the inverse document frequency is a measure of the general importance of the term (obtained by dividing the total number of documents by the number of documents containing the term, and then taking the logarithm of that quotient).

The *inverse document frequency*

$$\text{idf}(t) = \log \frac{|D|}{|\{d : t \in d\}|}$$

with

- $|D|$: the total number of documents in the document set.

- $|\{d : t \in d\}|$: number of documents where the term t appears.

A high weight in tf-idf is reached by a high term frequency (in the given document) and a low document frequency of the term in the whole collection of documents; the weights hence tend to filter out common terms. The tf-idf value for a term will be greater than zero if and only if the ratio inside the idf's log function is greater than 1. Depending on whether a 1 is added to the denominator, a term in all documents will have either a zero or negative idf, and if the 1 is added to the denominator a term that occurs in all but one document will have an idf equal to zero.

3.4 X^2 statistic Chi-square

X^2 statistic (CHI) measures the lack of independence between term and category. Then it compares with χ^2 distribution to measure the degree of freedom to judge extremeness.

Term goodness measure is expressed by:

$$X^2(t, c) = \frac{N \times (AD - CB)}{(A + C) \times (B + D) \times (A + B) \times (C + D)}$$

Where,

T: term

C: category

A: number of times t and c co-occur

B: number of time the t occur without c

C: number of times c occur without t

D: number of times neither C nor t occur

N: the total number of documents.

If t and c are independent then X^2 statistic equal 0. The X^2 statistic is computed for each category between each unique term and a training corpus and that category. Then combine the category-specific scores of each term into two scores:

$$X_{avg}^2(t) = \sum_{i=1}^m P_r(C_i) X^2(t, C_i)$$

$$X_{max}^2(t) = \max_{i=1}^m \{X^2(t, C_i)\}$$

3.5 Information Gain:

IG measures the number of bits of information obtained for category prediction by knowing the presence or absence of a term in a document [5]. Giving a corpus of training text, we compute the information gain of each term, and then remove those features whose information gain was less than some pre-determined threshold.

$\{C_i\}_{i=1}^m$ is the set of categories in the target space.
 IG of term t is:

$$G(t) = - \sum_{i=1}^m P_r(C_i) \log P_r(C_i) \\ + P_r(t) \sum_{i=1}^m P_r(C_i|t) \log P_r(C_i|t) \\ + P_r(t) \sum_{i=1}^m P_r(C_i|t) \log P_r(C_i|t)$$

After computing IG, remove terms whose information gain less than predefines threshold.

3.6 Mutual Information (MI)

MI is commonly used in statistical language modeling of word associations and related applications [12].

MI between c and t is defined as:

$$I(t, c) = \log \frac{P_r(t \wedge c)}{P_r(t)P_r(c)}$$

Then,

$$I(t, c) = \log \frac{A \times N}{(A + C) \times (A + B)}$$

Where,

A: number of times t and c co-occur

B: number of times t occurs without c

C: number of times c occur without t.

N: total number of documents.

If t and c are independent then $I=0$.

To measure the goodness of a term two measures are defined:

$$I_{avg}(t) = \sum_{i=1}^m P_r(C_i) I(t, C_i)$$

$$I_{max}(t) = \max_{i=1}^m \{I(t, C_i)\}$$

4. Synonym Merge Reduction

The main idea behind synonym merge is to preserve important terms from being excluded. Term-document matrix is constructed to apply merge over all training set. A dictionary of synonyms [25] is used, were terms with similar meaning are giving one group code. Checking for terms synonym then merging terms with same group id into one feature. Word weights are extracted After a text is being processed and, tokenized [26].

Algorithm (construct synonym tree) Figure 1.

A synonyms tree is constructed from list of synonyms in such a way:

- Null node having 28 branches constituting the Arabic alphabet.
- Each of the 28 node have 28 branches and so on until a word is constructed from root node to the n-1 node (n is the number of letters in a word)
- A leaf node contains the group ID.

Algorithm (synonym check)

Do for all terms

- From root select branch which contains the next word letter.

Continue until last term letter.

- If word does not belong to a synonym group the return null.
- If group found then return group ID.
- Merge all features with the same synonyms group id.

The result of the synonym check algorithm is a semantically reduced feature space. The resulting features are processed with a feature selection method to produce the final feature space that are used for later processing such as classification and clustering.

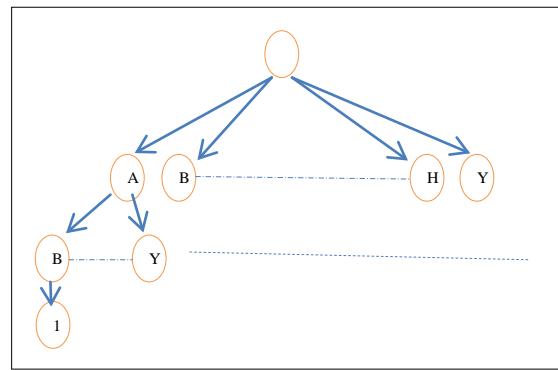


Fig.1 Synonym Tree

5. Dewy based Categorization

The Dewey Decimal Classification (DDC) is a proprietary system of library classification. The DDC attempts to organize all knowledge into 10 main classes [27] (table.1). The ten main classes are each further subdivided into ten divisions, and each division into ten sections, giving ten main classes, 100 divisions and 1000 sections. The system is made up of seven tables and ten main classes, each of which is divided into ten secondary classes or subcategories, each of which contain ten subdivisions.

Table.1: The main 10 classes

Class ID	Class Name
000	Computer science, information and general works
100	Philosophy and psychology
200	Religion
300	Social sciences
400	Language
500	Science (including mathematics)
600	Technology and applied Science
700	Arts and recreation
800	Literature
900	History, geography, and biography

Three levels of classes were used in this research and, the indexes were filtered to match the feature format, which is the morphological weight of the terms.

Algorithm (Dewey Classification)

For all terms in feature space do

- Search the term in the index list.
- If found set class ID to term.

End

Check for terms class ID:

- Choose class(s) which has the majority of terms belonging to.
- A document may belong to more than one class.
- Label document with class label.

An example of the classifying financial documents in Figure.2.

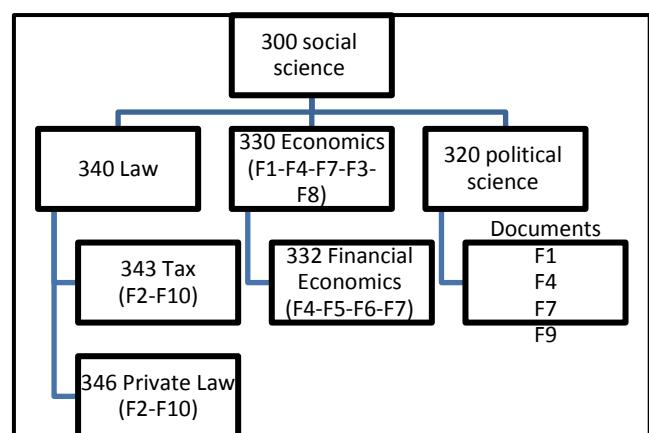


Fig.2 Example of Financial Documents Classification

6. Experiment and results

6.1 Data set:

An in house Arabic documents data set was used in this experiment. Documents represent various categories, news, finance, sport, culture, and science. The documents were tokenized to produce bag of words by eliminating unwanted characters. Then stop words were removed from the resulting documents. Removing stop words reduced the size of the features by 25% in average. Table 1. Shows the number of documents divided into the categories.

Table.2: The Date Set

Category name	Number of documents
News	50
Finance	30
Sport	30
Culture	60
Science	40
Total	210

Several reduction techniques were applied to the data which are shown in table 2.

6.2 Evaluation Criteria:

Text classifiers performance was evaluated using the F1 measure. This measure combines recall and precision in the following way:

Precision:

$$P = \frac{\text{number of correct classes found}}{\text{number of classes found}}$$

Recall:

$$P = \frac{\text{number of correct classes found}}{\text{number of correct classes}}$$

F-measure:

$$F = \frac{2 \cdot R \cdot P}{R + P}$$

6.3 Results:

Figure.1 displays the performance curve for classification of the training set after term selection using IG, DF, MI, TFIDF, CHI respectively. Figure.2 shows the performance curve for classification of the training set for term selection using IG, DF, MI, TFIDF, CHI respectively after applying the synonym reduction processes discussed earlier.

Table.2: The Date Set

Dimension Reduction Technique	Average Percentage of reduction	Average Percentage of reduction w/ syn merge
DF	83%	87%
Tfidf	78%	82%
CHI	90%	93%
IG	89%	91%
MI	60%	72%

Comparison between fig.1, and fig.2 result shows improvement in both number of reduced features and classification performance measure F1-measure.

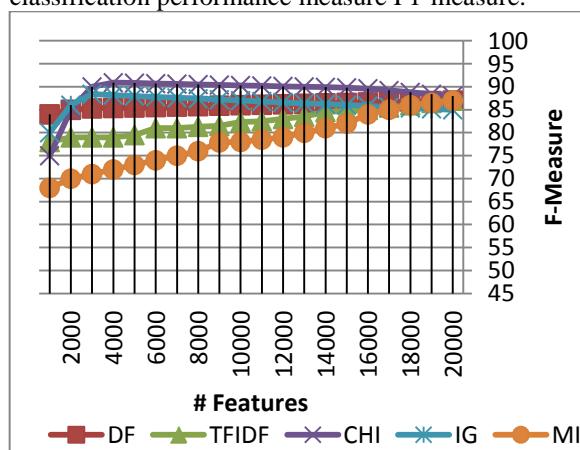


Fig.3 Classification performance without synonym merge

The computation of CHI, DF, and MI are similar to that of IG. The differences are the approaches to rank features. However, MI is not comparable with IG, DF, and CHI on text categorization. CHI and IG gives the best performance overall discussed selection methods.

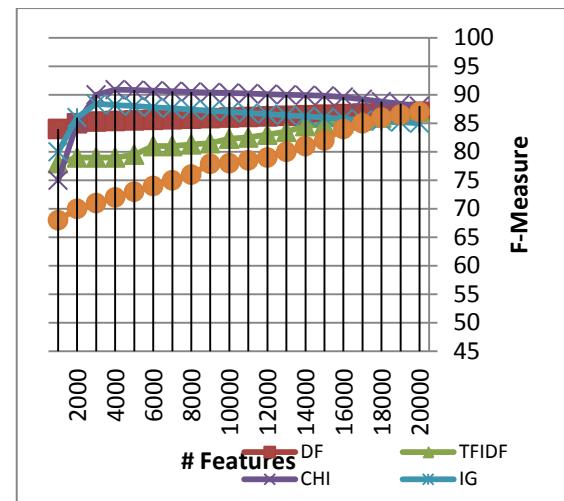


Fig.2 Classification performance after synonym merge

7. Conclusion:

In this paper a semantic feature reduction approach was presented with a Dewey based classification algorithm. The reduction is based on synonyms merge to overcome the problem of feature synonyms excluded during feature selection process. Terms with similar meaning are merged into one group then the resulting groups are used as the new features which results in two advantages, one is reducing the feature space, and the other is preserving the semantic of the feature without relying into complex methods. Five feature selection methods were applied after synonym merges, DF, TFIDF, CHI, IG, and MI to produce a more compact feature space. Experiment on using those methods with and without the synonym merge results in improvement of the feature reduction and the classification performance presented by the F-measure. Furthermore, a new approach based on Dewey indexing is presented. It classifies documents based on filtered version of Dewey indexes, and uses a hierarchy structure of three levels to produce labeled overlapped classes. Over the five features selection methods used CHI, and IG shows the best performance.

As a future word feature extraction methods will be experimented on with the new classification technique to decide the best performance yielding method.

8. References:

- [1] G. Salton, and M.I McGill, "An Introduction to Modern Information Retrieval", McGraw-Hill, 1983.
- [2] YI WANG, and XIAO-JING WANG, "A NEW APPROACH TO FEATURE SELECTION IN TEXT CLASSIFICATION", Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, 18-21 August 2005

- [3] [3] Muhua ZHU, Jingbo ZHU, and Wenliang CHEN, "Effect Analysis of Dimension Reduction on Support Vector Machines", Proceeding of NLP-KE'05
- [4] [4] K. Mugunthadevi, S.C. Punitha, M. Punithavalli, and K. Mugunthadevi, "Survey on Feature Selection in Document Clustering", International Journal on Computer Science and Engineering (IJCSE), 2011
- [5] [5] Cui Zifeng, Xu Baowen , Zhang Weifeng, Jiang Dawei, and Xu Junling, "CLDA: Feature Selection for Text Categorization Based on Constrained LDA", International Conference on Semantic Computing, 2007
- [6] [6] Jun Yan, Ning Liu, Qiang Yang, Weiguo Fan, and Zheng Chen, "TOFA: Trace Oriented Feature Analysis in Text Categorization", Eighth IEEE International Conference on Data Mining, 2008.
- [7] [7] Nguyen Duc Thang, Lihui Chen, and Chee Keong Chan, "Feature Reduction using Mixture Model of Directional Distributions", 10th Intl. Conf. on Control, Automation, Robotics and Vision, Hanoi, Vietnam, 17–20 December 2008
- [8] [8] Liu Haifeng, Su Zhan, Yao Zeqing, and Zhang Xueren, "A Method of Text Feature Extraction Based on Weighted Scatter Difference", Second WRI Global Congress on Intelligent Systems, 2010.
- [9] [9] LI Xi, DAI Hang, and WANG Mingwen, "Two-stage Feature Selection Method for Text Classification", International Conference on Multimedia Information Networking and Security, 2009.
- [10][10] Liu, T., Liu, S., Chen, Z. and Ma, W.Y. An Evaluation on Feature Selection for Text Clustering. In Proceedings of the Twentieth International Conference on Machine Learning (ICML'03), 2003
- [11][11] Yaohong JIN, Wen XIONG, and Cong WANG, "Feature Selection for Chinese Text Categorization Based on Improved Particle Swarm Optimization", IEEE, 2010.
- [12][12] Yiming Yang, and Jan O. Pedersen, "A Comparative Study on Feature Selection in Text Categorization", Proceedings of 14th International Conference on Machine Learning, San Francisco, pp.412-420, 1997
- [13][13] Richard Jensen, and Qiang Shen "Semantics-Preserving Dimensionality Reduction: Rough and Fuzzy-Rough-Based Approaches", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 16, NO. 12, DECEMBER 2004
- [14][14] Yan Lps,I Mon Chi-Keung Shiu, Sankar Kumar Pal, And James Nga-Kwok Liu, "A Rough Set-Based Cbr Approach For Feature And Document Reduction In Text Categorization", Proceedings of the Third International Conference on Machine Learning and Cybernetics, Shanghai, 26-29 August 2004
- [15][15] Yang Song1, Ding Zhou, Jian Huang, Isaac G. Councill, Hongyuan Zha, and C. Lee Giles "Boosting the Feature Space: Text Classification for Unstructured Data on the Web", Proceedings of the Sixth International Conference on Data Mining (ICDM'06)
- [16][16] Ren WangI, Amr M. Youssef , and Ahmed K. Elhakeem, "On Some Feature Selection Strategies for Spam Filter Design", IEEE CCECE/CCGEI, Ottawa, May 2006
- [17][17] Mahdi Shafiei, Singer Wang, Roger Zhang, Evangelos Milios, Bin Tang, Jane Tougas, and Ray Spiteri, "Document Representation and Dimension Reduction for Text Clustering", 2007 IEEE
- [18][18] Yiyuan Cheng, Ruiling Zhang, Xiufeng Wang, and Qiushuang Chen "Text Feature Extraction Based on Rough Set", Fifth International Conference on Fuzzy Systems and Knowledge Discovery, 2008
- [19][19] Cheong Hee Park, "Dimension Reduction Using Least Squares Regression in Multi-labeled Text Categorization", IEEE, 2008
- [20][20] Comparing Dimension Reduction Techniques for arabic text classification using BPNN algorithm. First international conference on integrated intelligent computing, 2010
- [21][21] Savio L. Y. Lam, and Dik Lun Lee, "Feature Reduction for Neural Network Based Text Categorization" , 1999
- [22][22] Yong-gong Ren, Nan Lin, and Yu-qi Sun, "An Improved LAM Feature Selection Algorithm", Seventh Web Information Systems and Applications Conference, 2010.
- [23][23] A. Chouchoula and Q. Shen, "Rough Set-Aided Keyword Reduction for Text Categorisation," Applied Artificial Intelligence, vol. 15, no. 9, pp. 843-873, 2001.
- [24][24] Guiying Wei, Xuedong Gao, and Sen Wu, "Study of text classification methods for data sets with huge features", 2nd International Conference on Industrial and Information Systems, 2010.
- [25][25] A. F. Alajmi, E. M. Saad and M. H. Awadalla, "Hidden markov model based Arabic morphological analyzer", International Journal of Computer Engineering Research Vol. 2(2), pp. 28-33, March 2011.
- [26][26] Mass'ad Abur-Rijaal, A pocket Dictionary of Synonyms and Antonyms, Librairie du Liban Publishers,2007.
- [27]"The Alphabetic Dewey decimal index", Mohammed Sherief, Jomhoria Publishing, 1997.

A Novel Architecture for Quantum-Dot Cellular Automata Multiplexer

Arman Roohi¹, Hossein Khademolhosseini², Samira Sayedsalehi³, Keivan Navi⁴

^{1,2,3} Department of Computer Engineering, Science and Research Branch, Islamic Azad University
Tehran, Iran

⁴ Faculty of Electrical and Computer Engineering, Shahid Beheshti University, G.C.
Tehran, Iran

Abstract

Quantum-dot Cellular Automata (QCA) technology is attractive due to its low power consumption, fast speed and small dimension; therefore it is a promising alternative to CMOS technology. Additionally, multiplexer is a useful part in many important circuits. In this paper we propose a novel design of 2:1 MUX in QCA. Moreover, a 4:1 multiplexer, an XOR gate and a latch are proposed based on our 2:1 multiplexer design. The simulation results have been verified using the QCADesigner.

Keywords: Quantum-Dot Cellular Automata (QCA), Nanotechnology, Circuit Design, Multiplexer, Circuit Simulation.

1. Introduction

In VLSI technology, researchers face the physical limits of conventional CMOS technology. Due to this failure, they have switched to the novel nanotechnologies such as Single Electron Transistor (SET), Carbon nanotube (CNT) and Quantum-Dot Cellular Automata (QCA) [1]. QCA functions are based on Columbic interaction instead of current used in CMOS, so there is no leakage current. Additionally, it has major advantages such as low power consumption, high speed and small space consumption. QCA was presented in [2] for the first time and many circuits and designs have been introduced so far [3]-[5]. The basic structure in QCA is a cell that has four dots positioned at the corners of the squared cell and two free electrons. Each dot can be occupied by one of the two hopping electrons. Since the mutual behavior of the electrons is based on the Columbic interaction, they arrange themselves diagonally in order to reach to the maximum distance. Electrons can tunnel between dots through the barriers but cannot leave the cell; hence, there is no current flow. As shown in Figure 1 two stable polarization (p) states might occur, which represent the binary values "0" and "1".

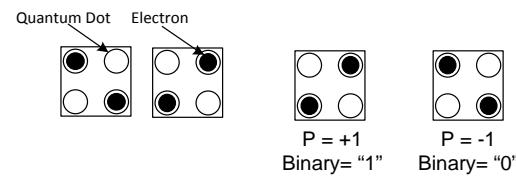


Fig. 1 QCA cell and the two stable polarizations.

The rest of this paper is organized as follows. Section 2 is dedicated to a brief review of previously introduced multiplexer designs. The proposed multiplexer is represented in section 2, as well. In section 3 we demonstrate simulation results and comparisons. Finally this paper is concluded in section 4.

2. QCA Review

In order to implement gates and circuits, QCA benefits from Columbic interaction between cells. An array of cells that are aligned can construct a QCA wire which is shown in Figure 2. The polarization of each cell in a QCA wire is directly affected by the polarization of its neighboring cells on account of electrostatic force. Accordingly, QCA wires can be used to propagate information from one end to another [6].

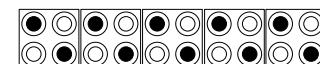


Fig. 2 QCA wire.

Two fundamental QCA gates are the inverter and the majority gate, (see Figure 3 and Figure 4). Many structures are implemented based on these two gates like the AOI [7], the complex gate [8] and one bit QCA full adder [9], [14] and [16].

2.1 Inverter Gate

Figure 3 shows three types of inverter gate; however, since the last one operates properly in all various circuits, it is used more in different designs compared to the two other types. This inverter is made of four QCA wires. The input polarization is split into two polarizations and in the end, two wires join and make the reverse polarization.

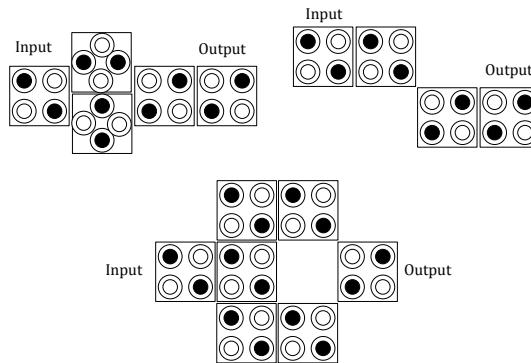
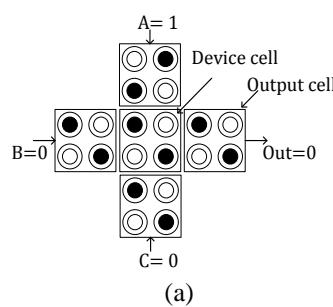


Fig. 3 Three types of inverter gate.

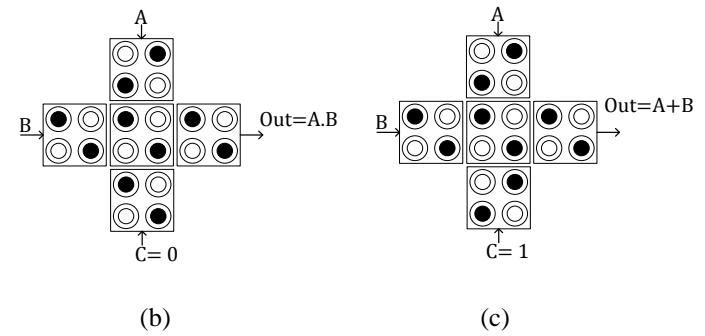
2.2 Majority Gate

Majority gate consists of five cells, three inputs, one output and a middle cell. The middle cell named device cell by reason of its function, switches to major polarization and determines the stable output. Majority gate can be programmed such that it functions as a 2-input AND or a 2-input OR by fixing one of the three input cells to $p = -1$ or $p = +1$, respectively. The Boolean expression of majority gate is as follows:

$$M(A, B, C) = AB + AC + BC \quad (1)$$



(a)



(b)

(c)

Fig. 4 (a) The QCA majority gate. It can be programmed to function as (b) the AND gate or (c) the OR gate.

3. Multiplexer Design

Multiplexer is an important part in implementation of signal control systems and memory circuits, since it allows us to choose one of the inputs and transfer it to the output. The functionality of multiplexer is shown in (2). A and B are the two inputs and Sel is used to select one between the two inputs.

$$Out = A.Sel + B.\overline{Sel} \quad (2)$$

3.1 Previously Presented Multiplexers

Various formerly suggested implementations of a 2:1 multiplexer in QCA, have been studied and a novel efficient design is proposed. The proposed multiplexer is compared with the recent designs presented in [9]-[12] and [15] in terms of area, speed and complexity. Figure 5 shows the previously introduced multiplexer implementations. The Figure 5(a) depicts the multiplexer presented in [9]. As mentioned earlier, many QCA designs including multiplexer can be implemented based on majority gate. The equivalent expression based on majority function is as (3):

$$Out = Maj(Maj(Out1, \overline{Sel}, 1), Maj(Out1, Sel, 0), In0) \quad (3)$$

Two proposed multiplexers based on majority gate presented in [10], [11] are shown in Figure 5(b) and 5(c), respectively. Both of them are constructed in three layers, however, the latter is smaller and therefore more efficient.

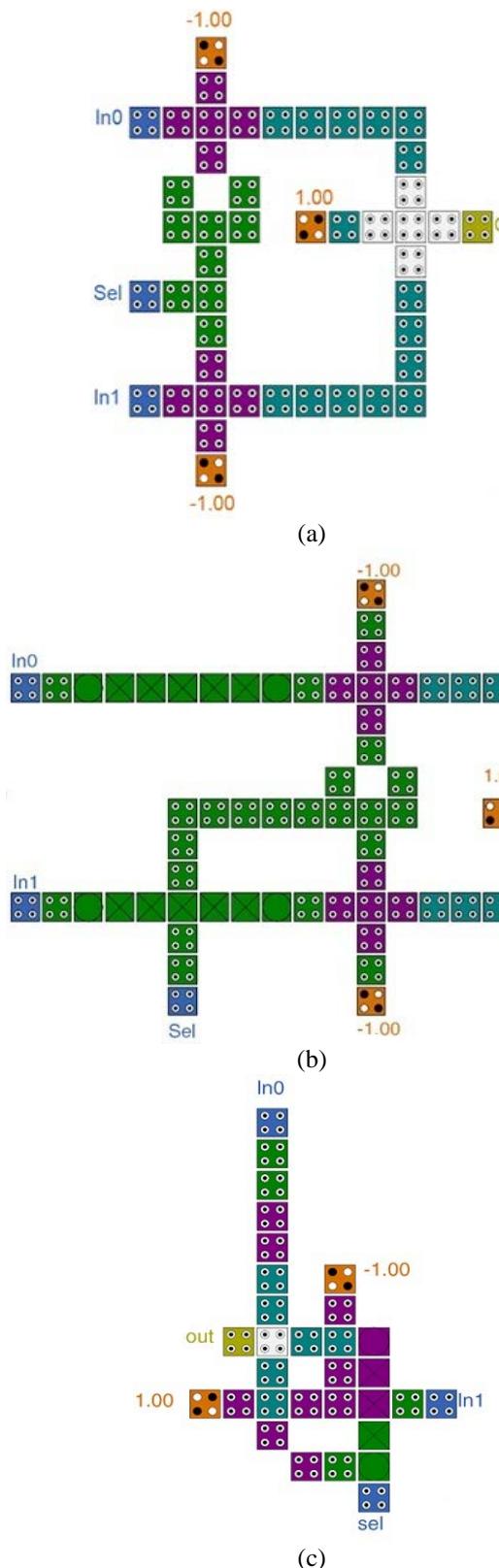


Fig. 5 Presented multiplexer (a) in [9], (b) in [10] and (c) in [11].

3.2 Proposed Multiplexer Design

Our 2:1 QCA multiplexer design is shown in Figure 6(a). The design consists of 27 cells covering an area of $0.03 \mu\text{m}^2$. The proposed QCA multiplexer has been designed and simulated using the QCADesigner tool [12], [13]. A 4:1 multiplexer based on the proposed 2:1 multiplexer is suggested in Figure 6(b). This 4:1 multiplexer is implemented in two stages. We can construct larger multiplexers (8:1, 16:1 and so forth) using our 2:1 multiplexer design.

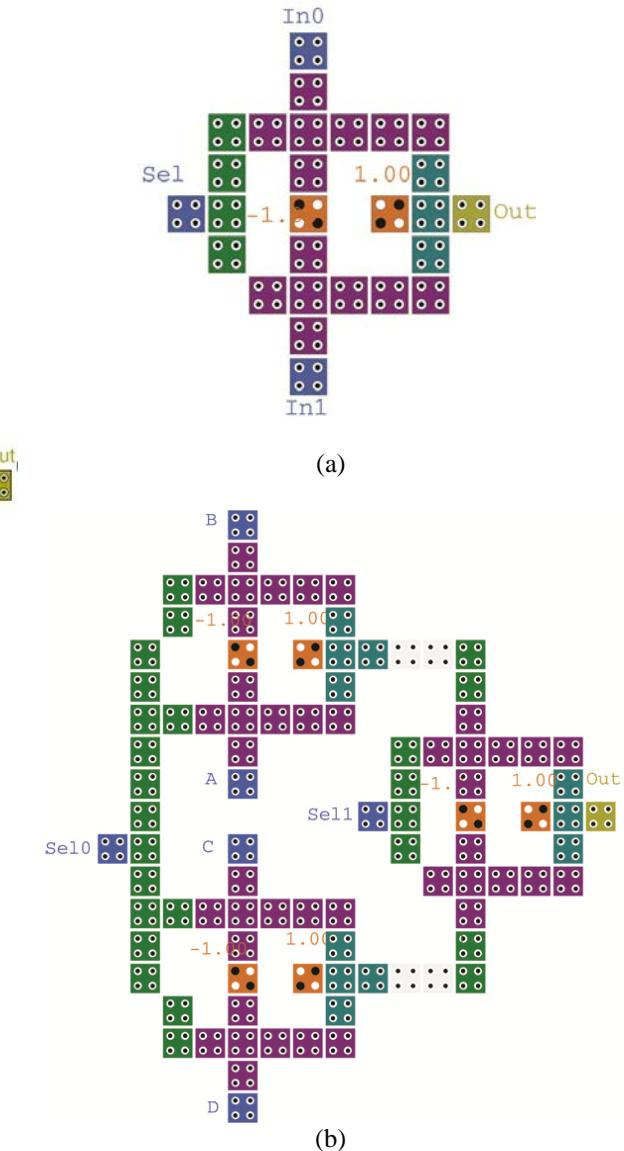


Fig. 6 (a) Schematic of proposed 2:1 multiplexer design, (b) Schematic of a 4:1 multiplexer.

The expression (4) is used for the 4:1 multiplexer. Its truth table is as Table 1.

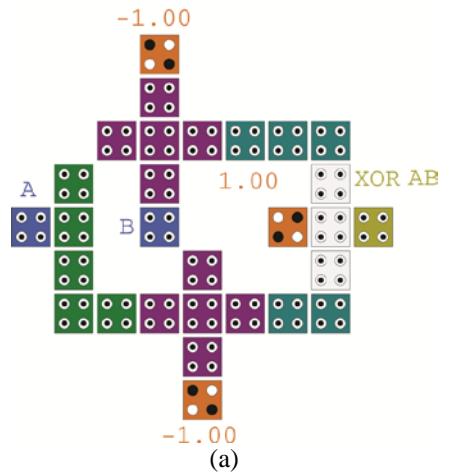
$$Out = AS_1S_0 + BS_1\bar{S}_0 + C\bar{S}_1S_0 + D\bar{S}_1\bar{S}_0 \quad (4)$$

Table 1: 4:1 Multiplexer truth table

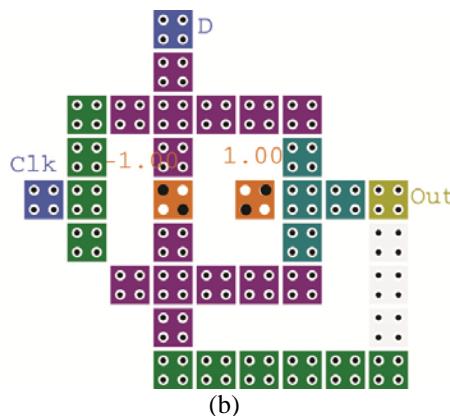
<i>S₁</i>	<i>S₀</i>	<i>Out</i>
0	0	D
0	1	C
1	0	B
1	1	A

3.3 Sample Gates Based on Proposed Multiplexer

Different gates can be designed using our proposed 2:1 multiplexer. Two novel designs which are derived from the proposed multiplexer, are shown in Figure 7(a) and Figure 7(b). The former is an XOR and the latter is a latch.



(a)

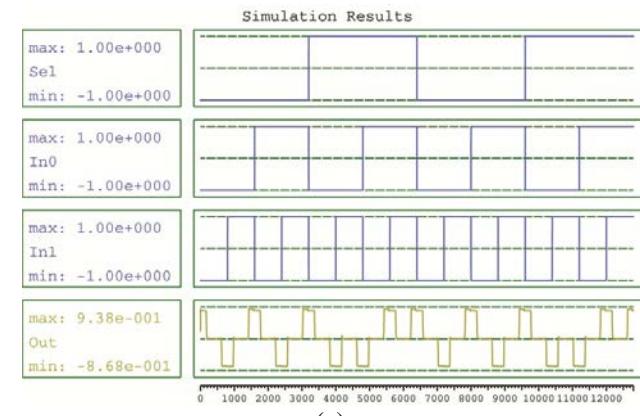


(b)

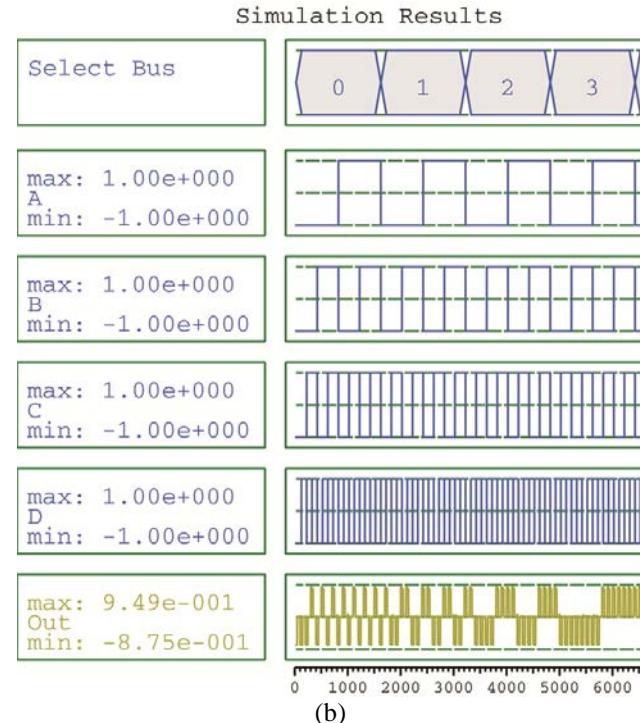
Fig. 7 (a) An XOR and (b) A latch based on proposed 2:1 multiplexer.

3.4 Simulation Result

The following parameters are used for a bistable approximation: cell size=18nm, number of samples=50000, convergence tolerance=0.0000100, radius of effect=65.000000nm, relative permittivity=12.900000, clock high=9.800000e-022 J, clock low=3.800000e-023 J, clock shift=0, clock amplitude factor=2.000000, layer separation=11.500000 and maximum iterations per sample=100. Most of the above mentioned parameters are default values in QCADesigner. Proposed 2:1 multiplexer and 4:1 multiplexer simulation results are provided with the input and output waveforms as shown in Figure 8.



(a)



(b)

Fig. 8 (a) Simulation results of proposed 2:1 multiplexer, (b) Simulation result of 4:1 multiplexer based on proposed 2:1 multiplexer.

Figure 9 shows simulation results of an XOR gate and a latch.

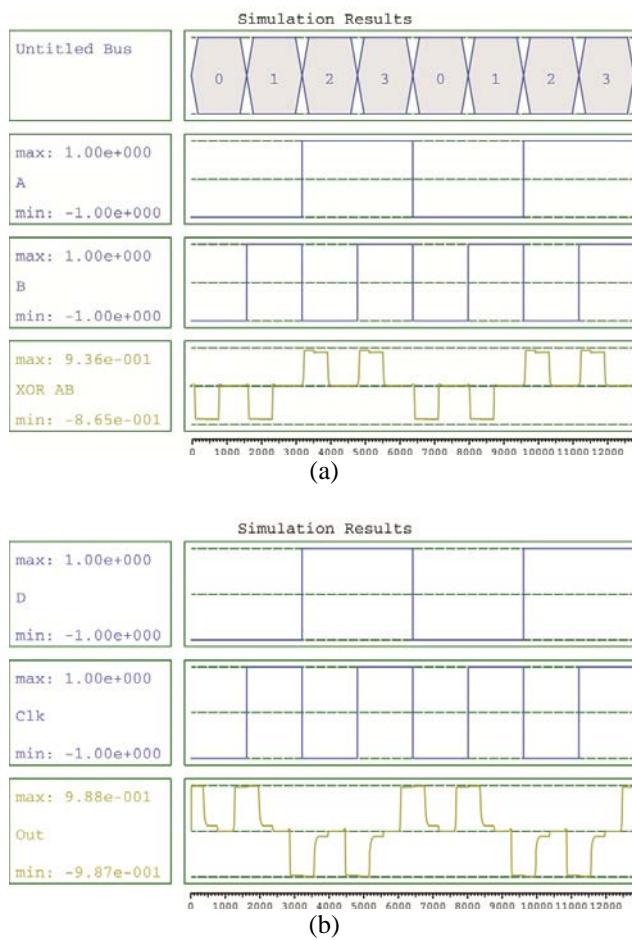


Fig. 9 Simulation results of (a) An XOR and (b) A latch based on 2:1 proposed multiplexer.

It is inferable from simulation results that the proposed multiplexer has achieved significant improvements in QCA circuits.

Table 2: Comparison of recent 2:1 multiplexer designs

2:1 Multiplexer	Cell count	Area μm^2
Handmade presented in [9] Figure 6 (a)	88	0.14
Multiplexer presented in [10] Figure 6 (b)	46	0.08
Multiplexer presented in [11] Figure 6 (c)	36	0.06
Proposed Multiplexer Figure 7	27	0.03

4. Conclusion

Multiplexer is an important and fundamental element in most commonly used circuits. This paper presented a novel and efficient design of 2:1 QCA multiplexer. The proposed multiplexer gate and the other suggested gates which are structured by use of it, have been simulated using QCADesigner and tested in terms of complexity (cell count) and area. As it was apparent in simulation results, the proposed multiplexer has some superiority over the previous common designs in QCA and the comparisons evidently showed significant improvements.

References

- [1] Jing Huang; Momenzadeh, M.; Lombardi, F.; , "An Overview of Nanoscale Devices and Circuits," Design & Test of Computers, IEEE , vol.24, no.4, pp.304-311, July-Aug. 2007.
- [2] C.S.Lent, P.D.Tougaw, W.Porod, and G.H.Bernstein, "Quantum cellular automata," Nanotechnology, vol.4, no.1, pp.49-57, 1993.
- [3] W. Porod, "Quantum-dot devices and quantum-dot cellular automata", Inter. J. Bifurcation and Chaos, vol. 7, no. 10 pp. 2199-2218, 1997.
- [4] Lent, C.S.; Tougaw, P.D.; , "A device architecture for computing with quantum dots," Proceedings of the IEEE , vol.85, no.4, pp.541-557, Apr 1997.
- [5] Tougaw, P. Douglas; Lent, Craig S.; , "Logical devices implemented using quantum cellular automata," Journal of Applied Physics , vol.75, no.3, pp.1818-1825, Feb 1994.
- [6] C.S. Lent, P.D. Tougaw, "Lines of interacting quantum-dot cells: a binary wire", Journal of Applied Physics (1993) 6227-6233.
- [7] Huang, J., M. Momenzadeh, et al. (2004). Design and characterization of an and-or-inverter (AOI) gate for QCA implementation, ACM: 429.
- [8] Townsend, W.J.; Abraham, J.A.; , "Complex gate implementations for quantum dot cellular automata," Nanotechnology, 2004. 4th IEEE Conference on , vol., no., pp. 625- 627, 16-19 Aug. 2004.
- [9] Kim, K.; Wu, K.; Karri, R.; , "The Robust QCA Adder Designs Using Composable QCA Building Blocks," Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on , vol.26, no.1, pp.176-183, Jan. 2007
- [10] Teodosio, T.; Sousa, L.; , "QCA-LG: A tool for the automatic layout generation of QCA combinational circuits," Norchip, 2007 , vol., no., pp.1-5, 19-20 Nov. 2007
- [11] Hashemi, S.; Azghadi, M.R.; Zakerolhosseini, A.; , "A novel QCA multiplexer design," Telecommunications, 2008. IST 2008. International Symposium on , vol., no., pp.692-696, 27-28 Aug. 2008
- [12] QCADesigner Home Page <www.atips.ca/projects/qcadesigner>.
- [13] Walus, K.; Dysart, T.J.; Jullien, G.A.; Budiman, R.A.; , "QCADesigner: a rapid design and Simulation tool for

- quantum-dot cellular automata, "Nanotechnology, IEEE Transactions on , vol.3, no.1, pp. 26- 31, March 2004.
- [14] Swartzlander, E.E.; Heumpil Cho; Inwook Kong; Seong-Wan Kim; , "Computer arithmetic implemented with QCA: A progress report," Signals, Systems and Computers (ASILOMAR), 2010 Conference Record of the Forty Fourth Asilomar Conference on , vol., no., pp.1392-1398, 7-10 Nov. 2010.
- [15] M.A. Tehrani, F. Safaei, M.H. Moaiyeri, and K. Navi, "Design and implementation of Multistage Interconnection Networks using Quantum-dot Cellular Automata", presented at Microelectronics Journal, 2011, pp.913-922.
- [16] S. Sayedsalehi, M. H. Moaiyeri and K. Navi, "Novel Efficient Adder Circuits for Quantum-Dot Cellular Automata", to be published in Journal of Computational and Theoretical Nanoscience, 2011.

Arman Roohi received B.S. degree in computer engineering in 2008 from Shiraz University, Shiraz, Iran. Currently, he is working toward his M.S. degree in computer architecture at Department of Computer Engineering, Science and Research Branch of Islamic Azad University, Tehran, Iran. His research interests are Computer Arithmetic and electronics with emphasis on CNFET, QCA and SET. He is a student member of IEEE. He is also a member of IEEE Computer Society.

Hossein Khademolhosseini received B.S. degree in computer engineering in 2008 from Shiraz University, Shiraz, Iran. Currently, he is working toward his M.S. degree in computer architecture at Department of Computer Engineering, Science and Research Branch of Islamic Azad University, Tehran, Iran. His research interests are computer arithmetic and electronics with emphasis on QCA and VLSI. He is a student member of IEEE. He is also a member of IEEE Computer Society.

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 computer architecture engineering at the Science and Research Branch of IAU. 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.

Keivan Navi received his M.Sc. degree in electronics engineering from Sharif University of Technology, Tehran, Iran in 1990. 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.

A Collaborative Algorithm for Ontological Matching in E-Learning Courseware Domain Knowledge Network

Akanbi Caleb¹ and Adagunodo Rotimi.² Omotosho Lawrence¹ and Adigun Adepeju¹.

¹ Department of Information and Communication Technology, Osun State University, Osogbo, Osun State, 23436, Nigeria

² Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-ifé Osun State ,Nigeria

Abstract

Domain Knowledge is the content repository of a courseware system consisting of a series of learning objects. However, the unstructured and inconsistent naming of domain knowledge components does not permit knowledge transfer across diverse collaborative systems due to differences in architecture, format and representations. To address this identified problem, we formulate an ontological matching algorithm that provides a sharable knowledge in collaborative learning environment in this paper. The Algorithm employs Hybrid Similarity Measure to compute both Concept and Relational similarity values of the various input graphs-learning objects.

Keywords: Courseware Systems, Domain Knowledge Network, Similarity Measure, Ontological Matching.

1.0 Introduction

e-Learning Courseware System (e-LCS) is an educational software platform for implementing e-learning programmes in tertiary institutions and cooperate organisations. It can organize, present, manage, evaluate the contents of courses and teaching activities, to promote the interaction between students and teachers[1]. e-LCS simulates the traditional learning classroom environment by using syllabus, schedule, course notes, examples, assignments, etc. In addition, it has on-line facilities such as online assessment and multimedia course delivery over the networks which are additional advantages over lecture only classes. Courseware major components are learner's model, on-line facilities such as online assessment, multimedia course delivery tools/environment for collaboration, learners model, facilitators models, pedagogical contents and domain knowledge [2].

Learning object is a smallest digital reproducible and addressable resources of a learning contents stored in various Knowledge Base of Learning Management Systems[3]). Domain Knowledge is a learning materials or content repository of a courseware system. It consists of a series of learning objects arranged in hierarchical format [4]). The knowledge network represents the content to be covered in e-learning system. Usually, it is always in hierarchical, tree or graph forms. The knowledge network represents concepts at different abstraction levels and combines multiple relationships such as *is-a*, *part-of*, *contained-in*, *associate -with*, *related-to*, *example-of*, *applicable-to*, *easier-than*, etc. in a single intuitive hierarchy[5,6,7,8]

The high rate of e-learning platforms implies that increasing complex and dynamic web based

infrastructures need to be managed more efficiently [9]. However, the unstructured and inconsistent format of domain knowledge components does not permit knowledge transfer across diverse collaborative systems due to differences in architecture, format and representations. To address this identified problem, we formulate an ontological matching algorithm in e-learning courseware systems that provide a sharable knowledge in collaborative learning environment in this paper.

1.1 Structure of a Domain knowledge

Domain knowledge is composed of other knowledge using some operations called associations or aggregations. Associations are some relationships defined on "simpler knowledge". Two basic categories of knowledge are ; *unit and aggregated* knowledge.). Learner/ Teacher treats it as indivisible item in the given context, unit cannot be decomposed into smaller parts . Aggregated knowledge is composed of either of atomic knowledge units or of other "lower-level" aggregated knowledge. Any kind of knowledge can be grouped and sequenced into clusters. Such a cluster is aggregated knowledge too.

In the view of [10], Learning object (LO) model consists of three basic parts: name, knowledge-based interface and knowledge-based body. The first, such as the topic/theme name, is for identification and referencing or the learning objective statement (e.g., as the context of the name). Interface is for communicating and transferring knowledge to the LO and from it. Some learning objects (knowledge) in that part may be teacher-oriented (e.g., tasks for test, answers, etc.) while the other objects are student-oriented. What is common to all the above mentioned structural units of the model is that within each section the information is strongly clustered into blocks.

As teacher and learner communicate knowledge, interface can be seen also as a media for the teacher/learner interaction, consists of two kinds of knowledge: input and output knowledge. From the learner's perspective, output knowledge is shift in time with respect to input knowledge within a given learning process when it is initiated. The conceptually input knowledge is called 'prerequisites' in pedagogy theories [11]). A typical structure of learning object(LO) model is consists of knowledge-based interface, knowledge-based body, Declarative part, Procedural part, Visibility layers, Contextual part and Managerial part.

1.2 Ontology Matching in Knowledge Network.

The benefits of ontological representation of domain knowledge lie in its capabilities of explicitly defining concepts and their attributes and relationships. Coupled with new information technologies, such representation can be encoded in ways that allow for direct conversion into implementation models. This in turn requires ontological modeling of domain knowledge not only to cover the content but also to take into consideration of how the content is to be used and interacted with users or other courseware systems . Ontological modeling for learning objects may be divided into three broad areas: content, presentation, and application.

Matching is a critical operation in many application domains, such as semantic web, schema/ontology integration, data warehouses, e-commerce, query, mediation, etc. It takes as input, two schemas/ontologies, each consisting of a set of discrete entities (e.g., tables, XML elements, classes, properties, rules, predicates) and determines as output the relationships (e.g., equivalence, subsumption) holding between these entities [12]

According to Shvaiko (2004), A *matching element* is a 5-tuple: $(id, e, n, R, P(t, s))$ where
 id - is a unique identifier of the given mapping element;
 e - are the entities (XML elements) of the first ontology e' and the second ontology e'' respectively.
 n - is a *confidence measure* in some mathematical structure typically in the $[0,1]$ range holding for the correspondence between the entities e' and e''
 R - is a relation (e.g., *equivalence*; *disjointness*; *overlapping*) holding between the entities e and e'
 r - is an *alignment matching* parameter which is an external resource(s) used by the matching process to extend the definition of the matching process (synonyms).
 p - represents the matching parameters (such as weights and thresholds)

The match operation is defined as a function that takes two schemas S_1 and S_2 as input graph and returns a mapping between those two schemas as output. Figure 2 shows the match operation where p represents the matching parameters (such as weights and thresholds) and r represents the external resources used by the matching process, e.g. thesauri and synonyms, etc.

1.3 Graph Similarity in a Knowledge Network

Graphs are widely used to represent complex structures that are difficult to model. In a labeled graph, vertices and edges are associated with attributes, called labels. In a Knowledge Network, the attributes could be tags in XML documents, atoms and bonds in chemical compounds, genes in biological networks, and object descriptors in images. Using labeled graphs or unlabeled graphs depends on the application need [13],[14].

The vertex set of a graph G is denoted by $V(G)$ and the edge set by $E(G)$. A label function, l maps a vertex or an edge to a label. The size of a graph is

defined by the number of edges it has, written as $|G|$. A graph G is a subgraph of G' if there exists a subgraph isomorphism from G to G' , denoted by $G \subseteq G'$. G' is called a super graph of G . Given a graph database and a query graph, we may not find a graph (or a few graphs) in the database that contains the whole query graph. Thus, it would be interesting to find graphs that contain the query graph approximately, which is a substructure similarity search problem.

Definition 1: Substructure similarity search.

Given a graph database $D = \{G_1, G_2, \dots, G_n\}$ and a query graph Q , similarity search is to discover all the graphs that approximately contain query graph Q . Reverse similarity search is to discover all the graphs that are approximately contained by this query graph. To distinguish a query graph from the graphs in a database, we call the latter, target graphs.

The structure-based similarity measure directly compares the topology of two graphs. However, since this measure takes structure connectivity fully into consideration, it is more accurate than the feature-based measure [15].

2.0 Literature Review

An ontology according to [14] is an explicit formal specification of a conceptualization and abstract and simplified vision of the world to be represented. Thus, an ontology permits the capturing of knowledge regarding a concrete domain.

Many ontological structure methods have been employed in the courseware domain knowledge modelling. [4] proposed four layer of educational meta model ontology consisting of four interconnected model types in a hierarchical form. These are: Course layer, Module layer, Concept layer (structured as conceptual network), Instruction Microfunction layer. Matching approaches could be Element matching, Structure matching, Hybrid matching approach combines different approaches together, using a hybrid matcher or a composite matcher. A hybrid matcher integrates several approaches based on multiple matching criteria, while a composite matcher combines multiple match results produced by different matchers, including hybrid matchers

A number of graph similarity algorithms have been proposed in the literature [15], [16] These are:

- (i) Three Edit Distance Algorithm which finds the similarity of two trees by computing the minimum total number of edit operations to transform one tree into the other by applying three elementary operations, namely insertion, deletion and relabeling. However, the tree edit distance algorithm has the problem of a “scattering effect” and mainly focuses on node-labeled.
- (ii) The Weighted Tree Similarity Algorithm computes the similarity between node-labeled, arc-labeled and arc-weighted trees. The algorithm can be applied to various environments where weighted trees are used. The algorithm calculates the edit distances between Undirected Acyclic Graphs (UAGs), which is a natural extension of the edit distance for strings and trees.
- (iii) Similarity Flooding Algorithm Similarity Flooding Algorithm is a graph matching algorithm in which the input schemas are converted into directed acyclic labeled graphs. The algorithm uses iterative fix point computations called *similarity measures* to determine the

similarity of nodes, relying on the intuition that whenever any two elements are found to be similar, the similarity of their adjacent elements increases [4]

During a match operation, similarity measures are used to determine the level of similarity of ontologies O_1 with O_2 . In the view of [16]), a similarity measure for comparison of ontology entities could be object equality, explicit equality, string equality and dice coefficient [17].

3.0. MACA Collaboration Algorithm

The objective of this paper is to formulate a collaboration algorithm for the mediating agent architecture (MACA). The interaction protocol between the requesting courseware, provider courseware and MACA in a collaborative environment is represented by the UML sequence diagram in figure 3.9.

The Multi-Agent System Architecture is shown below consisting of Courseware Mediator Interphase(CMI), Collaboration Agent Wrapper Module (CAWM), Collaboration Agent Match Module(CAMM), Recommender Engine. The details has been discussed in [18]. The architecture consists of two main processes-Match and Filtering processes.

The System Architecture adopts the following protocol interaction.

- (i) Filter agent receives ALQ and creates collaboration agent to execute
- (ii) MACA collaboration agent migrates to other courseware in the collaborative environment, to search for similar learning objects (LO_j) one after the other.
- (iii) Collaboration Agent Wrapper Module (CAWM) generates the XML file format for the active learner query file (ALQ_i), and the learning object files $LO_{j(1,...,n)}$ found in P-courseware.
- (iv) Collaboration Agent Match Module(CAMM) computes the similarity values(MatchSimVal) between ALQ_i and $LO_{j(1,...,n)}$ within confidence measure of interval[0,1] in order to determine the degree of similarity. This is achieved by the Hybrid Graph Similarity Flooding Algorithm (HGSFA) discussed in section
- (v) Filter agent searches for similar cases within the knowledge base Recommended Learning Object Repository (RLOR) that matches the ALQ and stores them and their similarity value in MLOR
- (vi) filter agent receives the computed learning objects $LO_{j(1,...,n)}$ and their computed MatchSimVal returned by the collaboration agent and stores them in Match learning Object Repository (MLOR).
- (vii) Filter agent computes and recommends to the learners, learning objects with 5-top similarity values. This is achieved by:
 - generating a ranked list of returned $LO_{j(1,...,m)}$ based on their MatchSimVal using sort algorithm.
 - selecting the $(LO_{j(1,...,n)})$ with MatchSimVal \geq threshold. In this study the researcher set the threshold at 0.5.

(viii) Active Learner can select any of $LO_{j(1,...,m)}$ he wishes to study out of the maximum 5-top learning objects recommended .

(ix) Collaboration process is established between the R-courseware and the P-courseware with the selected similar learning object.

3.1 Ontological Similarity Model

The graph similarity flooding scheme is employed in this study to model the proposed collaboration algorithm. The algorithm compares both the concept and structural matching using a dice coefficient similarity measure. The measures for similarity computation can be divided into two generic groups namely *Concept/Lexical Measures* and *Relational/Structural Measures* (Vargas-Vera et al., 2004). Lexical measures are based on surface similarities such as that of the title, label of entities(learning objects). The main idea in using such measures is that usually, similar learning objects have similar names and descriptions across different ontologies.

On the other hand, structural measures try to recognize similarities by considering the kinship of the components and structures residing in the ontology graphs of the knowledge base. Leveraging other available information in two ontologies, they hope to recognize related entities outside the site of the lexical measures. In this study similarities our algorithm rely on the intuition that *elements of two distinct models are similar when their adjacent elements are similar*.

In this study, the collaboration algorithm proposed(called MACA_ Algorithm) employs hybrid graph similarity flooding algorithm(HGSFA) equation in 3.1 to compute both *Concept* and *Relational values* for the input graphs. HGSFA builds the graph for the entities (ALQ and P-courseware learning object) and compute both concept and relational similarities (all in XML.format).

For concept similarity, if the nodes are the same, it is represented as 1, if the nodes are not, it is represented as 0. The relational similarity is determined by dice coefficient equation in 3.2 .

$$\text{Dice Coefficient} = \frac{2 * \sum a_i b_i}{(\sum a_i^2 + \sum b_i^2)} \quad (1)$$

where vectors a_i and b_i are the numbers of concept nodes of graph G1 and G2 respectively. The model for the Hybrid Graph Similarity Flooding Algorithm HGSFA is represented as;

$$H = \begin{cases} 0 & \text{no common concepts} \\ 1 & \text{same set of concepts, otherwise} \end{cases} \quad (2)$$

$$\text{sim_dice}(A,B) = \frac{2 * \sum a_i b_i}{(\sum a_i^2 + \sum b_i^2)}$$

The similarity value is therefore the average of the computed concept similarity value and relational similarity value. i.e. MACA_Similarity Value = (Concept_SimVal + Relational_SimVal)/2 (3)

3.3 System Algorithm

The focus of this paper is to present the algorithm of the collaboration algorithm for the architecture. The collaboration algorithm uses Hybrid Graph Similarity Flooding(HGSF) scheme that employs Dice Coefficient Similarity Measures. This Hybrid algorithm has two

main procedures – Filtering and Match. The two are used for ontological matching of domain knowledge network. The input and output automaton of the proposed collaboration algorithm is shown in figure 3

4.0 Conclusion

One of the goals of collaboration in e-learning system is to provide sharing of learning objects across diverse courseware domain knowledge. This paper proposed an ontological matching algorithm for learning object sharability across collaborative courseware e-learning objects. The algorithm employed a dice coefficient measure to compute the Concept and Relational values of the objects

References

- [1] J.H. Li, Y. Zhao, On the Course Management System (CMS), Modern Educational Technology, 9(18) (2008) 64-71
- [2] Monari M.(2005): Evaluation of Collaborative Tools in Web-Based E-Learning Systems M.Sc thesis, Department of Numerical Analysis and Computer Science, Royal Institute of Technology Stockholm, Sweden.
- [3] E.J.Koper, E.J.(2003).: Combining re- usable learning resources and services to pedagogical purposeful units of learning. In A.Littlejohn (Ed.), Reusing Online Resources: A Sustainable Approach to eLearning London Kogan Pgs. 46–59
- [4] Kanongchaiyos P.(2007): Calculation of Document Similarity using Cellular Structured Space Template. Proceedings of the third/ASTED International Conference Advance in Computer Science and Technology, Phuket, Thailand.
- [5] **P. Fournier-Viger, A. Mayers, M. Najjar, R. Nkambou.** A Cognitive and Logic Based Model for Building Glass-Box Learning Objects. Interdisciplinary Journal of Knowledge and Learning Objects, Vol. 2, 2006, 77-94.
- [56] **V. Ponkratius.** Aspect-Oriented Learning Objects. Proc. of the IASTED International Conference on WEB-BASED EDUCATION, February, Grindelwald, Switzerland, 2005, 21-23.
- [7] **A. Berlanga, F.J. Garcia.** A Proposal to Define Adaptive Learning Designs. Proceedings of Workshop on Applications of Semantic Web Technologies for Educational Adaptive Hypermedia (SW-EL 2004) in AH 2004, 23 August, 2004, Eindhoven, The Nether-lands. TUE Computer Science-Reports 04-19 AH2004: Workshop Proceedings Part II.
- [8] Dolog P. Hese N., Nejdl W. and Sintek M., (2004): Personalisation in distributed e-learning Environments . In Proceeding of the 13th International Word Wide Web Confrence New York, USA.
- [9] Štuikys V. and Damaševičius R. (2007): Towards Knowledge- Based GenerativeLearning Objects Learning Objects, Information Technology and Control Vol. 36 No 2
- [10] Merrill M.D.(2000) : Instructional Transaction Theory (ITT): Instructional Design Based on Knowledge Objects. C.M. Reigeluth (Ed.), Instructional-Design Theories and Models: A New Paradigm of Instructional Theory, Mahwah, NJ: Lawrence Erlbaum Associates
- [11] Raymond J., Gardiner E. and Rasca P. (2002): Calculation of graph similarity using maximum common edge subgraphs. The Computer Journal, Vol 45:pp 631–644.
- [12] Kalfoglou Y. and Schorlemmer M.(2003): Ontology Mapping: The State of The Art. The Knowledge Engineering Review Journal (KER), Vol18 No. 1, pp1–31.
- [13] Shasha D. Wang J, and Giugno R. (2002): Algorithmics and applications of tree and graph searching. In Proc. 21st ACM Symposium on Principles of Database Systems (PODS'02), pages 39–52.
- [14] Melnik, S., Molina, H.G. and Rahm, E.(2002): Similarity flooding: A versatile graph matching algorithm, Proceedings of Eighteenth International Conference on Data engineering San Jose, California, pp. 117-128.
- [15] Bhavsar, V.C. Boley, H. and Yang, L.(2004): A Weighted-Tree Similarity Algorithm for Multi-Agent Systems in e-Business Environments, Computational Intelligence, Vol 20 No 4, pp.584-602.
- [16] Qin, J. & Finneran, C. (2002). Ontologicalrepresentation of learning objects. In Proceedings of the Workshop on Document Search Interface Design and Intelligent Access in Large-Scale Collections, July 18, Portland, OR. <http://xtasy.slis.indiana.edu/jcdlui/uiws.html>
- [17] C.O. Akanbi C.O., E.R. Adagunodo. and O.Omotosho.(2009) : An Intelligent model for Web-based learning management systems International Journal of computer science and Issues Vol 8 issues 4
- [18] Vargas-Vera M. and Motta E. (2004): AQUOntology-based Question Answering System. Third International Mexican Conference on Artificial Intelligence (MICAI-2004), LNC2972, Springer Verlag.

Authors

First Authors : Dr. Akanbi C. Olufisoye.

Department of Information and Communication Technology, Osun State University, Osogbo is a lecturer I in the above named Dept. He obtained his first degree (B. Sc. Mathematics) in 1991, from University of Ilorin , He holds M.Sc. and Ph.D degrees in Computer Science in 1999 and 2010 respectively from Obafemi Awolowo University, Ile-Ife. He is the Acting Director of Information Management and Technology Centre Osun State University Osogbo. His area of specialization is

electronic learning (e-learning), collaborative Learning , Multi-Agent System, and Artificial Intelligence .

Second Author : Prof. Adagunodo E. Rotimi.

Department of Computer Science and Engineering ,Obafemi Awolowo University, Ile-ifé is a Prof. of computer science. He obtained his B.Sc, M.Sc and Ph.D in Computer Science from the same University. He is currently the Head of department of Computer Science and Engineering, He has supervised many M.Sc and Ph.D students and published in many international and local journals. Prof. Adagunodo area of research works include operating system, e-learning and soft computing.

Third Author Mr Omotosho Lawrence O

Mr Omotosho is a Lecturer in the department of Information and Communication Technology, Osun State University, Osogbo, Nigeria. *He obtained his first degree*

(B. Agric. Soil Science) in 1990, PGD and M.Sc. in Computer Science in 1995 and 2001 respectively from Obafemi Awolowo University, Ile-Ifé. He is currently a PhD student in the Department of Computer Science & Engineering, Obafemi Awolowo University, Ile-Ifé. His areas of specializations are Health Informatics, Mobile Computing and Software Engineering.

Fourth Author: Mrs. Adigun Adepeju, A.

Department of Information and Communication Technology, Osun State University, Osogbo is a lecturer in the above named department. She obtained the degrees of B. Sc. (Computer Science) from University of Ilorin, M. Sc. (Computer Science) and M.Pil (Computer Science) both from University of Ibadan. Her area of specialization is Human Computer Interaction (HCI), Operating System (OS) and Collaborative system.

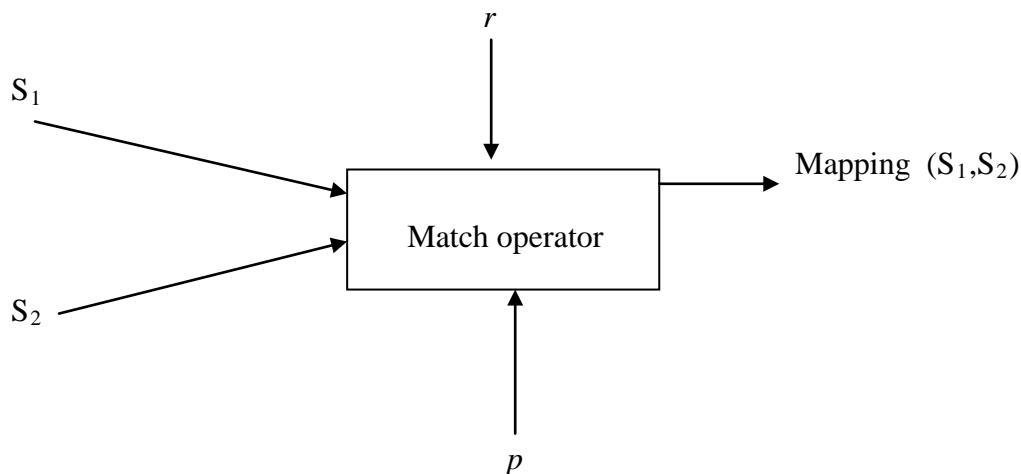


Figure 2: The Match Operation

Input ((Req_Mediating_Agent-CMI (reqCourswID, ALQ)) Accept learner Query

Precondition:

Effect: delQ: append(delQ (reqCourswID, ALQ))

Search (Simcase(RMOR, ALQ))

Precondition:

Effect: SearchSimCase = append (MLOR,(MatchCase,MatchSimVal)

Deploy((Collab_agent(ProvCourswID, delQ))) Collaboration Agent migration

Precondition:

Effect: SearchProvCoursw = append(ProvCourswID, reqCourswID, ALQ)

Generate(ProvCoursew(Wrapper(delQ.Xml, ProvCoursewKnowlBaseOntol.Xml)))

Conversion to XML formats

Precondition:

Effect: WdelQ:Append(delQ.Xml, ProvKnowlBaseOntol.Xml)

Compute(MatchSimVal(WdelQ)_i) for all ProvCoursw_i (i=1 to n) visited

Precondition: Call (MACA_ Similarity Algorithm)

Effect: T_MROR_i = Append (MatchSimAlgorithm(WdelQ)_i)

Send{MLOR (T_MROR)} for all ProvCoursw_i (i=1 to n)

Precondition

Effect: MLOR: Append (T_MROR, ProvCourswID)_i

Perform (Filtering _ Algorithm(Sort(MLOR)

Precondition: Call (Procedure Filtering & Procedure Quick sort)

Effect: RLOR =append(Filtering_Algorithm(SortAlgorithm(MLOR))

Send {(result (Max_5-top(RLOR),ReqCoursew)

Precondition:

Effect: RecQ=Append 5-top(RLOR, ProvCourswID (ReqCoursew)

Figure 3: Collaborative Algorithm for Ontological Matching

Procedure MACA_Similarity Algorithm;

Compute Concept_SimVal:

$$\text{Concept_similarity} = \begin{cases} 0 & \text{no common concepts} \\ 1 & \text{same set of concepts , otherwise} \end{cases}$$

Where vectors A and B are filled with the number of concept nodes of graph G1 and G2 respectively.

Compute Relational_SimVal:

$$\text{Ontol_similarity} = di = \text{sim_dice}(A,B) = 2 * \sum \ln a_i b_i / (\sum \ln a_i^2 + \sum \ln b_i^2)$$

If ontol_relation $\neq 0$

Then evaluate_query(ontol_relation(β_1, β_2))

else

Obtain synonyms for L_query using FODOC thesaurus

Ask active learner to select choice from FODOC thesaurus provided

Call evaluate_queryl(selected_sense(β_1, β_2))

Compute MACA_SimilarityVal = (Concept_SimVal + Relational_SimVal)/2

Procedure Filtering

Input all Match_SimVal

Rank Match_SimVal perform quick sort algorithm

Threshold_Val = 0.5

Display all Match_SimVal \geq Thresold_Val

Recommend maximum of 5-top ranked Match_SimVal, learning objects and the collaborative courseware for the learner

If 5-top ranked Match_SimVal, learning objects not found threshold Val refined to 0.2 and list learning objects

Blind Recognition Algorithm of Turbo Codes for Communication Intelligence Systems

Ali Naseri, Omid Azmoon and Samad Fazeli¹

¹ Department of Information and Communications Technology, Imam Hossein University
 Tehran, Iran

Abstract

Turbo codes are widely used in land and space radio communication systems, and because of complexity of structure, are custom in military communication systems. In electronic warfare, COMINT systems make attempt to recognize codes by blind ways. In this Paper, the algorithm is proposed for blind recognition of turbo code parameters like code kind, code-word length, code rate, length of interleaver and delay blocks number of convolution code. The algorithm calculations volume is $0.5L^3 + 1.25L$, therefore it is suitable for real time systems.

Keywords: Blind Recognition, Turbo Code, Convolution Code, Recursive Systematic Convolution (RSC) Code, Interleaver.

1. Introduction

Communication Intelligence systems (COMINT) play important role in electronic warfare. COMINT systems receive unknown signals and make attempt to demodulate, decode and decipher by blind ways.

Channel coding is used in communication systems to correct errors. Several methods have represented for channel coding by various aims and conditions. Because of complexity of decoding, turbo codes are custom in military communication systems. COMINT systems must be able to blind recognition of turbo codes. In this paper, an algorithm is proposed to finding turbo code parameters like code rate, length of interleaver and delay blocks number of convolution code. In continue turbo code is explained and then the proposed algorithm is described, simulated and evaluated [1].

Turbo decoder and encoder are illustrated in figure 1, 2 [2] [3]. In turbo encoder, RSC & interleave are used. RSC codes at first was suggested by Berrou[4]. The code rate change number of RSC and interleaver blocks. For example for 1/3 code rate two encoder RSC blocks and one interleaver block is needed. If X is the string of input bits, output of branches of encoder will be:

$$v^{(0)} = X = (v_0^{(0)}, v_1^{(0)}, \dots, v_{k-1}^{(0)})$$

$$\begin{aligned} v^{(1)} &= (v_0^{(1)}, v_1^{(1)}, \dots, v_{k-1}^{(1)}) \\ v^{(2)} &= (v_0^{(2)}, v_1^{(2)}, \dots, v_{k-1}^{(2)}) \\ &\dots \\ v^{(n)} &= (v_0^{(n)}, v_1^{(n)}, \dots, v_{k-1}^{(n)}) \end{aligned} \quad (1)$$

And string of output bits will be:

$$V = (v_0^{(0)}, v_0^{(1)}, \dots, v_0^{(n)}, v_1^{(0)}, v_1^{(1)}, \dots, v_1^{(n)}, \dots, v_{k-1}^{(0)}, v_{k-1}^{(1)}, \dots, v_{k-1}^{(n)}) \quad (2)$$

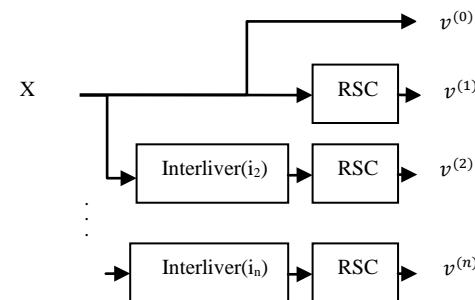


Fig. 1: Outline of a turbo code encoder [2] [3].

In turbo decoder, the noisy systematic and parity bits decode in two stages and message bits are estimated. Decoding stages uses different BCJR (Bahl, Cocke, Jelinek and Raviv) algorithm to solve detection problem by MAP (Maximum A Posteriori probability).

2. Proposed Algorithm

Flowchart of algorithm is illustrated in figure 3. Input coded string bits are named L. Matrix H contains encoded string bits and have N_c columns and N_r rows, that in first step $N_c=1$, $N_r=L$. Matrix H is divided to two matrices Z ($N_c * N_c$) and Y ($((N_r - N_c) * N_c)$):

$$H = \begin{bmatrix} [Z] \\ [Y] \end{bmatrix}$$

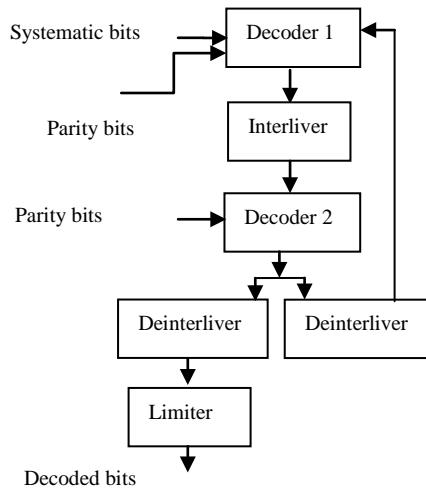


Fig. 2: Outline of a turbo code decoder [3].

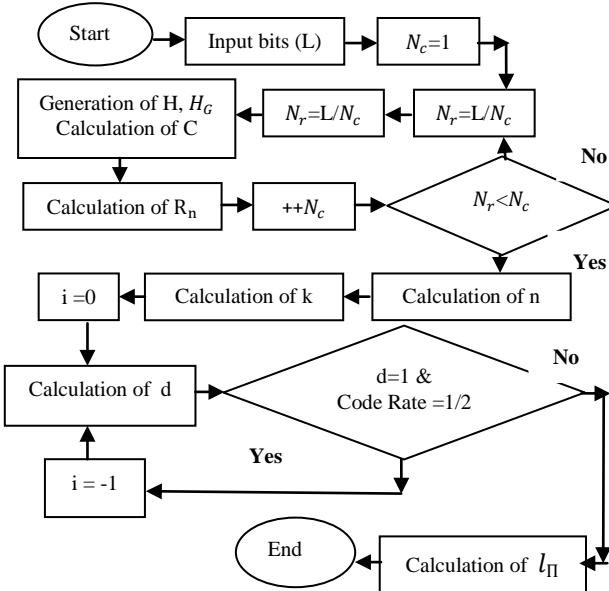


Fig. 3: Flow chart of blind recognition algorithm of turbo code.

Z converts to lower triangular matrix by Gauss-Jordan method and result is named Z_1 . Matrix H_G defines as follows:

$$H_G = \begin{bmatrix} [Z_1] \\ [Y] \end{bmatrix}$$

Algorithm runs until $N_r < N_c$, then R_n calculated:

$$R_n = \frac{N_c - c}{N_c} = \begin{cases} 1 & \text{if } N_c \neq pn \\ < 1 & \text{if } N_c = pn \end{cases} \quad (3)$$

That C is number of zero columns of matrix H_G , n is the length of the code word and p is a positive integer.

From $R_n - N_c$ diagram, turbo code parameters extract:

* When $R_n < 1$, the distance between two samples is equal to n for all code rates, except when code rate is $1/2$, then distance will be equal to $(2n)$.

* The integer part of $n * (\text{absolute (minimum } (R_n)})$ is equal to k .

* The delay blocks number is:

$$d = \begin{cases} \frac{X_n}{2n} - (1 + N_{CTC1}) & \text{for code rate} = \frac{1}{2} \\ \frac{X_n}{n} - (1 + N_{CTC1}) & \text{for other} \end{cases}$$

$$N_{CTC1} = 0, 1, 2, \dots \quad (4)$$

That X_n is block length or samples on horizontal axis, n is output bits number, d is delay blocks number and N_{CTC1} is counter turbo code number. N_{CTC1} counts samples that is not equal to one. If the code rate is $1/2$ and after calculation $d=1$, d must recalculate by $N_{CTC1} = -1, 0, 1, \dots$

* The size of interleaver is:

$$l_{II} = \frac{X_n}{nN_{CTC2}}, \quad N_{CTC2} = 1, 2, 3, \dots \quad (5)$$

l_{II} is size of interleaver and N_{CTC2} is counter of samples of l_{II} that starts from first non-contiguous sample and continues.

3. Simulation and Evaluation of Algorithm

Input data and proposed algorithm are simulated by MATLAB. Inputs are generated by changing code rates, delay blocks number and interleaver size. String bits of turbo code have specifications that are described in table 1 and results of simulation are illustrated in figures 4 to 8.

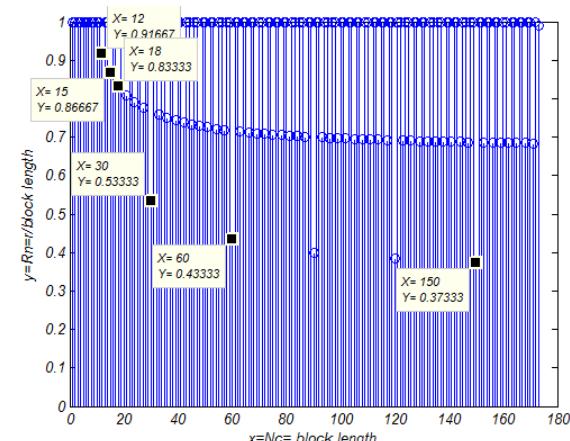


Fig. 4: Result of algorithm for code strings NO.1 of table 1.

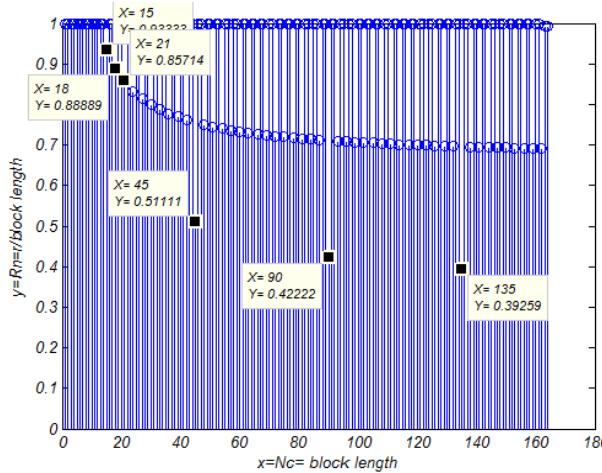


Fig. 5: Result of algorithm for code strings NO.2 of table 1

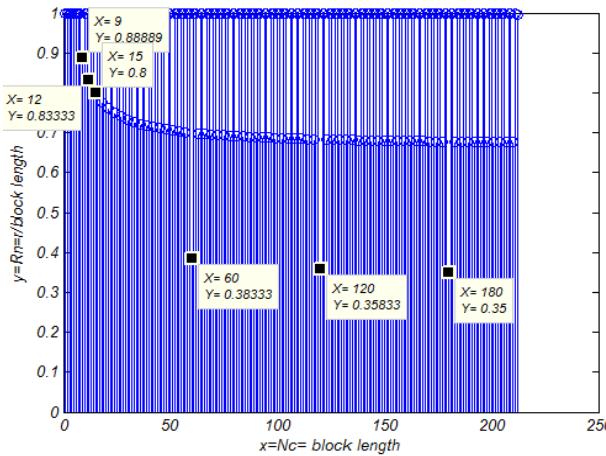


Fig. 6: result of algorithm for code strings NO.3 of table 1.

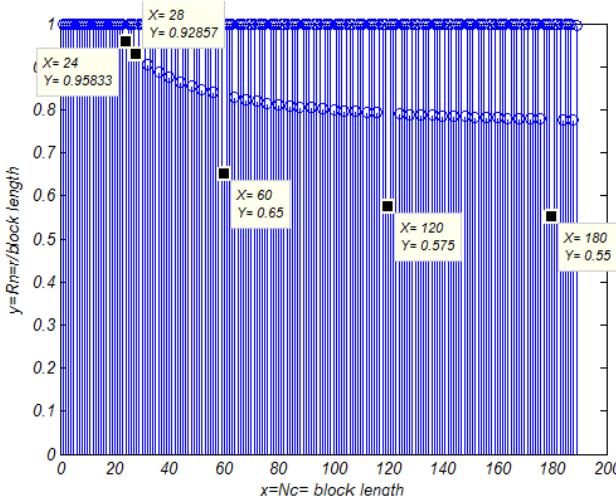


Fig. 7: result of algorithm for code strings NO.4 of table 1.

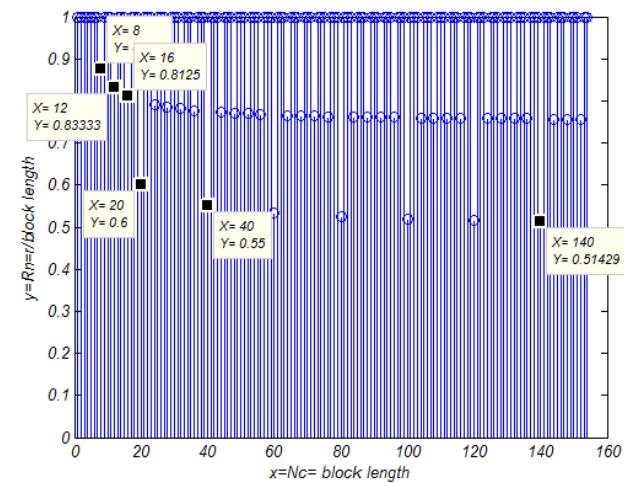


Fig. 8: result of algorithm for code strings NO.5 of table 1.

The results of calculating code rate, delay blocks number and size of interleaver are summarized in table2.

For evaluation of algorithm in various bit error rates (BER) from 10^{-1} to 10^{-6} , string bits is generated with code rates 1/2 and 1/3.

Table 1: Specification of input data

<i>Code string</i>	<i>Size of interleaver</i>	<i>Delay blocks number</i>	<i>Code rate</i>	<i>Length of string bits</i>
1	10	3	1/3	10000
2	15	4	1/3	9000
3	20	2	1/3	15000
4	30	5	1/2	18000
5	10	2	1/2	12000

In this condition, average of turbo code parameters is estimated. Results of statistical analysis, when algorithm runs 200 times, are illustrated in figures 9, 10, 11. Any information extract when BER is more than 10^{-2} , when BER is 10^{-2} to 10^{-4} percentage of recognition of turbo code increase, and when BER is 10^{-4} and less, turbo code is recognized completely. Results declare that algorithm has suitable performance in acceptable BER.

4. Conclusions

In this paper, the algorithm represented for blind recognition of turbo code. The proposed algorithm analyzes encoded string bits with new equations and

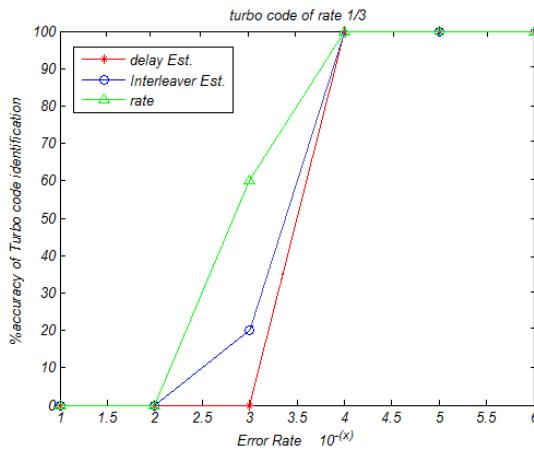


Fig. 9: Prediction value of correctly recognition with code rate 1/3.

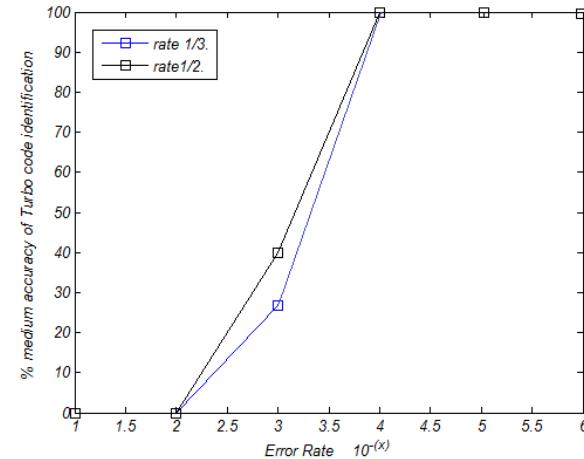


Fig. 11: Average of prediction values of correct recognition with code rates 1/3, 1/2.

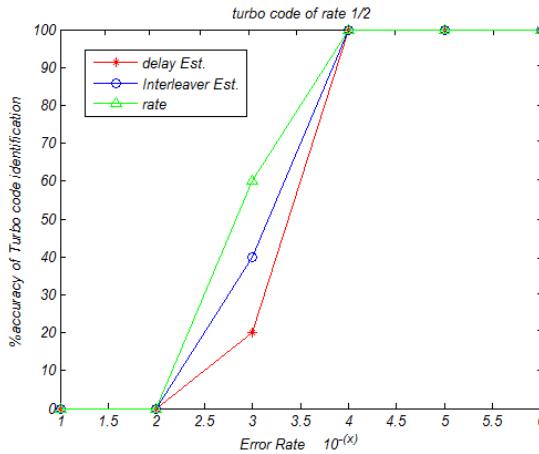


Fig. 10: Prediction value of correctly recognition with code rate 1/2.

extracts code rate, delay blocks number and size of interleaver. The volume of calculations of the proposed algorithm is $0.5L^3 + 1.25L$. Therefor the executing time of algorithm is suitable for practical issue and real time systems.

Table 2: result of analyzing

Parameters	Fig8	Fig7	Fig6	Fig5	Fig4
N_c for first $R_n \neq 1$	8	24	9	15	12
N_c for second $R_n \neq 1$	12	28	12	18	15
n	2	2	3	3	2
Absolute (minimum (R_n))	0.514	0.555	0.350	0.392	0.373
$k = [\min(n * R_n)]$	2	1	1	1	1
k/n	1/2	1/2	1/3	1/3	1/3
First sample of N_{CTC1}	-1	0	0	0	0
Second sample of N_{CTC1}	0	1	1	1	1
$X_n (= N_c \text{ in first sample of } N_{CTC1})$	8	24	9	15	12
$X_n (= N_c \text{ in second sample of } N_{CTC1})$	12	28	12	18	15
d	2	5	2	4	3
First sample of N_{CTC2}	1	1	1	1	1
Second sample of N_{CTC2}	2	2	2	2	2
$X_n (= N_c \text{ in first sample of } N_{CTC2})$	20	60	60	45	30
$X_n (= N_c \text{ in second sample of } N_{CTC2})$	40	120	120	90	60
	10	30	20	15	10

References

- [1] P.Grant, J.Collins."Introduction to Electronic Warfare",IEEE Processing, Vol.129, No.3, 1982.
- [2] S.Lin, D.Costello, "Error Control Coding Fundamentals and Applications", Pearson Education, Inc, 2005.
- [3] Todd k, Moon, "Error Correction Coding Mathematical Methods and Algorithms", John Wiley& Sons, 2005.
- [4] P.Sweeney, "Error Control Coding from Theory to Practice", John Wiley & Sons, 2002.
- [5] L.Bahi, J.Cocke, F.Jlinek, J.Raviv, "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate", IEEE Transaction, 1974.
- [6] G.Burel, R.Gautier, "Blind Estimation of Encoder and Interleaver Characteristics in an Non-Cooperative Context", International Conference of Communications, Internet and Information Technology, 2003.
- [7] J.Barbier, G.Sicot, S.Houcke, "Algebraic Approach for the Reconstruction of Linear and Convolutional Error Correcting Codes", International Journal of Applied Mathematics and Computer Sciences, 2006.
- [8] J.Barbier, "Analyse de Canaux de Communication Dansun Context Non-Cooperatif. Application aux Codes Corecteurs d'Erreurs et a la Steganalyse" PHD Dissertation, Ecole Polytechnique, 2007.
- [9] L.Lu, K.Li, Y.Guan, "Blind Detection of Interleaver Parameters for Non-Binary Coded Data Streams", School of Electrical and Electronic Engineering Nanyang, 2009.
- [10] G.Sicot, S.Houcke, "Blind Detection of Interleaver Parameters", IEEE International Conference of Acoustics, Speech and Signal Processing, 2005.
- [11] A.Valembois, "Detection and Recognition of a Binary Linear Code", Discrete Applied Mathematics, 2001.
- [12] J.Barbier, "Reconstruction of Turbo Codes Encoders", Space Communication Technologies Symposium, 2005.

Ali Naseri is Professor of Department of Information and Communications Technology, and Chair for Faculty of Information Processing of Imam Hossein University. He has twenty years research experience in processing and computing fields and radar processing. Dr Naseri is also an advisor of Information and Communications Technology Ministry of Iran.

Omid Azmoon is Master engineer of Department of Information and Communications Technology of Imam Hossein University. He has ten years research experience. His researches have focused on data fusion in radars network. He currently works in Iran Communications Industries.

Samad Fazeli is Master engineer of Department of Information and Communications Technology of Imam Hossein University. He has Five years research experience. His researches have focused on ELINT and COMINT systems

BPISG: A Batching Heuristic Scheduling Algorithm With Taking Index Parameters for Mapping Independent Tasks on Heterogenous Computing Environment

Arash Ghorbannia Delavar¹, Ali Reza Khalili Boroujeni² and Javad Bayrampoor³

¹ faculty of engineering, Payam Noor University
Tehran, Iran

²Payam Noor University
Tehran, Iran

³Payam Noor University
Tehran, Iran

Abstract

In this paper we consider the problem of allocating independent heterogeneous tasks on grid environment in which A new batching heuristic scheduling algorithm with Taking index parameters will be presented. Reference to tasks compute scheduling, several methods are evaluated. With compare Min-mean and Improved Min-mean algorithms, we can create the specific situation that BPISG algorithm can be more efficient and more dependable than similar than previous algorithms. With taking the round time, consist acknowledgement and return time parameters, specific functionality has been created.

With implementation these parameters in the simulate environment, we can create situation that scheduling task will be done with better position and achieve high performance on computational grids. Finally the experiment and simulated results will show that the BPISG heuristic scheduling algorithm performs significantly to ensure high throughput, reduced makespan and more efficient in the grid environment.

Keywords: Grid Computing, Job Scheduling, Heuristic Algorithm, Load Balancing.

1. Introduction

Grid computing environment includes of different interconnected machines by interface networks to execute different tasks that have diverse computational requirements. In distributed systems such as grids, in which there are various sources, deciding about regularity of task and selecting the computing machines is an important problem in grid environments. The main purpose of grid systems is optimize using sources and maximizing the efficiency of the system. Managing various resources and task scheduling in grid environment are challenging and indispensable works [1].

Grid environments have tried to study various scheduled algorithms for reaching to above purposes. Scheduling is an important tool for this purpose. Tasks scheduling is an NP-complete problem and finding the absolute optimal solution is too hard. So many heuristics have been developed to solve this hard problem. The heuristic scheduling can be classified into two categories: on-line mode and batch-mode heuristics. In the on-line mode heuristics, a task is mapped on to a machine as soon as it arrives at the scheduler. In the batch-mode heuristics, tasks are not mapped on to machines as they arrive; instead, they are collected into the buffer and then it is scheduled at prescheduled time [3, 4].

In this paper, the study is based on the batch-mode heuristics and presents a new heuristic scheduling algorithm named BPISG to acquire efficient static mapping of meta-tasks to the machines in heterogeneous computing systems. The primary objects of the paper are the throughput maximization and reduced makespan (measure of the throughput) of the heterogeneous grid computing systems. With taking the round time index parameter in simulated environment, BPISG heuristic scheduling algorithm performs more efficiently in the grid environment. The proposed heuristic uses the benchmark model of Braun et al. [2, 3].

2. Related Works

Many heuristics algorithms have been designed and developed to solve meta-tasks optimal scheduling In distributed heterogeneous computing systems. Two sample of these heuristic algorithms are defined as follows:

2.1 Min-min

Min-min algorithm starts with a set of all unmapped tasks. The completion time for each task on each machine is calculated. The machine that has the minimum completion time for each job is selected. Then the job with the overall minimum completion time is selected and mapped to the machine. Again, this process is repeated with the remaining unmapped tasks [2, 3].

2.2 Min-mean

The heuristic scheduling algorithm Min-mean works in two phases.

- In phase 1, the task allocation is done based on the Min-min algorithm.
- In phase 2, the mean of all machines completion time is taken. The machine whose completion time is greater than the mean value is selected. The tasks allocated to the selected machines are reallocated to the machines whose completion time is less than the mean value [3].

2.3 Improved Min-mean Algorithm

Like to Min-mean Algorithm, Improved Min-mean Algorithm works also in two phases.

- In phase 1, the job allocation is done based on the Min-min algorithm.
- In phase 2, the mean of all machines completion time is taken. Then a small constant value $\Delta t=7\%$ is added to the mean completion time ($k = MeanCT + \Delta t$). This causes an improvement over Min-mean heuristic scheduling algorithm. After that the machine whose completion time is greater than the k value, is selected and The tasks allocated to the selected machines are reallocated to the machines whose completion time is less than the k value [4].

3. Problem Definition

The problem of task scheduling, will be studied in heterogeneous computing environment. The experimental study is based on a benchmark simulation model by Braun et al. [2].

In this model, there is number of independent tasks to allocation and number of machines to execute these tasks. Because of Nonpreemptive scheduling, each machine executes one task at a time. For static mapping, number of tasks, the size of the tasks and the number of machines in the heterogeneous computing environment

are known a priori. Since there are static heuristics, the accurate estimate of the expected execution time for each task on each machine is known to execution and is contained within an ETC (expected time to compute) matrix where $ETC(t_i, m_j)$ is an estimated execution time of task i on machine j . The size of ETC is $t * m$, where t , represents the number of the tasks and m , represents the number of the machines [3, 4].

ct_{ij} – Expected completion time of the task t_i on the machine m_j .

et_{ij} – Expected execution time of the task t_i on the machine m_j . Suppose that m_j has no load when task t_i is assigned.

r_j – Ready time of the machine m_j (the time when machine m_j becomes ready to execute task t_i).

at_{ij} – acknowledgement time of the task t_i from the machine m_j (the wait time needed to mapping task t_i to the machine m_j).

rt_{ij} – Return time of the task t_i from the machine m_j (the wait time needed to return results of task t_i from machine m_j).

$$ct_{ij} = et_{ij} + r_j + at_{ij} + rt_{ij} \quad (1)$$

The primary object of BPISG heuristic scheduling algorithm is to minimize the makespan which is defined as completion time of the system:

$$\text{makespan} = \max(ct_{ij}). \quad (2)$$

3.1 BPISG Algorithm

The model of BPISG scheduling algorithm is based on the expected time to compute (ETC) matrix of Braun et al. [2]. The BPISG heuristic scheduling algorithm is defined as follows:

First:

do until all tasks t_i in meat-tasks M_t are scheduled

- for each task in M_t compute the earliest completion time and the machine that obtains it
 - find the task t_k with the minimum earliest completion time
 - assign each task t_k to the machine m_k giving the earliest completion time
 - delete task t_k from M_t
 - update r_j
 - update ct_{ij} for all i
- enddo

www.IJCSI.org

Second:

Sort all machine on minimum earliest completion time

Third:

for all machine m_j with the minimum earliest completion time

- for all task t_i selected with this machine m_j in phase 1
 - find the machine m_k with the minimum earlier completion time than m_j
 - assign task t_i to the machine m_k and delete task t_i from m_j
 - reschedule task t_i on machine m_k

• endfor

endfor

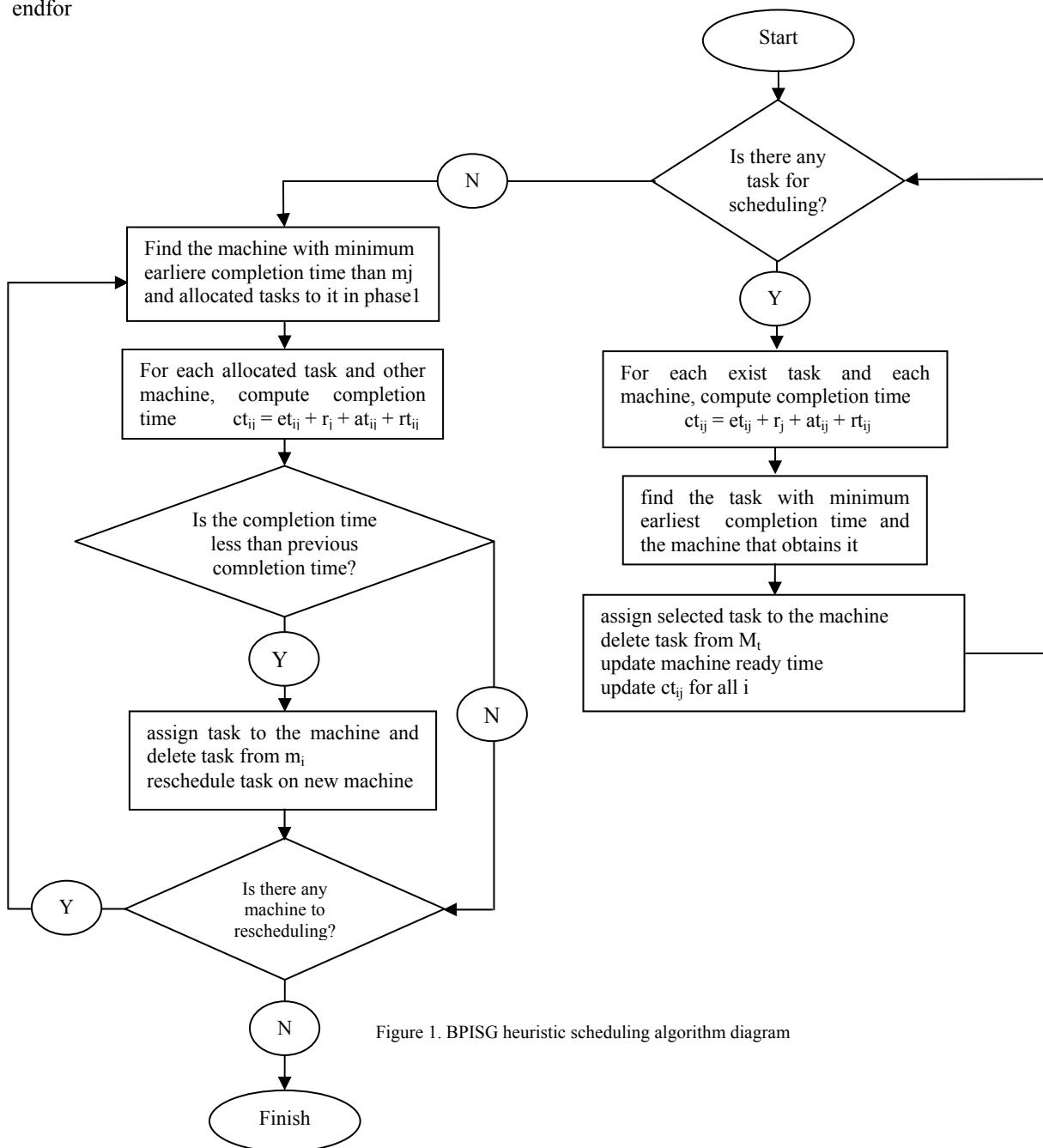


Figure 1. BPISG heuristic scheduling algorithm diagram

4. Benchmark Descriptions

To better evaluate the behavior of mapping heuristics, a model of the execution times of the tasks on the machines is needed; so that the parameters of this model can be changed to investigate the performance of the heuristics under different heterogeneous computing systems and under different types of tasks to be mapped. One such model consists of an expected time to compute (ETC) matrix, where contains the estimates for the expected execution times of a task on all machines, for all the tasks that are expected to arrive for service over a given interval of time.

The ETC model presented here can be characterized by three parameters: machine heterogeneity, task heterogeneity, and consistency. Four categories were proposed for the ETC matrix in: (a) high task heterogeneity and high machine heterogeneity, (b) high task heterogeneity and low machine heterogeneity, (c) low task heterogeneity and high machine heterogeneity, and (d) low task heterogeneity and low machine heterogeneity.

The ETC matrix can be further classified into three categories, consistent, inconsistent and partially-consistent. For a consistent ETC matrix, if a machine m_x has a lower execution time than a machine m_y for a task t_k , then the same is true for any task t_i .

In inconsistent ETC matrices, the relationships among the task computational requirements and machine capabilities are such that no structure as that in the consistent case is enforced.

A combination of these two cases, which may be more realistic in many environments, is the partially-consistent ETC matrix, which is an inconsistent matrix with a consistent sub-matrix [5].

The experimental results are classified into 12 different types of ETC matrices includes three groups of four instances each. Every instance consists of 64 tasks and 16 machines. The first group relates to the inconsistent ETC matrices. The second and third group relates to the consistent and partially-consistent ETC matrices. All instances based on the three metrics: task heterogeneity, machine heterogeneity and consistency.

the round time, consist acknowledgement and return time parameters for each of the instances with a discrete uniform random function which are created as the 32 byte packets size. the acknowledgement time of each packets is between 1 to 200 ms and the return time of the same packet is vary between 0 to 0.4 of acknowledgement time of that packet.

5. Performance Analysis

To evaluate the efficiency of propose algorithm , the BPISG heuristic scheduling algorithm is compared with Min-mean and Improved Min-mean heuristic algorithm in all the three group of four instances.

The follow tables and diagrams show the improvement in percentage of the BPISG heuristic scheduling algorithm over Min-mean and Improved Min-mean heuristic scheduling algorithm.

Table 1: Improvement of BPISG algorithm over Min-mean and Improved Min-mean

Consistency	Improved Over Min-mean		Improved Over Improved Min-mean	
Inconsistent	High-High	15.24 %	High-High	15.24 %
	High-Low	3.99 %	High-Low	0.95 %
	Low-High	19.13 %	Low-High	9.14 %
	Low-Low	19.54 %	Low-Low	19.54 %
Consistent	High-High	1.27 %	High-High	1.27 %
	High-Low	11.37 %	High-Low	2.92 %
	Low-High	0.00 %	Low-High	0.00 %
	Low-Low	9.93 %	Low-Low	9.93 %
Partially Consistent	High-High	20.87 %	High-High	20.87 %
	High-Low	14.49 %	High-Low	14.49 %
	Low-High	21.26 %	Low-High	21.26 %
	Low-Low	8.33 %	Low-Low	8.33 %

Table 2: Comparison of makespan value obtained by min-Mean, Improved Min-mean and BPISG algorithm for high task,high machine in Inconsistent type

Instance in Inconsistent	Min-mean	Improved Min-mean	BPISG Algorithm
High-High	5458450	5458450	4736396

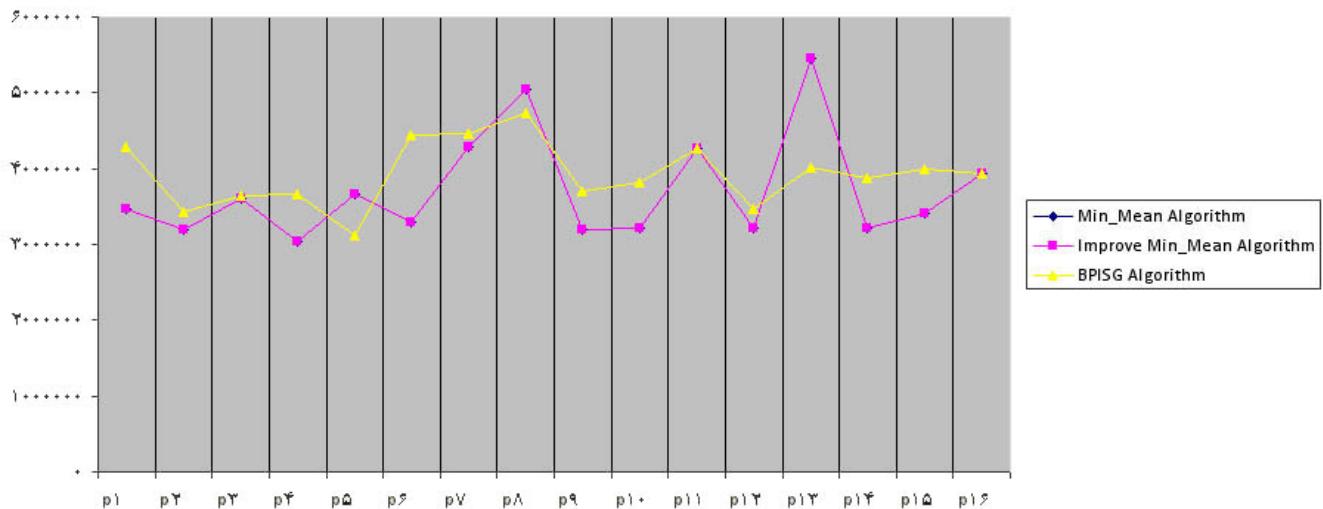


Figure 2. Comparison of makespan value obtained by Min-mean, Improved Min-mean and BPISG algorithm for high task & high machine in Inconsistent type with 15.24% improvement more than Min-mean and Improved Min-mean

Table 3: Comparison of makespan value obtained by min-Mean, Improved Min-mean and BPISG algorithm for high task,low machine in Inconsistent type

Instance in Inconsistent	Min-mean	Improved Min-mean	BPISG Algorithm
High-Low	1930096	1873779	1856117

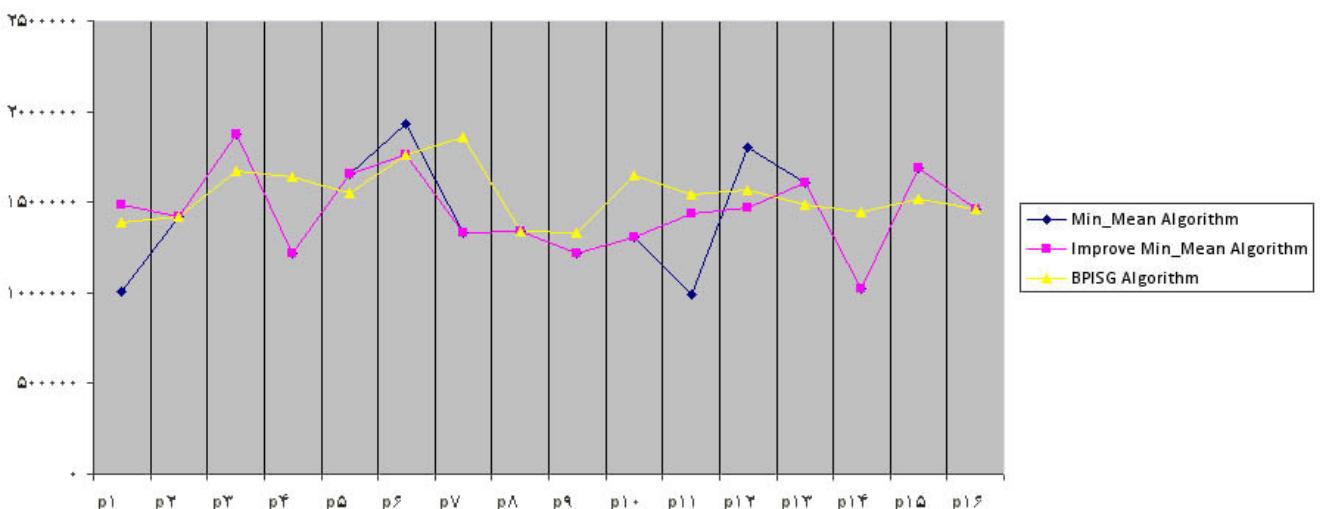


Figure 3. Comparison of makespan value obtained by Min-mean, Improved Min-mean and BPISG algorithm for high task & low machine in Inconsistent type with 3.99 % improvement more than Min-mean and 0.95 % in compare with Improved Min-mean

Table 4: Comparison of makespan value obtained by min-Mean, Improved Min-mean and BPISG algorithm for low task,high machine in Inconsistent type

Instance in Inconsistent	Min-mean	Improved Min-mean	BPISG Algorithm
Low-High	1084290	993399	910166

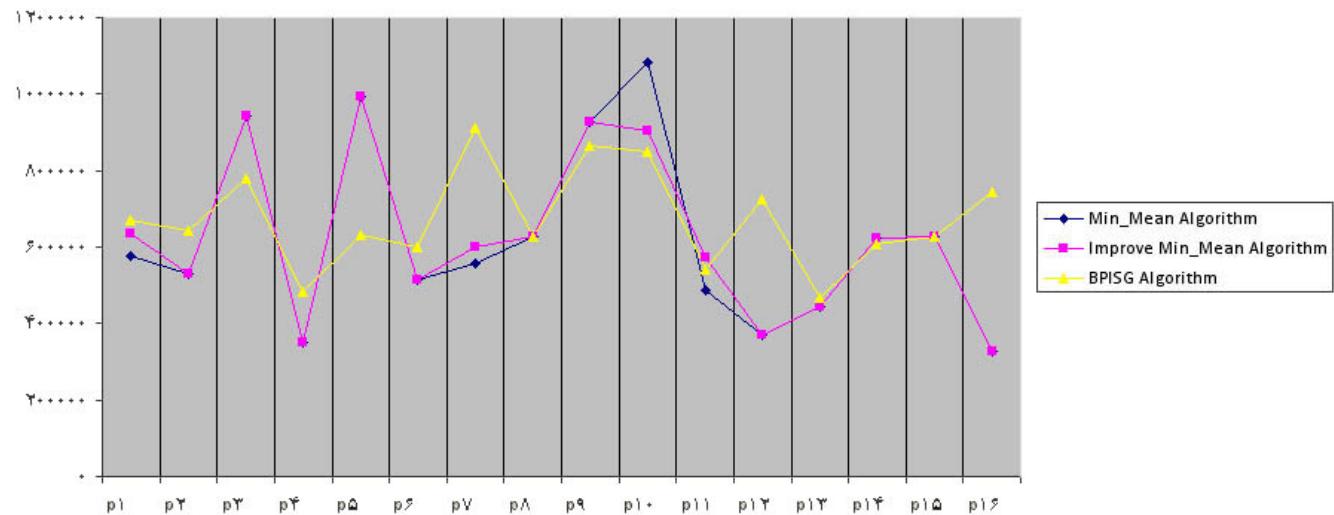


Figure 4. Comparison of makespan value obtained by Min-mean, Improved Min-mean and BPISG algorithm for low task & high machine in Inconsistent type with 19.13 % improvement more than Min-mean and 9.14 % in compare with Improved Min-mean

Table 5: Comparison of makespan value obtained by min-Mean, Improved Min-mean and BPISG algorithm for low task,low machine in Inconsistent type

Instance in Inconsistent	Min-mean	Improved Min-mean	BPISG Algorithm
Low-Low	1389099	1389099	1162040

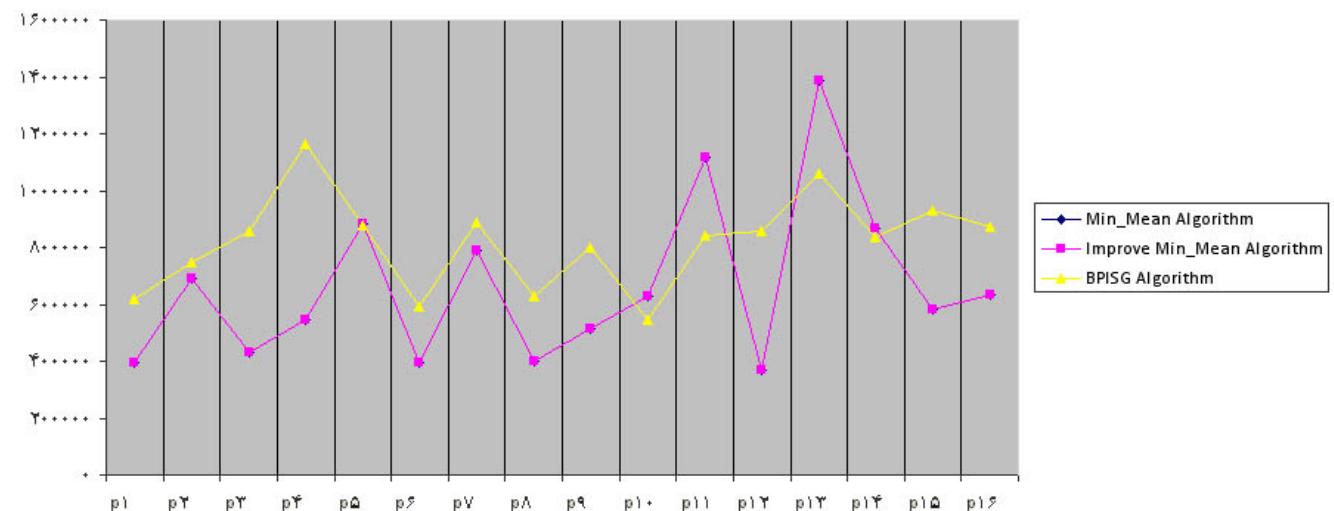


Figure 5. Comparison of makespan value obtained by Min-mean, Improved Min-mean and BPISG algorithm for low & task low machine in Inconsistent type with 19.54 % improvement more than Min-mean and Improved Min-mean

Table 6: Comparison of makespan value obtained by min-Mean, Improved Min-mean and BPISG algorithm for high task,high machine in Consistent type

Instance in Consistent	Min-mean	Improved Min-mean	BPISG Algorithm
High-High	10644873	10644873	10511821

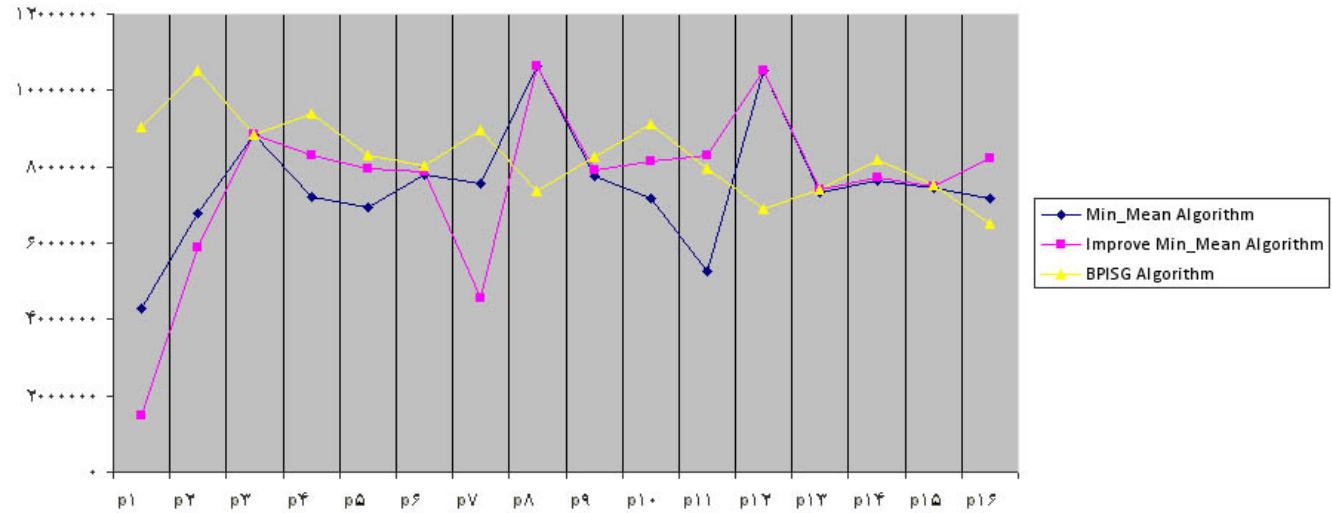


Figure 6. Comparison of makespan value obtained by Min-mean, Improved Min-mean and BPISG algorithm for high task & high machine in Consistent type with 1.27 % improvement more than Min-mean and Improved Min-mean

Table 7: Comparison of makespan value obtained by min-Mean, Improved Min-mean and BPISG algorithm for high task,low machine in Consistent type

Instance in Consistent	Min-mean	Improved Min-mean	BPISG Algorithm
High-Low	3723622	3441148	3343366

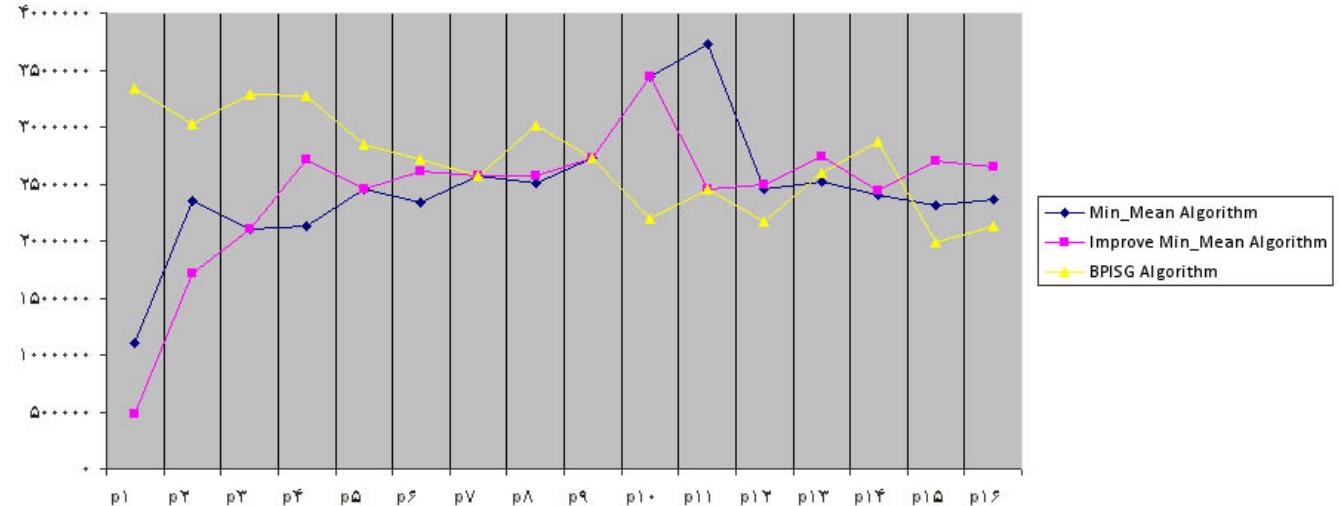


Figure 7. Comparison of makespan value obtained by Min-mean, Improved Min-mean and BPISG algorithm for high task & low machine in Consistent type with 11.37 % improvement more than Min-mean and 2.92 % in compare with Improved Min-mean

Table 8: Comparison of makespan value obtained by min-Mean, Improved Min-mean and BPISG algorithm for low task,high machine in Consistent type

Instance in Consistent	Min-mean	Improved Min-mean	BPISG Algorithm
Low-High	1799023	1799023	1799023

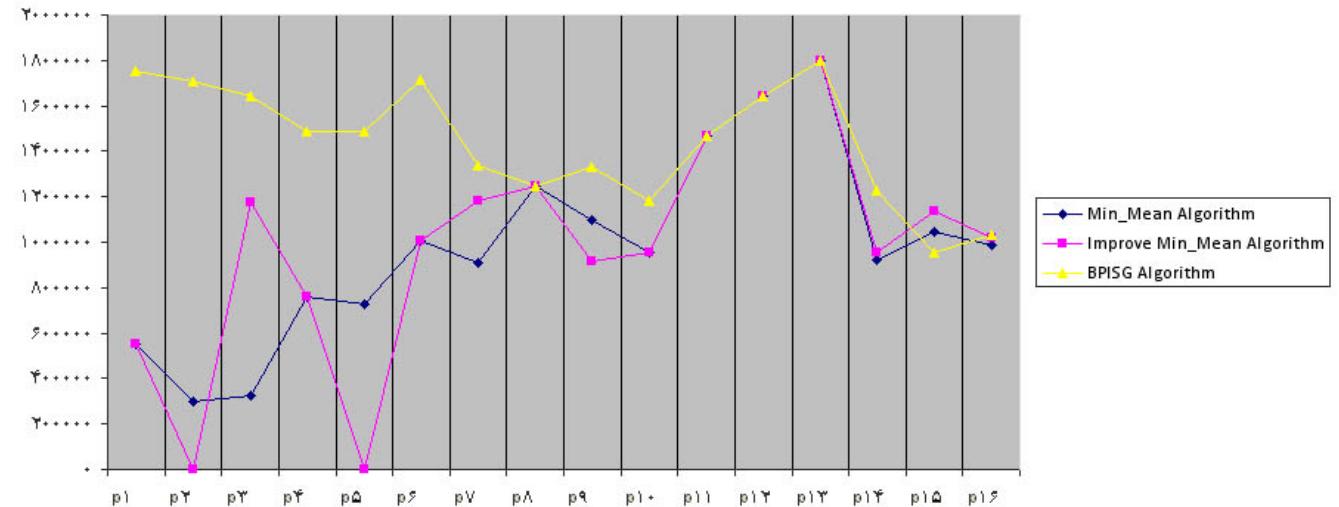


Figure 8. Comparison of makespan value obtained by Min-mean, Improved Min-mean and BPISG algorithm for low task & high machine in Consistent type with 0.00 % improvement more than Min-mean and Improved Min-mean

Table 9: Comparison of makespan value obtained by min-Mean, Improved Min-mean and BPISG algorithm for low task,low machine in Consistent type

Instance in Consistent	Min-mean	Improved Min-mean	BPISG Algorithm
Low-Low	1945199	1945199	1769463

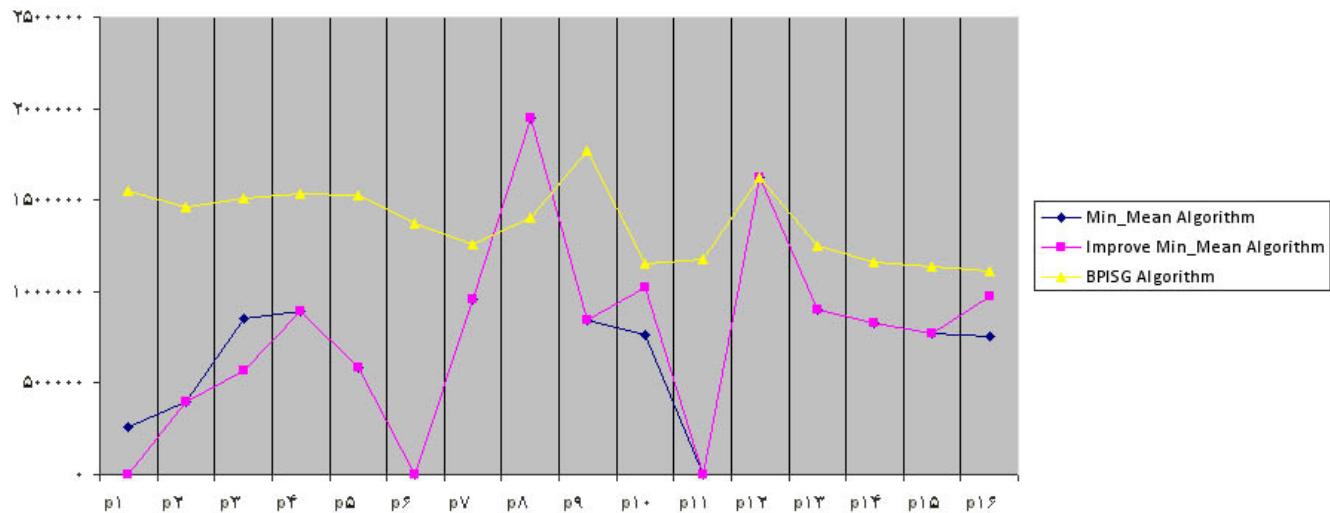


Figure 9. Comparison of makespan value obtained by Min-mean, Improved Min-mean and BPISG algorithm for low task & low machine in Consistent type with 9.93 % improvement more than Min-mean and Improved Min-mean

Table 10: Comparison of makespan value obtained by min-Mean, Improved Min-mean and BPISG algorithm for high task,high machine in partially-Consistent type

Instance in Partially-Consistent	Min-mean	Improved Min-mean	BPISG Algorithm
High-High	7198299	7198299	5955583

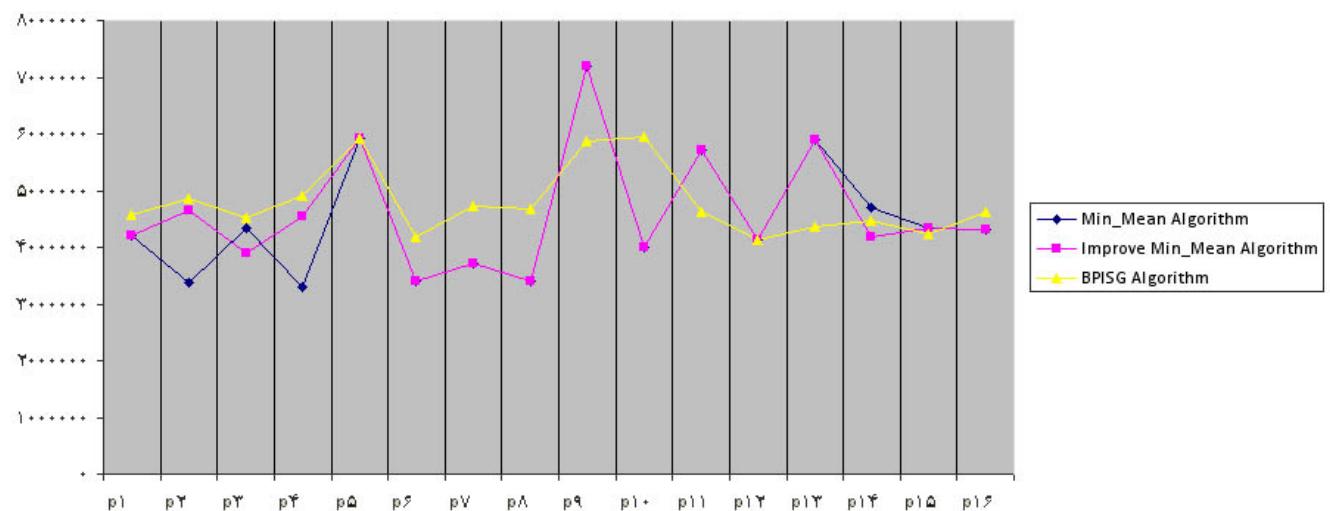


Figure 10. Comparison of makespan value obtained by Min-mean, Improved Min-mean and BPISG algorithm for high task & high machine in partially-Consistent type with 9.93 % improvement more than Min-mean and Improved Min-mean

Table 11: Comparison of makespan value obtained by min-Mean, Improved Min-mean and BPISG algorithm for high task,low machine in partially-Consistent type

Instance in Partially-Consistent	Min-mean	Improved Min-mean	BPISG Algorithm
High-Low	1967366	1967366	1718439

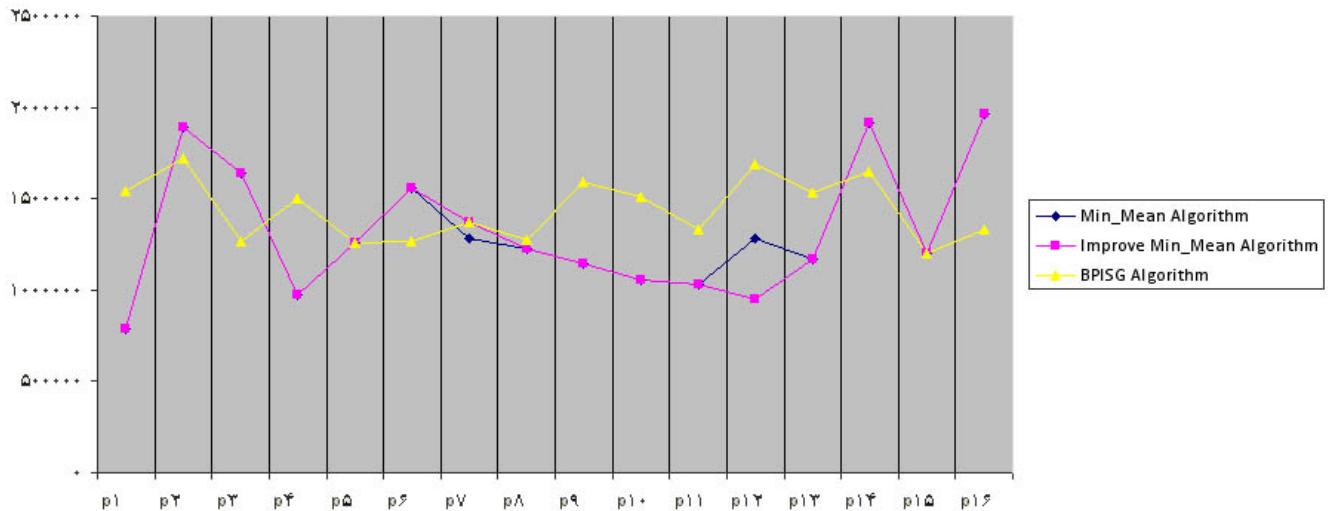


Figure 11. Comparison of makespan value obtained by Min-mean, Improved Min-mean and BPISG algorithm for high task & low machine in partially-Consistent type with 14.49 % improvement more than Min-mean and Improved Min-mean

Table 12: Comparison of makespan value obtained by min-Mean, Improved Min-mean and BPISG algorithm for low task,high machine in partially-Consistent type

Instance in Partially-Consistent	Min-mean	Improved Min-mean	BPISG Algorithm
Low-High	828352	828352	683121

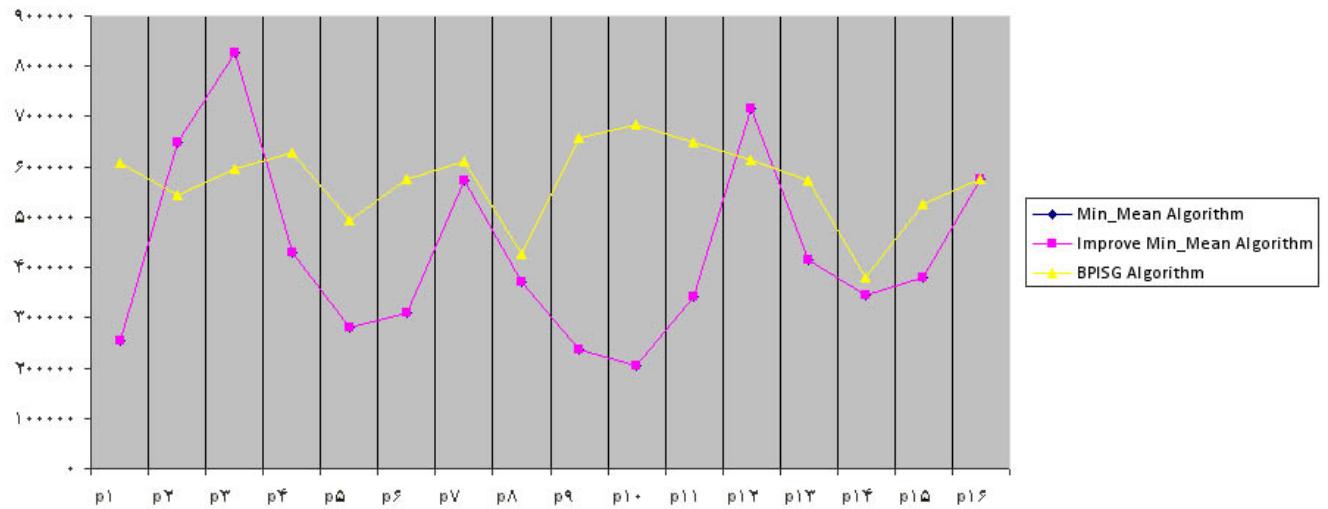


Figure 12. Comparison of makespan value obtained by Min-mean, Improved Min-mean and BPISG algorithm for low task & high machine in partially-Consistent type with 21.26 % improvement more than Min-mean and Improved Min-mean

Table 13: Comparison of makespan value obtained by min-Mean, Improved Min-mean and BPISG algorithm for low task,low machine in partially-Consistent type

Instance in Partially-Consistent	Min-mean	Improved Min-mean	BPISG Algorithm
Low-Low	1339542	1339542	1236585

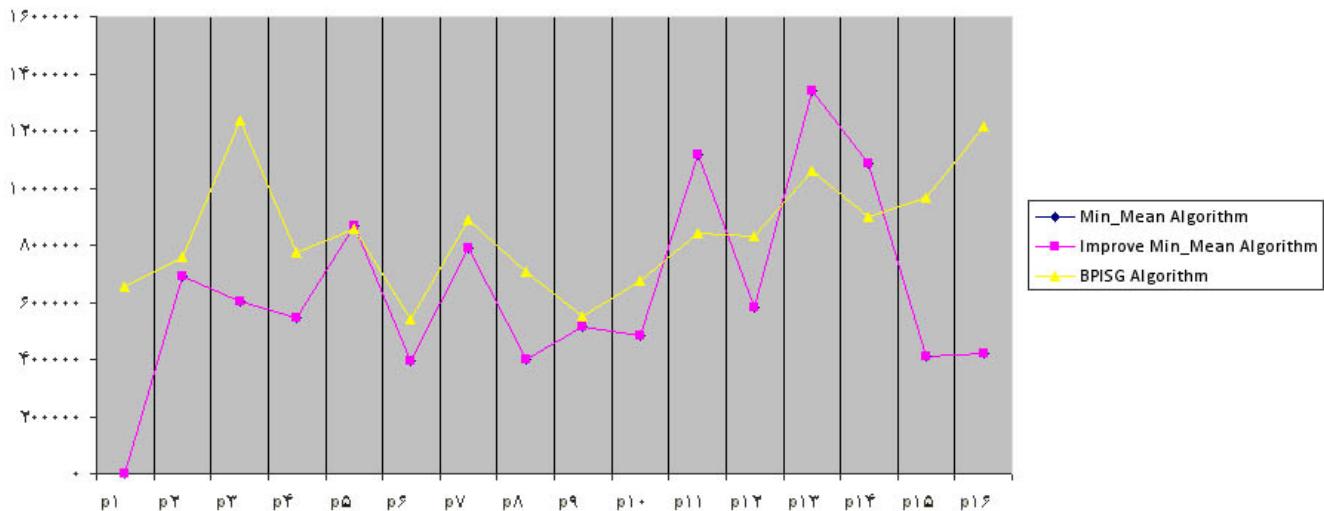


Figure 13. Comparison of makespan value obtained by Min-mean, Improved Min-mean and BPISG algorithm for low task & low machine in partially-Consistent type with 8.33 % improvement more than Min-mean and Improved Min-mean

6. Conclusions And Future Work

experimental results show the BPISG algorithm has average 12.12 and 10.33 % improvement more than Min-mean and Improved Min-mean algorithms. BPISG also reduced makespan, increase efficiency and achieve better mapping than previous algorithms in the grid environment. The future research will be focused on other index parameters, such as bandwidth and delay times in grid environment.

References

- [1] A.Ghorbannia Delavar, M.Nejadkheirallah and M.Motalleb, "A New Scheduling Algorithm for Dynamic Task and Fault Tolerant in Heterogeneous Grid Systems Using Genetic Algorithm", IEEE 2010.
- [2] Tracy D.Braun, Howard Jay Siegel and Noah Beck, "A Comparison of Eleven Static Heuristics for Mapping a Class of Independent Tasks onto Heterogeneous Distributed Computing Systems", Journal of Parallel and Distributed Computing 61, 2001, pp.810-837.
- [3] Kamalam.G.K and Murali bhaskaran.V, "A New Heuristic Approach:Min-Mean Algorithm For Scheduling Meta-Tasks On Heterogenous Computing Systems", Journal of Computer Science and Network Security, January 2010.
- [4] G. K. Kamalam and V. Murali Bhaskaran, "An Improved Min-Mean Heuristic Scheduling Algorithm for Mapping Independent Tasks on Heterogenous Computing Environment", Journal of Computational cognition, december 2010.
- [5] Shoukat Ali, Howard Jay Siegel and Muthucumaru Maheswaran, "Task Execution Time Modeling for

Heterogeneous Computing Systems", IEEE Computer, 2000.

Arash Ghorbannia Delavar received the MSc and Ph.D. degrees in computer engineering from Sciences and Research University, Tehran, IRAN, in 2002 and 2007. He obtained the top student award in Ph.D. course. He is currently an assistant professor in the Department of Computer Science, Payam Noor University, Tehran, IRAN. He is also the Director of Virtual University and Multimedia Training Department of Payam Noor University in IRAN. Dr.Arash Ghorbannia Delavar is currently editor of many computer science journals in IRAN. His research interests are in the areas of computer networks, microprocessors, data mining, Information Technology, and E-Learning.

Ali Reza Khalili Boroujeni received the BS, in 2000 and now, he is a Student the MS degree in the department of Computer Engineering and Information Technology in Payam Noor University, Tehran, IRAN. His research interests include computer networks, grid and scheduling algorithm.

Javad Bayrampoor received the BS, in 2007 and now, he is a Student the MS degree in the department of Computer Engineering and Information Technology in Payam Noor University, Tehran, IRAN. His research interests include computer networks, grid and scheduling algorithm.

Design Approaches to Enhance Usability for E-Commerce Sites

Mohiuddin Ahmed¹, Jonayed Kaysar² and Nafis Rahman³

¹ Department of Computer Science and Information Technology, Islamic University of Technology
Board Bazar, Gazipur-1704, Bangladesh

² Department of Computer Science and Information Technology, Islamic University of Technology
Board Bazar, Gazipur-1704, Bangladesh

² Department of Computer Science and Information Technology, Islamic University of Technology
Board Bazar, Gazipur-1704, Bangladesh

Abstract

Human computer interaction has a great collaboration with World Wide Web. The fastest growing web technology and interaction issues are compelling web designers to think for quality and user friendly designs in the web. Websites are expected to be designed in a way which will allure the visitors who are looking for particular information. E-commerce sites are one of the fastest growing sites where consumers or users shop things without any burden of being physically present at the shop and receive products at home. We are proposing novel design approaches which will ameliorate the interaction styles and will help the users to access information and do shopping efficiently.

Keywords: Ethnography; Usability; Problem Space; Design Implication; Prototypes; Evaluations.

1. Introduction

E-commerce site [1] is defined as any Web Site offering pre-sale support, products for sale or after sales service and backup. The discerning thing is that the consumers don't have to be physically present at the store rather they can buy anything online by credit card. But there is still lack of a good number of sites where customers are happy with service and support. As the world is facing a huge change in perspective of information technology so this aspect should also be considered. So we need more and more good number of websites where all types of feasible services and user friendly interaction will be present to make customers interested and make their life comfortable with so many privileges.

We are proposing an optimized design for a Mobile Phone Accessories e-commerce websites which will ensure the customer interaction with no complexity and access to the information will be much easier than any other sites present in the web now-a-days. Initially we focus on identifying user needs, understanding the problem space, social context, environment, describing tasks, and user and design implications. We have got some requirements using web survey and ethnography for our mobile phone accessories e-commerce site and we constructed low fidelity prototypes using sketching. As design issues are related with our project so it is better to use sketching for prototyping.

Finally we used evaluation techniques to evaluate the implemented design of the website for mobile phone accessories and according to our survey through Google docs we found our design is more efficient and user friendly. Subsequent parts contain the background study, proposed methodology, exemplary design, conclusion and references.

2. Background Study

2.1 Identifying User Needs & Establishing Requirements

First, For a Mobile phone accessories website a particular consumer will basically look for the following things

- Products and its classification
- Price
- Condition: HD Photo & Streaming Video
- Transactions
- Money back guarantee

Figure 1. From a large pool of products the customer should be able to find out the stipulated product and its type momentarily. Relevant information like prices, conditions, transactions and terms & conditions will have to be discerned very meticulously. According to these needs we have to ensure the website design. And most importantly the information access is the key issues to be considered. Here ethnography is used to find out the regular interactions and those will be analyzed to enhance the understanding of user needs.

The first issue [2] in problem space is what we want to create and the answer is pretty simple. We want to develop an e-commerce site for mobile phone accessories which will eradicate the existing problems and will enhance user interaction style.

The second issue [2] is assumptions. According to the user needs and requirements we assume that current e-commerce sites are way too complex for the ordinary non-tech users but it is expected to develop a site which can be used by all. More importantly we are considering the transaction system is still too complex and requires many procedures to complete the total process. And the visibility issues are still not handled carefully due to bad design in some places. The third issue [2] is claims. We are claiming that we are going to implement a novel approach which will focus more on the complexities. Basically the complexity arises in case of such sites when the customer finds difficulty to understand the actual condition of a product and transaction procedures. We are going to implement such design which will eradicate all these. The final issue [2] in problem space is whether our approach is going to be successful or not. We have studied the design of various e-commerce sites which will help us to understand the existing lacking and to develop far better sites than that and to burgeon the superb usability engineering, experience etc.

The main task of this site will be to meet the customers need. As it is an e-commerce site and the site deals with mobile phone accessories, the main task will be serving the customers with state-of-the-art accessories as well as normal accessories.

Now, we need to define the users of this site. This site will be a marketplace of mobile phone accessories for all types of people. A person who wants to buy a brand new iPhone4 and the person who wants to buy the simple micromaxX210; both are the user of this site. Also, people who run this site and administer all type of transactions and inclusion of all the products are also termed as the users. It is a website and like all other websites, its environment will be the World Wide Web (www). It will use the power of www to gather the general users. This site has a long range of customers. From a high end user to

a low end user, this site will meet all their expectations. So, this site will cover all the possible customer classification and will be helpful for all the people. Also, this site will use the positive power of social networking for its marketing; it will help the internet aware people to know about the site.

2.2 Prototyping

Prototypes [5] are a useful aid when discussing ideas with stakeholders; they are a communication device among team members, and are an effective way to test out ideas for you. The activities of building prototypes encourage reflection in design, as described by Schon (1983) and as recognized by designers from many disciplines as an important aspect of the design process. Liddle (1996), talking about software design, recommends that prototyping should always precede any writing of code. Prototypes answer questions and support designers in choosing between alternatives. Hence, they serve a variety of purposes: for example, to test out the technical feasibility of an idea, to clarify some vague requirements, to do some user testing and evaluation, or to check that a certain design direction is compatible with the rest of the system development. Which of these is your purpose will influence the kind of prototype you build. So, for example, if you are trying to clarify how users might perform a set of tasks and whether your proposed device would support them in this, you might produce a paper-based mockup.

Different kinds of prototypes

ii. Low fidelity

A low-fidelity prototype is one that does not look very much like the final product. For example, it uses materials that are very different from the intended final version, such as paper and cardboard rather than electronic screens and metal. The lump of wood used to prototype the Palm Pilot described above is a low-fidelity prototype, as is the cardboard-box laser printer.

ii. High fidelity

High-fidelity prototyping uses materials that you would expect to be in the final product and produces a prototype that looks much more like the final thing. For example, a prototype of a software system developed in Visual Basic is higher fidelity than a paper-based mockup; a molded

piece of plastic with a dummy keyboard is a higher-fidelity prototype of the PalmPilot than the lump of wood.

Prototype Construction

When the design has been around the iteration cycle enough times to feel confident that it fits requirements, everything that has been learned through the iterated steps of prototyping and evaluation [6] must be integrated to produce the final product. Although prototypes will have undergone extensive user evaluation, they will not necessarily have been subjected to rigorous quality testing for other characteristics such as robustness and error-free operation. Constructing a product to be used by thousands or millions of people running on various platforms and under a wide range of circumstances requires a different testing regime than producing a quick prototype to answer specific questions.

2.2 Evaluation Approaches

Using evaluation, designers make sure that their software is usable and is what users want. Evaluation [6] is the process of systematically collecting data that informs us about what it is like for a particular user or group of users to use a product for a particular task in a certain type of environment. The basic premise of user-centered design is that users' needs are taken into account throughout design and development. This is achieved by evaluating the design at various stages as it develops and by amending it to suit user's needs. The design, therefore, progresses in iterative cycles of design evaluate redesign.

Iterative design & evaluation is a continuous process that examines:

- Why: to check those users can use the product and that they like it.
- What: Conceptual model, early prototypes of a new system and later, more complete prototypes.
- Where: in natural and laboratory settings.
- When: throughout design; finished products can be evaluated to collect information to inform new products.

Designers need to check that they understand users' requirements.

Evaluation approaches

- Usability testing
- Field studies
- Analytical evaluation

Usability testing

Usability testing [5] involves measuring typical users' performance on carefully prepared tasks that are typical of those for which the system was designed. Users' performance is generally measured in terms of number of errors and time to complete the task. As the users perform these tasks, they are watched and recorded on video and by logging their interactions with software. This observational data is used to calculate performance times, identify errors, and help explain why the users did what they did. User satisfaction questionnaires and interviews are also used to elicit users' opinions.

Field studies

The distinguishing feature of field studies [6] is that they are done in natural settings with the aim of increasing understanding about what users do naturally and how technology impacts them. In product design, field studies can be used to

- help identify opportunities for new technology;
- determine requirements for design;
- facilitate the introduction of technology; and
- evaluate technology

Analytical evaluations

In analytical evaluations [5] experts apply their knowledge of typical users, often guided by heuristics, to predict usability problems. The key feature of analytical evaluation is that users need not be present, which makes the process quick, relatively inexpensive, and thus attractive to companies; but it has limitations.

Usability Inspections

There are several kinds of usability inspections [6]. Experts use their knowledge of users & technology to review software usability. Expert critiques can be formal or informal reports. Heuristic evaluation is a review guided by a set of heuristics. Walkthroughs involve stepping through a pre-planned scenario noting potential problems.

Heuristic evaluation

Developed Jacob Nielsen in the early 1990s. Based on heuristics [5] distilled from an empirical analysis of 249

usability problems. These heuristics have been revised for current technology. Heuristics being developed for mobile devices, wearable's, virtual worlds, etc. Design guidelines form a basis for developing heuristics. 3 stages for doing heuristic evaluation:

- Briefing session to tell experts what to do.
- Evaluation period of 1-2 hours in which:
 - Each expert works separately;
 - Take one pass to get a feel for the product;
 - Take a second pass to focus on specific features.
- Debriefing session in which experts work together to prioritize problems.

Cognitive walkthroughs

Focus on ease of learning. Designer presents an aspect of the design & usage scenarios [7]. Expert is told the assumptions about user population, context of use, task details. One or more experts walks through the design prototype with the scenario. Experts are guided by 3 questions. These are:

1. Will the correct action be sufficiently evident to the user?
2. Will the user notice that the correct action is available?
3. Will the user associate and interpret the response from the action correctly?

3. Proposed Methodology

3.1 Rational Description for selecting low fidelity prototype

Web design is to be done efficiently so that visitors can access information easily and interact with the website very easily. So to make that happen designer has to go through so many alternative designs. The more alternatives we make we have to keep in mind that we have to think about our financial support and time according to user requirement. That's why we decided to use low fidelity prototype.

3.2 Usability Testing

From our study, we have tested the approaches which suit the common users. Typical users don't want some heavy, colorful, difficult to understand design. As we have mentioned earlier, our design has the clear and easily understandable abstraction for the typical user class. From the survey results, our users give positive results about our proposed design. The survey enables the users to express their valuable opinions and that opinion shows that, from other e-commerce sites, our site has more points about usability. This observational data is then used to determine performance of the design. We have found that the design has some simple errors like small scroll bar etc. But obtained data from the survey gives us the answers from users. With the help of these answers we can differentiate the basic and advance user needs. Basic user needs are already implemented in the design because most of the users are classified as basic users. If we can satisfy the need of most of the users, we can then concentrate on the advance users. Our simple design elaborates the user's expectation for e-commerce websites and related the usability [5, 6] part of the site. Our design consists of the goal which is termed as easily understandable design. Users can view the necessary information and product condition in the most possible easiest way from this design. Users can get the contract information as well as terms of service and unlike other sites, users find the terms of services in an easily findable location. This option made our design unique from the other designs.

4. Design

Here in this section we are presenting some exemplary design of e-commerce sites. Our proposed approaches are perfect for achieving usability goals and with better user experience than other state-of-the-art e-commerce sites [8]. Some screen shots of our designed e-commerce site for mobile phone accessories are shown below:

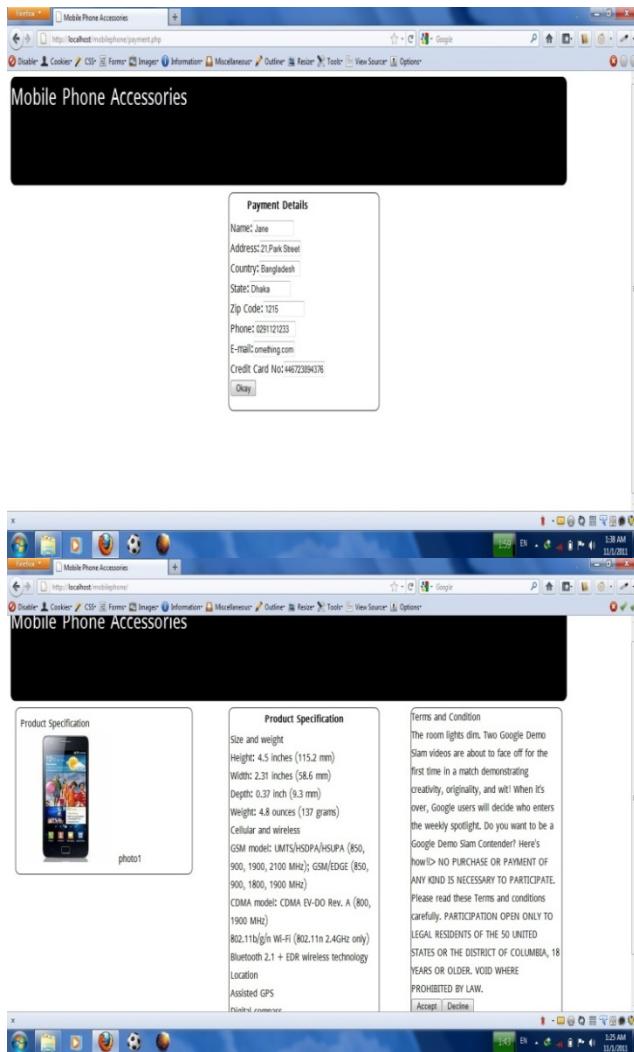


Fig: Two screenshots of our designed approaches

5. Conclusion & Future Works

We have discussed the prototype we used for designing and developing an e-commerce site for mobile phone accessories. And the necessary data and information is elucidated in the corresponding sections. Throughout our work we have tried to emphasize on the interaction issues corresponding to electronic commerce sites. And finally we have utilized the knowledge on various perspectives in HCI to accomplish our goal to ease the usage of e-commerce sites. Our future plan is to work thoroughly with this particular issue and with performance evaluation

of our proposed design with the state-of-the-art designs and to help novice users to accomplish the task of online transactions smoothly without any vulnerability of phishing and other malicious disorders.

References

- [1] <http://www.networksolutions.com/education/what-is-e-commerce/>
- [2] Bridging the Web Accessibility Divide. IV Ramakrishnan, J Mahmud, Y Borodin, M A Islam, F Ahmed, Electronic Notes in Theoretical Computer Science, Elsevier, 2009.
- [3] <http://en.wikipedia.org/wiki/Ethnography>
- [4] Accessibility challenges and tool features: an IBM Web developer perspective. Shari Trewin, Brian Cragun, Cal Swart, Jonathan Brezin, John Richards, W4A '10: Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A), ACM.
- [5] Interaction Design-beyond Human Computer Interaction, by Yvonne Rogers, Helen Sharp, Jenny Preece, Wiley, 2nd edition, 2007.
- [6] Human-Computer Interaction, by Alan Dix, Janet Finlay, Gregory Abowd, and Russell Beale. Prentice Hall, 2004.
- [7] Improving Accessibility of Transaction-centric Web Objects.. M A Islam, F Ahmed, Y Borodin, J Mahmud, I. V. Ramakrishnan, SIAM International Conference on Data Mining (SDM), 2010.
- [8] <http://www.usabilityinstitute.com/resources/stateOfTheArt.htm>

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, Wireless Network, Artificial Intelligence and Cloud Computing.

Jonayed Kaysar has completed Bachelor of Science in Computer Science and Information Technology from Islamic University of Technology, OIC. Research interest includes Human Computer Interaction, Peer to Peer and Overlay Network, Data Mining and Web Engineering.

Nafis Rahman has completed Bachelor of Science in Computer Science and Information Technology from Islamic University of Technology, OIC. Research interest includes Human Computer Interaction, Distributed Systems, Artificial Intelligence and Computational Algebra.

DTN Routing Protocols for VANETs: Issues and Approaches

Ramin Karimi, Norafida Ithnin ,Shukor Abd Razak , Sara Najafzadeh

Faculty of Computer Science and Information system
University technology Malaysia
Johor, Malaysia

Abstract

Delay Tolerant Networks (DTNs) are a class of networks that enable communication where connectivity issues like sparse connectivity, intermittent connectivity, high latency, long delay, high error rates, asymmetric data rate, and even no end-to-end connectivity exist. The DTN architecture model has been applied for vehicular Networks called Vehicular Delay Tolerant Networks (VDTN). In these networks vehicles use a message relaying service by their mobility in the network and collect messages from source nodes. In this paper, we will review the routing protocol in delay tolerant networks and compare the routing protocols that have been proposed specially for Vehicular Delay Tolerant Networks.

Keywords: *Delay tolerant networks; Vehicular Delay tolerant networks; Routing protocols.*

1. Introduction

Wireless networks have enabled a wide range of devices to be connected over vast distances. For example, today it is possible to connect from a cell phone to millions of powerful servers around the world [1]. Despite successfulness of these networks, they still cannot reach everywhere, and for some applications their cost is preventing. The main reason for these limitations is that current networking technology relies on a set of fundamental assumptions that are not true in all environments. The first and most important assumption is that an end-to-end connection exists from the source to the destination, possibly via multiple intermediaries [1]. These assumptions are not always true due to mobility, or unreliable networks. For example, when a wireless device is out of range of the network, it can use none of the application that requires network communication.

Delay-tolerant networks (DTN) try to extend the reach of networks. It promises to enable communication

between challenged networks, which includes space networks, mobile ad-hoc networks, and low-cost networks. The core idea is that these networks can be connected if protocols are designed to accommodate disconnection.

Vehicular Delay-Tolerant Networks have the potential to interconnect Vehicles in regions that current networking technology cannot reach. The main core is that an end-to-end connection may never be reached. To make end-to-end communication possible, relay nodes take the advantage of mobility for the data being transferred and forward it as the opportunity arises.

The DTN architecture implements a store-and forward paradigm by overlaying a protocol layer, called bundle layer that it is meant to provide internetworking on heterogeneous networks operating on different transmission media [2]. At the edge of each remote area network, a border system has an application layer gateway to terminate applications and produce data bundles. The DTN architecture concept was also extended to transit networks, called Vehicular DTN (VDTN). In these networks vehicles (e.g., cars, buses, and boats) are exploited to offer a message relaying service by moving around the network and collecting messages from source nodes. A number of projects have been based on this general concept. For example, the Message Ferry project to develop a data delivery system in disconnected areas [3, 4]. Another example is the DakNet project proposed to provide low-cost connectivity to the Internet to rural villages in India [5]. Vehicular Networks have also been used for traffic condition notifications, accident warnings, automatic tolling, free parking spots information, advertisements, and for example to gather data collected by vehicles like road pavement defects [6-7]. The rest of this paper is organized as follows. Section 2 reviews the routing protocols in Delay Tolerant Networks. Then, in Section 3, we have an overview on routing protocols proposed for Vehicular Ad Hoc Networks. Section 4 provides a comparative study of DTN routing protocols

VANETs. Finally, in Section 5, we draw some conclusions.

2. Routing Protocols in DTN.

First, we present two routing protocols, Epidemic Routing and Spray and Wait, which do not need any information about the network state. Then we review two other protocols named, Prophet, Mobyspace .In these routing protocols, nodes can memorize contact history and use it to make more informed forwarding decisions.

2.1 Epidemic Routing

Epidemic Routing protocol is for message delivery in a mostly disconnected network with mobile nodes. The protocol is basically a flooding mechanism especially for mobile wireless networks. It relies on exchanges of messages between nodes whenever they get in contact with each other to deliver the messages to their destinations. Each node have a buffer containing messages that have been generated at the current node as well as messages that has been generated by other nodes and relayed to this node.

When two nodes initiate a contact, they exchange their summary vectors in the anti-entropy session. Comparing message IDs, each node decides what messages it has not already received that it needs to pull from other nodes.

The second phase of a contact includes nodes exchanging messages. Each message has a time-to-live (TTL) field that limits the number of contacts they can pass through. Messages with TTL = 1 are forwarded only to the destination. In epidemic routing the messages are flooded in the whole network to reach just one destination. This creates contentions for buffer space and transmission time [8].

2.2 Spray and Wait.

Spyropoulos et al. present Spray and Wait, a zero-knowledge routing protocol introduced to reduce the wasteful flooding of redundant messages in a DTN [9]. Spray and wait like epidemic routing, forwards copy of messages to other nodes met randomly during connection in a mobile network. Spray and Wait disseminates a number of copies of the packet to other nodes in the network, and then waits until one of these copies meets the destination [9]. Spray and Wait includes two phases. In the first phase, the source node generates L copies of the message it holds, and then spreads these copies to other nodes for delivery to the destination node. The spreading

process works as follows: When an active node holding $n > 1$ copies meets another node, it hands off to it $F(n)$ copies and keeps for itself the remaining $n - F(n)$ copies and so forth until a copy of the message reaches the destination. F is the function that defines the spreading process [10].

The main difference from epidemic routing is that Spray and Wait limits the total number of disseminated copies of the same message to a constant number L [11].

2.3 PROPHET.

In [12] propose PROPHET (Probabilistic Routing Protocol using History of Encounters and Transitivity), a single copy history-based routing algorithm for DTNs. Each node in PROPHET estimates a delivery predictability vector containing an entry for each other node. A probabilistic metric called *delivery predictability* estimates the probability that node A will be able to deliver a message to node B . The delivery predictability vectors are maintained at each node A for every possible destination B . Predictability vectors will be used to decide on packet forwarding. When two nodes contact each other, node if the delivery predictability for the destination of the message is higher at the other node, a message is forwarded to the other. In addition to the predictability vector, a summary vector of stored packets will be also exchanged upon contact.

The information in the summary vector is used to decide on which messages to request from the other node. The entry update process happens whenever each contact and works as follows. Nodes that are often within mutual ranges have high delivery predictability for each other, and as they increase their corresponding delivery predictability entries. Nodes that rarely connect are less likely to be good forwarders of messages to each other, therefore they will reduce their corresponding delivery predictability entries [13].

2.4 MobySpace.

In [14] introduce a virtual location routing scheme which makes use of the frequency of visit of nodes to a discrete set of locations in the network area in order to decide on packet forwarding. Authors, define a virtual Euclidean space named MobySpace. In MobySpace, two nodes with a small distance between them are more likely to have a contact than two nodes that are further apart. Therefore the forwarding algorithm makes decision to forward a message during a contact to a node that has a shorter distance to the message destination. Messages take paths through the MobySpace to bring become to near to

the destination. Different distance functions have been proposed to measure node's mobility.

The MobyPoint of a node is not related to its physical GPS coordinate. The acquisition of the visit frequencies of the nodes to the location set is obtained by computing the respective fraction of time of being in a given location [10].

3. Delay-Tolerant Routing in VANETs.

Although most of the existing work on vehicle networks is limited to 1-hop or short range multihop communication, vehicular delay-tolerant networks are useful to other scenarios. For example, without Internet connection, a moving vehicle may want to query a data center ten miles away through a VANET. The widely deployed wireless LANs or infostations [11, 15, 16] can also be considered.

Vehicle delay-tolerant networks have many applications, such as delivering advertisements and announcements regarding sale information or remaining stocks at a department store. Information such as the available parking spaces in a parking lot, the meeting schedule at a conference room, and the estimated bus arrival time at a bus stop can also be delivered by vehicle delay-tolerant networks.

For the limited transmission range, only clients around the access point can directly receive the data. However, this data may be beneficial to people in moving vehicles far away, as people driving may want to query several department stores to decide where to go. A driver may query the traffic cameras or parking lot information to make a better travel plan. A passenger on a bus may query several bus stops to choose the best stop for bus transfer. All these queries may be issued miles away from the broadcast site. With a vehicular delay-tolerant network, the requester can send the query to the broadcast site and get a reply from it. In these applications, the users can tolerate up to a minute of delay as long as the reply eventually returns. In this section we will review the delay tolerant protocols specially proposed for Vehicular Networks.

3.1. VADD- Vehicle- Assisted Data Delivery

Zhao and Cao [17] proposed several vehicle-assisted data delivery (VADD) protocols. All of them share the idea of storing and forwarding data packets. That is, nodes can decide to keep the message until a more promising neighbor appears on their coverage range, but trying always to forward them as soon as possible. Additionally, decisions about which streets must be followed by the packet are made using vehicle and road information such as current speed, distance to the next junction, and maximum speed allowed. These routing

decisions are dynamically taken at junctions because the authors state that pre-computed optimal paths used by other protocols might rapidly lose their optimality due to the unpredictable nature of VANETs.

In VADD the main goal is to select the path with the smallest packet delivery delay. The behavior of the protocol depends on the location of the node holding the message. Two cases are considered: when nodes routing the message are located in the middle of a road and when they are located in a junction. The first case (also called routing in straight way) presents less alternatives: forwarding the packet toward the next junction or to the previous one. However, the second case (also called routing in intersections) is much more complicated because at junctions, the routing decision must consider the different roads, so that the number of options is higher.

3.2 GeOpps- Geographical Opportunistic Routing

GeOpps for vehicular networks [18] is a trajectory-based protocol that uses both the opportunistic nature of vehicular mobility patterns and the geographic information provided by navigation systems. This protocol assumes each vehicle is assumed to know its complete trajectory.

GeOpps applies a delay-tolerant mechanism, therefore a vehicle store data packets until a suitable next hop for them is found later on. To choose the next hop, each node computes the closest point in their trajectory in direction of the destination of the packet.

GeOpps is a protocol for delay-tolerant data, but the performance depends heavily on accurate the trajectory information. Therefore if the driver does not follow the route suggested by the GPS, the routing decision taken might be erroneous [19].

3.3 GeoDTN+Nav- Geographic DTN Routing with Navigator.

Traditionally, geo-routing routes packets in two modes: the first mode is the greedy mode, and the second mode is the perimeter mode. In greedy mode, a packet is forwarded to destination greedily by choosing a neighbor which has a bigger progress to destination among all the neighbors. However, due to obstacles the packet can arrive at a local maximum where there is no neighbor closer to the destination than itself. In this case, the perimeter mode is applied to extract packets from local maxima and to eventually return to the greedy mode. After a planarization process, packets are forwarded around the obstacle towards destination. In this way, the packet delivery is guaranteed as long as the network is connected. However, the assumption that the network is connected may not always be true. Due to the mobile characteristics of VANET, it is common that the network is disconnected or

partitioned, particularly in sparse networks. The greedy and perimeter modes are not sufficient in VANET. Therefore, they introduce the third mode: DTN (Delay Tolerant Network) mode, which can deliver packets even if the network is disconnected or partitioned by taking advantage of the mobility of vehicles in VANET. Unlike the common belief that mobility harms routing in VANET, this protocol specifically counts on it to improve the routing [20].

4. Comparison of Geographic Routing Protocols in VANETs.

Table1 indicates comparison study of routing protocol in vehicular delay tolerant networks. All of these protocols in table1 are position-based, using knowledge of vehicles' positions and velocities to route messages. In this table, the term traffic-aware use of traffic information for choosing a suitable route based on some information about the routes in the streets. All these protocols depend on digital map to determine positions of vehicle. In this table, recovery means using strategy to recover local optimum or repair broken routes. All protocols in table 1 use greedy

forwarding by choosing farthest neighbours and in sparse network. They use store-carry forward mechanism .They buffer the packet until they find a chance to forward it to other vehicle.

5. Conclusions

In this paper, we have reviewed Delay tolerant routing protocols and then the routing protocols specially proposed for Vehicular Ad hoc Networks. Finally, we give a comparative study of DTN protocols for Vehicular Ad hoc Networks in terms of map required, traffic data required and traffic aware and recovery.

Acknowledgments

This work was supported by grant from Universiti Teknologi Malaysia, the GUP grant Vote No.Q.J130000.7182.01J52,K-economy.

TABLE 1- QUALITATIVE COMPARISON OF VDTN ROUTING PROTOCOLS

characteristics \ Routing Protocols	Delay/tolerant	Traffic aware	Map required	Recovery	Greedy required	Buffering(carry-and-forward)
VADD	✓	✓	✓		✓	✓
GeOpps	✓		✓		✓	✓
GeoDTN+Nav	✓	✓	✓	✓	✓	✓

References

- [1] K. Fall, "A delay-tolerant network architecture for challenged internets," in Proceedings of ACM SIGCOMM, August 2003, pp. 27–34.[Online].Available: <http://doi.acm.org/10.1145/863955.863960>
- [2] V. Cerf, S. Burleigh, A. Hooke, L. Torgerson, R. Durst, K. Scott, K. Fall, and H. Weiss, "Delay-Tolerant Networking Architecture," RFC 4838, April 2007, [Online]. Available: <ftp://ftp.rfc-editor.org/innotes/rfc4838.txt>
- [3] W. Zhao, M. Ammar, and E. Zegura, "A Message Ferrying Approach for Data Delivery in Sparse Mobile Ad Hoc Networks," in *The Fifth ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc 2004)*, Roppongi Hills, Tokyo, Japan, May 24-26, 2004, pp. 187–198.
- [4] H. Ammar, "Message Ferrying: Proactive Routing in Highly-Partitioned Wireless Ad Hoc Networks," in *The Ninth IEEE Workshop on Future Trends of Distributed Computing Systems*, San Juan, Puerto Rico, May 28-30, 2003, pp. 308-314.
- [5] A. Pentland, R. Fletcher, and A. Hasson, "DakNet: Rethinking Connectivity in Developing Nations," in *IEEE Computer*, vol. 37, 2004, pp. 78-83.
- [6] R. Tatchikou, S. Biswas, and F. Dion, "Cooperative Vehicle Collision Avoidance using Inter-vehicle Packet Forwarding," in *IEEE Global Telecommunications Conference (IEEE GLOBECOM 2005)*, St. Louis, MO, USA, 28 Nov. - 2 Dec., 2005.
- [7] L. Franck and F. Gil-Castineira, "Using Delay Tolerant Networks for Car2Car Communications," in *IEEE International Symposium on Industrial Electronics 2007 (ISIE 2007)*, Vigo, Spain, 4-7 June, 2007, pp. 2573-2578.
- [8] Vahdat, A. and Becker, D., Epidemic Routing for Partially Connected Ad Hoc Networks, in Technical Report, April 2000.
- [9] Spyropoulos, T., Psounis, K., and Raghavendra, C.S., Spray and Wait: An Efficient Routing Scheme for Intermittently Connected Mobile Networks, in Proc. ACM SIGCOMM Workshop on Delay-Tolerant Networking, Philadelphia, PA, USA, 2005.

- [10] A.AHanbali,M.Ibrahim,V.Simon,A Survey of message Diffusion protocols in Mobile Ad Hoc Networks ,2008
- [11] S.Olariu,M.C.Weigle , Vehicular Networks From Theory to Practice,CRC Press,2009
- [12] A. Lindgren, A. Doria, and O. Scheln. Probabilistic routing in intermittently connected networks. In *Proc. of ACM MobiHoc (poster session)*, Maryland, USA, June 2003.
- [13] Lindgren, A., Doria, A., and SchelSn, O., Probabilistic Routing in Intermittently Connected Networks, in Poster of ACM MOBIHOC, Annapolis, MD, USA, 2003.
- [14] J. Leguay, T. Friedman, and V. Conan. Evaluating mobility pattern space routing for DTNs. In *Proc. Of IEEE INFOCOM*, Barcelona, Spain, Apr. 2006.
- [15] Frenkiel, R.H., Badrinath, B.R., Borras, J., and Yates, R., The Infostations Challenge: Balancing Cost and Uiquity in Delivering Wireless Data, *IEEE Personal Communications Magazine*, 7(2), 66–71, April 2000.
- [16] Goodman, D., Borras, J., Mandayam, N., and Yates, R., INFOSTATIONS: A New System Model for Data and Messaging Services, in Proc. of IEEE VTC, Braunschweig, Germany, May 1997.
- [17] Zhao, J., and Cao, G., VADD: Vehicle-Assisted Data Delivery in Vehicular Ad Hoc Networks, in Proc. of IEEE INFOCOM, Barcelona, Spain, 2006.
- [18] Leontiadis, I., Mascolo, C. (2007), “GeOpps: Geographical Opportunistic Routing for Vehicular Networks,” World of Wireless, Mobile and Multimedia Networks, 2007. WoWMoM 2007. IEEE International Symposium on a , vol., no., pp.1-6, 18-21 June 2007.
- [19] H.Moustafa,Y.Zhang,Vehicular Networks ,Techniques, Standards, and Applications,CRC Press,2009.
- [20] Cheng, P.-C., Weng, J.-T., Tung, L.-C., Lee, K. C., Gerla M., and Härr J. (2010), "GeoDTN+NAV: Geographic DTN Routing with NavigatorPrediction for Urban Vehicular Environments ,", 2010.

Shukor Abd Razak is a senior lecturer at Universiti Teknologi Malaysia. His research interests are on the security issues for the Mobile Ad Hoc Networks, Mobile IPv6 networks, Vehicular Ad Hoc Network and network security. He is the author and co-author for many journal and conference proceedings at national and international levels.



Sara Najafzadeh is currently a Ph.D candidate in Department of Computer Science and Information Technology at Universiti Teknologi Malaysia, Johor, Malaysia. She received M.Sc degree in computer engineering from Iran University of Science and Technology in 2006. Her research interests include Vehicular Ad Hoc Networks, Mobile ad-hoc networks and communication Networks.

Ramin Karimi is currently a Ph.D candidate in Department of Computer Science and Information Technology at Universiti Teknologi Malaysia, Johor, Malaysia. He received M.Sc degree in computer engineering from Iran University of Science and Technology in 2006. His research interests include Vehicular Ad Hoc Networks, Mobile ad-hoc networks, security and communication Networks.

Norafida Ithnin is a senior lecturer at Universiti Teknologi Malaysia. She received her B.Sc degree in computer science from Universiti Teknologi Malaysia in 1995, her MSc degree in Information Teknologi from University Kebangsaan Malaysia in 1998 and her PHD degree in computer science from UMIST, Manchester in 2004. Her primary research interests are in security, networks, Mobile ad-hoc networks, Vehicular Ad Hoc Networks.



Theoretical Analysis of VoIP Transmission Through Different Wireless Access Technologies

Emma CHARFI, Lamia CHAARI and Lotfi KAMOUN

National School Engineering of SFAX, Laboratory of Electronics and Information Technologies,
University of SFAX, SFAX

Abstract

One of the challenges in today's wireless networks is to provide appropriate throughput for data and multimedia application. The physical data rate enhancements can be achieved through new physical capabilities, however to achieve high efficiency and to improve the throughput at the medium access control (MAC) layer, new and innovative MAC mechanisms are required. The disgraceful overhead occurs at the MAC layer prevents the WLANs from achieving desirable performance, this problem becomes more severe in the very high-speed WLAN. We will consider the potential benefits of frame aggregation in order to enhance the throughput. In this paper, the latest MAC layer mechanisms of IEEE 802.11, IEEE 802.11e and 802.11n standards are explained in details. Besides a theoretical analysis is used to evaluate and analyze the theoretical capacity of VoIP over different emerging wireless access technologies, ranging from IEEE 802.11 to IEEE 802.11n. Results confirm our expectation.

Keywords: Access technologies, frame Aggregation, throughput enhancement, VoIP, WLAN.

1. Introduction

Recently, IEEE 802.11 WLAN has gained a widespread position in the market for wireless networking. The 802.11 standard defines both the medium access control (MAC) layer and the physical layer (PHY) specifications [1]. The mandatory part of the 802.11 MAC is called the Distributed Coordination Function (DCF), which is based on Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) and the optional one is the Contention Free Period (PCF). The Enhance the current 802.11 MAC to expand support for applications with Quality of Service requirements such as VOIP, and in the capabilities and efficiency of the protocol requires new concepts and solutions in MAC mechanisms most of them are focused on overhead reduction. The overhead comes primarily from packet preambles, acknowledgements, contention windows and various interframe-spacing parameters. Through the IEEE 802.11e[2] and the IEEE802.11n [3][4] amendments, new channel access

techniques have been introduced to enhance the throughput and to reduce the overhead. The most known techniques are:

Enhanced Distributed Coordination Access (EDCA)[5][6]: which is an extension of the basic DCF introduced in the 802.11e amendment to support prioritized quality of service.

Block Acknowledgement protocol (BA)[7][8][15], was introduced with the 802.11e amendment to improve efficiency by allowing for the transfer of a block of data frames that are acknowledged with a single Block Acknowledgement (BA) frame instead of an ACK for each data frame.

Reduced inter-frame space (RIFS) [8].

Frame aggregation at MAC Service Data Unit (MSDU)[4][9][10], the principle of MSDU aggregation is to allow multiple MSDUs to be sent to the same receiver concatenated in a single MPDU.

Frame aggregation at MAC Protocol Data Unit (MPDU) [10][11], the principle of MPDU aggregation is to join multiple MPDUs to be sent with a single PHY header.

All these enhancements improve 802.11e MAC efficiency while retaining the reliability, simplicity, interoperability and QoS support of 802.11/802.11e MAC.

Many others mechanisms have been suggested in the literature to enhance the capacity of wireless technologies. In [12] the authors develop a new wireless link quality metric, ECOT that enables a high throughput route setup in wireless mesh networks. The key feature of ECOT is being applicable to diverse mesh network environments where IEEE 802.11 MAC (Medium Access Control) variants are used. They take into account the following features (EDCA with Block Acknowledgment and 802.11n A-MPDU Aggregation).

In [13] the authors propose an adaptive delayed channel access that outperforms the current 802.11n specification. They introduce an adaptive aggregate size threshold that gradually increases or decreases until the number of segments in the sender's buffer size of the TCP flow is reached.

In the literature there also some works that evaluates via simulations or analytical models the wireless access protocols efficiency.

In [14] the authors analyze the MAC protocols performance in imperfect wireless channel and develop an analytical model to evaluate the unsaturated throughput performance of the frame-burst-based CSMA/CA protocol. In this article we mainly focus on the application of IEEE802.11 WLAN mechanisms to real-time services such as VoIP. To conduct this study thoroughly, the theoretical capacity and the efficiency of the studied approach are computed related to the physical rate. This paper is organized as follows: In Section 2 the theoretical capacity of VoIP transmission through basic MAC protocols (CSMA/CA, PCF, RTS/CTS) that have been adopted in the IEEE 802.11a/b/g is analyzed. Section 3 deals with determining the performance for some extension concepts that have been introduced in the 802.11e amendment to support prioritized quality of service. We mainly focus on the EDCA and Immediate Block Acknowledgement with SIFS or RIFS mode. Section 4 concentrates on the latest work of MAC 802.11n enhancements. Frame aggregation methods that TGn has proposed in the latest 802.11n standard draft are evaluated for VoIP traffic., and how these can improve WLANs throughput and maximize efficiency is discussed. For each section, we begin with a brief outline of the studied MAC concept, followed by a discussion of its maximal throughput limitations because of overhead. Section 6 concludes the paper by summarizing this article's findings.

2. Theoretical Capacity analysis of VoIP transmission through basic MAC protocols

2.1 Carrier Sense Multiple Access with Collision Avoidance: CSMA/CA (DCF)

The principle of the mechanism CSMA/CA is based on listening to the channel to see if it is occupied. The node must insure that the medium is free for a certain length (DIFS) before emitting. If the medium is idle and continues to be idle for a period of time set in DIFS, then the station gains access to the medium and can start sending the pending frame. However, if the medium become busy during the time the station is monitoring the medium, a random backoff procedure will start. A random backoff time is chosen in the interval $[CW_{min}, CW_{max}]$ where, CW in the contention window. The timer will decrease the backoff time when the medium is idle. If the medium is busy during a station's backoff procedure, the backoff timer will be suspended. When many stations are competing for the medium, the station that chose the

shortest backoff time will gain access to the medium first. DCF uses positive acknowledgment. When a frame is successfully received by the destination station, an ACK frame is sent back to the source station after a SIFS period. The principle of the CSMA/CA mechanism is illustrated by fig.1.

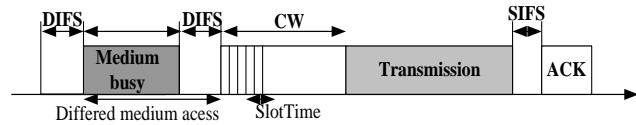


Fig. 1 CSMA/CA mechanism

In the continuation, we considered the scenario described by fig.2. The network comprises a single IEEE 802.11 basic service set (BSS) with one access point (AP), and a number of wireless users. Temporal multiplexing between N stations [16][17] is illustrated by fig.3.



Fig. 2 Network topology

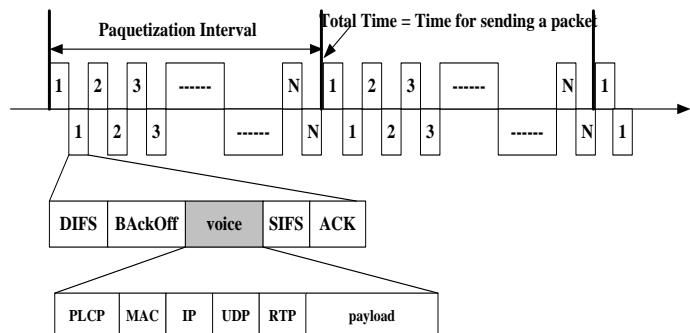


Fig.3 Temporal multiplexing

We focus on the capacity of the wireless network as the principal metric of interest. We define capacity in this context to be the maximum number of simultaneous, bidirectional calls that can be supported. Voice traffic is generated by packetizing the output of a voice encoder (we consider G.711, G723 and G.729 schemes, without the use of silence suppression), creating packets each containing a fixed amount of voice data; we consider 20 ms. of voice data per packet.

In wireless networks, G.711 is applied for encoding telephone audio signal at a rate of 64 kbps with a sample rate of 8 kHz and 8 bits per sample. In an IP network, voice is converted into packets with durations of 5, 10 or

20 ms of sampled voice, and these samples are encapsulated in a VoIP packet. G.723.1 codec belongs to the Algebraic Code Excited Linear Prediction (ACELP) family of codec and has two bit rates associated with it: 5.3 kbps and 6.3 kbps. The coder operates on speech frames of 30 ms corresponding to 240 samples at a sampling rate of 8000 samples/s. G.729 codec belongs to the Code Excited Linear Prediction coding (CELP) model speech coders and uses Conjugate Structure - Algebraic Code Excited Linear Prediction (CS-ACELP). This coder was originally designed for wireless applications at fixed 8 kbit/s output rate. The coder works on a frame of 80 speech samples (10 ms). In [18] factors affecting QoS such as packet loss, jitter, throughput, and delay for various capacity networks are studied for several codec's.

VoIP packets are transmitted over the network using RTP over UDP/IP. We present an upper bound on the network capacity, by making certain assumptions about the performance of the network.

The number of calls given by Eq(1) is based on temporal multiplexing provided by fig.3.

$$N = \frac{\text{Packet}_{\text{Ineterval}}}{2T_{\text{total}}} \quad (1)$$

Where the total duration of one packet is equal to:

$$T_{\text{total}} = T_{\text{DIFS}} + T_{\text{CW}} + T_{\text{SIFS}} + T_{\text{ACK}} + T_{\text{voice-pkt}}$$

The duration necessary for transmission one voice packet is given by Eq.2.

$$T_{\text{voice-pkt}} = T_{\text{PLCP}} + \frac{L_{\text{voice-pkt}}}{\text{Bit rate}} \quad (2)$$

$$L_{\text{voice-pkt}} = L_{\text{MAC}} + L_{\text{IP}} + L_{\text{UDP}} + L_{\text{RTP}} + L_{\text{voice}}$$

Thus, the total number of calls will be given by Eq.3:

$$N = \frac{\text{Packet}_{\text{Ineterval}}}{2(T_{\text{PLCP}} + T_{\text{DIFS}} + T_{\text{CW}} + T_{\text{SIFS}} + T_{\text{ACK}} + \frac{L_{\text{voice-pkt}}}{\text{Bit rate}})} \quad (3)$$

Fig.4 shows the capacity of the wireless network (number of calls) for different voice encoder. Parameters used to compute the numbers of calls are regrouped in table1.

The performance of the G.711, G.723.1 and G.729 codec are shown in Fig4. With all codec, there is a saturation of the capacity of the capacity with physical rate. With G.711, going from 802.11b to 802.11g enhances the capacity from 10 VoIP calls to 14calls. For each voice frame, a RTP/UDP/IP header has to be added. The proportion of this overhead is particularly high for small data packets.

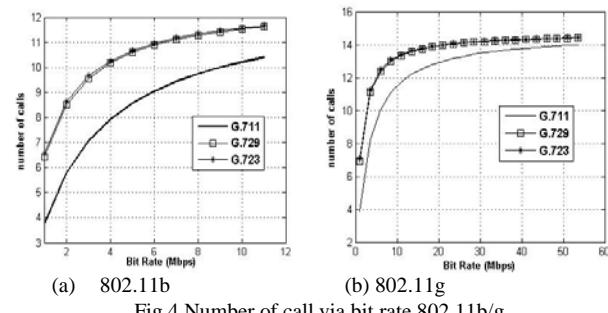


Fig.4 Number of call via bit rate 802.11b/g

2.2 RTS/CTS (Request To Send/ Clear To Send)

In order to eliminate hidden node problem, the IEEE802.11 MAC protocol defines an optional mechanism known as "Request To Send / Clear To Send" RTS/CTS. The node must insure that the medium is free for a certain length (DIFS) before emitting. If the medium is idle and continues to be idle this period, then the station gains access to the medium and can start sending the pending frame. Otherwise, the node enters in a backoff period, and after that, it transmits a short frame called RTS to reserve the channel. When the receiving node receives the RTS frame, it responds, after SIFS period with a clear to send (CTS) frame. The transmitting node is allowed to transmit only if the CTS frame is correctly received. Thus, the channel is reserved for the length of the transmission. When a frame is successfully received by the destination station, an ACK frame is sent back to the source station after a SIFS period.

As indicated in figure.5, RTS/CTS mechanism will change the timing diagram of successful data transmission and the VoIP capacity will be computed as below:

Table 1: Used Parameters

	Time (μs)	Length (Byte)
PLCP preamble	144.00	18
PLCP Header	48.00	6
PLCP	192.00	24
IP		20
UDP		8
RTP		12
Data voice	116.36	160 (G711), 20 (G729), 16 (G723)
ACK	10.18	14
Slot Time	9 (802.11g, et n) 10 (802.11b)	
SIFS	16 (802.11g et n) 20 (802.11b)	
DIFS	34 (802.11g et n) 50 (802.11b)	
PIFS	25 (802.11g et n) 30 (802.11b)	

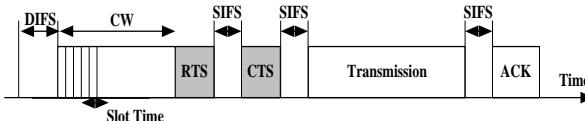


Fig.5 RTS/CTS mechanism

The total duration of one packet is equal to:

$$T_{total} = T_{DIFS} + T_{CW} + T_{RTS} + T_{CTS} + 3T_{SIFS} + T_{ACK} + T_{voice-pkt}$$

Thus, the total number of calls will be:

$$N = \frac{Packet_{Inerval}}{2(T_{PLCP} + T_{DIFS} + T_{CW} + T_{RTS} + T_{CTS} + 3T_{SIFS} + T_{ACK} + \frac{L_{voice-pkt}}{Bit_rate})} \quad (3)$$

Fig.6 shows the capacity of the wireless network (number of calls) for different voice encoder when RTS/CTS mechanism is implemented.

The RTS/CTS handshake leads to more overhead. As can be seen by comparing figures 4a, 4b and figures 6a,6b , the VoIP capacity is better when CSMA/CA is used. CSMA/CA performs better than RTS/CTS since it uses less control frames.

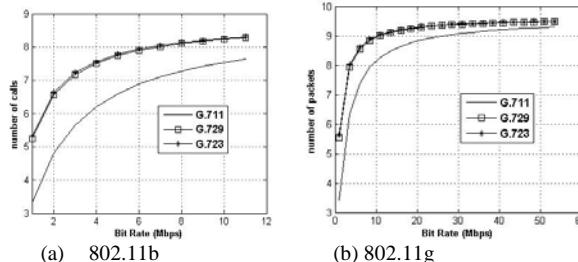


Fig.6 Number of call via bit rate 802.11b/g for RTS/CTS

2.3 Contention Free period: CFP

PCF mode may be an alternative way to convey real-time traffic over IEEE 802.11 WLANs. The idea is that by using a centralized controller, it is easier to realize QoS assurance in the central controller. With PCF, the point coordinator (PC), which resides in the AP, establishes a periodic contention free period (CFP) during which contention free access to the wireless medium is coordinated by the PC. The CFP period is initialized by the transmission of a Beacon frame. During the CFP the NAV of all nearby stations is set to the maximum expected duration of the CFP. In addition, all frame transfers during the CFP use an inter frame spacing that is less than that used to access the medium under DCF, preventing stations from gaining access to the medium using contention-based mechanisms. At the end of the CFP, the PC transmits a CF-End frame. The transmission in PCF period is shown by figure.7.

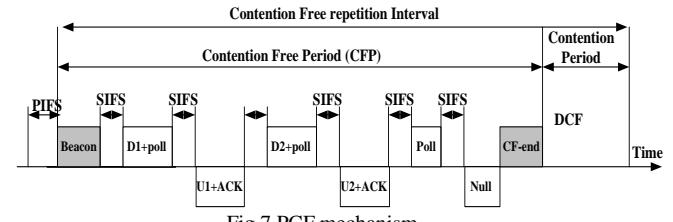


Fig.7 PCF mechanism

The total duration of transmission is given by:

$$T_{total} = T_{SIFS} + T_{voice-pkt} + T_{CFACK} + T_{CFPol}$$

Thus, the total number of calls will be:

$$N = \frac{T_{CFP} - T_{Beacon} - T_{SIFS} + T_{CFend} + T_{PIFS}}{2(T_{PLCP} + T_{SIFS} + \frac{L_{voice-pkt}}{Bit_rate} + T_{CFPoll} + T_{CFAck})}$$

Parameters used to compute the numbers of calls in PCF mode are regrouped in table 2.

Table 2: PCF Parameters

Parameter	Length (Byte)
Beacon	40
data+CF-Poll,data+CF-ACK	28+Payload
CF-End,CF-End+CF-ACK	29

Fig.8 compares the capacity of the wireless network (802.11b, 802.11g) for different transmission mechanism (CSMA/CA, RTS/CTS and PCF) when G.711 voice encoder is used.

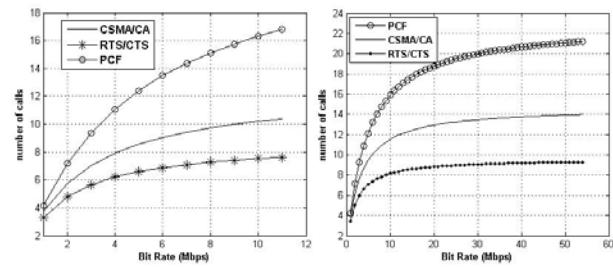


Fig.8 Number of call via bit rate 802.11b/g Access mechanisms

PCF achieves higher capacity than using DCF when polled stations have always packet to send. This is can be credited to the higher overhead needs when using DCF such as performing Backoff before transmitting. Moreover PCF allows the uplink and downlink to use piggy-packed frames.

2.4 Throughput and efficiency analysis

The objective for this subsection is to see how overheads can affect the system throughput and system efficiency through a numerical analysis. For this analysis, Packet errors rate is

equal to 10%. The throughput “T” is given by equations 5, 6, 7 respectively for CSMA/CA, RTS/CTS and PCF modes:
 For CSMA/CA, Eq(5)

$$T = \frac{\text{Payload} \cdot (1 - \text{PER})}{(T_{PLCP} + T_{DIFS} + T_{CW} + T_{SIFS} + T_{ACK} + \frac{L_{\text{voice-pkt}}}{\text{Bit rate}})} \quad (5)$$

For RTS/CTS, Eq(6)

$$T = \frac{\text{Payload} \cdot (1 - \text{PER})}{(T_{PLCP} + T_{DIFS} + T_{CW} + T_{RTS} + T_{CTS} + 3T_{SIFS} + T_{ACK} + \frac{L_{\text{voice-pkt}}}{\text{Bit rate}})} \quad (6)$$

For PCF, Eq(7)

$$T = \frac{\text{Payload} \cdot (1 - \text{PER})}{(T_{PLCP} + T_{SIFS} + \frac{L_{\text{voice-pkt}}}{\text{Bit rate}} + T_{CFPoll} + T_{CFAck})} \quad (7)$$

Fig.9 compares the throughput of the wireless network (802.11b, 802.11g) for different transmission mechanism (CSMA/CA, RTS/CTS and PCF).

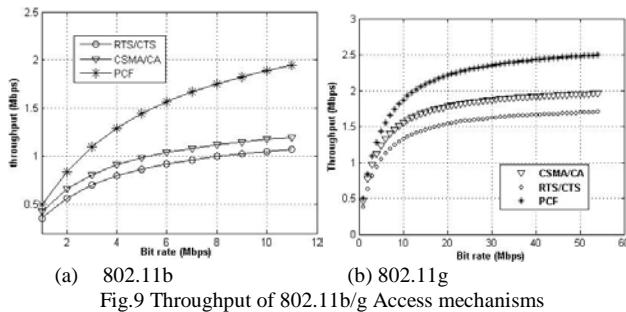


Fig.9 Throughput of 802.11b/g Access mechanisms

The throughput, for the two cases, is increasing when the physical rate increases. For IEEE802.11g, the throughput reaches more important values than these for IEEE802.11b. For example, at 11Mbps for IEEE802.11b, the throughput is equal to 1.1, 1.4, and 1.9 Mbps respectively for RTS/CTS, CSMA/CA, and PCF. While for IEEE802.11g, at 54Mbps physical rate, these values are equal to 1.7, 2, and 2.5Mbps for the previous mechanism.

The efficiency is given by Eq (8):

$$\eta = \frac{\text{Throughput}}{\text{Bit rate}} \quad (8)$$

The efficiency, for the two standards, is represented by fig.10.a for 802.11b, and fig.10.b for 802.11g. It decreases when the physical rate increases. Since the throughput of PCF is the best one, then the efficiency is better for the Contention Free Period. The efficiency decreases immensely when the physical rate increases. Indeed, for IEEE802.11g, it was equal to 48% at 1Mbps

for CSMA/CA, and it decrease to 16% at 10Mbps, and it decrease again at 54Mbps and achieves 5%.

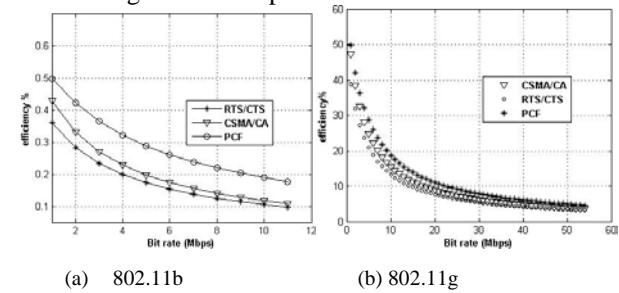


Fig.10 Efficiency of 802.11b/g Access mechanisms

3. Theoretical Capacity analysis of VoIP transmission through Enhanced Distributed Channel Access

This scheme consists on grouping the frame and sharing the access time in the channel between several frames possessing the same destination. Thus, the frames are sent in a burst during the period of a transmit opportunity (TXOP). As defined by figure.11, each frame is acknowledged by an ACK frame, SIFS after the transmission of the data. In this scheme, a station is able to transmit after an AIFS period followed by a counter backoff if the medium is busy.

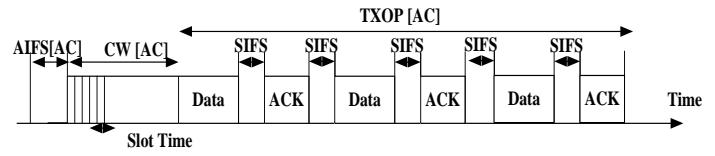


Fig.11 EDCA mechanism

The parameters of the types of data, voice and video are listed in table3.

Table 3: EDCA parameters

	CWmin	CWmax	AIFSN	TXOPLimit
voice	7	15	2	1.504(ms)
video	15	31	2	3.008(ms)

The total duration of one packet is equal to:

$$T_{\text{total}} = T_{\text{AIFS}} + T_{\text{Backoff}} + T_{\text{ACK}} + T_{\text{SIFS}} + T_{\text{voicepkt}}$$

The total number of calls is:

$$N = \frac{\text{Packet}_{\text{Ininterval}}}{2(T_{PLCP} + T_{SIFS} + T_{CW} + T_{AIFS} + T_{ACK} + \frac{L_{\text{voice-pkt}}}{\text{Bit rate}})}$$

The number of calls, fig.12, increases when the physical rate increases. It is more important for packets of voice,

than video packets. At 54Mbps, it is equal to 17 for voice packets while it is limited to 14 for video packets.

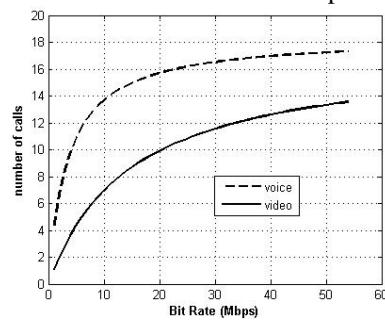


Fig.12 Number of calls for EDCA

The throughput is represented by [fig.13](#), and it is equal to « T »:

$$T = \frac{\text{Payload} \cdot (1 - \text{PER})}{(T_{PLCP} + T_{AIFS} + T_{CW} + T_{SIFS} + T_{ACK} + \frac{L_{voice-pkt}}{\text{Bit rate}})}$$

The throughput is greater when we are transmitting video packets. Indeed, the throughput reaches 5Mbps for video packet transmitted, and it is limited to 1Mbps for voice packet. When we want transmitting voice packets, it is more beneficial if we use the mechanism DCF then using EDCA.

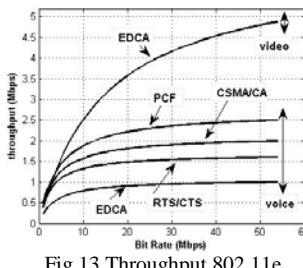


Fig.13 Throughput 802.11e

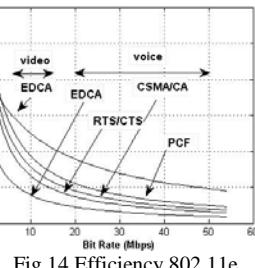


Fig.14 Efficiency 802.11e

Thus, the efficiency, which is presented by [fig.14](#), is better when we are transmitting video packets. Similarly, the transmission of voice packets is more efficient by using the mechanism CSMA/CA, then proceeding to EDCA model. The problem here is that the efficiency decreases immensely when the physical rate increases. Indeed, for the transmission of voice packets using the EDCA mode, the efficiency is equal to 25% at 1Mbps and is decreased to 5% at 20Mbps, and 2% at 54Mbps.

4. Theoretical Capacity analysis of VoIP transmission through Enhanced Distributed Channel Access

This scheme consists on grouping the frame and sharing the access time in the channel between several frames possessing the same destination. Thus, the frames are sent

in a burst, separated with SIFS, during the period of a transmit opportunity (TXOP). All the frames sent are acknowledged by a unique Block Acknowledgement (BA) instead of an ACK frame for each frame transmitted, as it is presented by [fig.15](#).

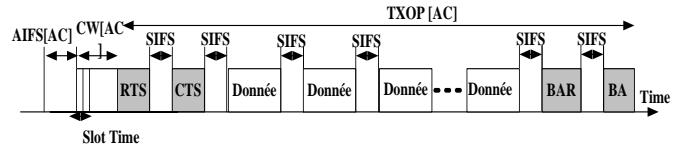


Fig.15 Immediate BA principle

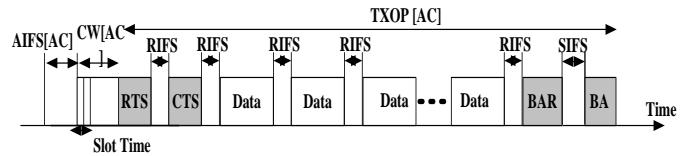


Fig.16 Immediate BA with Reduced Interface RIFS

The station transmits a request for BA (BAR), and the receiver respond with BA after a SIFS period. The transmission takes place during a TXOP period. The Immediate BA with RIFS mode is similar to Immediate BA SIFS mode, but the frames are separated with RIFS which is less than SIFS, as it is shown by [fig.16](#). The total duration for transmission of one packet is:

$$T_{total} = T_{AIFS} + T_{CW} + T_{RTS} + T_{CTS} + 3T_{SIFS} + a * (T_{voicepkt} + T_{SIFS}) + T_{BAR} + T_{BA}$$

Where “a” is the number of packet per TXOP period, which is defined as:

$$TXOP_{Limit} = T_{RTS} + T_{CTS} + 3T_{SIFS} + a * (T_{voicepkt} + T_{SIFS}) + T_{BAR} + T_{BA}$$

Thus “a” is :

$$a = \frac{TXOP_{Limit} - T_{RTS} - T_{CTS} - 3T_{SIFS} - T_{BAR} - T_{BA}}{(T_{PLCP} + T_{SIFS} + \frac{L_{voice-pkt}}{\text{Bit rate}})}$$

The number of video packets per TXOP is shown by [fig.17](#), and it is more important than voice packets for the two model IBA SIFS mode or IBA RIFS mode.

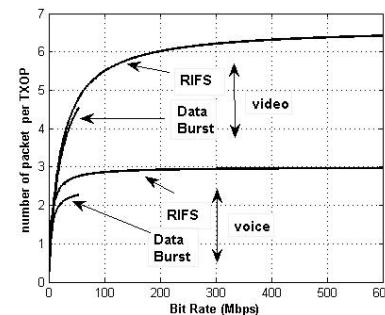


Fig.17 Number of packets per TXOP

Since the introduction of the RIFS mode is appeared with IEEE802.11n, the physical rate is spread to 540Mbps (Very High Throughput). For the mode RIFS, the number of packets per TXOP is better than the SIFS mode, and continues to increase and keeps a constant values at VHT. For Block Acknowledgement using RIFS, we replace

$$\frac{L_{voice-pkt}}{Bit_{rate}} + T_{SIFS} \quad \text{by} \quad \frac{L_{voice-pkt}}{Bit_{rate}} + T_{RIFS}$$

Let A denote the total number of packets for all N stations:
 $A = N * a$, Where N is the number of stations, and it is

$$\text{equal to: } N = \frac{\text{Packet}_{Ineterval}}{2T_{total}}$$

Since T_{total} has a fixed value, N is a constant number.

$$T_{total} = T_{AIFS} + T_{CW} + TXOP_{lim_it}$$

The total number of packets, as shown by fig.18, is more important when using the IBA mode RIFS, but at VHT, and when we use reduce inter frame space, the total number of packet increase and is nearly equal for voice and video parquets.

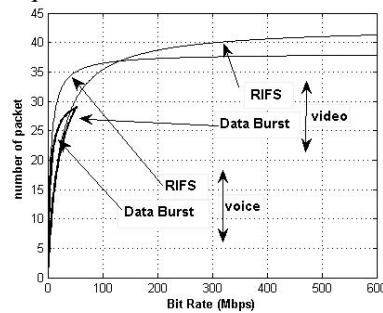


Fig.18 Total number of packets

The throughput is defined by:

$$T = \frac{\text{Payload} * a * (1 - PER)}{[T_{AIFS} + T_{CW} + 3T_{SIFS} + T_{RTS} + T_{CTS} + T_{BAR} + T_{BA} + T_{ACK} + a \left(T_{PLCP} + \frac{L_{VOICE}pkt}{Bit_{rate}} + T_{SIFS} \right)]}$$

Fig.19 and fig.20 represent the throughput and the efficiency of voice and video respectively. The throughput is more important by using the RIFS mode than the SIFS mode especially at Very High Throughput. It increases as the number of packets per TXOP increases. The efficiency presents the same limitation as previously. It decreases immensely, and it is near to zero, by increasing the physical rate.

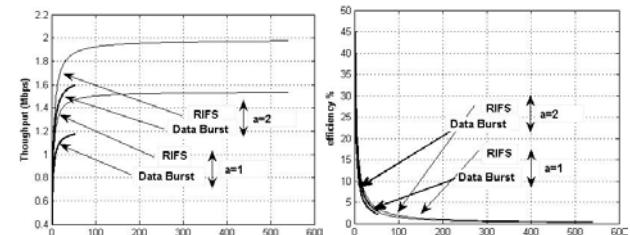


Fig.19 Throughput and efficiency for voice packets

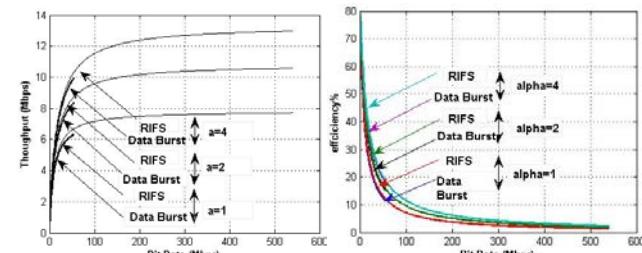


Fig.20 Throughput and efficiency for video packets

5. Theoretical Capacity analysis of VoIP transmission through Aggregation mechanism

The standard High Throughput IEEE 802.11n adapts two approaches for the aggregation data. The first one is the Aggregated MAC Service Data Unit (A-MSDU), and the second is Aggregated MAC Protocol Data Unit (A-MPDU). The transmission of aggregated data is shown by fig.21.

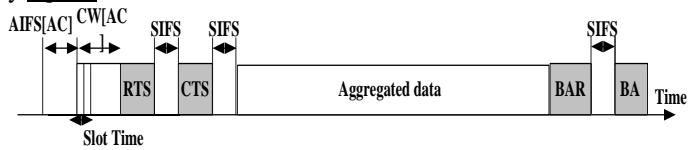


Fig.21 Transmission of aggregated data

5.1 A-MSDU

With A-MSDU, MAC service data units (MSDUs) received from the LLC and destined for the same receiver and of the same service category (same traffic identifier or TID) may be accumulated and encapsulated in a single MAC protocol data unit (MPDU).

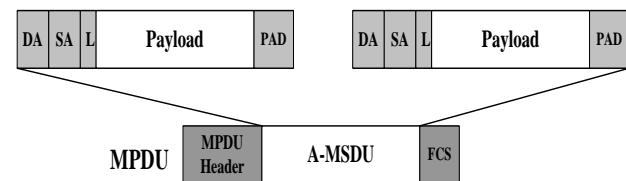


Fig.22 A-MSDU data

As shown by [fig.22](#), the MSDU as received from the LLC is prefixed with a 14 byte subframe header consisting of the destination address (DA), source address (SA), and a length field giving the length of the SDU in bytes. The header together with the SDU is padded with 0 to 3 bytes to round the subframe to a 32-bit word boundary. Multiple such subframes may be concatenated together to form the payload of the QoS Data frame, provided the total length of the data frame does not exceed the maximum MPDU size. The maximum length A-MSDU that a station can receive is either 3839 bytes or 7935 bytes. The total Duration of transmission is:

$$T_{total} = T_{AIFS} + T_{CW} + T_{RTS} + T_{CTS} + 3T_{SIFS} + T_{MPDU-hdr} + \theta T_{A-MSDU} + T_{BAR} + T_{BA}$$

Where θ is the number of packet MSDU per station, and it is dependent with the length of one A-MSDU:

- For $L_{A-MSDU} (\max_1) = 7963 \Rightarrow \theta = 31$
- For $L_{A-MSDU} (\max_2) = 3839 \Rightarrow \theta = 15$

The time for sending one A-MSDU is:

$$T_{A-MSDU} = \frac{L_{DA} + L_{SA} + L_{length} + L_{MSDU} + L_{FCS} + L_{PAD}}{Bit_rate}$$

And the length of one MSDU frame is defined by:

$$L_{MSDU} = L_{MAC} + L_{IP} + L_{UDP} + L_{RTP} + L_{voice}$$

Let A is the total number of packets for all the stations:

$$A = N * \theta \quad \text{Thus } A \text{ will be: } A = \frac{Packet_Ineterval}{2T_{total}} \theta$$

The total number of packet represented by [fig.23](#) is more important for voice packets. When the number of packets per MPDU increases, the throughput increases in the same way. Indeed, for voice packets, at 500Mbps physical rate, the total number of packet achieved for $L_{A-MSDU} (\max_1)$ is more than 400, where it is equal to 220 for $L_{A-MSDU} (\max_2)$. The throughput T is given by :

$$T = \frac{Payload * \theta * (1 - PER)}{[T_{PLCP} + T_{AIFS} + T_{Backoff} + 3 * T_{SIFS} + T_{RTS} + T_{CTS} + T_{BAR} + T_{BA} + T_{ACK} + \alpha * \left(\frac{L_{A-MSDU}}{Bit_rate} \right)]}$$

The throughput is shown by [fig.24](#), and it is more important for video packets than voice packets. It increases when we increase the size of MPDU frame. Similarly, the transmission of video packets is more efficient than voice packets. As shown by [fig.25](#), the efficiency in this mode of

transmission looks more important, indeed, the efficiency for packets voice for example doesn't fall many as the previous cases. It is equal to 80% at 1Mbps, and it will be 38% for the largest video MPDU size.

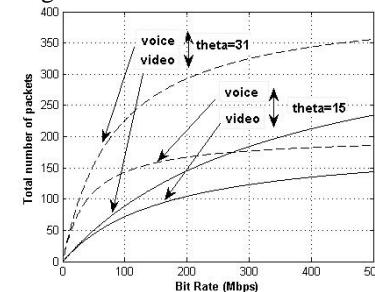


Fig.23 Total number of transmitted packets

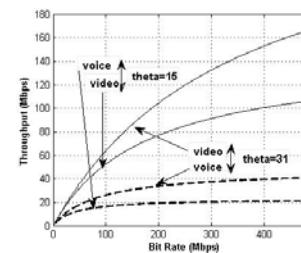


Fig.24 Throughput of A-MSDU

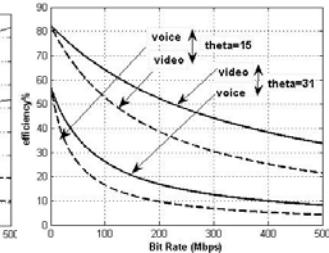


Fig.25 Efficiency of A-MSDU

5.2 A-MPDU

With A-MPDU, fully formed MAC PDUs are logically aggregated at the bottom of the MAC. A short MPDU delimiter is pretended to each MPDU and the aggregate presented to the PHY as the PSDU for transmission in a single PPDU. [Fig.26](#) shows the format of an A-MPDU.

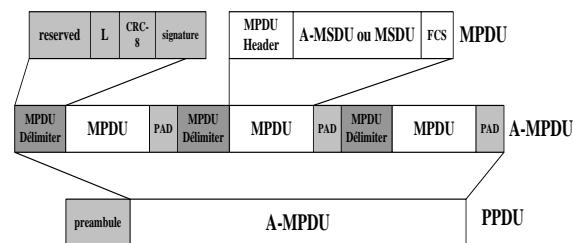


Fig.26 A-MPDU data

The MPDU delimiter is 32 bits in length and consists of a 4-bit reserved field, a 12-bit MPDU length field, an 8-bit CRC field, and an 8-bit signature field. The 8-bit CRC covers the 4-bit reserved and 12-bit length fields and validates the integrity of the header. The MPDU is padded with 0–3 bytes to round it up to a 32-bit word boundary.

A station advertises the maximum A-MPDU length that it can receive in its HT Capabilities element. The advertised maximum length may be one of the following: 8191, 16383, 32767, or 65 535 bytes. The sending station must

not send an A-MPDU of greater length. The total Duration of transmission is:

$$T_{total} = T_{AIFS} + T_{CW} + T_{RTS} + T_{CTS} + 3T_{SIFS} + T_{preamble} + \xi T_{A-MPDU} + T_{BAR} + T_{BA}$$

Where ξ is the number of packet A- MPDU per station and it is dependent with the length of one A-MPDU:

$$T_{A-MPDU} = \frac{L_{MPDU-Delimiter} + \theta L_{A-MSDU} + L_{PAD}}{Bit_rate}$$

Where:

$$L_{A-MSDU} = L_{DA} + L_{SA} + L_{length} + L_{MSDU} + L_{PAD}$$

$$L_{MSDU} = L_{MAC} + L_{IP} + L_{UDP} + L_{RTP} + L_{voice}$$

Thus, the number of packet per MPDU is given by the following ε combination:

$$L_{A-MPDU} (\max_1) = 32767 \Rightarrow (\theta, \xi) = (15,8) or (31,4)$$

$$L_{A-MPDU} (\max_2) = 16383 \Rightarrow (\theta, \xi) = (15,4) or (31,2)$$

$$L_{A-MPDU} (\max_3) = 8191 \Rightarrow (\theta, \xi) = (15,2) or (31,1)$$

Let B is the total number of packets for all the stations:

$$B = N * \theta * \xi$$

The total number of packets for voice and video data is shown by fig.27 and fig.28 respectively. The total number of packets for voice is more than video packets. It increases when we use the larger frame PPDU. Indeed, for the largest size of PPDU frame, the total number of voice packets is equal to 1180 at 500Mbps while it is equal to 450 for video packets. The throughput is :

$$T = \frac{payload * \theta * \varepsilon * 0.9}{[T_{PLCP} + T_{AIFS} + T_{Backoff} + 3 * T_{SIFS} + T_{RTS} + T_{CTS} + T_{BAR} + T_{BA} + T_{ACK} + \varepsilon * \left(\frac{L_{A-MPDU}}{Bit_rate} \right)]}$$

Where:

$$L_{A-MPDU} = L_{MPDU-Delimiter} + \theta L_{A-MSDU} + L_{PAD}$$

As shown in fig.29, the throughput is more important for video packets than voice packets. It increases immensely when the physical rate increases.

The efficiency is represented by fig.30. As similarly, it is more important for video packets. It decrease when the physical rate increase, but it still efficient for Very Hight Throughput. For example, for voice packets, it is between 55% at 1Mbps and 35% at 500Mbps for the largest size of PPDU.

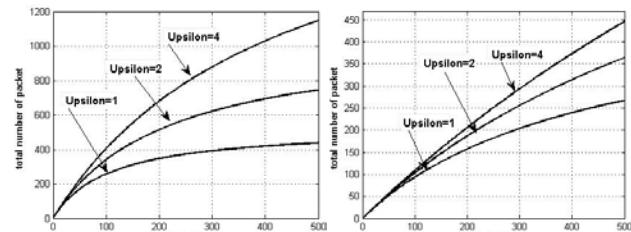


Fig.27 Number of calls (voice) Fig.28 Number of calls (video)

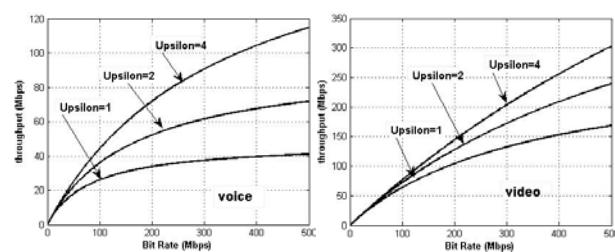


Fig.29 Throughput of A-MPDU data

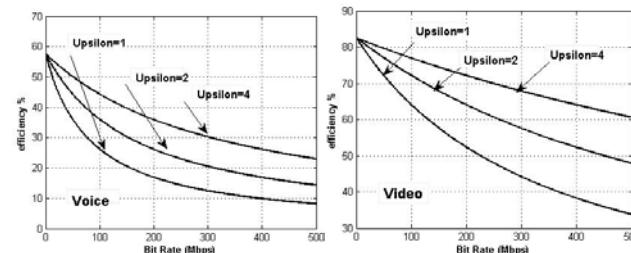


Fig.30 Efficiency of A-MPDU data

6. Conclusions

For legacy IEEE802.11, the problem of the access methods resides on the immense decrease of the efficiency for Hight throughput. The second amelioration was for supporting QoS, by introducing EDCA, Immediate Block Acknowledgment, but these mechanisms have the same limitation of efficiency. The most recently mechanism was introduced to Very Hight Throughput. The use of aggregation seems the most efficient. Indeed, the efficiency is maintained at a good level at VHT.

References

- [1] IEEE, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications, IEEE Std 802.11-1999, 1999.
- [2] IEEE, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications Amendment 8: Medium Access Control MAC Quality of Service Enhancements, November 2005.
- [3] IEEE P802.11n™/D11.0 raft STANDARD for Information Technology, IEEE Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications Amendment 5: Enhancements for Higher Throughput, 2009 .

- [4] E.Perahia, R.Stacey, Next Generation Wireless LANs Throughput, robustness, and reliability in 802.11n, Cambridge University Press 2008,
- [5] Y. GE, H. Jennifer, C. Sunghyun, "An analytic study of tuning systems parameters in IEEE 802.11e enhanced distributed channel access", Computer networks journal, vol. 51, no8, 2007.
- [6] I. Inan, F. Keceli, E. Ayanoglu, "Analysis of the 802.11e Enhanced distributed channel access Function", technical report
- [7] Tinnirello, I. S. Choi, "Efficiency analysis of burst transmissions with block ACK in contention-based 802.11e WLANs", in ICC, 2005 5 ,pp 3455 - 3460
- [8] R. Roy , Y. Chen, S. Emeott, "Performance Analysis of data Aggregation Techniques for wireless LAN", IEEE Globecom, 2006.
- [9] D. Skordoulis, Q. Ni, U. Ali, H. Marios "Analysis of Concatenation and Packing Mechanisms in IEEE 802.11n", PGNET 2007, Liverpool, UK, June 2007.
- [10] Y. Lin, W.S. Wong "Frame Aggregation and Optimal Frame Size Adaptation for IEEE 802.11n WLANs", Proc. IEEE GLOBECOM-2006, pp. 1-6
- [11] D Skordoulis, N. Qiang, H. Chen, A.P. DRIAN, P. TEPHENS, L. CHANGWEN, ABBES JAMALIPOUR "IEEE 802.11 n MAC frame aggregation mechanisms for next-generation", IEEE Wireless Communications, February 2008
- [12] S. Kim, O. Lee, S. Choi, Sung-Ju Lee, "MAC-Aware Routing Metric for 802.11 Wireless Mesh Networks", IEEE PIMRC 2009, September 2009.
- [13] D. Skordoulis, N. Qiang, G. Min, K. Borg, " Adaptive Delayed Channel Access for IEEE 802.11n WLANs", IEEE international conference on circuits and systems for communications, 26-28 may 2008.
- [14] Y. ZHENG, K. LU, W. Dapeng, F. Yuguang, "Performance Analysis Of Frame-Burst-based Medium Access Control Protocols Under Imperfect Wireless Channels", International journal of intelligent control and systems, (2005) VOL. 10, NO. 1, 2005, 43-51
- [15] L. Tianji, N. Qiang, Y. Xiao, "Investigation of the block ACK scheme in wireless ad hoc networks": Research Articles Wireless Communications & Mobile Computing Volume 6 , Issue 6 (2006) pp 877 - 888
- [16] W. Wang, C. Soungh Liew, V.O.K. Li, "Solutions to Performance Problems in VoIP over 802.11 Wireless LAN", Vehicular Technology, IEEE Transactions on, v (54), issue 1, pp 366-384
- [17] P. David Hole, A. Fouad Tobagi "Capacity of an IEEE 802.11b Wireless LAN supporting VoIP", Conference on Communications (ICC) 2004
- [18] A. Mohd, O.L. loon, "performance of voice over ip (voip) over a wireless lan (wlan) for different audio/voice codecs", *Journal Teknologi*, 47(D) Dis. 2007: 39–60

First Author Emna CHARFI was born in Monastir, Tunisia, in 1985. She received her degree in Electrical Engineering and her Masters in Electronic Engineering from Sfax National School of Engineering (ENIS), Tunisia, in 2009 and 2010, respectively. In 2010, she joined the Sfax High Institute of Electronics and Communication, Tunisia, as an assistant. She is a member of Sfax Laboratory of Electronics and Information Technology. Her current research interests are communications, networking and are specially related to wireless local area network.

Second Author Dr Lamia CHAARI was born in Sfax, Tunisia, in 1972. She received the engineering and PhD degrees in electrical and electronic engineering from Sfax national engineering school (ENIS) in TUNISIA. Actually she is an associate professor in multimedia and informatics higher institute in SFAX She is also a researcher in electronic and technology information laboratory (LETI). Her scope of research are communications, networking and signal processing which are specially related to wireless and new generation networks.

Third Author Pr Lotfi Kamoun was born in Sfax Tunisia, 25 January. 1957. He received the electrical engineering degree from the Sciences and Techniques Faculty in Tunisia. Actually he is a Professor in Sfax national engineering school (ENIS) in TUNISIA. He is the director of electronic and technology information laboratory (LETI). His scope of research are communications, networking, Software radio and signal processing which are specially related to wireless and new generation networks.

Comparing Methods for segmentation of Microcalcification Clusters in Digitized Mammograms

Hajar Moradmand¹, Saeed Setayeshi ² and Hossein Khazaei Targhi³

¹ Islamic Azad University-Khorasgan Branch, Khorasgan, Iran

² Department of Biomedical Radiation Engineering, Amirkabir University of Technology
Tehran, Iran

³ Islamic Azad University-Khorasgan Branch, Khorasgan, Iran

Abstract

The appearance of microcalcifications in mammograms is one of the early signs of breast cancer. So, early detection of microcalcification clusters (MCCs) in mammograms can be helpful for cancer diagnosis and better treatment of breast cancer. In this paper a computer method has been proposed to support radiologists in detection MCCs in digital mammography. First, in order to facilitate and improve the detection step, mammogram images have been enhanced with wavelet transformation and morphology operation. Then for segmentation of suspicious MCCs, two methods have been investigated. The considered methods are: adaptive threshold and watershed segmentation. Finally, the detected MCCs areas in different algorithms will be compared to find out which segmentation method is more appropriate for extracting MCCs in mammograms.

Keywords: Mammograms, Microcalcification Clusters, Segmentation, Compare.

1. Introduction

Breast cancer is one of the most important factors of mortality in women over the world. In the United States alone, a most recent survey estimated 207,090 new cases of breast cancer and 39,840 deaths in women during 2010 [1]. Causes of this disease still remain unknown so; there is no sure way to prevent breast cancer. Early detection is the key to improve breast cancer treatment. Up to now, mammography remains the most effective diagnostic technique for early breast cancer detection. Radiologists are capable for detection the abnormalities of cancerous cells such as calcification, masses, architectural distortion, and asymmetry between breasts, breast edema and lymphadenopathy from mammogram images. Two major of mammographic abnormalities in breast cancer is calcifications and masses. Calcifications are small mineral deposits within the breast tissue, which look like small white spots on the films. They are often important and common findings on a mammogram. They can be

produced from necrotic cellular debris or from cell secretion. They may be intramammary, within and around the ducts, within the lobules, in vascular structures, in interlobular connective tissue or fat. Alternatively, they may be found in the skin. Calcifications can appear with or without an associated lesion, and their morphologies and distribution provide clues as to their etiology as well as whether they can be associated with a benign or malignant process. There are two types of calcifications: macrocalcifications and microcalcifications. Accurate segmentation of individual microcalcifications is challenged by microcalcifications size and shape variability, superimposed surrounding tissues and high frequency noise [2]. Elder and Horsch a thorough review of the proposed methods for microcalcifications segmentation is provided [3]. Segmentation of individual microcalcifications has been achieved by grey-level based methods with empirically defined parameters such as region growing [4] and grey-level thresholding on pre-processed regions of interest (ROIs) [5]. To fulfill requirements for real-time behavior and parameter-free segmentation, more sophisticated techniques have been proposed such as morphologic operations [6], watershed algorithms [7], Bayesian pixel classification combined with Markov Random Field models [8] and radial gradient-based methods [9]. Furthermore, the wavelet transform has been used for the segmentation of microcalcifications, due to its ability to spatially localize high frequency components [10]. Recently, a segmentation method based on multiscale active rays was proposed to deal with microcalcification size variability [11]. Although the spectrum of computational methods has been proposed for detection of MCCs is wide, automated interpretation of microcalcifications still remains very difficult. This paper has been proposed two methods for MCCs segmentation. Although, these methods had been used but the way that they have been implemented in this paper is new and

achieved very good segmentation result. The rest of paper is organized as follows. The proposed approach for segmentation of MCs with two methods: Morphology operation and watershed segmentation has been described in Section 2. The subsequent section discusses the obtained results (section3), and Section 4 presents the conclusions and future research directions.

2. The proposed method

This paper presented two methods for MCCs segmentation in digital mammograms. A block diagram of the proposed method is shown in Fig. 1.

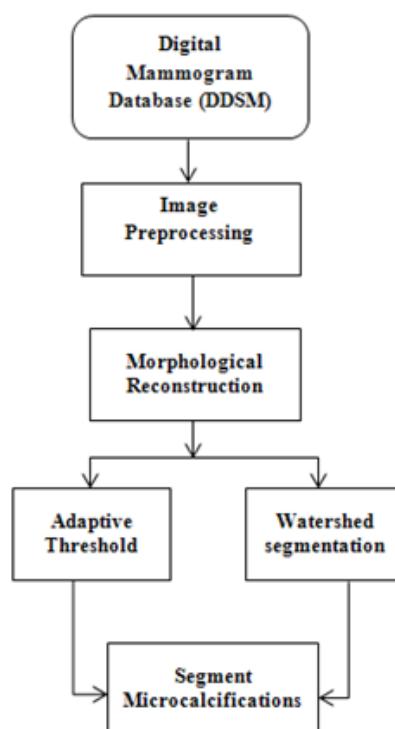


Fig. 1 Block diagram of the proposed method.

2.1 Databank

In this research Digital Database for Screening Mammography (DDSM) from University of South Florida is used for experiments. It was downloaded from <http://marathon.csee.usf.edu/Mammography/DDSM.html>.

Images containing suspicious areas have associated pixel-level ground truth information about the locations and types of suspicious regions. Also provided is software both for accessing the mammogram and truth images and for calculating performance figures for automated image analysis algorithms [12]. The use of such a database aids in comparison of CAD algorithms through evaluation using a common database. Our experimental data contains

126 images, from only craniocaudal (CC) view. The sampling rates of mammogram images are digits between 42 to 50 microns.

2.2 Preprocessing

A preprocessing step is performed in order to facilitate the subsequent MC segmentation task. In our algorithm this step includes three phases: masking, denoising, and image enhancement.

2.2.1 Mask Generation

First, breast tissue is separated from the other parts of image which are completely dark. This allows processing of only the tissue region in further steps and eliminating the artifacts out of the mammogram. To accomplish this, a segmentation mask is used to separate the tissue region from the film region. The mask template is a binary matrix of size equal to that of the original image. Morphology operation and Otsu algorithm had been used to create appropriate mask.

2.2.2 Denoising

Digital mammography images often contain significant amounts of noise which, accommodation with salt and pepper type noise. This noise should be removed for avoiding deterioration of the contrast enhancement algorithm and negative influence on the whole detection procedure. Therefore, de-noise the image with a 4×4 median filter. Median filter is a non-linear filter having the ability to remove salt and pepper noise without blur edges especially is a powerful tool for improving mammogram images.

2.2.3 Image Enhancement

The contrast of a mammogram image is often poor, especially for dense, glandular tissues. In these cases distinguish MCs is quite hard. To overcome this obstacle contrast enhancement algorithm had been used. The aim of contrast enhancement is to increase the contrast of MC over the threshold. Our system performs automatic contrast enhancement, using a method based on unsharp masking and 2D discrete wavelet transform. To egregious edges and small details in the mammogram image Unsharp masking filter is very useful. The unsharp masking can be expressed as:

$$D_p(x, y) = D_o(x, y) + K(x, y) \times \left[D_o(x, y) - \frac{1}{mn} \sum_{j=1}^n \sum_{i=1}^m D_o(x_i, y_j) \right] \quad (1)$$

Where $D_o(x, y)$ and $D_p(x, y)$ are the densities of the original and the processed mammograms, respectively. The last term is the unsharp term with an $m \times n$ area centered at pixel (x, y) , $k(x, y)$ is a weight factor [2]. The mask size and the weight factor determine the frequency range and the degree of enhancement. The unsharp masking method reduces the low-frequency information while amplified the high-frequency detail. However, these processes could change the images dramatically to be applied to the mammograms.

Wavelet transform is a powerful tool for filtering that represents images hierarchically on the basis of scale and resolution, analyzing high-spatial frequency phenomena localized in space, and, thus can effectively extract information derived from localized high-frequency signals, such as those emitted by microcalcification. The two dimensional discrete wavelet transform (DWT) decomposes the approximation coefficients at level j into approximation coefficients at level $j+1$ and three sets of detailed coefficients, i.e., horizontal, vertical and diagonal Fig. 2 describes the basic decomposition step for image.

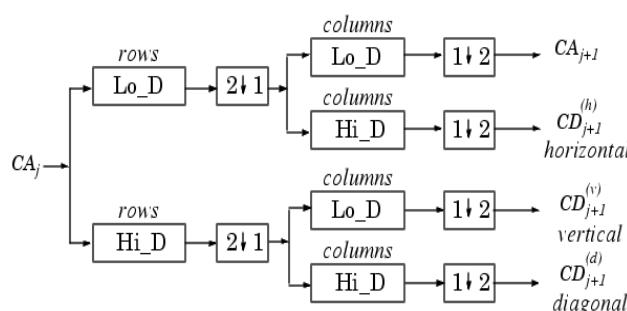


Fig. 2 Wavelet decomposition of an image. Two-dimensional wavelet transform leads to a decomposition of approximation coefficients at level j in four components: the approximation at level $j+1$, and the details in three orientations (horizontal, vertical, and diagonal).

The detailed coefficients contain small-scale components of the image. In frequency domain analysis, high frequency coefficients are detailed. Microcalcification often form on the mammogram image of fine grains appear bright in the breast tissue. So, we can assume that a wavelet decomposition of the mammogram image will contain the MC mostly within the detailed coefficients [10]. Five-level discrete wavelet decomposition had been employed by using Asymmetric Daubechies of order 8 since it accumulates more energy corresponding to the details of the wavelet transform and moreover it is characterized by symmetry and finite length to enhance mammograms. Due to these features, they can achieve high correlation with the clustered MC, and, therefore, they can effectively enhance MC. So, the filtered image is subjected to five-level discrete wavelet decomposition. This produces an approximation and five sets of

horizontal, vertical and diagonal detailed coefficients. Afterwards, the contrast enhanced mammogram is obtained by reverse wavelet transform.

2.3 Segment Microcalcifications

For reasons such as; characteristics of breast tissue varies in texture and small size Microcalcification their detection and isolation of tissue still remains difficult. Select appropriate threshold is a sensible step, as a too high threshold can neglect the MCs which present less contrast, while a too low threshold makes that brilliant points which are selected do not correspond to some microcalcification; these points are caused by the oscillations in the grey levels of the background, due to the noise which contaminates the image. To overcome this problem first morphological reconstruction (opening and then closing by-reconstruction) had been used to clean up the image and smoothing so the foreground that contains breast tissue would be more recognizable. Because, comparing reconstruction-based opening and closing to standard opening and closing are more effective at removing small blemishes without affecting the overall shapes of the objects. Reconstruction is a morphological transformation involving two images and a structuring element (instead of a single image and structure element). One image, the marker, is the starting point for the transformation. The other image, the mask contains the transformation. The structure element used defined connectivity. If g is the mask and f is the marker, the reconstruction of g from f , denoted; $R_g(f)$. The high points, or peaks, in the marker image specify where processing begins. The processing continues until the image values stop changing.

The following steps for morphologically reconstruct of mammogram images has been performed. First a marker image has been created. The characteristics of the marker image determine the processing performed in morphological reconstruction. The peaks in the marker image should identify the location of objects in the mask image. To create a marker image, first compute the background by imopen, and then subtract it from the mask image by imsubtract as indicated in Eq. (2) and Eq. (3).

$$\text{Background} = \text{imopen}(A, \text{SE}) \quad (2)$$

$$\text{Marker} = \text{imsubtract}(A, \text{background}) \quad (3)$$

Next the opening-by-reconstruction computed by imreconstruct as shown in Eq. (4). In the output image, all the intensity fluctuations except the intensity peak have been removed.

$$\text{Recon} = \text{imreconstruct}(\text{marker}, \text{mask}) \quad (4)$$

Following the opening with a closing can remove the dark spots and stem marks, so try imclose followed by imreconstruct.

2.3.1 Adaptive Thresholding

Whereas, there are large variations threshold from one image to the next, so a constant threshold will not be good enough. To find a threshold for a particular image adaptive thresholding has been implemented. Local adaptive thresholding selects an individual threshold for each pixel based on the range of intensity values in its local neighborhood. This allows for thresholding of an image whose global intensity histogram doesn't contain distinctive peaks. The result is a binary image representing the suspicious areas that contains MCs. To achieve adaptive thresholding, first determined the image's histogram. As be illustrated in Fig. 3 the gray level of all histogram contain two Gaussian distribution function.

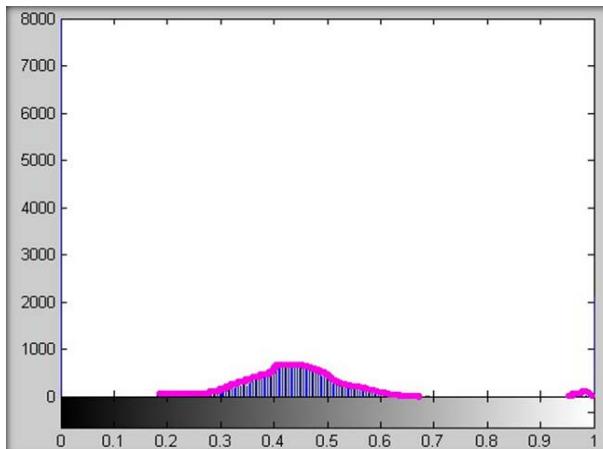


Fig. 3 An example of a mammogram's histogram.

Therefore, two gaussian functions have been adapted over histograms then mean and variance of them has been computed. In order to select a more accurate threshold Eq. (5) has been considered

$$Threshold = \frac{\frac{c_1 + c_2}{\sigma_1 + \sigma_2}}{\frac{1}{\sigma_1} + \frac{1}{\sigma_2}} \quad (5)$$

Where c is the mean and σ is the variance.

2.3.2 Watershed Segmentation

The watershed transform finds catchment basins and watershed ridge lines in an image by treating it as a surface where light pixels are high and dark pixels are low. The watershed transformation considers the gradient magnitude of an image as a topographic surface. Pixels having the

highest gradient magnitude intensities (GMIs) correspond to watershed lines, which represent the region boundaries. Water placed on any pixel enclosed by a common watershed line flows downhill to a common local intensity minimum (LIM). Pixels draining to a common minimum form a catch basin, which represents a segment. Sobel filter is used often for approximates the maximum gradient of the image by giving more weight to the pixels nearest to (i, j) . It is equal to Eq. (6).

$$\sqrt{(\frac{\partial f_{ij}}{\partial x})^2 + (\frac{\partial f_{ij}}{\partial y})^2} \quad (6)$$

Direct application of the watershed segmentation to a gradient image can be lead to over-segmentation due to serious noise or image abnormality. Segmentation using the watershed transforms works better if, foreground objects and background would be marked whose goal is to detect the presence of homogeneous regions from the image. Internal markers are inside each of the objects of interest and external markers are contained within the back-ground. For segmentation with Marker-controlled watershed follow this basic procedure:

1. Compute foreground markers. These are connected blobs of pixels within each of the objects.
2. Compute background markers. These are pixels that are not part of any object.
3. Modify the segmentation function so that it only has minima at the foreground and background marker locations.
4. Compute the watershed transform of the modified segmentation function.

3. Results

The proposed algorithm operates on 126 mammograms image that contain MCCs from DDSM. The experimental result demonstrated that watershed segmentation is more accurate than adaptive thresholding but is more time consuming. Fig. 4 shows an example of our method on a mammogram image with adaptive thresholding method and in Fig. 5 with watershed segmentation.

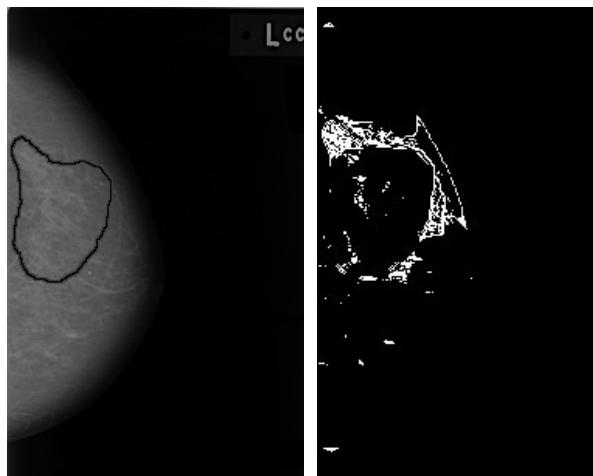
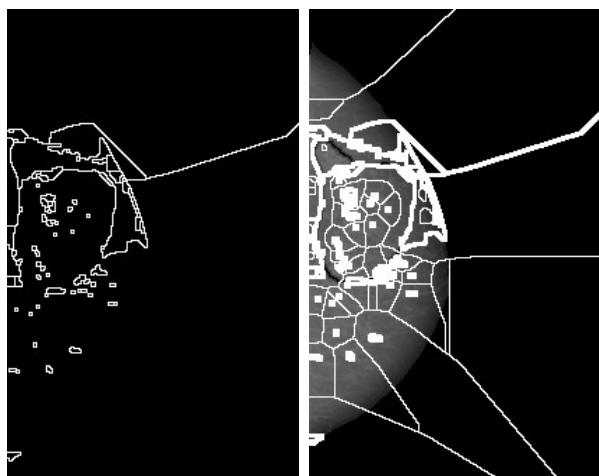


Fig. 4 Microcalcification segmentation with adaptive threshold.



a) Watershed ridge lines

b) Markers and object boundaries superimposed on original image

Fig. 5 Microcalcification segmentation with watershed segmentation.

4. Conclusions

In this paper a computer system devised that support a radiologist in small field of digital mammography has been proposed. As noted earlier, MCs has very small and ubiquitous nature. So, we focused on the detection of clusters of microcalcifications. For this reason two method of segmentation had been implement on 126 mammograms. Result shows that watershed segmentation is more accurate than adaptive thresholding but more time consuming. All code required to accomplish these tasks was written in MATLAB. The system was evaluated by two radiologists. Their results show that the system introduces an improvement in the breast cancer detection. In our future research, we would like to evaluate the

proposed method on more mammograms from clinical images and other database. We would also like to extend this research for detect and classification of other factors of breast cancer such as mass.

Acknowledgments

The authors would like to thank Dr. Mohammad Esmail Akbari master of Iranian Cancer Research Center for his sincere contributions.

References

- [1] The American College of Radiology (ACR), <http://www.acr.org/>.
- [2] HD. Cheng, X. Cai, X. Chen, L. Hu and X. Lou, "Computer-aided detection and classification of microcalcifications in mammograms: a survey", *Pattern Recognition*, Vol. 36, No. 12, 2003, pp. 2967–91.
- [3] M. Elter and A. Horsch, "CADx of mammographic masses and clustered microcalcifications: a review", *Medical Physics*, Vol. 36, No. 6, 2009, pp. 2052–68.
- [4] HP. Chan, B. Sahiner, KL. Lam, N. Petrick, MA. Helvie, MM. Goodsitt, et. al., "Computerized analysis of a mammographic microcalcifications in morphological and texture feature spaces", *Medical Physics*, Vol. 25, No. 10, 1998, pp. 2007–19.
- [5] I. Leichter, R. Lederman, S. Buchbinder, P. Bamberger, B. Novak and S. Fields, "Optimizing parameters for computer-aided diagnosis of microcalcifications at mammography", *Acad. Radiol.*, Vol. 7, No. 6, 2000, pp. 406–12.
- [6] O. Tsujii, MT. Freedman and SK. Mun, "Classification of microcalcifications in digital mammograms using trend-oriented radial basis function neural network", *Pattern Recognition*, Vol. 32, No. 5, 1999, pp. 891–903.
- [7] D. Betal, N. Roberts and GH. Whitehouse, "Segmentation and numerical analysis of microcalcifications on mammograms using mathematical morphology", *Br J Radiol.*, Vol. 70, No. 837, 1997, pp. 903–17.
- [8] S. Paquerault, LM. Yarussi, J. Papaioannou, Y. Jiang and RM. Nishikawa, "Radial gradient-based segmentation of mammographic microcalcifications: observer evaluation and effect on CAD performance", *Medical Physics*, Vol. 31, No. 9, 2004, pp. 2648–57.
- [9] RN. Strickland and HI. Hahn, "Wavelet transforms for detecting microcalcifications in mammograms", *IEEE Trans. on Medical Imaging*, Vol. 15, No. 2, 1996, pp. 218–28.
- [10] P. Heinlein, J. Drexl and W. Schneider, "Integrated wavelets for enhancement of microcalcifications in digital mammography", *IEEE Trans. on Medical Imaging*, Vol. 22, No. 3, 2003, pp. 402–13.
- [11] G. Lemaur, K. Drouiche, J. DeConinck, "Highly regular wavelets for the detection of clustered microcalcifications in mammograms", *IEEE Trans. on Medical Imaging*, Vol. 22, No. 3, 2003, pp. 393–401.
- [12] M. Heath, K. Bowyer, D. Kopans, R. Moore and P. Kegelmeyer, "The digital database for screening mammography", *Medical Physics Publishing*, IWDM-2000.

Grayscale Image Compression Based on Min Max Block Truncating Coding

Mr.Hilal Almara'beh¹, Mr.Khalil Barhoum ², Mr.Majdi Ahed³

¹ Computer Science Department, Isra University
 Amman,11622, Jordan

² Computer Science Department, Isra University
 Amman,11622, Jordan

³ Software Engineering Department, Isra University
 Amman, 11622, Jordan

Abstract

This paper presents an image compression techniques based on block truncating coding. In this work, a min max block truncating coding (MM_BTC) is presented for grayscale image compression relies on applying dividing image into non-overlapping blocks. MM_BTC differ from other block truncating coding such as block truncating coding (BTC) in the way of selecting the quantization level in order to remove redundancy. Objectives measures such as: Bit Rate (BR), Mean Square Error (MSE), Peak Signal to Noise Ratio (PSNR), and Redundancy (R), were used to present a detailed evaluation of MM_BTC of image quality.

Keywords: BTC, MM_BTC, Bit Rate, MSE, PSNR.

1. INTRODUCTION

Image compression is an important tool to reduce the bandwidth and storage requirements of practical image systems such as multimedia, communication, medical images, education and business etc. The image compression techniques are categorized into two main classification namely lossy compression techniques and lossless compression techniques [1]. Lossy compression techniques lead to loss of data with higher compression [2] whereas the lossless compression ratio gives good quality of compressed images, but yield only less compression. BTC [3] is a lossy image compression technique, which reduce bit rate and preserving the low computational complexity [4].

The original algorithm of BTC preserves the mean and the standard deviation [5]. The truncated block of the BTC is the one-bit output of the quantizer for every pixel in the block.

In this paper we proposed a min max block truncating coding using the block diagonal for grayscale image compression which provides a better image quality.

2. BTC ALGORITHM

Block truncating coding (BTC) was introduced by Delp and Mitchell in 1979 [6]. In the BTC algorithm the input image is divided into non-overlapping block of size $m \times n$ pixels. Figure.1 shows the diagram of BTC for grayscale image compression.

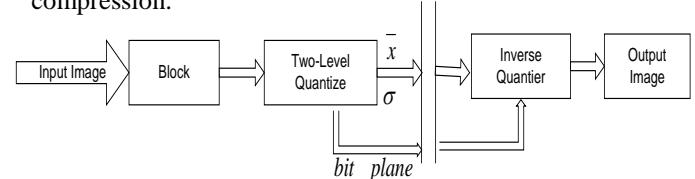


Figure 1.The original BTC System.

The BTC algorithm involves the following steps:

- The input image is divided into non-overlapping block of size $m \times n$ pixels.
- For a two level (1 bit) quantizer select mean (\bar{x}) and standard deviation (σ) values to represent each pixel in the block.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (2)$$

Where x_i represents the i^{th} pixel value of the image block and n is the total number of pixels in that block.

- The two values \bar{x} and σ are termed as quantizers of BTC. Taking \bar{x} as the threshold value a two level bit plane is obtained by comparing each pixel value x_i with the threshold.

$$B = \begin{cases} 1 & x_i \geq \bar{x} \\ 0 & x_i < \bar{x} \end{cases} \quad (3)$$

Where B is a binary block used to represent the pixels.

- d. Inverse quantizer of an image is reconstructed by replacing '1's in the bit plane with U and the '0's with L, which are given by:

$$U = \bar{x} + \sigma \sqrt{\frac{k}{y}} \quad (4)$$

$$L = \bar{x} - \sigma \sqrt{\frac{k}{y}} \quad (5)$$

Where k and y are the number of 0's and 1's in the compressed bit plane respectively.

3. MM_BTC ALGORITHM

The proposed method improves not only bit rate but also the performance of BTC. The MM_BTC simple as BTC but uses a maximum and minimum diagonal of a block. The MM_BTC algorithm involves the following steps:

1. The input image is divided into non-overlapping block of size $m \times n$ pixels.
2. For a two level (1 bit) quantizer select the maximum and the minimum diagonal values to represent each pixel in the block.

$$R = \frac{\max(x_{i=j}) + \min(x_{i=j})}{\sqrt{n}} \quad (6)$$

Where x represents the i^{th} pixel value of the image block and n is the number of divided blocks (2, 4, 8, etc). And R is termed as quantizer of MM_BTC.

3. Binary block denoted by B is also used to represents the pixels. We can use "1" to represent a pixel whose

gray level is greater than or equal to R and "0" to represents a pixel whose gray level is less than R.

$$B = \begin{cases} 1 & x_i \geq R \\ 0 & x_i < R \end{cases} \quad (7)$$

4. Inverse quantizer of an image is reconstructed by replacing '1's in the bit plane with U and the '0's with L, which are given by:

$$U = R \sqrt{\frac{k}{y}} \quad (8)$$

$$L = R \sqrt{\frac{k}{y}} \quad (9)$$

Where k and y are the number of 0's and 1's in the compressed bit plane respectively.

MM_BTC has two advantages over BTC; the first is in the case that the quantizer is used to transmit an image from sender to a receiver, it is necessary to compute at the sender the two quantities, the mean and the standard deviation for BTC while it needs to compute one quantity for MM_BTC. Second when we compare the necessary computation for deviation information, it is necessary to compute a sum of m values and each of them will be squared while in case of MM_BTC it is only necessary to compute the maximum and minimum diagonal of these m values.

4. OBJECTIVE QUALITY MEASURMETNS

Image quality measures are important roles in various images processing application. Once image compression system has been designed and implemented, it is important to evaluate the compression performance and perceptual quality. Objective measures [7, 8] of image quality metrics, some statistical indices are calculated to indicate the image quality, such as mean square error (MSE), peak signal to noise ratio (PSNR) and bit rate (BR).

4.1 Mean Square Error (MSE)

The mean square error represents the mean squared error between the compressed and the original image. If the reconstructed image is close to the original image, then MSE is small. This means the lower value of MSE, the lower error. MSE given by:

$$MSE = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N [x(i, j) - y(i, j)]^2 \quad (10)$$

4.2 Peak Signal to Noise Ratio (PSNR)

The PSNR is an interactive measure and it is most commonly used as measure of quality of reconstruction of loss compression. If the reconstructed image is close to the original image, then PSNR is large value. This means the higher the PSNR, the better the quality of the compressed or reconstructed image. PSNR given by:

$$PSNR = 10 \cdot \log_{10} \left(\frac{255 * 255}{MSE} \right) \quad (11)$$

4.3 Bit Rate (BR)

The performance indicator of image compression can be specified in terms of compression efficiency. Compression efficiency is measured by the compression ratio or by the bit rate. The compression ratio CR , which means that the compressed image is stored using only $CR\%$ of the initial storage size; the bit rate is the number of bits per pixel required by the compressed image. Compression ratio given by:

$$CR = \frac{Size(x)}{Size(y)} \quad (12)$$

Where x is the original image and y is the compressed image.

And the bit rate given by:

$$BR = \frac{b}{CR} \quad (13)$$

Where b : the number of bits per pixel (bit depth) of the uncompressed image.

5. EXPERIMENTAL RESULTS

We have taken the results of 6 images (512*512, 8 bit per pixels) Lena, Cameraman, Woman, Peppers, baboon and Lifting-body, obtained by running both the proposed method and BTC. Results are obtained in terms of bit rate, MSE, and PSNR when a block size 8*8. These values are listed in Table.1.

Table.1 Comparison results between MM_BTC and BTC

Image	Method	Bit Rate	MSE	PSNR
Lena	MM_BTC	1.04	15.7975	36.1449
	BTC	1.04	18.1300	35.5468
Cameraman	MM_BTC	1.04	15.6503	36.1856
	BTC	1.04	17.3780	35.7308
Woman	MM_BTC	1.04	8.6558	38.7564
	BTC	1.04	9.1936	38.4959
Peppers	MM_BTC	1.04	14.7380	36.4464
	BTC	1.04	17.0919	35.8029
Baboon	MM_BTC	1.04	24.2043	35.2014
	BTC	1.04	25.6533	34.0394
Lifting-body	MM_BTC	1.04	5.0843	41.0685
	BTC	1.04	6.3547	40.0998

The table assures that the image compression using Figure.3 Compressed images using MM_BTC

MM_BTC provides better image quality than image compression using BTC. Figure.2 shows the original image, figure.3 shows the compressed image using MM_BTC and figure.4 shows the compressed image using BTC. Figure 5 and 6 shows a comparison chart of MSE and PSNR results.

Figure.2 The original image

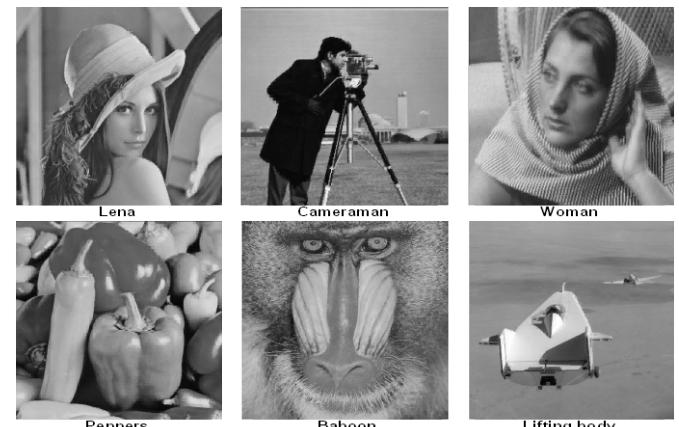
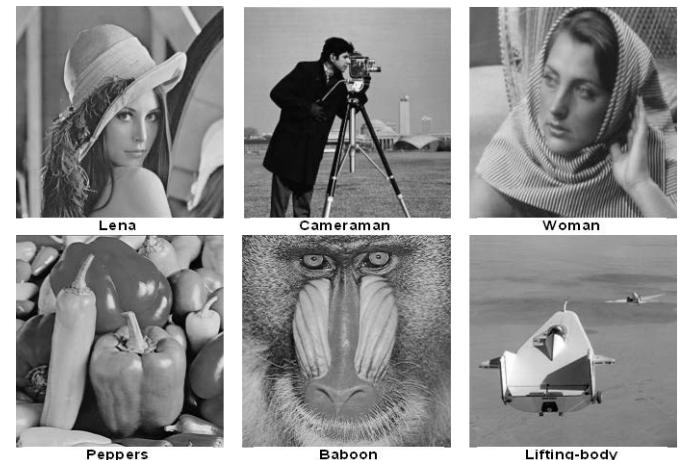


Figure.3 Compressed image using MM_BTC

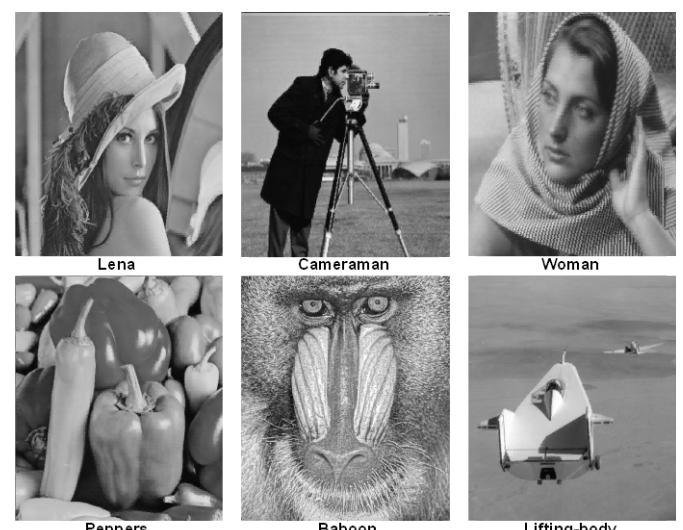


Figure.4 Compressed image using BTC

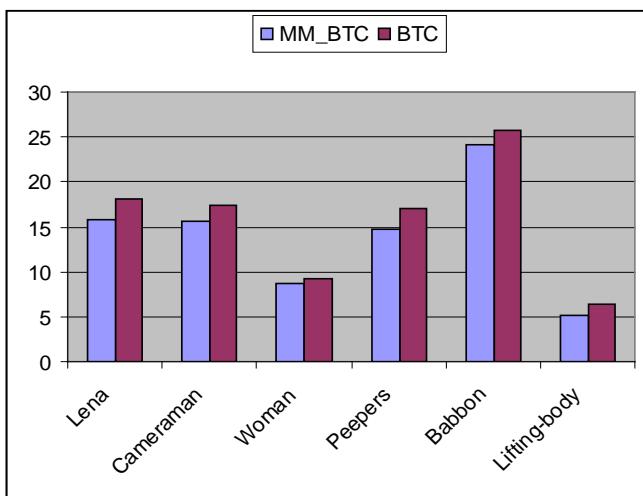


Figure.5 Comparison of the MSE results

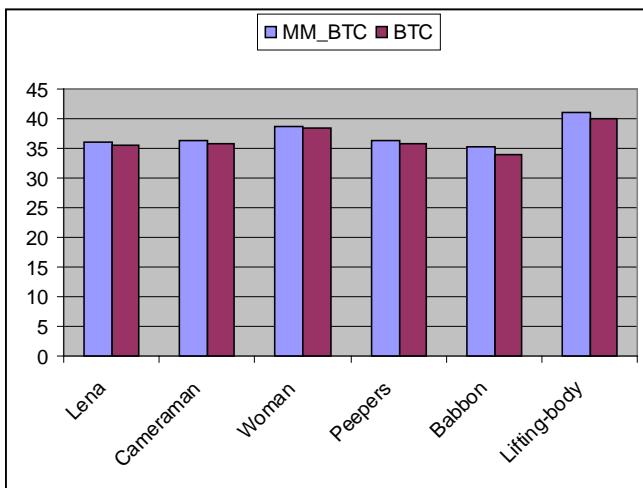


Figure.6 Comparison of the PSNR results

6. CONCLUSION

We have presented a new algorithm for compression of loss grayscale images which achieves good results than BTC. Image measures (bit rate, MSE and PSNR) were used to evaluate the image quality. The result shows that the proposed method provides better image compression than BTC.

REFERENCES

- [1] Rafael C. Gonzalez, Richard Eugene; "Digital image processing", Edition 3, 2008, page 466.
- [2] M. Ghanbari "Standard Codecs: Image Compression to Advanced Video Coding" Institution Electrical Engineers , ISBN: 0852967101,2003 , CHM , 430 pages.
- [3] Delp, E. J., Saenz, M., and Salama, P., 2000, Block Truncation Coding (BTC), Handbook of Image and Video Processing, edited by Bovik A. C., Academic Press, pp. 176-181.
- [4] C.Huang and Y.Lin,"Hybrid block truncating ",IEEE Signal Processing Letter, Vol. 4, No. 12,pp.328-330,1997.
- [5] A. M. Eskicioglu and P.S. Fisher, "Image quality measures and their performance," IEEE Trans. Communications, vol. 34, pp. 2959-2965, Dec. 1995.
- [6] T.M. Amarunnishad, V.K. Govidan, and T.M. Abraham, "Block truncating coding using a set of predefined bit plans", In proc. Int. Conf. Computational Intelligence and Multimedia Applications (ICCIMA 2007), Dec, 2007, Sevikasi, Tamil Nadu, India, pp.73-78, IEEE,2007.
- [7] Somasundaram, K. and I.Kaspar Raj."Low Computational Image Compression Scheme based on Absolute Moment Block Truncation Coding". May 2006. Vol. 13.
- [8] W. J. Chen and S. C. Tai, "Postprocessing Techniques for Absolute Moment Block Truncation Coding" Proc. of ICS'98, Workshop on Image Processing and Character Recognition, pp. 125-130, Dec. 17-19, 1998, Tainan, Taiwan.

Hilal Almara'beh received the B.S degree in Computer Science from Isra University, Amman, Jordan, in 2000. He received the M.S. degree in Computer Science from Amman Arab University for Graduate Studies, Amman, Jordan, in 2005. He is currently a lecturer of Computer Science Department, Faculty of Information Technology, Isra University, Jordan. His research interests are image processing and video segmentation, semantic web, grid and cloud computing, and database systems.

Khalil Barhoum received the B.S degree in Computer Engineering from Technical University of Budapest, Hungary , in 1992. He received the M.S. degree in Computer Science from Newyork Institute of Technology, Newyork, USA, in 2005. He worked as an instructor at Heald College, USA, for five years, and he is currently a lecturer of Computer Science Department, Faculty of Information Technology, Isra University. His research interests are security in the software engineering life cycle, image compression, and semantic web.

Majdi Ahed received the B.S and M.S. degree in Computer Science from Jordan University, Amman, Jordan, in 1999, 2001 respectively. He worked as lecturer at Philadelphia University, Computer Science Department, Jordan, for five years, and he is currently a lecturer of Computer Science Department, Faculty of Information Technology, Isra University. His research interests are quality assurance, image processing, and semantic web.

The Conception and the Study of a Triple-band Planar Inverted-F Antenna

Saida Ibnyaich¹, Abdelilah Ghammaz², Moha M'rabet Hassani¹

¹Department of Physics
Electronics and Instrumentation Laboratory, Faculty of Sciences Semlalia, Cadi Ayyad University,
P.O. Box 2390/40000, Marrakesh, Morocco

²Department of Physics
Laboratory of Electrical Systems and Telecommunications, Faculty of Sciences and Technology, Cadi Ayyad university,
P.O. Box 618/ 40000, Marrakesh, Morocco

Abstract

The development of small integrated antennas plays a significant role in the progress of rapidly expanding mobile communications systems.

Although the resonance frequency, gain and bandwidth performance of an antenna are directly related to its dimensions, the idea is not only to decrease the overall size of the antenna, but also to find the best geometry and structure, for a specific problem. The advantage of planar inverted-F antenna (PIFA) makes them popular in many applications requiring a low profile antenna. PIFA antenna is promising to be a good candidate for the future technology due to the flexibility of the structure as it can be easily incorporated into the communication equipments.

In this paper, a triple-band Planar Inverted-F Antenna (PIFA) with Slots and a parasitic element is introduced. The proposed antenna is compact with a single feed and a single short strip. It operates at the frequency bands of 800 MHz, 1.6 GHz and 2.9 GHZ. So the antenna has been successfully researched and well optimized.

Keywords: Planar inverted-F antenna, PIFA, Slots, triple-band antennas, parasitic element.

1. Introduction

Nowadays, several novel multiband antenna designs have been introduced in the literature. The objective is to find the best geometry and structure giving the best performance while maintaining the overall size of the antenna small. The integration of different radio modules into the same piece of equipment has given a need for triple-band antennas. If three separate antennas are used, this will be inefficient in terms of space usage, which is an important factor in small hand-held devices. At present, triple-band planar inverted-F antennas have gained a lot of interest, due to their small size and good applicability for portable device because of its low profile and built-in

structure. In recent years; many methods can be used to achieve a compact multiband antenna. In [1], it is shown that a capacitive load between the short and the feeding probe can considerably enhance the impedance bandwidth of a small size PIFA. A capacitive load has also been employed in [2] and [5] to increase the electrical length of the radiating element without increasing the antenna volume. In [2], three coplanar parasitic patches were used to enhance the impedance bandwidth of the dual-resonant driven element. In [3], a meandered multiband planar inverted-F antenna with two coplanar parasitic patches is presented operating in the GSM band with high radiation efficiency. The multiband antennas introduced in [2-6] have rather complicated geometries, which makes them less attractive for mass production. A slot also has been used to decrease the size of antenna, to achieve multi band and alter the polarization characteristic as a means of conventional technique [7-10].

2. Antenna Geometry and Parametric Study

2.1 Planar inverted-F antenna (PIFA)

The inverted-F antenna is evolved from a quarter-wavelength monopole antenna. It is basically a modification of the inverted F antenna IFA which consists of a short vertical monopole wire.

To increase the bandwidth of the IFA, a modification is made by replacing the wires with a horizontal plate and a vertical short circuit plate to obtain a PIFA antenna.

The PIFA antenna has the advantages of having small and multiband resonant properties, a simple design, a light weight, a low cost, a conformal nature, an attractive radiation pattern, and a reliable performance [11]. These

characteristics make the PIFA a suitable antenna candidate for mobile phones.

The conventional PIFA consists of a top patch, a shorting pin and a feeding pin. The top patch is mounted above the ground plane, which is also connected to the shorting pin and feeding pin at proper positions. They have the same length as the distance between the top patch and the ground plane.

The standard design formula for a PIFA antenna is [12]:

$$f = \frac{C}{4(L+W)} \quad (1)$$

Where f is the resonant frequency of the main mode, C is the speed of light in the free space; W and L are the width and length of the radiation patch respectively.

2.2 PIFA antenna Configuration

The configuration of the studied PIFA antenna consists of a radiating top plate with the dimensions $40 \times 22 \text{ mm}^2$, and the thickness of the copper used is 0.8 mm. The ground plane dimensions are $40 \times 100 \text{ mm}^2$. The dielectric material used above the rectangular ground plane is FR-4 having a thickness $t=1 \text{ mm}$, a loss tangent of 0.02 and a relative permittivity of 4.4, this is meant for the application when the antenna is integrated with the printed circuit board (PCB). The antenna height is $h=8 \text{ mm}$, and the space between the top plate and the substrate is filled with air (free space).

Figure 1 shows the structure of the proposed antenna in top view, a tri-dimensional view and a side view. The detailed dimensions are given in Table I.

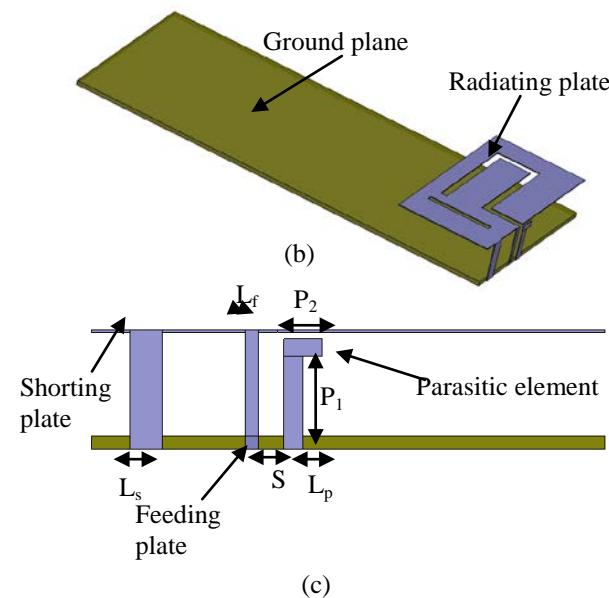
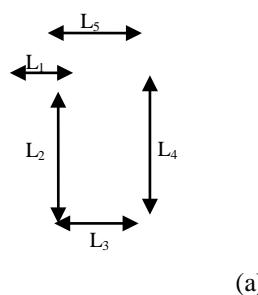


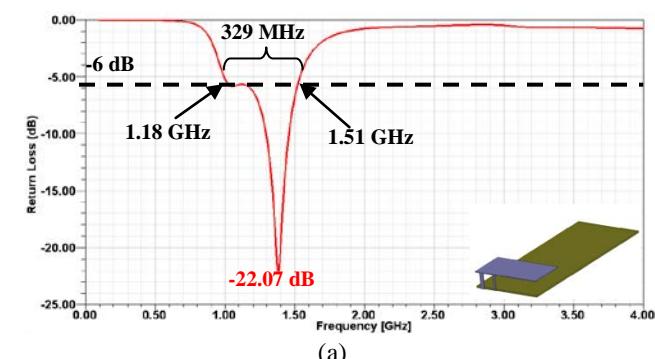
Fig. 1 Geometry of the proposed PIFA antenna, (a).Top view, (b).3D view and (c).Side view.

Table 1: The overall dimensions

L₁	L₂	L₃	L₄	L₅	P₁
8 mm	20 mm	10.5 mm	23.5 mm	15 mm	7.2 mm
P₂	L_f	L_s	L_p	S	
6 mm	1 mm	2.5 mm	1.5 mm	2 mm	

3. Results and discussions

3.1 The return loss



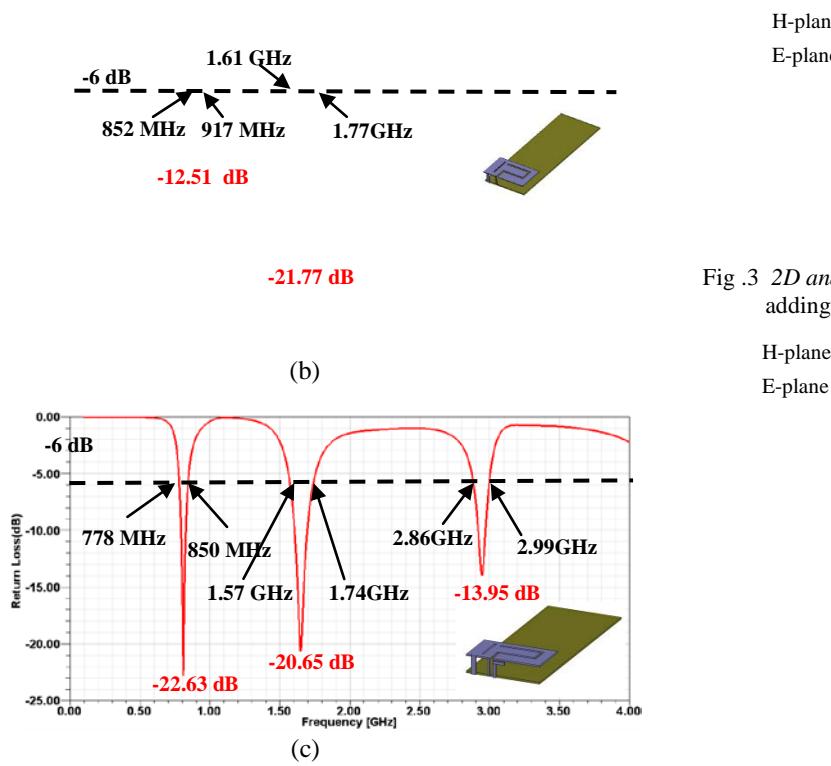


Fig .2 The simulated return loss for the proposed antenna (a) without slots and without parasitic element, (b) with slots and without parasitic element (c) with slots and with parasitic element.

As shown in Figure 2(a), we notice that the maximum return loss is -22.07 dB at 1.38GHz, The upper and lower band frequencies are 1.18 GHz and 1.51 GHz respectively and the absolute impedance bandwidth is 329 MHz and the bandwidth obtained for the proposed PIFA is 13.24%. It can be seen in Figure 2(b) that a new resonance frequency is added due to the slots. These slots force the surface currents to meander, thus artificially increasing the antenna's electrical length without modifying its global dimensions, which results in a new resonant frequency. The figure 2 (c), demonstrates that by adding a parasitic element, we obtain an odder resonance frequency. We conclude that by adding slots and a parasitic element to a PIFA antenna we have obtained a triple-band antenna.

3.2 The radiation pattern

In this section, we have computed the far field radiation patterns at the three operating frequencies for which the proposed antenna have been analyzed.

Simulated cuts of the studied antenna are shown in the Figure 3 before adding the slots and a parasitic element, then after adding slots (Fig. 4) and after adding the slots and a parasitic element (Fig. 5).

Fig .3 2D and 3D radiation patterns for the PIFA antenna before adding slots and parasitic element for $f_r=1400\text{MHz}$

H-plane
E-plane

(a)
H-plane
E-plane

(b)

Fig .4 2D and 3D radiation patterns for the PIFA antenna with slots and without parasitic element for (a) $f_r=880\text{MHz}$ and (b) $f_r=1700\text{MHz}$.

H-plane
E-plane

(a)

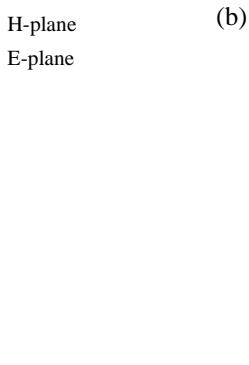


Fig .5 2D and 3D Radiation patterns for the proposed PIFA antenna with slots and with parasitic element for (a) $f = 809 \text{ MHz}$, (b) $f = 1.6 \text{ GHz}$ abd (c) $f = 2.9 \text{ GHz}$

From these results, we can see the gain variation of the antenna in the xz plane (the E-plane of the antenna) and the xy plane (the H-plane of the antenna) before and after adding slots and a parasitic element at the different resonance frequencies. The patterns of our proposed PIFA antenna are omni-directional in the three resonance frequencies with a maximum gain of about 5.11 dB for 809 MHz, 6.94 dB for 1.6 GHz and 6.72 dB for 2.9 GHz (Fig.6).

4.conclusion

This paper has focused on the development of triple-band planar inverted-F antennas, operating in the 809 MHz, 1.6 GHz and 2.9 GHz. The obtained results demonstrate that by adding slots on the top plate of the planar inverted-F antenna and a parasitic element we obtain a triple-band antenna.

Moreover, good radiation characteristics for frequencies within the operating bands have been obtained.

References

- [1]S. Villeger, P. Le Thuc, R. Staraj and G. Kossiavas, "Dual-band planar inverted-F antenna", *Microw. Opt. Technol. Lett.*, vol. 38, no. 1, pp.40–42, 2003.
- [2]P. Ciais, R. Staraj, G. Kossiavas and C. Luxey, "Design of an internal quad-band antenna for mobile phones", *IEEE Microw.Wireless Compon. Lett.*, vol. 14, no. 4, pp. 148–150, Apr. 2004.
- [3]Y.-X. Guo, I. Ang and M. Y. W. Chia, "Compact internal multiband antennas for mobile handsets", *IEEE Antennas Wireless Propag. Lett.*, vol. 2, pp. 143–146, 2003.
- [4]Y.-X. Guo and H. S. Tan, "New compact six-band antenna", *IEEE Antennas Wireless Propag. Lett.*, vol. 3, pp. 295–297, 2004.
- [5]J. Ollikainen, O. Kivekäs, A. Toropainen, and P. Vainikainen, "Internal dual-band patch antenna for mobile phones", in *Proc. AP2000 Millennium Conf. Antennas Propagation*, Davos, Switzerland, Apr. 9–14, 2000, p. 1111.
- [6]P.Salonen.; M. Keskilammi; M.Kivikoski; "Dual-band and wide-bandPIFA with U- and meanderline-shaped slots" ,*Antennas and Propagation Society International Symposium*, 2001. IEEE Vol. 2, 8-13 July 2001, pp.116 – 119.
- [7]M.Jung, M.; Yunghee Kim; Lee, B.; "Dual frequency meandered PIFA for Bluetooth and WLAN applications" ,*Antennas and Propagation Society International Symposium*, 2003. IEEE Vol. 2, 22-27 June 2003, pp.958 – 961.
- [8]S. Ibnyaich, A. Ghannaz, and M. M. Hassani, "Development of Wideband Planar Inverted-F Antennas for Wireless Application", *Journal of Computer Science*, vol. 7, pp. 1172-1177, 2011.
- [9]S. Ibnyaich, A. Ghannaz, and M. M. Hassani, " Triple-band Planar Inverted-F antenna PIFA for mobile phone ", 11th mediterranean microwave symposium", *MMS'2011*, Tunisia-Hammamet .pp 67 September 08-10, 2011.,
- [10]Y. Belhadjef, N. Boukli-Hacene, "Design of New Multiband Slotted PIFA Antennas, "IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 4, No 1, July 2011 ISSN (Online): 1694-0814.
- [11]C. A. Balanis, (1997) *Antenna Theory: Analysis and Design*, Second, edit. John Wiley & Sons.
- [12]Y. Huang and K. Boyle, *Antennas: From Theory to Practice*. Hoboken, NJ: Wiley, 2008.

Saida Ibnyaich was born in Morocco, in 1979. She received the Bachelor of sciences (4 years study after the baccalaureate) in electronic engineering from Cadi Ayyad University, Marrakesh Morocco, in 2002. Received the DESA (equivalent to M.A.) in Electrical Engineering, Power Electronics and Industrial Control from Cadi Ayyad University; Marrakesh Morocco, in 2005. She is currently working toward the Ph.D. degree at the Department of Physics, Electronics and Instrumentation Laboratory, Cadi Ayyad University of Marrakesh, Morocco. Her research interest includes telecommunications and antennas.

Abdelilah Ghannaz received the Doctor of Electronic degree from the National polytechnic Institut (ENSEEIHT) of Toulouse, France, in 1993. In 1994 he went back to Cadi Ayyad University of Marrakech – Morroco. Since 2003, he has been a Professor at the Faculty of Sciences and technology, Marrakech, Morroco. His

research interests in the field of electromagnetic compatibility and antennas.

Moha M. Hassani was born in Morocco, in 1954. He received the Third Cycle Doctorate in automatic and signal processing from The University of Nice, France, in 1982, and the Ph.D. degree in electrical engineering from the University of Sherbrooke, Canada, in 1992. He joined Cadi Ayyad University, Marrakesh, Morocco in 1982 as an Assistant Professor. Since 1993, he has been a Professor at the Faculty of Sciences, Semlalia, Marrakesh, Morocco. His research interests are mainly nonlinear process modeling and identification applied to telecommunication and solar energy fields.

HASoC for Developing a Software System

Salah Eldin Abdelrahman¹, and Mohammed Badawy²

¹ Computer Science and Engineering Dept., Faculty of Electronic Engineering, Menoufia University
Menouf, 32952, Egypt

² Computer Science and Engineering Dept., Faculty of Electronic Engineering, Menoufia University
Menouf, 32952, Egypt

Abstract

We present an object-oriented development of a software system. The development is based on the customization of the lifecycle of a novel developing method called HASoC (Hardware and Software Objects on Chip). The development is presented in order to evaluate HASoC and establish a complete A-to-Z object-oriented teaching example for developing software systems. The evaluation is performed through an object-oriented development process of a software system and is aimed to prove the HASoC practicability and reveal its limitations. The HASoC method was originally aimed for developing embedded systems that are targeted at system-on-a-chip (SoC) implementations.

Keywords: Object-Oriented Approaches, UML, Object-Oriented Database, Software Development, Lifecycle Modeling.

1. Introduction

Object-oriented technology offers several specific benefits to software development. Good software engineering stresses modularity, especially in system architecture, detailed design, and implementation phases [1]. Therefore, the benefits of object-orientation are realized throughout the entire development lifecycle, which is typically supported by a set of system models. The system models, which are necessary to communicate the required information that users understand, can be translated into implementation code. object-oriented technology can be consistently applied in the analysis, design, implementation, and test phases of software systems. This means that an object-oriented approach is model-based and covers the development lifecycle in an integrated way.

The basic concept of object-oriented development approaches is that systems should be built from a collection of reusable interacting objects. Various object-oriented methods have been used for systems analysis and design. Therefore, the necessity to standardize these methods has become important. The contemporary modeling language called the Unified Modeling Language (UML) [2] has emerged from the various object-oriented

analysis and design notations of these methods in order to meet this necessity.

The developers of UML have attempted to unify past experience of software modeling techniques and incorporate a number of the current software practices into a standard language. UML combines the commonly accepted concepts, notation, and terminology from many object-oriented development methods and selects a clear definition of them. The UML is process, domain, and media independent. It represents the conceptual and physical elements of systems as a set of views, which permit systems to be understood for different purposes. UML views are divided into Structural Classifications, Dynamic Behaviors, and Model Management as well as Extension Mechanisms [3, 4].

According to [5, 6], object-oriented development approaches provide architects, designers, and developers with the conceptual framework, i.e. object model, to attack the rapid increase of systems complexity, and enhance system flexibility. The object model encompasses the principles of abstraction, encapsulation, modularity, hierarchy, typing, concurrency, and persistence, and it provides a conceptual foundation of object-oriented development approaches. Therefore, such approaches offer several potential benefits for system developers and users. Since almost nothing is perfect and without cost, the benefits of object-oriented development are associated with potential limitations, problems, and drawbacks. As [7] has pointed out, most object-oriented development costs are concerned with performance and the difficulties of moving from non-object-oriented to object-oriented development techniques.

A set of case studies are driven to illustrate and demonstrate the developing stages of object-oriented approaches. As mentioned in [5, 12, 15-18], not only one use case study is driven but also different case studies are driven. Each of such case studies illustrates and

demonstrates one stage of the development process. This leads to miss modeling mapping among the development stages. Therefore, the relation between different system models is not clear. Driving only one case study for all stages of the development process makes the relationship between system's models more clear and the transfer from one model to another straightforward.

This paper firstly aims to illustrate and demonstrate the design approach of the HASoC (Hardware and Software Objects on Chip) method presented in [19] for a software system to evaluate its performance against these systems. The evaluation is performed through the customization of HASoC lifecycle and object-oriented development process. In addition to the evaluation of the novel developing method HASoC for software systems development, a complete A-to-Z object-oriented teaching example for developing such systems is established. We chose a database system as an example of a software system to apply the HASoC method. This software system will be explained later in this paper.

This paper is organized as five sections. In the following section, an overview of the HASoC method as used in this paper is provided. The third section considers and indicates the customization of the HASoC lifecycle to target it specifically at the development of pure software systems. The development process of a software system is demonstrated and illustrated in section four. In the fifth section, the paper is concluded with final remarks.

2. HASoC Overview

HASoC is a method for developing 'complete' embedded systems that are targeted at system-on-chip (SoC) implementations. HASoC aims to develop the entire system not its components that meets functional and non-functional requirements. It supports concurrent development of computer systems software and hardware components. Since the use of contemporary object-oriented analysis and design techniques is helpful for developing embedded systems, HASoC is an object-oriented method. HASoC utilizes object-oriented principles, in general, and the generalized UML-RT profile in particular for system modeling. The generalized UML-RT profile is an extension of the standard UML. This profile includes alternative forms of behavioral description and a new communications mechanism. So, HASoC has a richer behavioral modeling capability than other embedded systems development methods such as POLIS [8], MOOSE [9, 10], and COSY [11].

A set of positive characteristics and development techniques from a number of existing well-defined development approaches have been merged within HASoC. The existence of such techniques is helpful to meet or reduce the effect of many of the challenges that face developers of embedded systems. These challenges include the ever increasing time-to-market pressure, the productivity gap, and the complexity of embedded systems coupled with the multi-disciplinary nature of the development task. Iterative, incremental, and use-case-driven techniques have been merged within the HASoC lifecycle (see Figure 1) to make it responsive to changing system requirements. The system functionality is gathered and specified in a set of use cases that are depicted in a use case model of the system. If one or more requirements of a system are changed, the HASoC lifecycle provides developers with the ability to update the system functionality by editing, adding, or removing one or more use cases. As well as the ability to accommodate changes in system requirements, the adoption of these techniques leads to a simplification of the development task, as it is achieved in a set of iterations rather than a single pass through the lifecycle. The adoption of the use-case driven technique also provides designers with the ability to manage the system development process, and trace requirements.

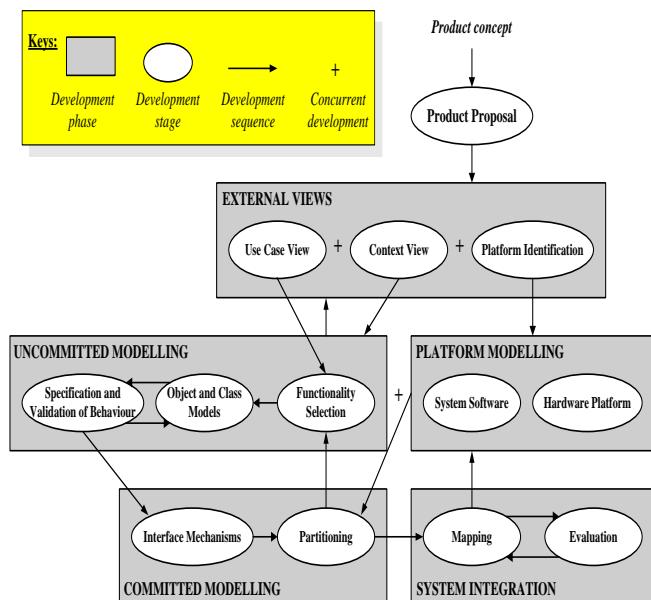


Fig. 1 The HASoC lifecycle.

The platform-based design approach has been integrated into the HASoC lifecycle. This integration, coupled with the above techniques provides a concurrent development facility within HASoC. The application and platform

models of a system can be concurrently constructed and evaluated at the early stages of development. This enables different specialists within the development groups to work in parallel but always in the context of a homogenous model of the complete system. Therefore, the development time of a system may be reduced, which is helpful in reducing the time-to-market of the product. The concurrent development capability of the HASoC method supports the reuse of pre-existing platforms, and the early exploration of system trade-offs. It provides consistency between the development of the application and platform models of a system.

3. HASoC Lifecycle for Software Systems

Although the basic HASoC lifecycle has been identified as shown in Figure 1, it is possible to operate different lifecycles in the context of the method. This could allow developers to adopt UML or HASoC-based modeling whilst still using a variant of their original lifecycle. It also offers the possibility of customizing the lifecycle to suit a particular project or type of project, or even specific application domains. However, we investigate this kind of use of the HASoC method to develop software systems.

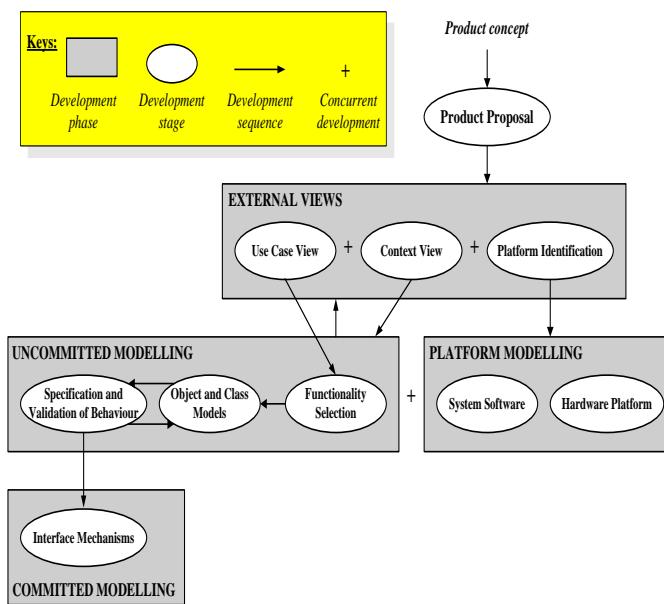


Fig. 2 The customized HASoC lifecycle for software systems.

Some parts of the HASoC lifecycle may be cancelled, in particular, the partition-map-evaluate cycle. Figure 2 illustrate the customization of the HASoC lifecycle for software systems development. This development starts

with the basic concept definition of the application and progress through external views and uncommitted stages. We drive the University Registration System (URS) as a case study application.

4. The Software Development Process

The software development process partially follows the HASoC development phases (see Figure 2). These phases include product proposal, external views of the product, and the uncommitted modeling. In the next sections, we explain the steps of each phase in details.

4.1 Product Proposal

The application as a case study in this paper is a University Registration System (URS). The URS has enough features to partially illustrate and demonstrate the design approach adopted in HASoC to develop a software system. It is acknowledged that the development discussed in this paper may not be optimal for a URS. However, it was chosen to illustrate and evaluate as many aspects of the HASoC method.

Students enroll in courses which are offered by instructors. The registrar's office is in charge of maintaining a schedule of courses in a course catalog. They have the authority to add and delete courses and to add schedule changes. They also set enrolment limits on courses (such as, number of students in a course and studying the prerequisite of that course). The financial aid office is in charge of processing student's aid applications for which the students have to apply. The instructors are in charge of setting course offerings, receiving their courses' rosters, and entering the grades of courses they are teaching. Whereas, the students are in charge of being able to register for courses and apply for the financial aid. Assume that we have to design a database that maintains the data about students, Instructors, courses, aid, etc. We also want to design the application that enables us to do the course registration, financial-aid application processing, and maintaining of the university-wide course catalog by the registrar's office.

4.2 External Views of the URS

In the previous phase, the URS requirements have been recorded as a textual report that documents, classifies and clarifies (in an ad hoc way) what is initially known about the system. In this phase, the URS requirements are analyzed and categorized. The external views are developed to express the externally observable behavior that is required of the URS, and its interface with the surrounding environment, see Figure 3.

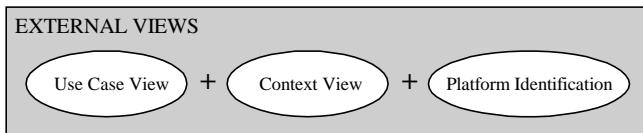


Fig. 3 External views of the HASoC lifecycle

4.2.1 The Context View

The URS Context View aims to state precisely but at a high-level of abstraction, the communications between the URS and its environment. The design and construction of this view are based on the analysis of the above statement that reports the URS requirements. The URS Context View, which is illustrated in Figure 4, depicts the complete URS under development as a single object (the system object) surrounded by external objects (actors) with which it has to communicate or interface. We describe the URS actors as in Table 1 based on an analysis of the written statement of the URS. The actors send a set of messages to the URS, which issues a set of responses. The set of messages passed between the URS and its actors defines the interactions between the URS and its surroundings. Table 2 names and describes the messages and the URS responses.

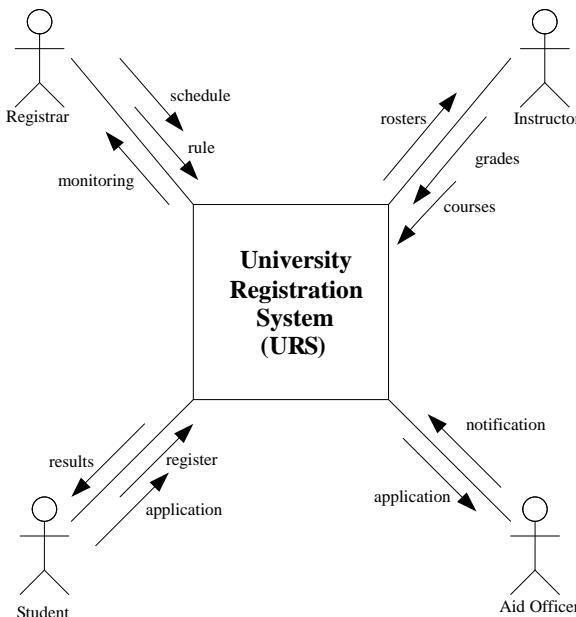


Fig. 4 A URS context diagram

Table 1 The description of the URS actors

Actor	Description
Registrar	An employee who maintains the schedule, sets the enrollment limits, and monitors the registration process.
Instructor	A person who offers the courses, receives the rosters, and enters the grades of the students.
Student	A person, who registers for a course, applies for a financial aid.
Aid Officer	An employee who receives the application for an aid and return a notification of acceptance.

Table 2: The description of the external interacting messages of the URS with its actors

Message	Description	The URS response
schedule	Courses are added or deleted and the schedule is changed by the registrar.	It accepts the commands and updates the catalog.
rules	The registrar sets the enrolment limits on courses.	It accepts or refuses the registrations based on those limits.
monitoring	The registrar monitors the use of the system.	It sends messages about the use of the system to the registrar.
register	The student requests a registration for courses.	It accepts all or part of his/her request based on the rules.
application	The student applies for a financial aid.	It accepts or refuses the aid. It sends the request to the aid officer.
results	Messages returned to the student about his/her registration or aid requests.	It sends a message to the student about accepting or refusing his/her requests.
notification	A notification of acceptance the aid is sent by the aid officer.	It sends the notification to the student.
courses	The instructor offers some courses.	It adds these courses to the catalog.
grades	The instructor enters the grades of students	It saves the grades in the system.
rosters	The instructor receives the rosters of his/her courses.	It sends rosters to the instructors for their courses.

4.2.2 The Use Case View

The Use Case View of the URS establishes the forces that will drive subsequent development stages. It can be developed concurrently with the URS Context View as a UML use case diagram (Figure 5) with the associated use case descriptions (examples are shown in Table 3). It

contains a set of use cases that define the functionality of the URS.

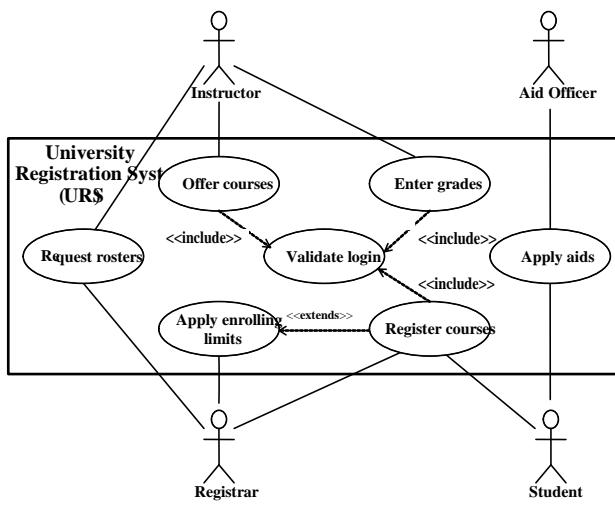


Fig. 5 A use case diagram for a URS

Table 3: Use cases descriptions

Name	Register courses
Summary	The URS receives requests for registration from students.
Actors	Student, Registrar
Description	The students send a registration requests for courses to the system. Registrar monitors the registration process.
Name	Validate login
Summary	The URS has to validate the using of the system.
Actors	Instructor, Student
Description	The instructor and the student must have the authority to use the URS for offering the courses and receiving the rosters in the case of the instructor and enrolling for courses and applying for an aid in the case of the student.
Name	Apply enrolling limits
Summary	The URS ensures the limits of enrolling for courses.
Actors	Registrar, Student
Description	The registrar has to ensure that the enrolling limits are satisfied when students registering for courses.
Name	Request rosters
Summary	The URS receives a request for a roster from the instructor.
Actors	Instructor, Registrar
Description	The instructor requests the names of students who are registered for his/her courses.

Name	Apply aids
Summary	The URS receives the requests for aids from the student.
Actors	Aid Officer, Student
Description	The aid officer receives the requests from the student and sends notifications by an acceptance to him/her.
Name	Offer courses
Summary	The URS receives course offerings from the instructor.
Actors	Instructor
Description	The instructor offers the courses he/she wants to add to the catalog for student registration.
Name	Enter grades
Summary	The URS receives the grades of courses from the instructor.
Actors	Instructor
Description	The instructor enters the grades of courses he/she teaches.

4.2.3 Platform Identification

We assume the initial platform of the URS exists. This at least consists of one microcomputer, and a set of software including a database management system such as Oracle [12, 20], an operating system e.g. Windows, etc. Therefore, based on a UML deployment diagram, the initial version of the URS platform is shown in Figure 6. The basic URS platform may be as that in the initial platform.

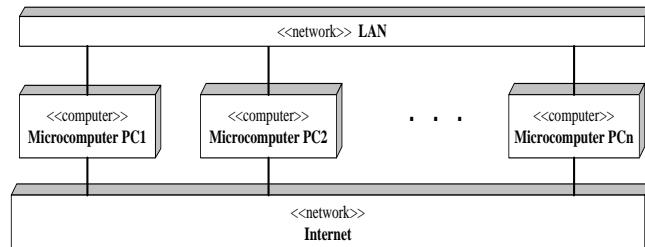


Fig. 6 A deployment diagram for the URS platform

4.3 Uncommitted Modeling: First Iteration

For the sake of simplicity, attention is restricted to the registration process. In other words, the intention is to consider the development of URS registration sub-system. The activities associated with uncommitted modeling are shown in Figure 7. Here, the concern is the development of

the core functionality of the URS, and assume that in the later stages additional functionality will be introduced.

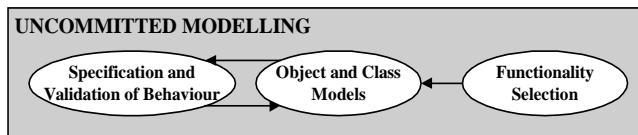


Fig. 7 Uncommitted modeling phase of the HASoC lifecycle

4.3.1 Functionality Selection

Since a typical URS has to register courses for students, the ‘Register courses’ use case is selected as a core functionality. We select the ‘Register courses’ use case from the URS use case model and then develop it within the first iteration.

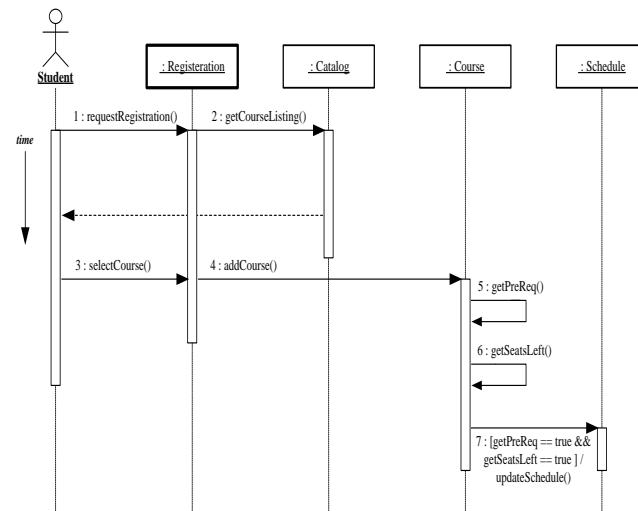


Fig. 8 A sequence diagram of the “Register courses” use case

4.3.2 Object and Class Models

A UML sequence diagram representing the objects and the interactions involved in a scenario that realizes the selected ‘Register courses’ use case is depicted in Figure 8. This scenario describes a typical course of events that might occur when the URS registers courses for students. The sequence diagram indicates that the :Registration object is active because it is capable of autonomous activities. Once the student asks for a course, the :Registration object calls the :Catalog object for bringing the course list. Once the student selects the requested course from the list, the :

Registration object calls the :Course object for adding the course to the student’s courses. Figure 9 illustrates the collaboration diagram.

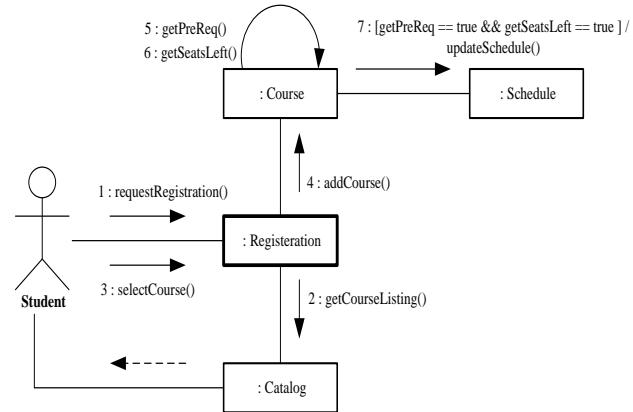


Fig. 9 A collaboration diagram of the “Register Courses” use case

From the above sequence/collaboration diagrams, the corresponding UML class diagram is illustrated in Figure 10. This diagram depicts the static structure of the application-specific part of the URS corresponding to this iteration.

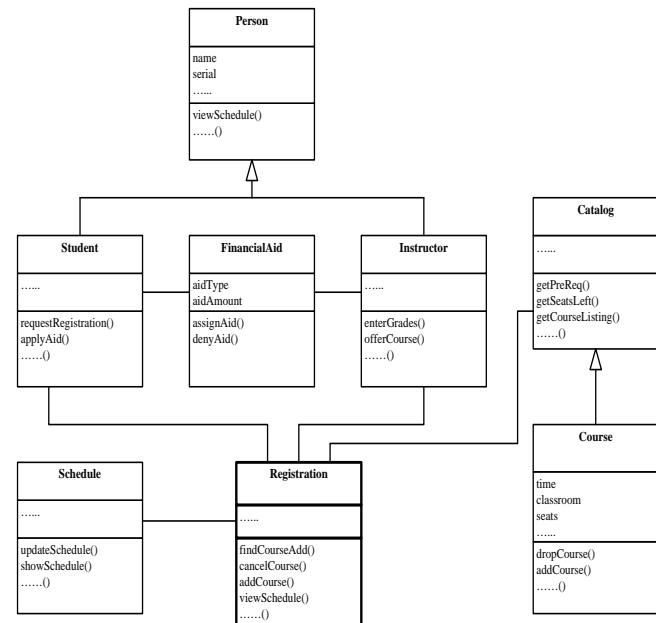


Fig. 10 A UML class diagram of the URS (first iteration)

4.3.3 The Specification and Validation of the URS Behavior

In this stage, the detailed implementation of the classes relevant in the current development iteration is specified. This can be involved in a convenient way through the synthesis of an executable model from the details of object interactions and internal behavior specified in this stage. In other words, model validation can be accomplished by considering the overall behavior of the current increment. The creation of an executable model is briefly considered below. Each development iteration extends the class and object models by introducing functionality into the system specified by the use cases under consideration in the current iteration. Hence, the use case view can be used to provide test specifications for the executable model.

Executable Modeling

The detailed model validation can be achieved through the execution of a full software model. The class model provides important information for synthesis of the executable model and supports the synthesis of interface code. Also, the sequence diagrams provide information about run-time inter-object communications. From these models a skeleton of the executable model can be synthesized. This basic model can be made executable with the addition of supporting generic objects that control the simulation process. Once the model is executable, then the validation of object-level communications can be performed. In order to provide a finer degree of behavioral validation, detailed code would be added to the skeleton executable model, via the addition of code to describe the behavior of passive objects.

4.4 Committed Modeling: First Iteration

The commitment of a system model aims to move the uncommitted model of the URS system towards an implementable specification. Once the executable and initial platform models of the system have been constructed and evaluated, uncommitted objects of the system can be committed. In case of a pure software system, the commitment is achieved as illustrated in Figure 11 through the only Interface Mechanisms development stage.



Fig. 11 The committed modeling phase of the modified HASoC lifecycle

Interface Mechanisms

The only stage in the commitment of a software system model considers the external interface of the system. The mechanisms by which the environment interacts with the software system will be fixed at the start of development, whereas in others they must be determined as part of the design process. Clearly, the decision that is taken leads to different sets of interface objects being added to the model. Once the structural and behavioral aspects of the selected 'Register courses' use case have been developed, we move towards the development of an implementable specification. With respect to this use case, there are no additional interface objects that need to be added to the URS model.

4.5 Platform Modeling: First Iteration

The aim of the platform modeling is to describe the overall execution environment in sufficient detail to facilitate the implementation of the complete system, application and platform. Therefore, the platform-modeling phase considers those components that play a supporting role in the operation of the URS. The hardware components including processor, buses, memories, etc. are described through the HAM (Hardware Platform in Figure 12), and the software components, including the operating system, DBMS, etc.

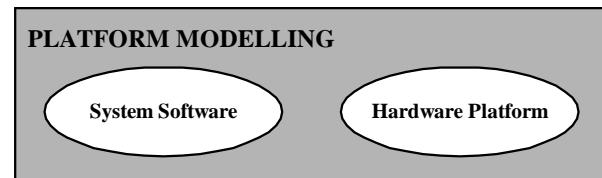


Fig. 12 The platform modeling phase of the modified HASoC lifecycle

At the completion of an iteration of the Platform Identification stage, a model of the implementation of one or more use cases exists in terms of a set of interacting objects. In order to construct a working system, however, further consideration must be given to those components that play a supporting role in the operation of system. These components including processors, buses, memories, DBMS, and an operating system provide the execution environment.

At one end of the spectrum, a pre-existing Hardware Platform may be used in its entirety for the implementation of a system. However, it is more likely that an existing platform would be customized or reconfigured for a particular system through the addition, removal, replacement, or modification of hardware and software IP objects. In each iteration, the capabilities of the platform

could be updated to support the additional functionality introduced into the system in that iteration. If a new platform were to be developed, the HASoC method may be used (in an iterative way) to develop those parts of the platform that are directly used by the application through the introduction of IP cores and supporting software.

4.6 Uncommitted Modeling: Second Iteration

By the same way, we can consider and complete the second iteration of the development of the URS. This development starts by selecting the '**Offer courses**' use case from the use case model and then develop it through the subsequent stages.

5 Conclusion

We have customized the lifecycle of the HASoC method for software systems development, although this method was aimed for developing embedded systems that are targeted at system-on-chip implementations. Not all stages of the HASoC lifecycle are applied for developing software systems. Such development is concerned with the evaluation of the HASoC method and established a complete A-to-Z object-oriented teaching example for developing software systems.

A set of specific benefits have been gained from object-oriented development of software systems using HASoC method. Iterative, incremental, and use-case-driven techniques that have been merged within the HASoC lifecycle make the development of software systems responsive to changing system requirements. The development process has reaped the benefits of using a widely understood notation of UML that has been successfully applied in the development of large-scale software systems. Indeed, the development process has a significant concurrency that makes the underlying platform model of the developed system can commence much earlier with its application model development. Moreover, object oriented development can lead to robust and extendible software systems.

As future work, the following topics deserve more attention as a result of using HASoC method. It is fundamental to investigate, at different aspects, what are the implications and consequences of software systems development. It is important to devise more rigorous supporting software tools for HASoC method, to bring out the benefits of object-oriented mechanisms like inheritance and polymorphism. Applying the HASoC method and demonstrating its perceived strengths in the development of other software systems would allow more solid

assessments about the worth and usefulness of this method. We believe that the HASoC is a realistic and feasible method for software systems development. Moreover, the current application areas under consideration include the development of pure hardware systems using HASoC method.

6 References

- [1] E. J. Braude, Software Engineering: an Object-Oriented Perspective, John Wiley & Sons Inc., 2001.
- [2] <http://www.uml.org>
- [3] G. Booch, J. Rumbaugh, and I. Jacobson, The Unified Modeling Language User Guide (The Second Edition), Addison Wesley Object Technology Series, 2005. ISBN: 0321267974.
- [4] J. Rumbaugh, I. Jacobson, and G. Booch, The Unified Modelling Language Reference Manual: Addison Wesley, 2004.
- [5] G. Booch, Object-Oriented Analysis and Design with Applications, Third Edition: Redwood City, CA, Benjamin/Cummings Publishing Company, Inc., 2007 ISBN: 9780201895513.
- [6] G. Booch, Object-Oriented Design with Applications: Redwood City, CA, Benjamin/Cummings Publishing Company, Inc., 1991.
- [7] I. Graham, Object-Oriented Methods Principles and Practice, Third Edition: Addison Wesley, 2001.
- [8] F. Balarin, E. Sentovich, M. Chiodo, P. Giusto, H. Hsieh, B. Tabbara, A. Jurecska, L. Laavangno, C. Passerone, K. Suzuki, and A. Sangiovanni-Vincentelli, Hardware-Software Co-Design of Embedded Systems: The POLIS Approach (The Springer International Series in Engineering and Computer Science) Kluwer Academic Publishers, 1997.
- [9] Computer Systems Engineering Group, "MOOSE: User Manual Version 5.01," <http://www.cl.co.umist.ac.uk/moose>, February 1, 1999.
- [10] Derrick Morris, David Evans, Peter Green, and Colin Theaker "Object Oriented Computer Systems Engineering (Applied Computing)", Springer, 1996 ISBN-13: 978-3540760207.
- [11] J.-Y. Brunel, W. Kruijzer, H. Kenter, F. Petrot, L. Pasquier, E. d. Kock, and W. Smits, "COSY Communication IP's," in Proceedings of Design Automation Conference (DAC), pp. 406-409, 2000.
- [12] Ramez Elmasri, and Shamkant Navathe, "Fundamentals of Database Systems (6th Edition)", Addison Wesley, (April 9, 2010) ISBN-13: 978-0136086208
- [13] <http://www.phindia.com/gupta/>
- [14] <http://www.webhostingsearch.com/articles/oracle-database-management-system.php>
- [15] Michael R. Blaha and William Premerlani, Object-oriented Modeling and Design for Database Application Using OMT and UML: Prentice Hall 1 edition, 1997.
- [16] C.S.R. Prabhu, Object-oriented Database Systems Approaches and Architectures, 2nd edition: Prentice-Hall, 2005

- [17] M. O'Docherty, Object-oriented Analysis and Design Understanding System Development Using UML 2.0: John Wiley and Sons, 2005
- [18] J. Harrington, Object-oriented Database Design: Morgan-Kaufman, 2000
- [19] A. Salah Eldin, "Object-Oriented Technology for System-Level Design," Ph.D. Thesis, University of Manchester Institute of Science and Technology, UMIST, Manchester, U.K, 2003
- [20] Oracle Corporation, Oracle 10g Release 2 (10gR2), July 2005.

Salah Eldin Abd Elrahman. got his B. Sc. in Industrial Electronics Engineering May 1988, and M. Sc. in Computer Science and Engineering 9th of July 1994; both degrees from the Faculty of Electronic Engineering, Menoufia University, Egypt. He got his Ph. D. in Computations 11th of July 2003, from the Dept. of Computation, University of Manchester Institution of Science and Technology (UMIST), U. K. From 15th of January, 1989 to 4th of September 1994, he was a demonstrator, from 5th of September 1994 to 22nd of September 1998, as a lecturer, and from 21st of September 2003 to 2005 as an assistant professor at the department of Computer Sciences and Engineering, Faculty of Electronic Engineering, Menoufia University, Egypt. From 2005 to July 2010, he was an assistant professor at the department of Computer Engineering, Faculty of Computers and Information Systems, Taif University, Saudi Arabia and from 27st of September 2010 to August 2011, as an assistant professor at the Faculty of Computers and Information Technology, Tabuk University, Saudi Arabia. he is interested to do some researches in Using Object-Oriented Technology for Developing Embedded Systems, Real-Time Systems, and Software Systems such as Database Systems. The best publication is P. N. Green, M. D. Edwards, and S. E. S. Essa, "HASoC-Towards a New Method for System-on-a-Chip Development," Design Automation for Embedded Systems, vol. 6, No. 4, pp. 333-353, 2002.

Mohammed Badawy received his B.Sc. and M.S. in computer science and engineering at Menoufia University (Egypt) and received his Ph.D. in computer science and engineering at Czech Technical University in Prague (Czech Republic). He worked as assistant professor in the department of Computer Science and Engineering at Menoufia University (Egypt) from 2002 to 2005. He worked as assistant professor in the department of Information Technology at Taif University (Saudi Arabia) from 2005 to 2010 (3 years of them as chairman of the department). Currently, he worked as a consultant in the deanship of Information Technology at Islamic University (Saudi Arabia) from 2010. His research interests includes databases, data stream systems, and software development. He is a member of Association of Computer Science and Information Technology (IACSIT) and a reviewer of the International Journal of Engineering and Technology (IJET). He has published about 13 papers in various scientific journals and refereed conferences.

Nurture IDR Segmentation and Multiple Instruction Queues in Superscalar Pipelining Processor

J.Nandini meera¹, N.Indhuja², S.Devi Abirami³ and K.Rathinakumar⁴

¹ Computer Science and Engineering, Narasu's Sarathy Institute of Technology
Salem, Tamil Nadu, India

² Computer Science and Engineering, Narasu's Sarathy Institute of Technology
Salem, Tamil Nadu, India

³ Computer Science and Engineering, Narasu's Sarathy Institute of Technology
Salem, Tamil Nadu, India

⁴ Electronics and Communication Engineering, Narasu's Sarathy Institute of Technology
Salem, Tamil Nadu, India

Abstract

This paper proposes a model which improves the speed of the pipelining mechanism therefore increasing the speed of the processor. Superscalar operation is used to get maximum throughput from the processor using the pipelining concept. This proposal can be considered as the advancement of the super scalar property in pipelining which presently exists. We introduce a concept, using multiple instruction queues and a new unit called as Identifier unit. The Identifier unit is designed as having the ability of identifying the type of the instruction which is being fetched and separating it based on its types thus creating separate segments of execution, which in turn increases the speed of the processor. Moreover we have implemented separate decoding and the executing unit for each type of the instruction segments.

Keywords: Pipelining, Superscalar property, Identifier unit and Multiple Instruction queue.

1. Introduction

In a computer system the work of the processor is to execute the instructions given by the user. The processor executes the instructions in different fashions. Olden days the machine level instructions were executed in serial fashion which took lot of time the Pipelining mechanism is used achieve high execution rate in a processor. Superscalar operation is used to improve the speed of the processor.

Nurture IDR segmentation and multiple instruction queues in superscalar pipelining processor, is a model which is the modification of superscalar property. This model can enhance the speed of the pipelining processor. The core idea of this process is segmentation of the machine instructions which is being fetched from the fetch unit based on its type and storing it into separate instruction queue for each type, thus forming multiple instruction queues.

Pipelining is the first concept which is to be known in detail to understand the working of the Nurture IDR segmentation and multiple instruction queues in superscalar pipelining processor. Moreover the superscalar operation and the processing of the instruction through it, is important because it is the base of the proposed system. The implementation of Nurture IDR segmentation and multiple instruction queues in superscalar pipelining processor is aimed at increasing the speed of the superscalar pipelined processor. The working of each unit of the proposed system such as fetch unit, identifier unit, decode unit, execute unit and write unit are explained with sufficient information of its design and working. The new concept of Nurture IDR segmentation and multiple instruction queues in superscalar pipelining processor will definitely create an evolution in the speed of the superscalar pipelined processor.

2. Pipelining

The olden processors used the sequential technique to execute the instructions. The instructions were fetched and executed one by one in serial fashion, which took lot of time. Once the instruction is being fetched by the fetch unit it is executed by the execution unit during the execution of the instruction the fetch unit is ideal.

The following diagram Fig: 1 shows the execution of the instructions in the sequential execution of instruction

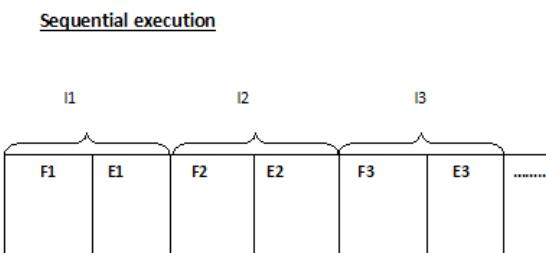


Fig: 1 sequential technique of execution

This order of execution had many disadvantages which lead to the invention of the Pipelining mechanism. Pipelining is the concept which has separate stages called as the fetch unit, decode unit and execute unit and the write unit [1]. By using the pipelining the time taken to execute an instruction is faster the sequential order of execution. This made the pipelining mechanism to a successful fashion of execution of the machine instruction in the processor. This mechanism is being used in present processors.

Pipelined instruction processing is a basic technique used in the design of micro-architectures [2]. Pipelining is method in which the instructions are executed in separate stages such as the fetch unit, decode unit and execute unit and the write unit. The fetch unit will actually read the instruction from the memory unit hence fetching it and after which the instruction reaches into the decoding unit where the instruction is decoded from high level language to low level language [1]. Then it enters into the execution unit where the ALU performs the arithmetic and the logical operations. Finally the executed instruction is sent to the write unit to write

the results. Consider the following example to understand the working mechanism of pipelining concept. In a PC manufacturing company, there are different units of people working to manufacture a single PC. First the design unit plans the design of the PC.

Then supplier unit supplies the hardware material needed, then the assembling unit do the assembling work of the hardware which is followed by the testing unit so on. Once the designing is over by the design unit for the first PC it proceeds to the next. Pipelining overlaps the execution of multiple instructions within a functional unit, much like an assembly line overlaps the steps in the construction of a product [3].

In the processor there are different units, which are listed as follows:

- Fetch unit
- Decode unit
- Execute unit
- Write unit

Pipelining is controlled by the clock whose period is such that the fetching, decoding, executing and write steps of each instruction can be completed in one clock cycle [1].

The following Fig: 2 show the process of pipelining mechanism in a processor with clock cycle.

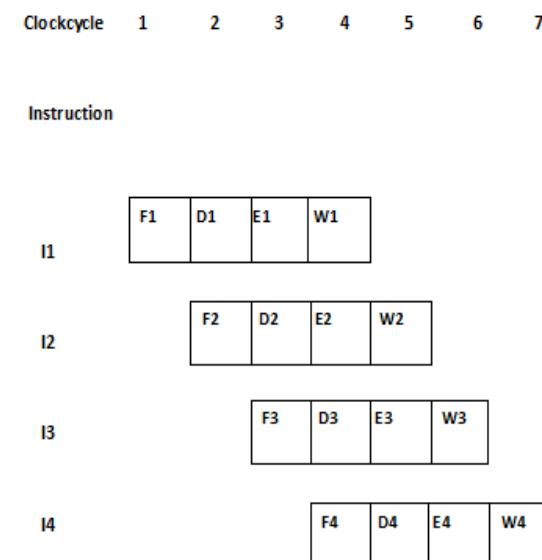


Fig: 2 Pipelined Instruction executions in four steps

3. Superscalar operation

Before the invention of the superscalar operation the execution of the instructions was done in sequential order which consumed a lot of time and ultimately resulted in the speed of the processor which was very slow in the execution process.

The following Fig: 3 show the execution of the instructions in without using the superscalar mechanism.

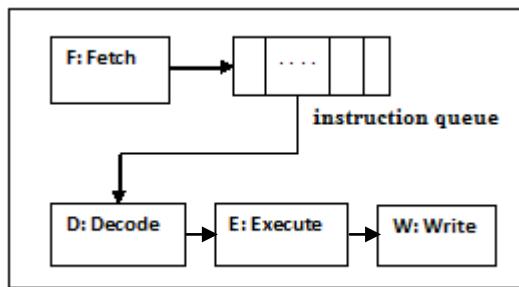


Fig: 3 normal execution of instruction

Superscalar processors include multiple functional units, such as arithmetic logic units and floating-point units. This enables the processor to exploit instruction-level parallelism (ILP), executing several independent instructions concurrently [3].

The key concept of superscalar operation os using of two separate execution units due to which the time of execution is fast than the normal pilelining processor. There buffers are used to store the instruction from the decode unit and the execution temporarily before they reach the next unit of processing mechanism.

Under the superscalar operation there are the following process [1].

- Out-of-order Execution.
- Execution Completion
- Dispatch Operation

Pipeline makes it possible to execute instruction concurrently [1]. Instructions are present in the pipelining at the same time, but they are in different stages of their execution [1]. While one instruction is performing an ALU operation, another instruction is being decoded and yet another is being fetched from the memory [1].

The front end of the processor is responsible for fetching instructions from memory and supplying them to the issue stage [4]. First the instruction is being fetched from the memory and in the instruction fetch block show the function of the fetch unit. It is followed by the instruction queue in which the instructions are stored in a queue waiting to be decoded. Then buffer are used to store the decoded instructions temporarily followed by two execution units namely,

- Floating point unit
- Integer point unit

The following Fig: 4 show the block diagram of the Superscalar property using the processor with two execution units. This gives the mechanism of the superscalar operation in a processor.

A processor with two execution units

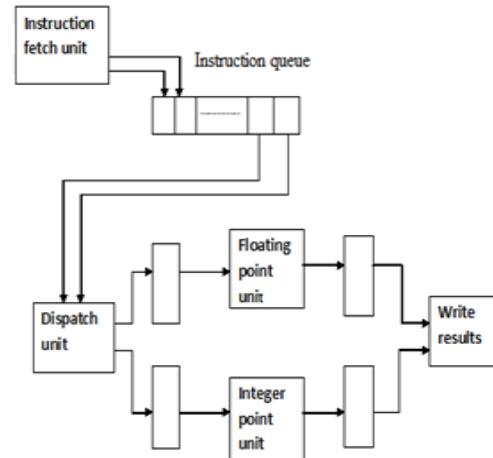


Fig: 4 Superscalar operations

By using separate execution improves the speed of the processor. Calculation of the instruction per cycle throughput (IPC) of superscalar processors [4] is very important to know the speed of the processor. High performance superscalar processors dynamically predict branches and execute instructions along the predicted control flow path [5]. Thus by using this mechanism the speed and the potential of the processor were increased.

4. Nurture IDR Segmentation and Multiple Instruction Queues in Superscalar Pipelining Processor

In Nurture IDR segmentation and multiple instruction queues in superscalar pipelining processor, we have introduced a new unit named as the Identifier unit. When the instruction is fetched from the memory by

The fetch unit it enters into the identifier unit which identifies the type of the instruction and separates the instructions by creating individual segments of executions. Once the instruction type is identified it enters into its corresponding instruction queue then followed by the decoding unit and the executing unit. When the execution is completed it reaches into the write unit.

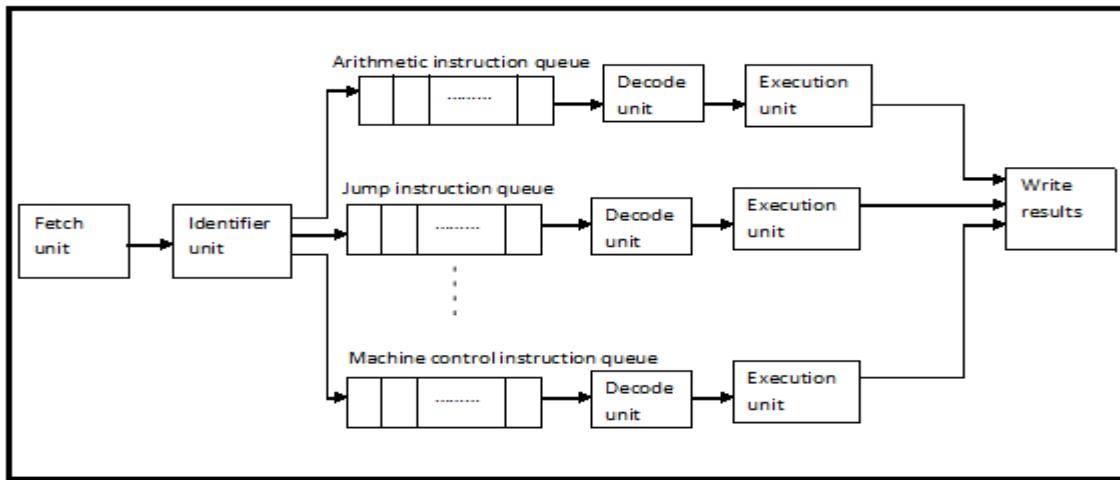


Fig: 5 Proposed Nurture IDR segmentation and multiple instruction queues

Identifier unit can be considered as the heart of the entire Nurture IDR segmentation and multiple instruction queues in superscalar pipelining processor shown in Fig: 5. Because when it separates the instruction and sends it corresponding instruction queue the time consumed is very less and ultimately the speed of the processor increases. While an instruction of one type is being processed, the next instruction is fetched and identified and separated and so on. By doing this the performances of the processor can be definitely improved. In Nurture IDR segmentation and multiple instruction queues, multiple instruction queues are used here to store different types of machine instructions in each queue. So lot of instructions are fetched and identified and stored in the corresponding queues.

The number of queue depends on the number of types of machine instruction. Each queue sends the instruction to the corresponding decode and execution units.

4.1 Fetch unit

Instruction Fetch unit fetch the instruction from the memory [1]. Every fetch operation sends one instruction to the Identifier unit through directly transfer. Only after sending an instruction to the identifier unit the fetch unit will fetch the next instruction. Fetch unit will fetch the instruction one by one from the memory. If pre-fetching is not used, the fetching and the decoding widths will be equal; the instruction fetcher has the responsibility of determining the new instruction counter [6]. Fetch throughput by addressing can be improved by three factors: fetch efficiency, by partitioning the fetch unit among threads; fetch effectiveness, by improving the quality of the instructions fetched; and fetch availability; by eliminating conditions that block the fetch unit [7]. Conventional instruction fetch mechanisms fetch contiguous blocks of instructions in each cycle. They are difficult to scale since taken branches make it hard to increase the size of these

blocks beyond eight instructions. The wider the fetch unit, the more likely it is that fetch slots will be wasted because of discontinuities in the instruction stream [8]. Only when the fetch unit functions without any interruption the instruction will be transferred to next unit through which the speed of the processor can be maintained. Here in this Nurture IDR segmentation and multiple instruction queues concept after fetching the instructions are forwarded to the identifier unit where the type is identified and stored in corresponding instruction queues. If the speed of the fetch unit is high obviously the speed of the processor will also be high.

4.2 Identifier unit

The identifier unit is the main important unit in which major operations takes place. In this proposal we have tried to throw light on the idea of segmenting the machine instructions before decoding by using the Identifier unit. Segmentation takes place on the bases of the types of the instructions.

The list of few instructions types commonly used are shown as follows:

- Arithmetic instructions
- Logical instructions
- Jump instructions
- Data transfer instructions
- Boolean variable instructions
- Machine control instructions

These are most common types of instructions used in the programming. The identifier unit identifies the instructions based on the above types and sends it to the corresponding instruction queue and decode and the execute unit finally to the write unit store the results. The identifier unit plays the important role in Nurture IDR segmentation and multiple instruction queues in superscalar pipelining processor. The process of segmentation takes place in the identifier unit. It must identify the type of the instruction and send in the respective type of the instruction queue by which we are increasing the speed and this type of operation can store a large number of instructions in the instruction queues because there are separate queue for each type of instructions.

4.3 Multiple Instruction Queues

In Nurture IDR segmentation and multiple instruction queues in superscalar pipelining processor we have implemented multiple instruction queues. The Identifier unit identifies the type of the instruction and stores it in the corresponding instruction queue. This give raise the concept of multiple instructions queues in the superscalar pipelining processor so large number of instructions can be stored in the processor at a time. Hence the processor witness high performances rate and an increase in the speed is made possible.

4.4 Decode unit

Decoder unit is used to convert high level language to machine understandable language. Here we have multiple decoders. For each type of instruction there are separate decoding unit. Instructions from the instruction queues are forwarded to the Decode unit based on the instruction types.

The segmentation of the instructions is done before the decoding. The decoder module decodes the VLIW passed by the memory pre fetch module and generates the control signals and register addresses for all the seven instructions forming the Very Long Instruction Word (VLIW) [9]. Thermometer-to-binary decoder can be implemented by various approaches, e.g., a ROM, Wallace-tree (oronescounter), multiplexer-based decoder, fat-tree decoder and logic-based decoder [10]. The main concept here is we are identifying the type of the instruction before the decoding process.

4.5 Execute unit

The execute stage performs ALU operations[11].The complexity of modern processors [12] has made the task of calculating or even bounding the execution time of a sequence of operations very difficult [13].The units in the execution engine are pipelined, though an EU can only issue one instruction per cycle[14].In Nurture IDR segmentation and multiple instruction queues in superscalar pipelining processor we have implemented multiple execution units, which have separate execution unit for each type of the instruction. The execution is done by the ALU individually and finally forwarded to the Write unit. The addition of execution units of the same type in order to leverage instruction level parallelism [15] .

4.6 Write unit

After the execution the instructions the results of the instructions are stored by the write unit. The write unit will store the results in the destination location [1]. Execution unit will send the executed results directly to the Write unit to process the write operation in the specified location. The process taking place in this unit is the final stage in Nurture IDR segmentation and multiple instruction queues in superscalar pipelining processor. This model is the advanced technique to improve the speed of the processor.

5. Conclusion

This paper has ultimately focused on the goal of increasing the speed of the pipelining processor mechanism and segmentation of instructions before decoding. The implementation of the identifier unit is believed to have an optimistic impact on the speed of the processor. The concurrent execution of instruction can have tremendous improvement. The implementation of Nurture IDR segmentation and multiple instruction queues in Superscalar pipelining processor will have a promising enhancement in the speed therefore reducing the execution time of the present pipelining processor. The implementation of multiple instruction queue increase the temporary storage so many instructions can be stored concurrently.

References

- [1] Carl Hamacher, Zvonko Vranesic, Safwat Zaky. Computer Architecture , Mc Graw Hill, 2002.
- [2] J.A. Bergstra and C.A. Middelburg, "Maurer Computers for Pipelined Instruction Processing", 1998.
- [3] Chris Stoltz, Robert Bosch, Pat Hanrahan, and Mendel Rosenblum , "Visualizing Application Behavior on Superscalar Processors", Stanford University.
- [4] Tarek M. Taha D. Scott Wills, "An Instruction Throughput Model of Superscalar Processors", 2003.
- [5] Walid J. Ghandour June, "Dynamic Control Independence Predictor for Speculative Multithreading Processors" , 2009.
- [6] Mojtaba Shojaei, Bahman Javadi*, Mohammad Kazem Akbari, Farnaz Irandejad , "Designing and Optimizing the Fetch Unit for a RISC Core".
- [7] Dean M. Tullsen, Susan J. Eggers, Joel S. Emery, Henry M. Levy, Jack L. Lo, and Rebecca L. Stamm, "Exploiting Choice: Instruction Fetch and Issue on an Implementable Simultaneous Multithreading Processor". Santa Margherita Ligure, Italy, June, 1995.
- [8] Paramjit Oberoi and Gurindar Sohi, "Out-of-Order Instruction Fetch using Multiple Sequencers" ,Computer Sciences Department, University of Wisconsin - Madison.
- [9] R. SESHASAYANAN, DR.S.K. SRIVATSA, "A novel architecture for VLIW processor", 2007.
- [10] E. Sall, M. Vesterbacka, and K. O. Andersson, "A study of digital decoders in flash analog-to-digital converters," in Proceedings of IEEE International Symposium on Circuits and Systems, vol. 1, pp. 129–132, May 2004.
- [11] Ben Lickly, Isaac Liu, Sungjum Kim, Hiren D. Patel, Stephen A. Edwards, Edward A. Lee, "Predictable Programming on a Precision Timed Architecture" ,2008.
- [12] J. L. Henessey and D. J. Patterson, "Computer Architecture: A Quantitative Approach". Morgan Kaufmann Publishers, third edition, 2003.
- [13] C. Ferdinand, R. Heckmann, and et. al. " Reliable and precise WCET determination for a real-life processor", International Conference on Embedded Software Oct. 2001.
- [14] "WiDGET: Wisconsin Decoupled Grid Execution", Tiles June 2010.
- [15] H. Peter Hofstee , " Power Efficient Processor Architecture and The Cell Processor ", Proceedings of the 11th Int'l Symposium on High-Performance Computer Architecture 2005.

J.Nandini Meeraa pursuing bachelor's degree in computer science and engineering third year at Narasu's Sarathy Institute of Technology, Salem. Approved by All India Council for Technical Education, New Delhi (AICTE) and is affiliated to Anna University of Technology, Coimbatore. I am a member of CSI computer society. My current research interest is on creating evolution in the speed of the processor.

N.Indhuja pursuing bachelor's degree in computer science and engineering third year at Narasu's Sarathy Institute of Technology, Salem. Approved by All India Council for Technical Education, New Delhi (AICTE) and is affiliated to Anna University of Technology, Coimbatore. I am a member of CSI computer society. My current research interest is on creating evolution in the speed of the processor.

S.Devi Abirami pursuing bachelor's degree in computer science and engineering third year at Narasu's Sarathy Institute of Technology, Salem. Approved by All India Council for Technical Education, New Delhi (AICTE) and is affiliated to Anna University of Technology, Coimbatore. I am a member of CSI computer society. My current research interest is on creating evolution in the speed of the processor.

K.RathinaKumar, Working as a lecturer at Narasu's Sarathy Institute of Technology, Salem in the Department of Electronics and Communication Engineering. Approved by All India Council for Technical Education, New Delhi (AICTE) and is affiliated to Anna University of Technology, Coimbatore.; published paper in the international journal in

the title of " Efficient method for escalating the performance of programmable router for network on chip by lightweight circuit switched approach" in International Journal of Communication ,Computation and Innovation[IJCSI] of volume 1, issue 2(Jan-July 2011) of ISSN 2229-6808; published paper in the international journal in the title of "Multi Machine power system stabilizer design using particle swarm optimization technique" in International Journal of Communication ,Computation and Innovation[IJCSI] of volume 1, issue 2(Jan-July 2011) of ISSN 2229-6808; Presented a paper in National level conference on significant challenges of smart antennas in ADHOC networks at Idhaya Engineering College on 26th March 2011 conducted by department of CSE, ECE & IT in the title of National Conference on Advanced Computing and Communication Systems; Presented a paper in International level conference on channel noise cancellation using blind adaptive equalization at SSM college of engineering between September 21-23, 2011 conducted by Department of ECE in the title of International Conference on Computer Communication & Signal Processing (IC³SP)-2011;

Design of a High Performance Reversible Multiplier

Md. Belayet Ali¹, Hosna Ara Rahman² and Md. Mizanur Rahman³

¹Department of Computer Science and Engineering,
Mawlana Bhashani Science And Technology University,
Tangail-1902, Bangladesh

²Department of Computer Science and Engineering,
Mawlana Bhashani Science And Technology University,
Tangail-1902, Bangladesh

³Department of Computer Science and Engineering,
Mawlana Bhashani Science And Technology University,
Tangail-1902, Bangladesh

Abstract

Reversible logic circuits are increasingly used in power minimization having applications such as low power CMOS design, optical information processing, DNA computing, bioinformatics, quantum computing and nanotechnology. The problem of minimizing the number of garbage outputs is an important issue in reversible logic design. In this paper we propose a new 4x4 universal reversible logic gate. The proposed reversible gate can be used to synthesize any given Boolean functions. The proposed reversible gate also can be used as a full adder circuit. In this paper we have used Peres gate and the proposed Modified HNG (MHNG) gate to construct the reversible fault tolerant multiplier circuit. We show that the proposed 4x4 reversible multiplier circuit has lower hardware complexity and it is much better and optimized in terms of number of reversible gates and number of garbage outputs with compared to the existing counterparts.

Keywords: Reversible logic circuit, reversible logic gates, reversible multiplier circuits, quantum computing, nanotechnology based systems.

1. Introduction

Irreversible hardware computation results in power dissipation due to information loss. As demonstrated by R.Landauer in the early 1960s, irreversible hardware computation, regardless of its realization technique, results in energy dissipation due to the information loss[1]. According to his research,

the combinational logic circuits dissipate heat in an order of $kT \ln 2$ joules of energy for every bit of information that is erased, where $k=1.3806505 \times 10^{-23} \text{ m}^2 \text{ kg}^{-2} \text{ K}^{-1}$ (joules Kelvin⁻¹) is the Boltzman constant and T is the operating temperature [1]. For room temperature T the amount of dissipating heat is small(i.e. 2.9×10^{-21} joule), but not negligible. The design that does not result in information loss is called reversible. It naturally takes care of heating generated due to the information loss. Such gates or circuits allow the reproduction of the inputs from observed outputs and we can determine the inputs from the outputs [3] [4] [5]. Thus reversibility will become an essential property in future circuit design. Reversible logic has applications in various research areas such as low power CMOS design, optical computing, quantum computing, bioinformatics, thermodynamic technology, DNA computing and nanotechnology.

The difference of reversible logic synthesis compared to binary logic synthesis can be summarized as follows[8]:

- 1) The gates used to implement the circuit have the equal number of inputs and outputs.
- 2) Every output of a gate, which is not used in the circuit, is a garbage signal. A good synthesis method minimizes the number of garbage signals.
- 3) The total number of constants at inputs of the gates is kept as low as possible.

- 4) A gate output can be used only once (the fanout count of each output is equal to one). If two copies of a signal are required, a copying circuit is used.
- 5) The resulting circuit is acyclic.

Any reversible gate performs the permutation of its input patterns only and realizes the functions that are reversible. If a reversible gate has k inputs, and therefore k outputs, then we call it a $k \times k$ reversible gate. Any reversible circuit design includes only the gates that are reversible. A reversible circuit should have the following features [4]:

- 1) Use minimum number of reversible logic gates.
- 2) Use minimum number of garbage outputs.
- 3) Use minimum constant inputs.

Every output of a gate, which is not used for further computations, is a garbage signal[8]. The input that is added to an $n \times k$ function to make it reversible is called constant input [12]. In many computational units multiplication is a heavily used arithmetic operation. For the processors to have high speed multipliers is very important. In this paper, we present a reversible multiplier circuit using reversible MHNG gate. We demonstrate that the proposed reversible multiplier circuit is better than the existing counterparts in terms of number of gates, number of garbage outputs, number of constant inputs and hardware complexity [20] [21] [22] [23] [24].

2. Reversible Logic

The n -input k -output Boolean function $f(x_1, x_2, \dots, x_n)$ (referred to as (n, k) function) is called reversible if:

- 1) The number of outputs is equal to the number of inputs;
- 2) Each input pattern maps to a unique output pattern[13].

In other words, reversible functions are those that perform permutations of the set of input vectors.

3. Reversible Logic Gates

An $n \times n$ reversible logic gate can be represented as:

$$I_v = (I_1, I_2, I_3, \dots, I_N)$$

$$O_v = (O_1, O_2, O_3, \dots, O_N)$$

Where I_v and O_v are input and output vectors respectively. Several reversible logic gates have been proposed in the past few decades. Some of them are: Feynman gate, FG [6], Toffoli gate, TG [7], Fredkin gate, FRG[15], Peres gate, PG [11], New Gate, NG [14], TSG gate, TSG [5], MKG gate, MKG [16] and HNG gate, HNG [18]. In this section we review these reversible logic gates. Some of them are presented to allow for comparison with existing studies.

3.1 Feynman gate (FG)

Feynman gate (FG), also known as controlled-not gate (1-CNOT), is a 2×2 gate that can be described by the equations:

$$I_v = (A, B)$$

$$O_v = (P = B, Q = A \oplus B)$$

where 'A' is control bit and 'B' is the data bit. It is shown in Fig. 1.

3.2 Fredkin gate (FRG)

Fredkin gate (FRG), also known as controlled permutation gate, is a 3×3 reversible logic gate. It can be represented as:

$$I_v = (A, B, C)$$

$$O_v = (P = A, Q = A'B \oplus AC, R = A'C \oplus AB)$$

Where I_v and O_v are input and output vectors. It is shown in Fig. 2. Fredkin Gate is a conservative gate, that is, the Hamming weight of its input vector is the same as the Hamming weight of its output vector.

3.3 Toffoli gate (TG)

Toffoli gate (TG), also known as controlled controlled-not (CCNOT), is a 3×3 reversible logic gate. The Toffoli gate can be represented as:

$$I_v = (A, B, C)$$

$$O_v = (P = A, Q = B, R = AB \oplus C)$$

Where I_v and O_v are input and output vectors. The Toffoli gate is shown in Fig. 3.

3.4 Peres gate (PG)

Peres gate (PG), also known as New Toffoli Gate(NTG), combining Toffoli gate and Feynman gate is a 3x3 reversible logic gate. It can be represented as:

$$I_v = (A, B, C)$$

$$O_v = (P = A, Q = A \oplus B, R = AB \oplus C)$$

Where I_v and O_v are the input and output vectors. The Peres gate is shown in Fig. 4. Peres gate is equal with the transformation produced by a Toffoli Gate followed by a Feynman Gate.

3.5 New gate (NG)

New gate (NG), is a 3x3 reversible gate. It can be represented as:

$$I_v = (A, B, C)$$

$$O_v = (P = A, Q = AB \oplus C, R = A'C' \oplus B')$$

Where I_v and O_v are the input and output vectors. The New gate is shown in Fig. 5.

3.6 TSG gate

TSG gate is a 4x4 reversible gate. The TSG gate is shown in Fig. 6., where each output is annotated with the corresponding logic expression:

$$I_v = (A, B, C, D)$$

$$O_v = (P = A, Q = A'C' \oplus B, R = (A'C' \oplus B') \oplus D,$$

$$S = (A'C' \oplus B') D \oplus (AB \oplus C))$$

3.7 MKG gate

MKG gate is a 4x4 reversible logic gate. The MKG gate can be represented as:

$$I_v = (A, B, C, D)$$

$$O_v = (P = A, Q = C, R = (A'D' \oplus B') \oplus C,$$

$$S = (A'D' \oplus B') \cdot C \oplus (AB \oplus D))$$

Where I_v and O_v are the input and output vectors. The MKG gate is shown in Fig. 7., where each

output is annotated with the corresponding logic expression.

3.8 HNG gate

The HNG gate is universal. The HNG gate is shown in Fig.8., where each output is annotated with the corresponding logic expression. The corresponding logic expression of HNG gate:

$$I_v = (A, B, C, D)$$

$$O_v = (P = A, Q = B, R = A \oplus B \oplus C,$$

$$S = (A \oplus B) \cdot C \oplus AB \oplus D)$$

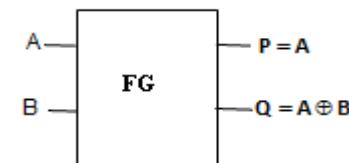


Fig. 1 Feynman Gate

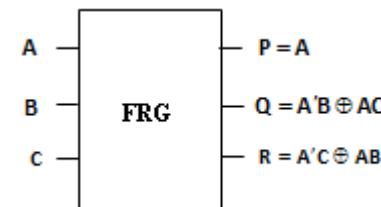


Fig. 2 Fredking gate

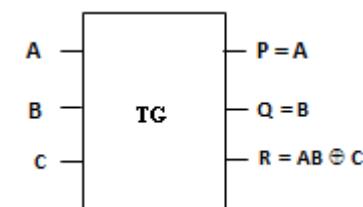


Fig. 3 Toffoli Gate

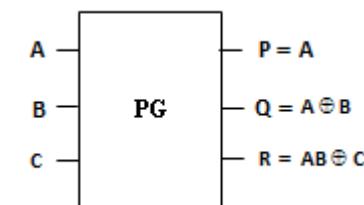


Fig. 4 Peres Gate

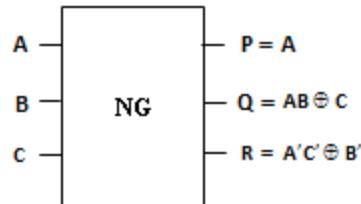


Fig. 5 New Gate

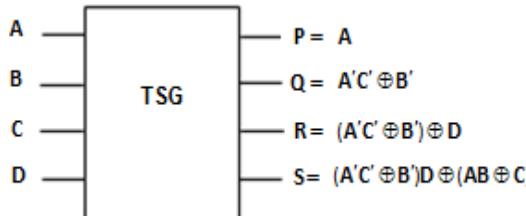


Fig. 6 TSG Gate

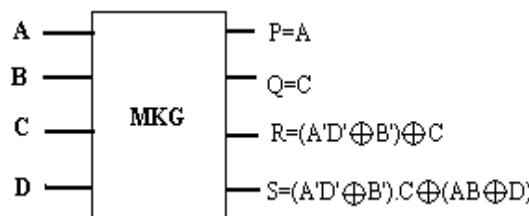


Fig. 7 Reversible MKG Gate

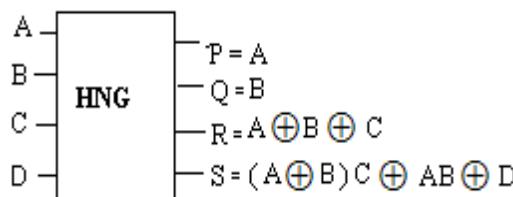


Fig. 8 Reversible HNG Gate

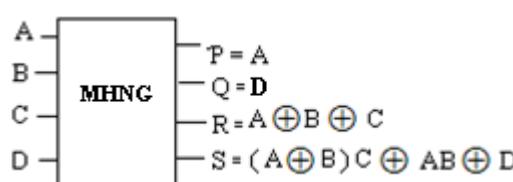


Fig. 9 Reversible MHNG Gate

4. Proposed Modified HNG Gate

Our proposed Modified HNG (MHNG) gate is a 4×4 reversible logic gate. The corresponding logic expression of MHNG gate:

$$\begin{aligned} I_v &= (A, B, C, D) \\ O_v &= (P = A, Q = D, R = A \oplus B \oplus C, \\ &S = (A \oplus B).C \oplus AB \oplus D) \end{aligned}$$

Where, I_v and O_v are the input and output vectors. The MHNG gate is shown in Fig. 9, where each output is annotated with the corresponding logic expression.

The corresponding truth table of the MHNG gate is depicted in Table 1.

Table 1: Truth Table for Proposed Reversible MHNG gate

A	B	C	D	P	Q	R	S
0	0	0	0	0	0	0	0
0	0	0	1	0	1	0	1
0	0	1	0	0	0	1	0
0	0	1	1	0	1	1	1
0	1	0	0	0	0	1	0
0	1	0	1	0	1	1	1
0	1	1	0	0	0	0	1
0	1	1	1	0	1	0	0
1	0	0	0	1	0	1	0
1	0	0	1	1	1	1	1
1	0	1	0	1	0	0	1
1	0	1	1	1	1	0	0
1	1	0	0	1	0	0	1
1	1	0	1	1	1	0	0
1	1	1	0	1	0	1	1
1	1	1	1	1	1	1	0

One of the prominent functionalities of the MHNG gate is that it can work singly as a reversible full adder unit. If $I_v = (A, B, C_{in}, 0)$, then the output vector becomes: $O_v = (P=A, Q=0, R=Sum, S=C_{out})$. Therefore, we have both of the required outputs. Implementation of the MHNG gate as the reversible full adder is shown in Fig. 10. The proposed reversible full adder circuit uses only one reversible logic gate. It produces only two garbage outputs. It requires only one constant input.

Let,

$$\begin{aligned} a &= A \text{ two input EX-OR gate calculation} \\ \beta &= A \text{ two input AND gate calculation} \end{aligned}$$

$$d = A \text{ NOT calculation}$$

$$T = \text{Total logical calculation}$$

For [5]: $T=6a+3\beta+3d$, for [16]: $T=5a+3\beta+3d$, for [18]: $T=5a+2\beta$ and for proposed MHNG gate : $T=5a+2\beta$. Thus, the proposed reversible full adder is better than the reversible full adder circuits in [5,16,18] in term of hardware complexity.

MHNG gate also performs as reversible half adder . If $I_v = (0, B, C_{in}, 0)$, then the output vector becomes: $O_v = (P = 0, Q = 0, R = \text{Sum}, S = C_{out})$. Implementation of the MHNG gate as the reversible half adder is shown in Fig. 11.

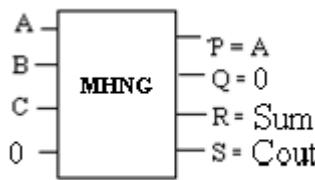


Fig. 10 MHNG Full Adder

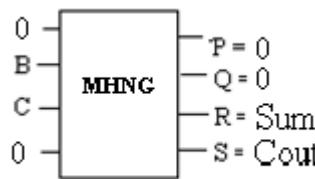


Fig. 11 MHNG half Adder

We will use MHNG gates to construct the novel reversible multiplier circuit.

5. Reversible Multiplier Circuit

The operation of the 4×4 multiplier is depicted in Fig. 12. It consists of 16 partial product bits of the form $x_i \cdot y_i$. The proposed reversible 4×4 multiplier circuit has two parts. First, the partial products are generated in parallel using Toffoli and Peres gates [23]. Then, the addition is performed as shown in Fig. 13 .

The basic cell for such a multiplier is a full adder (FA) accepting three bits. We use MHNG gates as reversible full adder which is depicted in Fig. 10. The proposed reversible multiplier circuit uses eight reversible MHNG full adders. In addition, it needs four reversible half adders. It is possible to use MHNG gate

as half adder, but we use Peres gate as reversible half adder because it has less hardware complexity compared to the MHNG gate.

6. Evaluation of Reversible Multiplier Circuit

The proposed reversible multiplier circuit is more efficient than the existing circuits presented in [19] [20] [21] [22] [23] [24]. Evaluation can be comprehended easily with the help of the comparative results in Table2. One of the main factors of a circuit is its hardware complexity. We can prove that our proposed circuits are better than the existing approaches in term of hardware complexity.

Let,

$$a = A \text{ two input EX-OR gate calculation}$$

$$\beta = A \text{ two input AND gate calculation}$$

$$d = A \text{ NOT calculation}$$

$$T = \text{Total logical calculation}$$

	x_3	x_2	x_1	x_0
x	y_3	y_2	y_1	y_0
	x_3y_0	x_2y_0	x_1y_0	x_0y_0
	x_3y_1	x_2y_1	x_1y_1	x_0y_1
	x_3y_2	x_2y_2	x_1y_2	x_0y_2
	x_3y_3	x_2y_3	x_1y_3	x_0y_3
P_7	P_6	P_5	P_4	P_3
P_2				P_1
P_0				

Fig. 12 Partial products in a 4×4 multiplication

For [19] the Total logical calculation is: $T=80a+100\beta+68d$, for [20] the Total logical calculation is: $T=110a+103\beta+71d$, for [21] the Total logical calculation is: $T=92a+52\beta+36d$,for [22] the Total logical calculation is: $T=80a+36\beta$, for [23] the Total logical calculation is: $T=71a+36\beta$, for [24] the Total logical calculation is: $T=71a+35\beta$, and for our proposed reversible multiplier circuit, the Total logical calculation is: $T=80a+36\beta$. Therefore, the proposed reversible multiplier circuit is better than the existing circuits[19][20][21] in term of complexity.

Our proposed reversible 4×4 multiplier circuit is shown in Fig.13. We used eight reversible MHNG full adders and four reversible half adders.

In the proposed reversible multiplier circuit we use MHNG gate as reversible full adder and it produces

zero (0) in second output, which we have used as the fourth input for the next MHNG full adder circuit. Thus this proposed reversible 4×4 multiplier circuit produces less garbage outputs than the existing counterparts in [19]-[24].

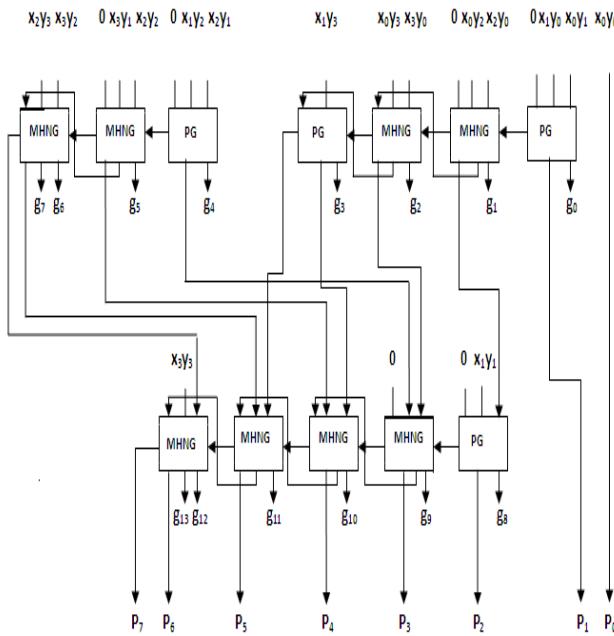


Fig. 13 Proposed 4×4 reversible multiplier circuit using MHNG gates and Peres gates.

Table 2: Comparative Experimental Results of Different Reversible Multiplier Circuits.

Paper No.	No. of Gates	No. of Garbage Outputs	No. of Constant Inputs	Total Logical Calculation
This Work	28	22	28	$80a+36\beta$
[24]	28	32	28	$71a+35\beta$
[23]	28	28	28	$71a+36\beta$
[22]	28	52	28	$80a+36\beta$
[21]	28	56	32	$92a+52\beta+36d$
[20]	29	58	34	$110a+103\beta+71d$
[19]	40	56	31	$80a+100\beta+68d$

7.Results and Discussion

The proposed reversible multiplier circuit is more efficient than the existing circuit presented in [19] – [24]. Evaluation of proposed circuit can be comprehended

easily with the help of the comparative results in Table 2. The difference between partial products generation design with the existing designs in [19, 20, 21, 22, 24] is the use of Toffoli gates and Peres gates. This structure is proposed in [23]. This design method of generating partial product has less hardware complexity. Therefore, the proposed reversible multiplier circuit is better than the existing circuits in terms of complexity.

One of the other major constraints in designing a reversible logic circuit is less number of garbage output, but the design in [19] produces 56 garbage outputs, the design in [20] produces 58 garbage outputs, the design in [21] produces 56 garbage outputs, the design in [22] produces 52 garbage outputs, the design in [23] produces 28 garbage outputs, the design in [24] produces 32 garbage So, we can state that our design approach is better than all the existing counterparts in term of number of garbage outputs.

Number of constant inputs is one of the other main factors in designing a reversible logic circuit. Our proposed reversible multiplier circuit requires 28 constant inputs, but the [19] requires 31 constant inputs, the design in [20] requires 34 constant inputs, the design in [21] requires 32 constant inputs, the design in [22] requires 28 constant inputs, the design in [23] requires 28 constant inputs, the design in [23] requires 28 constant inputs So, we can state that our design approach is better than all the existing designs in term of number of constant inputs.

Comparing our proposed reversible multiplier circuit with the existing circuits in [19]- [24], it is found that the proposed design approach requires 28 reversible logic gates but the existing design in [19] requires 40 reversible gates, the existing design in [20] requires 29 reversible gates, the existing design in [21] requires 28 reversible gates, the existing design in [22] requires 28 reversible gates, the existing design in [23] requires 28 reversible gates and the existing design in [24] requires 28 reversible gates. So, the proposed circuit is better than [19] [20] [21] [22][23][24] in term of number of reversible logic gates, which is one of the other main factors in reversible circuit design.

From the above discussion we can conclude that the proposed reversible multiplier circuit is better than all the existing design.

8. Conclusions

In this paper, we presented a 4x4 bit reversible multiplier circuit using MHNG gates, Toffoli gates and Peres gates. Table II demonstrates that the proposed reversible multiplier circuit is better than the existing designs in terms of hardware complexity, number of logic gates, garbage outputs, constant inputs and optimized in terms of area and delay time. Our proposed reversible multiplier circuit can be applied to the design of complex systems in nanotechnology.

References

- [1] Landauer, R., 1961. Irreversibility and heat generation in the computing process”, IBM J.Research and Development, 5(3) : 183-191.
- [2] C. H. Bennet, “Logical reversibility of computation”, IBM J. Res. Develop., vol. 17, no. 6, pp. 525-532, 1973.
- [3] M. Perkowski, A. Al-Rabadi, P. Kerntopf, A. Buller, M. Chrzanowska-Jeske, A. Mishchenko, M. Azad Khan, A. Coppola, S. Yanushkevich, V. Shmerko and L. Jozwiak, “A general decomposition for reversible logic”, Proc. RM 2001, Starkville, (2001) pp. 119-138.
- [4] Perkowski, M. and P. Kerntopf, 2001. “Reversible Logic. Invited tutorial”, Proc. EURO-MICRO, Sept 2001, Warsaw, Poland.
- [5] Thapliyal Himanshu and M. B. Srinivas, 2005. “Novel reversible TSG gate and its application for designing reversible carry look ahead adder and other adder architectures.” Proceedings of the 10th Asia-Pacific Computer Systems Architecture Conference (ACSAC 05), Lecture Notes of Computer Science, 3740: 775-786. Springer-Verlag.
- [6] Feynman, R., 1985. “Quantum mechanical computers”, Optics News, 11: 11-20.
- [7] Toffoli T., 1980. “Reversible computing”, TechMemo MIT/LCS/TM-151. MIT Lab for Computer Science.
- [8] Alan Mishchenko and Marek Perkowski, “Logic Synthesis of Reversible Wave Cascades”, Portland Quantum Logic Group, Department of Electrical and Computer Engineering, Portland State University,Portland, OR 97207, USA.
- [9] M. Haghparast and K. Navi, “A novel fault tolerant reversible gate for Nanotechnology based systems”, Am. J. of App. Sci., vol. 5, no.5, pp. 519-523,2008.
- [10] M. Haghparast and K. Navi, “Design of a novel fault tolerant reversible full adder for nanotechnology based systems”, World App. Sci. J., vol. 3, no. 1, pp. 114-118,2008.
- [11] Peres, A., 1985. “Reversible logic and quantum computers”, Physical Review: A, 32(6): 3266-3276.
- [12] Saiful Islam, M.D. and M.D. Rafiqul Islam, 2005. “Minimization of reversible adder circuits”. Asian J.Inform. Tech., 4 (12): 1146 1151.
- [13] Dmitri Maslov and Gerhard W. Dueck, “Garbage in Reversible Designs of Multiple Output Functions”, University of New Brunswick, Fredericton, N.B. E3B 5 A3,CANADA.
- [14] Azad Khan, Md.M.H., 2002. “ Design of full adder with reversible gate”, International Conference on Computer and Information Technology, Dhaka, Bangladesh, pp: 515-519.
- [15] E. Fredkin and T. Toffoli, “Conservative logic”, Intl. Journal of Theoretical Physics, pp. 219-253,1982.
- [16] M. Haghparast and K. Navi, “A Novel Reversible Full Adder Circuit for Nanotechnology Based Systems”, Journal of Applied Sciences, 7 (24) (2007) 3995-4000.
- [17] M. Haghparast and K. Navi, “A Novel Fault Tolerant Reversible Gate For Nanotechnology Based Systems”, American Journal of Applied Sciences, 5 (5) (2008) 519-523.
- [18] M. Haghparast and K. Navi, “A Novel reversible BCD adder for nanotechnology based systems”, American Journal of Applied Sciences, 5 (3) (2008) 282-288.
- [19] Thapliyal, H., M.B. Srinivas and H.R. Arabnia, 2005. “A Reversible Version of 4x4 Bit Array Multiplier With Minimum Gates and Garbage Outputs”, The 2005 International Conference on Embedded System and Applications (ESA'05), Las Vegas, USA, pp: 106-114.
- [20] Thapliyal, H. and M.B. Srinivas, 2006. “Novel Reversible Multiplier Architecture Using Reversible TSG gate”. IEEE international Conference on Computer Systems and Applications, pp: 100-103.
- [21] Shams, M., M. Haghparast and K. Navi, 2008. “Novel Reversible Multiplier Circuit in Nanotechnology”. World Appl. Sci. J., 3 (5): 806-810.
- [22] Haghparast.M, jafarali .S, Navi. K, Hashemipour. O,2008, “Design of a Novel reversible Multiplier circuit using HNG gate in Nanotechnology”, World Applied Sciences. J. , 3(6) : 974-978.
- [23] M. Haghparast , M. Mohammadi, K.Navi, M.Eshghi, “Optimized reversible multiplier circuit”, Journal of Circuits, Systems, and Computers, World Scientific Publishing Company.
- [24] Nidhi Syal, Dr. H.P. Sinha, ”High performance reversible parallel multiplier”, International Journal of VLSI & Signal processing applications, Vol.1, Issue 3, (21-26), ISSN 2231-3133.

Biographical notes:

Md. Belayet Ali. received his B.Sc degree in Computer Science and Engineering from Mawlana Bhashani Science and Technology University, Santosh, Tangail, Bangladesh, in 2008. He is working as Lecturer in Department of Computer Science and Engineering at Mawlana Bhashani Science and Technology University (MBSTU), Santosh, Tangail, Bangladesh. He has 1 year experience of working in MBSTU. He has published 03(three) papers in national, international journals. His area of interest includes reversible logic synthesis, multi-valued reversible logic synthesis, network business security, cloud computing.

Hosna Ara Rahman. is currently completing her B.Sc degree in Computer Science and Engineering at Mawlana Bhashani Science and Technology University, Santosh, Tsangail, Bangladesh. Her area of interest include reversible logic synthesis.

Md. Mizanur Rahman. is currently completing his B.Sc degree in Computer Science and Engineering at Mawlana Bhashani Science and Technology University, Santosh, Tsangail, Bangladesh. His area of interest include reversible logic synthesis.

A Block Cipher Using Rotation and Logical XOR Operations

¹D. Sravan Kumar

²CH. Suneetha

³A.Chandrasekhar

¹Reader in Physics, SVLNS Government College, Visakhapatnam, India

²Assistant Professor in Engineering Mathematics, GITAM University, Visakhapatnam, India

³Professor in Engineering Mathematics, GITAM University, Visakhapatnam, India

Abstract

Cryptography is the study of methods of sending messages in disguised form so that only the intended recipients can remove the disguise and read the messages. Information security has become a very critical aspect of modern communication systems. With the global acceptance of the Internet as a medium of communication, virtually every computer in the world is connected to every other. It has created a new risk for the users of the computers with a constant threat of being hacked and being victims of data theft. In this connection data encryption has become an essential part of secure communication of the messages. In the present paper we propose a new method of encryption of data in blocks using the operations Rotation and Logical XOR.

Keywords: Block Matrices, Rotation Operation, Logical XOR Operation, encryption, decryption.

1. Introduction

Encryption is the process of converting a plaintext message into cipher text which can be decoded back into the original message using the secret key. An encryption/decryption algorithm, along with a key is used in the encryption/decryption of data. There are several types of data encryption schemes which form the basis of network security. Encryption schemes are generally based on either block or stream ciphers. Historically the focus of encryption has been on the use of symmetric encryption to provide confidentiality. It is only in the last several decades that other considerations, such as authentication, integrity, digital signature have been included in the theory and practice of cryptology. The security of the message basically depends on two factors 1) confidentiality and 2) Authentication. One of the means of achieving confidentiality of the message is encrypting bulk digital data using block ciphers. Single round of encryption offers

inadequate security but multiple rounds offer increasing security. In the present paper we propose a new method of encryption of data block in 8 rounds using Rotation and Logical XOR operations with a one-time sub key derived for each round from the session key of that particular data block which will be generated from the master key(private key). In the key scheduled algorithm [1, 2] for encryption/decryption of the data proposed in this paper, the size of the data block is selected to be 64 characters. The characters of each data block are coded to 8 bit binary format using ASCII code table and are written as an 8x8 matrix. The binary digits of each element of the message matrix are rotated to new positions so that the outcome is a new element. The number of places an element rotated is different for different elements in each round of encryption i.e. rotation of the digits is not fixed, but depends on the sub keys derived for each round of encryption from the session key of each data block. The session key of each data block is generated [13, 14] from the master key (private key) agreed upon by the communicating parties. Between two successive rotation operations the logical XOR operation is performed on each element of the matrix with its nearest four neighbouring elements so that on completion of eight rounds of rotations and XOR operations good avalanche effect is achieved which is one of the desired properties of encryption algorithm. The procedure designed in this method ensures that the message is highly secure as long as the key selected by the communicating parties is secure. The encryption/decryption procedure further assures relatively low computation overhead. A Logical XOR gate is digital logical gate which performs a logical operation on one or more logic inputs and produces a single logic output. Several researchers of cryptography used the logical operation XOR [7, 8] in their encryption protocols. For describing the algorithm the following notation and definitions are adopted:-

2. Symbols and Notation

Symbol	Expression	Meaning
M,m,n,1	${}^l M_n^m$	The 8x8 matrix whose elements are in the 8 bit binary format where l,m,n take integer values, n $\in\mathbb{N}$, l, m $\in \{0,1,2,\dots,8\}$
K,m,n	K_n^m	The 8x8 matrix whose elements are the decimal digits from 0 to 7
R,E, \wedge	\hat{R}_E	The rotation operator used in the encryption process.
R,D, \wedge	\hat{R}_D	The rotation operator used in the decryption process.
X,E, \wedge	\hat{X}_E	The logical XOR operator used in the encryption process
X,D, \wedge	\hat{X}_D	The XOR logical operator used in the decryption process
S, \wedge	\hat{S}_n^m	The operator used for deriving the sub key for the m th round of encryption from the key used for the first round encryption of n th data block
G, \wedge	\hat{G}_n	The operator used for generating the session key for the encryption of the n th data block from the session key used in the encryption of the (n-1) th data block.
M,l,m,n,i,j	$[{}^l M_n^m]_{ij}$	Represents the element in the i th row and j th column of the matrix $[{}^l M_n^m]$

3. Definitions

1) ${}^l M_n^m$ denotes an 8x8 matrix whose elements are 8 bit binary numbers. The right superscript m denotes the number of rotation operations (\hat{R}_E) performed on the elements of the message matrix M. The left superscript l represents the number of times the logical XOR operator (\hat{X}_E) is applied on the message matrix M. The right

subscript n indicates the number of data block that is being encrypted.

2) K_n^m denotes an 8x8 matrix whose elements are the decimal digits $\in \{0,1,\dots,7\}$ which is called the key matrix used in the rotation operation of the (m+1)th round encryption of the nth data block matrix ${}^m M_n^m$. The right superscript m of K represents the number of round of encryption of the matrix ${}^m M_n^m$, where m ranges from 0 to 7. The right subscript indicates the data block which is being encrypted.

3). The Operator $\hat{R}_E [{}^m M_n^m, K_n^m]$ is the right rotation operator used in the encryption process. It has two arguments ${}^m M_n^m$ and K_n^m . The first argument is the matrix

on which the operation \hat{R}_E is applied. The second argument K_n^m is the key matrix used for performing the

operation \hat{R}_E in the (m+1)th round encryption of nth data block matrix. With this operation the right superscript m of the matrix ${}^m M_n^m$ increases by one unit. With this operation the digits of each element $[{}^m M_n^m]_{ij}$ of the matrix ${}^m M_n^m$ which is in the 8 bit binary format are right rotated $[K_n^m]_{ij}$ places.

Every element $[{}^m M_n^m]_{ij}$ consists of 8 bits

b₇ b₆ b₅ b₄ b₃ b₂ b₁ b₀

$$[{}^m M_n^m]_{ij} = [b_7 \ b_6 \ b_5 \ b_4 \ b_3 \ b_2 \ b_1 \ b_0] \text{ here}$$

$$b_i \in \{0,1\} \text{ and } i \in \{0,1,\dots,7\}$$

The binary digits b₇.....b₀ are right rotated the number of places equal to $[K_n^m]_{ij}$

Which results in the matrix ${}^m M_n^{m+1}$.

$$\hat{R}_E [{}^m M_n^m, K_n^m] = {}^m M_n^{m+1}$$

$$\left[K_n^m \right]_{ij} = p \text{ say}$$

Let the right rotation mapping by p times be denoted by

$\overset{\Lambda}{R}_E(p)$. Then the right rotation p times corresponds to the mapping

$$\overset{\Lambda}{R}_E(p) : b_j \rightarrow b_{[(j-p) \bmod 8]}$$

Example let M_{ij} is right rotated three times then

$$b_7 \rightarrow b_{(7-3) \bmod 8} = b_4$$

$$b_6 \rightarrow b_{(6-3) \bmod 8} = b_3$$

.....

.....

$$b_1 \rightarrow b_{(1-3) \bmod 8} = b_6$$

$$b_0 \rightarrow b_{(0-3) \bmod 8} = b_5$$

$$\overset{\Lambda}{X}_E [{}^m M_n^{m+1}] \rightarrow (((({}^m M_n^{m+1})_{ij} \text{XOR} {}^m M_n^{m+1}_{i-1,j})) \text{XOR} {}^m M_n^{m+1}_{i+1,j}) \text{XOR} {}^m M_n^{m+1}_{i+1,j+1}) \text{XOR} {}^m M_n^{m+1}_{i-1,j-1})$$

$$\overset{\Lambda}{X}_E [{}^m M_n^{m+1}] = {}^{m+1} M_n^{m+1}$$

Example

$$\left[{}^m M_n^{m+1} \right]_{24} = 10010100, \left[{}^m M_n^{m+1} \right]_{23} = 10100010$$

$$\left[{}^m M_n^{m+1} \right]_{34} = 00111000, \left[{}^m M_n^{m+1} \right]_{25} = 10010010,$$

$$\left[{}^m M_n^{m+1} \right]_{14} = 01100101 \text{ then}$$

$$\overset{\Lambda}{X}_E [{}^m M_n^{m+1}]_{24} = (((((10010100 \text{XOR} 10100010) \text{XOR} 00111000) \text{XOR} 10010010) \text{XOR} 01100101))$$

$$\left[{}^{m+1} M_n^{m+1} \right]_{24} = 11111001$$

5). $\overset{\Lambda}{X}_D [{}^m M_n^m]$ represents the XORing of each element of the matrix ${}^m M_n^m$ with the surrounding elements in the decryption process. With this operation the left superscript m decreases by one unit.

$$\overset{\Lambda}{X}_D [{}^m M_n^m] \rightarrow (((({}^m M_n^m)_{ij} \text{XOR} {}^m M_n^m_{i-1,j})) \text{XOR} {}^m M_n^m_{i,j+1}) \text{XOR} {}^m M_n^m_{i+1,j})$$

$$\cdot \overset{\Lambda}{X}_D [{}^m M_n^m] = {}^{m-1} M_n^m \text{ where } m \text{ ranges from 0 to 7.}$$

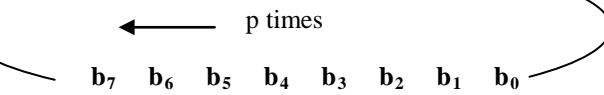
Example

$$\left[{}^m M_n^m \right]_{45} = 11111001 \quad \left[{}^m M_n^m \right]_{35} = 10100010$$

$$\left[{}^m M_n^m \right]_{46} = 00111000 \quad \left[{}^m M_n^m \right]_{55} = 10010010$$

$$\left[{}^m M_n^m \right]_{44} = 01100101 \text{ Then}$$

$$\overset{\Lambda}{X}_D [{}^m M_n^m]_{45} = (((((11111001 \text{XOR} 01100101) \text{XOR} 10010010) \text{XOR} 00111000) \text{XOR} 10100010))$$



Example:

Let $\left[{}^m M_n^m \right]_{24} = 10100100$ and $\left[K_n^m \right]_{24} = 3$ then

$$\left[{}^m M_n^{m+1} \right]_{24} = 10010100$$

4). $\overset{\Lambda}{X}_E [{}^m M_n^{m+1}]$ represents the XORing of each element of the matrix ${}^m M_n^{m+1}$ with the nearest four neighboring elements in the encryption process. With this operation the left superscript m increases by one unit. With this operation each element $\left[{}^m M_n^{m+1} \right]_{ij}$ of the matrix ${}^m M_n^{m+1}$

which is in the 8 bit binary format is XORed with the nearest four neighboring elements which results in the matrix ${}^{m+1} M_n^{m+1}$

$$\left[{}^{m-1}M_n^m \right]_{45} = 10010100$$

6). The Operator $\overset{\Lambda}{R}_D \left[{}^{m-1}M_n^m, K_n^m \right]$ is the left rotation operator used in the decryption process. It has two arguments ${}^{m-1}M_n^m$ and K_n^m . The first argument is the

matrix on which the operation $\overset{\Lambda}{R}_D$ is performed in the $(9-m)^{th}$ round decryption of the nth data block matrix. The second argument K_n^m is the key matrix used for

performing the operation $\overset{\Lambda}{R}_D$ in the m^{th} round decryption of the n^{th} data block matrix. With this operation the right superscript m of the matrix ${}^{m-1}M_n^m$ decreases by one unit. With this operation the digits of each element

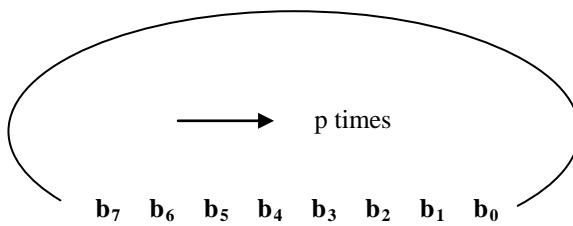
$$\left[{}^{m-1}M_n^m \right]_{ij}$$

is in the 8 bit binary format are right rotated $\left[K_n^m \right]_{ij}$ places.

i.e, every element $\left[{}^{m-1}M_n^m \right]_{ij}$ consists of 8 bits

$b_7 \quad b_6 \quad b_5 \quad b_4 \quad b_3 \quad b_2 \quad b_1 \quad b_0$

$$\left[{}^{m-1}M_n^m \right]_{ij} = [b_7 \ b_6 \ b_5 \ b_4 \ b_3 \ b_2 \ b_1 \ b_0]$$



the binary digits b_7, \dots, b_0 are left rotated the number of times equal to $\left[K_n^m \right]_{ij}$

which results in the matrix ${}^{m-1}M_n^{m-1}$.

$$\overset{\Lambda}{R}_D \left[{}^{m-1}M_n^m, K_n^m \right] = {}^{m-1}M_n^{m-1}$$

7). $\overset{\Lambda}{S}_n^m$ is the operator used for deriving the sub key for the $(m+1)^{th}$ round encryption of nth data block from the session key used for the first round operation i.e.

$$\overset{\Lambda}{S}_n^m \left[K_n^0 \right] = K_n^{m-1}$$

The key matrix K_n^{m-1} for the m^{th} round encryption of the nth data block is obtained from K_n^0 by shifting the columns of the matrix K_n^0 to the right by $(m-1)$ places.

$$\text{i.e. } \left[K_n^{m-1} \right]_{ij} = \left[K_n^0 \right]_{ij-m+1}$$

8). $\overset{\Lambda}{G}_n$ is the operator which defines the session key generation for the first round encryption of the nth data block from the session key used for the first round encryption of the $(n-1)^{th}$ data block.

$$\overset{\Lambda}{G}_n \left[K_{n-1}^0 \right] = K_n^0, \text{ where } \left[K_n^0 \right]_{ij} = \left[(K_{n-1}^0)_{ij} + (K_{n-1}^0)_{ij+1} \right]_{\text{mod}8}$$

i and j take values in all the definitions made above from 0 to 7. If $i+1$ or $i-1$ or $j+1$ or $j-1$ or any subscript fall out of the range {0,1,2,...,7} then modulo 8 of that number be considered. The session key of each data block is itself sub key for the first round of encryption of the data block.

Before communicating the messages both the sender and the receiver agree upon to use the secret key which is in the form of an 8x8 matrix K (master key) whose elements are the decimal digits from 0 to 7. This matrix K (master key) is denoted by K_1^0 in the encryption/decryption process i.e. the master key itself is the session key for the encryption/decryption of first data block. For implementing the algorithm the entire message is divided into data blocks of 64 characters each $D_1, D_2, D_3, \dots, D_n$ where n is a natural number. The characters in each message block are coded to 8 bit binary numbers using ASCII code table and are arranged in the form of 8x8 matrices ${}^0M_1^0, {}^0M_2^0, {}^0M_3^0, \dots, {}^0M_n^0$ row wise. The number of characters in the message always may not be the integral multiple of 64. Hence, at the end of the message the sender adds three # characters (###) and ensures that the message fills integer number of text blocks by adding random different characters after the three # characters.

4. Algorithm

4.1 Encryption-

K = Key agreed upon by the communicating parties

Set n = 1

m = 1

Step 1:- PRINT DATA BLOCK = n

If $n = 1$ $K_n^0 = K$
 else
 $\hat{G}_n[K_{n-1}^0] = K_n^0$, where $[K_n^0]_{ij} = [(K_{n-1}^0)_{ij} + (K_{n-1}^0)_{ij+1}]_{\text{mod } 8}$

Step2:- PRINT ENCRYPTION ROUND = m
 Step3:- $S_n^m[K_n^0] = K_n^{m-1}$
 where $[K_n^{m-1}]_{ij} = [K_n^0]_{ij-m+1}$

Step4:- $\hat{R}_E[m^{-1}M_n^{m-1}, K_n^{m-1}] = {}^{m-1}M_n^m$

Step5:- $\hat{X}_E[{}^{m-1}M_n^m] = {}^mM_n^m$

Step6:- If $m < 8$, increment m by one unit and go to Step 2
 Else set m = 1
 If $n < N$ increment n by one unit and go to Step 1.
 Else Stop

After 8 rounds of encryption using rotation operation and logical XOR operation the encrypted message matrices ${}^8M_1^8, {}^8M_2^8, {}^8M_3^8, \dots, {}^8M_n^8$ are obtained. Then all the elements of the matrices which are in 8 bit binary format are converted into equivalent text characters using ASCII code table to get the cipher data blocks $D_1^E, D_2^E, D_3^E, \dots, D_n^E$. This cipher text is communicated to the receiver through the public channel.

4.2 Decryption

The receiver after receiving the cipher text divides the message into data blocks $D_1^E, D_2^E, D_3^E, \dots, D_n^E$ of 64 characters each. All the 64 characters of each data block are converted into 8 bit binary numbers using ASCII code table and all the elements are written as 8x8 matrices ${}^8M_1^8, {}^8M_2^8, {}^8M_3^8, \dots, {}^8M_n^8$

K = Key agreed upon by the communicating parties

Set n = 1
 m = 8
 Step 1:- PRINT DATA BLOCK = n
 If $n = 1$ $K_n^0 = K$
 Else
 $\hat{G}_n[K_{n-1}^0] = K_n^0$, where $[K_n^0]_{ij} = [(K_{n-1}^0)_{ij} + (K_{n-1}^0)_{ij+1}]_{\text{mod } 8}$

Step2:- PRINT DECRYPTION ROUND = 9-m
 Step3:- $S_n^m[K_n^0] = K_n^{m-1}$,

Where $[K_n^{m-1}]_{ij} = [K_n^0]_{ij-m+1}$

Step4:- $\hat{X}_D[{}^mM_n^m] = {}^{m-1}M_n^m$

Step5:- $\hat{R}_D[{}^{m-1}M_n^m, K_n^{m-1}] = {}^{m-1}M_n^{m-1}$

Step6:- If $m > 0$, decrement m by one unit and go to Step 2
 Else set m = 8
 If $n < N$ increment n by one unit and go to Step 1
 Else Stop

After 8 rounds of decryption using logical XOR operation and rotation operation the original message matrices ${}^0M_1^0, {}^0M_2^0, {}^0M_3^0, \dots, {}^0M_n^0$ are obtained. Then all the elements of the matrices ${}^0M_1^0, {}^0M_2^0, {}^0M_3^0, \dots, {}^0M_n^0$ which are in 8 bit binary format are converted into text characters using ASCII code table to get the original message blocks $D_1, D_2, D_3, \dots, D_n$

5. Security Analysis

In the key scheduled algorithm proposed here different keys are used for encrypting different data blocks which are called session keys generated from the master key (secret key between the sender and the receiver) and the key used for the encryption of each round is different and is derived from the session key of the corresponding round which is called the sub key. As different keys are used for different data blocks cipher is less vulnerable to passive attacks. As each element of the message matrix M is rotated (not fixed rotation) and logical XOR operation is performed with its all nearest neighboring elements the same characters in the plain text space are mapped to different characters of the cipher text space even though they are in the same text block or different text blocks. So, cipher text is not easily amenable to cryptanalysis [6, 7]. Even the change of a single element of the message matrix changes almost the entire cipher block matrix, i.e., to say that the proposed algorithm has achieved a good avalanche effect [4, 5] which is one of the desired qualities of a good encryption algorithm.

If the same message is sent in I and II (or any subsequent) data block, they are mapped to different cipher texts, i.e., even if the same message is sent repeatedly in the same message block, the messages are enciphered to different cipher texts. Hence, active attacks such as chosen plain text attacks [6, 15], chosen cipher text attacks [9, 10, 16] are quite difficult to execute. Hence, the proposed algorithm is less vulnerable to active attacks. The present encryption algorithm is at most secure against man-in-middle attack [3, 11, 12] because the entire master key is

agreed upon by the sender and the receiver rather than the electronic exchange of the parts of the key.

The proposed key scheduled algorithm in this paper is less prone to timing attacks because the time required to encipher or decipher a data block is same for all data blocks since time for enciphering or deciphering is independent of characters in the data block. Even though the original message contains less than 64 characters the remaining characters are filled at random, so that each data block contains exactly 64 characters.

The size of the key is 64 decimal digits where each decimal digit takes values from 0 to 7. Hence 64^8 different keys are possible. It is estimated that on a 4GHz single core processor the time required to encipher/decipher a text block is 18 μ sec. Hence, the vulnerability to brute force attack is very less [the life time of a human being i.e., 100years is approximately equal to 3Gsec, the time required to try all possible keys to decipher a single cipher block by brute force method is roughly 5Gsec] It is estimated that the time required to encipher a text book containing 500 pages, each page having 40 lines and each lines having 40 characters is 7½minutes.

6. References

- [1] O. Acricmez, C.K. Koc, J.P. Seifert “Predicting secret keys via branch prediction” in Proc. RSA(CT-RSA)2007, Leterure Notes in Computer Sciences, Vol.4377 (Springer Berline, 2007) pp.225-242.
- [2] Adams C. “Simple and Effective key scheduling for symmetric ciphers” proceedings, workshop in selected areas of cryptography SAC’94,1994.
- [3] Anna M. Johnston, Peter S. Gemmell, “Authenticated key exchange Provably Secure Against the Man-in-Middle Attack”, Journal of Cryptology (2002) Vol. 15 Number 2 pages 139-148.
- [4] Carlisle Aams and Stafford Tavares, “The Structured Design of Cryptographically good s-boxes, Journal of Cryptology, 1990, Vol.3, No.1, Pages 27-41.
- [5] Eli Biham, “Cryptanalysis of multiples modes of operation”, Journal of Cryptology, 1998, Vol.11, No.1,pgs 45-58.
- [6] Denis X Charles, Kristin E. Lauter and Eyal Z. Goren, “Cryptographic Hash Functions from Expander Graphers”, Journal of Cryptology (2009), Vol. 22 Number 1 pages 93-113.
- [7] Ivan B. Damgard and Lars R. Kudsen, “Two-key Triple Encryption” Journal of Cryptology(1998), Vol. 11, Number 3, pages 209-218.
- [8] John Blaack and Phillip Rogaway “ CBC MACs for Arbitrary-Length Messages: The Three-key constructions” Journal of cryptology(2005) Vol. 18 pages 111-131.
- [9] J. Kelesey B. Schneir and D. Wagner, “key-schedule cryptanalysis of IDEA,G-DEA,GOST,SAFER and

triple-DES. In N. Koblitz editor”, Advances in Cryptology-Proc.CRYPTO’96, LNCS 1109, pages 237- 251.Springer-Verlag Berlin 1996.

- [10] Lorenz Minder and Alistair Sinclair, “The Extended k-tree Algorithm” Journal of cryptologyDOI:10.1007/s00145-011-9097-y.
- [11] Luke O’Connor, “An Analysis of a class of algorithms for s-box construction”, Journal o Cryptology(1994) 133-151.
- [12] Marc Fischlin and Roger Fischlin, “Efficient Non-Malleable Commitment Schemes”, Journal of Cryptology, 2011, Vol. 24, Number 1, pages203-244.
- [13] Minh-Huyen Nguyen and Salil Vadhan, “Simple Session Key-generation from the short random passwords”, Journal of Cryptology(2008), Vol. 21,Number 1, pages52-96.
- [14] Moses Loskov, Ronald L. Rivest and David Wagner, “Tweakable Block Ciphers”, Journal of Cryptology(2011), Vol.24, Number 3, pages 588- 613.
- [15] Sachar Paulus and Tsuyoshi Takagi, “A New Public-key Cryptosystem over a Quadratic Order with Quadratic Decryption Time”, Journal of Cryptology(2000), Vol 13,Number 2, pages 263-272.
- [16] Victor Shoup and Rosario Gennaro “Securing Threshold Cryptosystems against Chosen Cipher text Attack, Journal of Cryptology(2002) Vol15,Number2 pages75-96.

Dr. D. Sravana Kumar is a senior faculty in Physics in Government College, Visakhapatnam. He obtained his doctorate in Ultrasonics. His research interest includes Ultrasonics, Molecular Interactions and Cryptography. He has published 12 research papers in various international journals those published by Springer, Elsevier, etc. He is an ex-scientist in the Department of Atomic Energy, Government of India. He is very much interested in interdisciplinary research.

CH. Suneetha is Assistant Professor in Engineering Mathematics in GITAM University, Visakhapatnam. She obtained her master’s degree in applied mathematics, M.Phil. in algebra. At present she is pursuing her Ph.D under the guidance of Dr. A. Chandrasekhar. Her research interests include Cryptography, Linear Transformations and Algebraic Curves.

Dr. A. Chandrasekhar is Professor and Head of the Department of Engineering Mathematics in GITAM University, Visakhapatnam. He obtained his doctorate in Cryptography. His research interest includes Cryptography and Fixed Point Theory. He has published 14 research papers in various reputed national and international journals including IEEE.

Framework for Ethernet Network Functionality Testing

Mirza Aamir Mehmood¹, Ahthasham Sajid² and Amir Shahzad Khokhar³

Department of Computer Science, Balochistan University of Information Technology, Engineering & Management Sciences
Quetta, Pakistan

Abstract

Computer networks and telecommunication systems use a wide range of applications. Therefore, the power and complexity of computer networks are increasing every day which enhances the possibilities of the end user, but also makes harder the work of those who have to design, maintain and make a network efficient, optimized and secure. Ethernet functionality testing as a generic term used for checking connectivity, throughput and capability to transfer packets over the network. Especially in the packet-switch environment, Ethernet testing has become an essential part for deploying a reliable network. A platform and vendor independent framework is required to verify and test the functionality of the Ethernet network and to verify the functionality and performance of the TCP/IP stack. "NetBurst" is developed for Ethernet functionality testing.

Keywords: *Communication, Networking, Performance, Load, Portability, Throughput, TCP, IP, Protocol, Ethernet, Testing.*

1. Introduction

In the late 1990s, the spectacular growth of the Internet dramatically affected the evolution of computer networking. Numerous network techniques and technologies boomed, but quickly faded into oblivion. Others have stood the test of time. This growth is not possible without protocols and communications software. As a result, the end user is greatly empowered with numerous choices available to them. Contrary to this, the large number of transmission techniques and communication protocols had made it very difficult for the network engineers, who have to design, develop and maintain the performance and optimization of their network. Thus, with the power and complexity of computer networks raised, the need for tools to measure, analyze and test the functionality and performance of the network is greatly intensified.

2. Related Work

Number of researchers has worked on Ethernet and protocol testing. This section provides an insight of work done in relevant area.

There are two main methods to capture data from a network; the first is based on the use of dedicated hardware, while the second makes use of the hardware of a normal PC or workstation connected to the communication channel. In the second method, the network adapter of the computer is used to obtain the frames from the network, and the software performs packet capturing process. The software solution has usually the low performance as compare to hardware solution, particularly on slow machines, but it is cheaper, and easier to modify and upgrade. For this reason, it is widely adopted on the most used network architectures, where the performance of dedicated hardware is not needed. In network performance analyzing or Ethernet testing, ranges of different software and hardware tools and products are available for packet capturing, packet filtering and network monitoring [4, 5].

Following are the software and hardware products for capturing and analyzing network traffic.

2.1 Tcpdump

Tcpdump is a simple packet sniffer that uses command line interface, and allows a user to intercept and display TCP/IP and other packets being transmitted or received over a network. It is freeware software which works on most Unix-like operating systems (Linux, Solaris, BSD, Mac OS X, and HP-UX) and there is also a port of tcpdump for Windows called WinDump [6].

2.2 Wireshark

Wireshark is similar to tcpdump, but it provides graphical user interface (GUI) and many other options as well, like packet filtering and sorting. It is a free packet sniffing application. It works for different protocols with deep inspection of packets and uses decryption options for many protocols. It can run on Windows, Linux, Solaris, Mac OS and BSD [25]. Wireshark can be used for data analysis captured by NetBurst.

2.3 Snoop

Snoop is a freeware command line packet sniffer included as part of Sun Microsystems' Solaris Operating System. It is dedicated for the Solaris system, on a data-link or IP interface. Snoop captures packets and displays their contents. If the data-link or IP interface is not specified, snoop picks the first non-loopback data-link it finds [7].

2.4 TPTEST

TPTEST is another application used for measuring performance on Internet connection; TPTEST measures the throughput speed from various reference servers on the Internet. TPTEST measures the throughput of TCP/UDP incoming and outgoing packets and packet loss. The application is written in C++ and is portable for Windows, Mac OS, Linux and BSD [8].

2.5 Maxwell Network Emulator

Maxwell network emulator is a hardware appliance that helps network managers, software developers and testers learn how their products will perform in real-world production networks, including satellites and the Internet. It has both graphical and command line interface and also script driven interface for maximum flexibility. Moreover, it is fully customizable and programmable using C++ [9].

2.6 IxANVL™ - Automated Network Validation Library (ANVL)

Ixia's IxANVL (Automated Network Validation Library or ANVL) is another hardware device that is also known as the industry standard for automated network/protocol validation. Software developers and manufacturers of networking equipment and Internet devices use IxANVL to validate protocol compliance and interoperability. [10].

There are many software applications for packet sniffing for example TCP Dump, Wireshark but they can only analyze traffic. They do not offer any option to inject customized packets. Applications that can test TCP for example TTCP [22], can send TCP packets, but they do not support custom

packet creation. Applications like Snoop are platform dependent and can run on a specific platform.

The research work done so far also focused on certain environments or functionalities. Not a single research work provides us all the required functionalities that are sufficient to test an Ethernet node in our desired way. Since existing applications and research work do not fulfill all the requirements, a new application has to be developed that can fulfill all the requirements. We have adopted the technique that Douglas E. Comer and John C. Lin have demonstrated [21], i.e. active probing technique, to develop application named "NetBurst".

3. Application Development Environment

The application has been developed in ANSI C using Eclipse IDE on Fedora core 10. External libraries used are Libnet and Libpcap on Linux where, as on the Windows platform, Winpcap and Libnet for Win32 are used, which are ported versions of early stated libraries.

3.1 Libnet

Libnet is a library for C programming used for packet construction and injection. Libnet is used to control every field of every header of every packet; large number of programs goes through a high-level interface in order to send traffic on the network. Occasionally, for security or hacking reasons, a program needs to construct its own network headers. The existing TCP/IP stack is unable to build these headers, and it must bypass it and go directly to the hardware drivers. Libnet is a library that makes custom packet generation easier [23].

3.2 Libpcap

Libpcap gives a portable structure for low level network monitoring. Libpcap can provide network statistics collection, security monitoring and network debugging. Every system vendor provides a different interface for packet capture. To overcome this problem, Libpcap was developed. This system-independent API makes it easier to port and alleviates the need for several system-dependent packet capture modules in each application [6].

3.3 Eclipse

Eclipse is an open source, integrated development environment. It comprises extensible frameworks, tools and runtimes for building, deploying and managing software. It provides plugins in order to provide all of its functionality on top of the runtime system; this is in contrast to some other applications where functionality is typically hard coded. This plug-in mechanism is a lightweight software component framework. Moreover, Eclipse allows extension using other programming languages, such as C and Python.

This plug in setup allows Eclipse to work with networking applications such as Telnet, and database management systems. The plug-in architecture supports writing any desired extension to the environment, such as for configuration management [24].

4. NetBurst Architecture

NetBurst is developed for Ethernet functionality testing. It is a vendor independent, cross platform application. It can be used to test load, performance and functionality of underlying TCP/IP stack. Functionality testing includes throughput, packet loss, latency, fragmentation, option processing etc.

4.1 Application Overview

The following figure provides an overall working model of NetBurst. The application works by sending customized packets, capturing and observing the response by sniffing the packets. The packet generator is sending packets to the machine whose protocol stack needs to be tested. Sniffer is running on machine under test (MUT) to capture and observe the response of MUT's protocol stack.

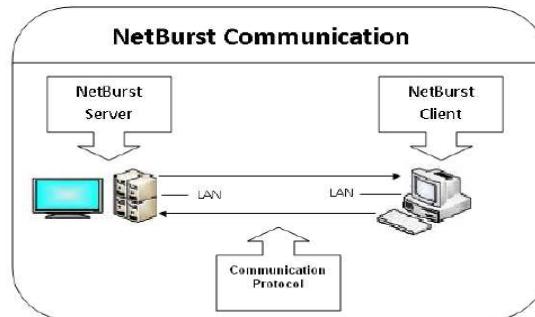


Figure 1. NetBurst Communication

Communicating nodes are connected in LAN and protocol being used is TCP/IP. Supported protocols are Internet Protocol and Transmission Control Protocol. NetBurst supports only Ethernet technology. An error will be generated if any other technology is being used.

4.2 Communication Mode

The application can be executed in two communication modes: synchronous and asynchronous. When executed in synchronous mode, the client is running on MUT, and the server is sending TCP or UDP packets. This mode is recommended only for test case designing, since each and every packet is processed to extract fields in the packet header. All incoming traffic displayed on the terminal. Since all incoming packets are processed, it may result in slower packet capturing and may lost many correctly received packets. Due to these reasons, this mode may produce false results. When executed in asynchronous mode, Packetizer

and sniffer run independently. All captured packets are stored in a file for future processing.

4.3 Application Architecture

NetBurst has two main components: Packetizer and Sniffer. Packetizer has been built over Libnet and is controlled through configuration files. Sniffer is built over pcap and Winpcap libraries and is also controlled through configuration files. These configuration files are used to control the packet transmission rate, packet construction, and various other properties.

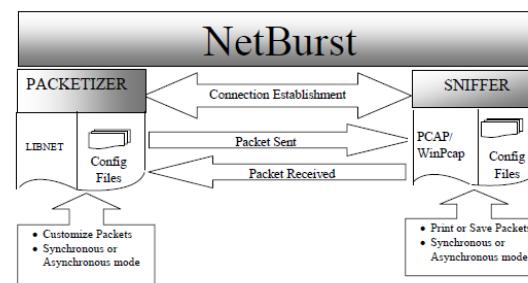


Figure 2. Application Architecture

As shown in the above figure, Packetizer can transmit random or fixed length packets. On the other hand, Sniffer records all incoming traffic and the response generated by MUT. The above figure depicts the application's execution mode.

4.3.1 NetBurst Configuration File

A configuration file is designed to control the application behavior. Following is an example of a configuration file containing different parameters. A user can interact with the application through this file to control application behavior. The structure of the configuration file is given in following figure.

```
time: 2
interval: 1
count: 3
ip_ver: 4
random: 1
cache: 0
protocol: 1
```

Figure 3. Configuration File

4.3.2 NetBurst Packetizer

This component is responsible for constructing and injecting packets. To inject a packet, NetBurst reads the packet constructed that is values of different fields of packet header from a specific file for each protocol header. The structure and purpose of these files are elaborated below.

4.3.3 TCP Configuration file

The following diagram depicts *tcp.opt* file which is used in Packetizer. The record parameter controls number of packet constructs provided in the file. The value of the source port and destination port is given in *source_port* and *destination_port* fields. The *window_size* takes the value for the window size, whereas *flag_urg* is used to take the value of the urgent flag which could be any numeric value. Sequence and acknowledge values are provided through *seq* and *ackg*, respectively. The user also has to provide the control flags (for example *th_syn*, TCP use this flag in its header to initiate communication) and this is done through providing the value in the *control_flag* parameter

```
record: 1
source_port: 1234
destination_port: 9876
window_size: 77
flag_urg: 9
seq: 11
ackg: 8
```

Figure 4. Configuration File

4.3.4 IP Configuration File

The structure of the *ipv4.opt* file is depicted in figure 5. The application uses this file to construct a custom IPv4 packet. In *id*, the value of ID field of IP header is provided. The user provides the source and destination address in *source_ip* and *dest_ip* respectively. *Time_to_liv* takes the value for time to live, whereas *tos* takes value of type of service field of IPv4 header

```
record: 1
id: 2
source_ip: 192.168.1.1
dest_ip: 192.168.1.11
time_to_liv: 12
tos: iptos_lowdelay |
     iptos_throughput |
     iptos_reliability |
     iptos_mincost!
```

Figure 5. IP Configuration File

5. Results

In this section, test cases and results are discussed to analyze the TCP functionality. Test cases are categorized in four groups: performance, load, socket settings and TCP functionality. For traffic analysis, Wireshark is used, which is standard traffic analysis tool.

5.1 Performance Testing

The performance of any communication system can be measured based on three main factors i.e. throughput, latency and packet loss. To analyze and measure these factors, a set of test cases are executed through NetBurster, and the performance of the underlying TCP Stack is observed. Following are the results of these test cases.

5.1.1 Test Case I: Throughput

To verify throughput, a burst of packets having packet size 1500 bytes is fired for 60 seconds. The following graph depicts the throughput.

Throughput is defined as number of bytes received divided by the transmission time. The figure 6 shows a normal test with no throughput problem. It is observed that the level of "fuzziness" of the throughput speed distribution is normal and may reflect a slight timing inaccuracy in the Ethernet Network Functionality Testing computer's clock. On analyzing the graph, a small gap is observed, which indicates a packet delay after a retransmission.

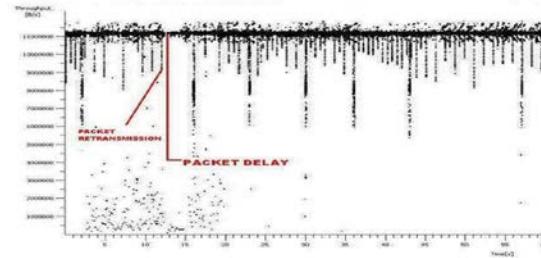


Figure 6. TCP Packet Throughput Graph

5.1.2 Test Case II: Latency

In continuation to test case I, a latency graph is also produced which is given below. In figure 7, packet number is given on x-axis and time (in seconds) given on y-axis

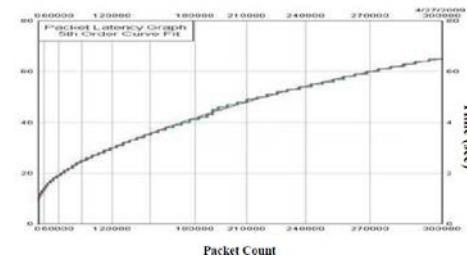


Figure 7. Packet Latency

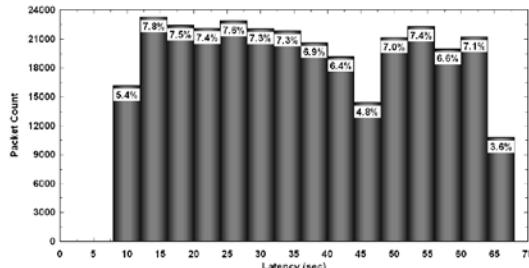


Figure 8. Packet Latency Histogram

The above graph shows latency during a burst of packet transmitted from NetBurst. We can deduce from the figure 7 that the latency gradually increases with the time. Likewise, in figure 8, latency histogram depicts the latency interval. It also shows that 7.8% packets were affected with highest latency. On the other hand, 3.6% of packets were affected by lowest packet latency, which occurred at the last packet transmitted. The variation of delay, known as jitter, occurred during transmission as depicted in figure 8.

5.1.3 Test Case III: Packet Loss

Packet loss occurs due to several reasons including the network conjunction. For analyzing packet loss, NetBurst has transmitted a total of 478861 packets. On the receiving node a total of 430550 packets were received, which shows that 11% of the packets were lost. For statistical analysis, the following graph shows the loss which occurred in the transmission.

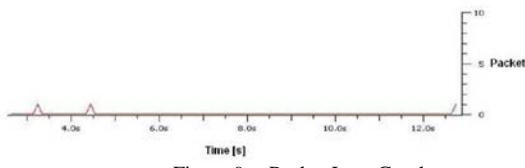


Figure 9. Packet Loss Graph

From the figure 9 it is clear that a small number of packets are lost during the whole period of packet transmission. In the above figure, small peaks depict packet loss with respect to packet sequence number.

5.2 TCP Functionality Testing

The objective of these test cases is to test the functionality of the TCP stack. A set of sample test cases were executed. These test cases, and their results, are discussed in this section.

TCP establishes connection using three way handshake mechanisms. This functionality is verified here.

5.2.1 Test Case I: Establishing Connection on Open Port

The following flow graph, shown in figure 7-6, describes how a “SYN” is acknowledged from the machine under test with a “SYN ACK”.



Figure 10. TCP Handshake

The above figure depicts that client has sent a packet with “SYN” flag. Server has acknowledged this “SYN” request through a TCP packet having control flags “SYN ACK”.

5.2.2 Test Case II: Establishing Connection from Closed Port

In this test case, a “SYN” is sent from the client and then the port is closed. The server has acknowledged this “SYN” with a “SYN-ACK” packet. Since the port is closed from the client, an “RST” will have been sent to the server to inform it that the port is closed. The following graph in figure 11 depicts the same scenario.



Figure 11. TCP Connections.

5.3 Packet Parameters

Once the connection is setup, a desirable test is to see how the TCP stack deals with legal and illegal sequence numbers. Following test case verifies this functionality.

5.3.1 Test Case I: Sending Illegal Sequence and Acknowledgment Numbers

As shown in the below figure 12, the TCP stack should mark and ignore illegal SEQ and ACK number packets as out-of-order packets.

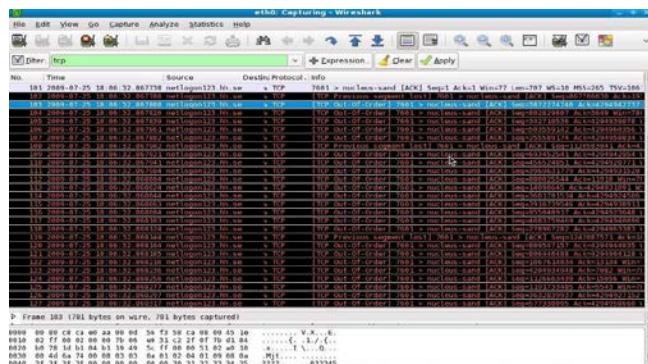


Figure 12. TCP Connection

5.4 TCP Option List

How the TCP stack processes the different TCP options is validated here. A set of test cases are executed to analyze different aspects of option processing

5.4.1 Test Case I: Valid options with legal values

An options byte string is sent with “SYN” packet. Option string is 20 bytes long and has the value “\003\003\012\001\002\004\001\011\010\012\077\077\077\000\000\000\000\000”

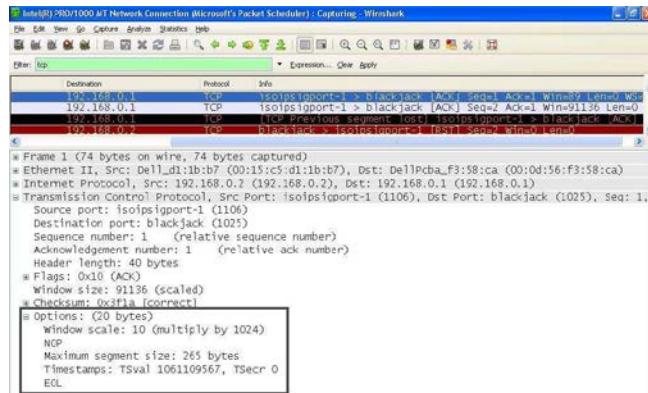


Figure 13. TCP Option Processing

5.4.2 Test Case II: Valid options with illegal and unusual values

An options byte string is sent with the “SYN” packet and the values “\003\003\012\001\002\004”. As shown in the figure 14 invalid options are detected and an error message “option goes past end of option” is displayed.

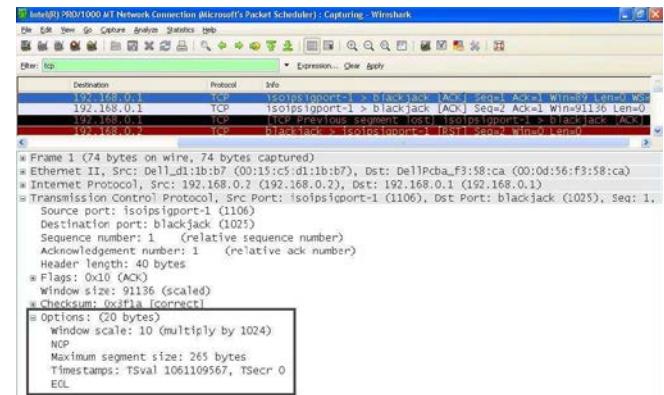


Figure 14. Bad Option Processing

In this section we have elaborated on TCP Functionality testing with different test cases for TCP handshake, Connection Establishment etc using “NetBurst” application. In the same way we can check the TCP functionality for TCP Packets, Payloads, and IP Fragmentation

6. NetBurst Portability

Due to resource constraints, the application portability is tested only for Windows and Linux systems. The application should work on VxWorks and Solaris, as the libraries used to develop this application claim to work on the above mentioned operating systems [23].

7. Conclusions

This paper introduces a cross platform and vendor independent farm work that uses active probing technique to test Ethernet functionality, comparatively, the NetBurst application gives an efficient result and a cost effective solution for network functionality testing.

There are many software applications to test the Ethernet functionality, like TCP Dump, Wireshark but they can only analyze traffic. They do not offer any option to inject customized packets. Applications that can test TCP for example TTCP can send TCP packets, but they do not support custom packet creation. Applications like Snoop are platform dependent and can run on a specific platform. All these drawbacks are resolved in “NetBurst”.

Active probing technique was introduced by Douglas E. Comer and John Lin. Douglas E. Comer and John C. Lin’s work is focused only on RTO (retransmission time-out) estimation, retransmission interval and keep-alive functionality, and does not test any other functionality available in TCP

Our test cases have tested TCP/IP stack functionality and results are discussed in this paper. It is also explained in the report that the number of packets per second, and the latency, are directly proportional to each other. Latency will increase gradually as the number of packets increase. NetBurst also gives an output result of the network performance testing, such as load, packet loss, latency, and throughput.

8. References

- [1] S. Bradner & J. McQuaid, "Benchmarking Methodology for Network Interconnect Devices", *RFC 2544*, Available at: <http://www.ietf.org/rfc/rfc2544.txt>. (Last accessed July 09, 2009).
- [2] Douglas E. Comer, Internetworking with TCP/IP: Principles, Protocols, and Architecture, Volume 1 (4th ed.), Upper Saddle River: Prentice Hall, 2000, ISBN: 0130183806 (International Ed.)
- [3] "Resolve IP Fragmentation, MTU, MSS, and PMTUD Issues with GRE and IPSEC", Available at: http://www.cisco.com/en/US/tech/tk827/tk69/technologies_white_paper_09186a00800d6979.shtml. (Last accessed July 1, 2009)
- [4] "Ethernet Testing Software's" Available at <http://www.allbusiness.com/technology/computer-networking/831290-1.html>. (Last accessed July 1, 2009)
- [5] "Ethernet Testing Hardware" Available at: <http://www.iol.unh.edu/services/testing/ethernet/tools/>. (Last accessed July 1, 2009).
- [6] "TCP Dump", Available at: <http://www.tcpdump.org/>. (Last accessed July 1, 2009).
- [7] "Snoop (1M) - capture and inspect network packets (man pages section 1M: System Administration Commands)", Available at <http://docs.sun.com/app/docs/doc/819-2240/snoop-1m?&a=view&q=snoop>. (Last accessed July 09, 2009).
- [8] "TPTEST", Available at <http://tpptest.sourceforge.net/about.php>. (Last accessed July 09, 2009).
- [9] "Maxwell Network Emulator Information", Available at <http://www.maxwelltester.com>. (Last accessed July 09, 2009).
- [10] "IxANVL™ - Automated Network Validation Library (ANVL)", Available at <http://www.ixiacom.com/products/display?key=ixanvl#note1#note1>, (Last accessed July 1, 2009).
- [11] W. Hengeveld, "Protocol Testing: Using hardware techniques for software", *Seventh International Conference on Software Engineering for Telecommunication Switching Systems*, 1989. SETSS 89, July 1989
- [12] Ethernet Network Functionality Testing Chanson, S.T. and Zhu, J, "A Unified Approach to Protocol Test Sequence Generation", *INFOCOM '93. Proceedings. Twelfth Annual Joint Conference of the IEEE Computer and Communications Societies. Networking: Foundation for the Future*. 28 March-1 April 1993 Page(s):106 - 114 vol.1
- [13] Huang,C.M., Lin, Y.C. and Jang, M.Y, "An Executable Protocol Test Sequence Generation Method for EFSM Specified Protocols". IWTCS 95, Evry, 4-6 September.
- [14] C.Bourhfir,R.dssouli,E. Aboulhamid, P.Rico, "Automatic executable test case generation for extended finite state machine protocols", 10th IFIP IWTCS, 1997.
- [15] Wu; Wang, "Internet protocol conformance testing by using a TTCN based protocol integrated test system," *Communications, 1999. ICC '99. 1999 IEEE International Conference*, vol.1, pp.646-650, 1999
- [16] K Shemyak , K Vehmanen , "Scalability of TCP servers, handling persistent Connections", *Proceedings of the Sixth International Conference on Networking*.22-28, April 2007 on page(s): 89-89
- [17] D Kassabian, A Albicki, "A Protocol Test System for the Study of Sliding Window Protocols on Networked UNIX Computers" *IEEE Transactions On Education*, Vol. 38, No. 4, November 1995
- [18] X Xie; M Zheng; Kassabian, D.; Albicki, A., "Design and testing of a sliding window protocol in a protocol testing system," *Circuits and Systems, 1993., Proceedings of the 36th Midwest Symposium on* , pp.1144-1147 vol.2, 16-18 Aug 1993
- [19] Y Lin, R E. Newman, H Latchman "A New TCP and UDP Network Benchmark Suite", *Proceeding of the 10th Communications and Networking Simulation Symposium (CNS'07)*, March 2007.
- [20] R J. Linn, Jr., "Conformance Evaluation Methodology And Protocol Testing", *IEEE Journal on Selected Areas In Communications*. Vol. 7. No. 7. September 1989
- [21] D E. Comer , J C. Lin, "Probing TCP implementations", *Proceedings of the USENIX Summer 1994 Technical Conference on USENIX Summer*, p.17-17, June 06-10, 1994, Boston, Massachusetts. Available at: <http://www1.bell-labs.com/user/johnlin/probing-TCP.pdf>, (last accessed July 1, 2009)
- [22] "Test TCP (TTCP) Benchmarking Tool for Measuring TCP and UDP Performance", Available at <http://www.pcausa.com/Utilities/pcattep.htm> . (Last accessed July 09,2009).
- [23] "Libnet", Available at: <http://libnet.sourceforge.net/#whatis> (Last accessed July 1, 2009)
- [24] "Eclipse", Available at: <http://www.eclipse.org/> (Last accessed July 1, 2009)
- [25] "Wireshark", Available at <http://www.wireshark.org>. (Last accessed July 1, 2009).
- [26] Eric Hall, Internet Core Protocols: The Definitive Guide Help for Network Administrators (1st ed.), 1005 Gravenstein Highway North Sebastopol, CA: O'Reilly Media, Inc., 2009, ISBN: 1565925726
- [27] Douglas E. Comer, Internetworking with TCP/IP: Client-Server Programming and Applications, Linux/Posix Sockets Version, Volume 3 (4th ed.), Upper Saddle River: Prentice Hall, 2000, ISBN: 0130320714 (Paperback Ed).

Incorporating Security in Embedded System – A critical analysis

Mirza Aamir Mehmood¹, Amir Shahzad Khokhar² and Mazhar Ali³

Department of Computer Science^{1,2}, Department of Information Technology³

Balochistan University of Information Technology, Engineering & Management Sciences
Quetta, Pakistan

Abstract

Security is becoming a major concern in embedded system designing and development. This paper surveys security requirements, attack techniques and will review countermeasures for these attacks

Keywords: component; embedded systems, security, software attacks, viruses

1. Introduction

Use of embedded systems in diverse and complex applications have relentless importance of secure and fault tolerant system. It is essential for embedded device's to secure sensitive information, ensuring availability and providing secure communication system. Reliability is directly coupled with security in embedded systems thus an unsecured system is also an unreliable system. Traditionally security concaved as an issue related to networks and cryptography and embedded system designers consider it as an additional feature. Growing number of security breaches has dictated that compromise on security can lead to consequences ranging from inconvenience to a complete disaster. Embedded systems are co-design of Software and Hardware. To make a secure and reliable system, it is essential that both components are secure.

2. Embedded System Security Requirements

The functions of embedded system is to access and process sensitive data and provide critical functionality. For example, let's assume that an embedded system has been implanted in a heart patient to monitor heartbeat, blood pressure and sugar level of that person. When it finds any anomaly in blood pressure or sugar level, it send an alarm signal to doctor and when it found that heart beat

is gone down a certain limit, it generates a mild shock to save life of that person. In this example, it is very obvious that system should be available twenty four hours a day and seven days a week. It is also important that when communicating with doctor, it should ensure privacy of patient. In general, an embedded system should ensure dependability, confidentiality, integrity and availability to consider as secure system [4] [5]. Figure 1 shows security requirements for embedded system (Taken from [1]).



Figure 1. Security requirements for embedded system

2.1. Confidentiality

To ensure that sensitive information is protected against deliberate or accidental disclosure.

2.2. Integrity

Ensuring that sensitive information is protected against deliberate or accidental corruption and data is protected against illegitimately changes.

2.3. Availability

The ability to protect against deliberate or accidental actions that cause automated information resources to be unavailable to users when needed.

2.4. Dependability

Embedded systems often reside in machines that are expected to run continuously for years without errors/faults and in some cases recover by them-selves if an error/fault occurs [3]. Thus an embedded system should be dependable. Dependability is combination of

2.4.1. Fault-Avoidance

Constructing / developing a mechanism to prevent a fault situation to occur. This is accomplished in designing phase.

2.4.2. Fault-Tolerance

Ensuring that system behaves in normal fashion even when a fault situation arises. This is accomplished through redundancy.

2.4.3. Fault-Removal

In verification phase, it is ensured that system is error/fault free.

2.4.4. Fault-Forecasting

How to estimate, by evaluation, the presence, the creation and the consequences of errors.

3. Challenges in Secure Software Development

Many external factors influence confidentiality and availability of system administrative controls, physical barriers, are few of them. Whereas integrity of the computer system depends on the degree to which vulnerabilities have been eliminated from the system [2]. In addition to the requirements, discussed earlier, embedded system should also be able to face threats like denial of service attacks, system tempering etc. which becomes more complicated in presence of modern techniques for breaking security, such as power analysis and fault analysis. Although Software solutions are not sufficient to keep up with the computational demands of security processing in embedded system but they are major source of security vulnerabilities. Three main factors make development of secure software a challenge i.e. Complexity, Extensibility and Connectivity [1]. Let's briefly examine each of these issues.

3.1. Complexity

With every passing day, software's are becoming more and more complex. With each new line of code, new bugs and security risks introduced in software. The complexity aggravate when unsafe programming languages (e.g., C or C++) are used, which do not protect against simple kind of attacks, such as buffer overflows.

3.2. Extensibility

New development platforms like .Net and Java developed mainly to provide extensibility. Embedded system now can get updates or extensions from internet for example new mobile phones by Sony Ericsson accept BIOS update through internet. Unfortunately, the very nature of extensible systems makes it hard to prevent software vulnerabilities from slipping in as an unwanted extension.

3.3. Connectivity

Most of modern embedded systems are coming with built in connectivity with internet. This internet connectivity has made it possible that a small problem can propagate and result in substantial security breaches. This also means that an intruder does not require a physical access to embedded system to launch attack and exploit vulnerable software. Thus this ubiquity of networking will result in increase in number of attacks and greater risks from par software security practices.

4. Attacks on Embedded System

Attacks on embedded systems can be categorized in three classes i.e. software attacks, physical attacks and side channel attacks. Software attacks have largest share in total number of attacks on embedded systems and it is most difficult to protect against such attacks. In this article we will focus on software attacks and countermeasure against these attacks. An overview of physical and side channel attacks will be provided.

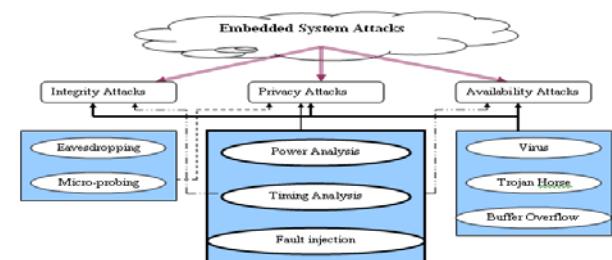


Figure 2. Embedded System attacks [6]

4.1. Physical Attacks

Embedded system can be divided in two categories

1. System on circuit board
2. System-on-chip

On circuit board embedded systems, attacks can be launched by probe to eavesdrop on inter-component communications. Whereas when launching attack on system-on-chip, micro-probing techniques are required. Physical attacks are relatively hard because they require expensive infrastructure and very complex techniques are used.

The most important part of any embedded system is microcontroller as it essentially controls all the operation of embedded system. The attacks on the microcontroller can be possible via JTAG, it is necessary to disable access to the microcontroller's internals via JTAG before fielding the finished product

4.2. Side Channel Attacks

Side channel attacks are based on observing system properties e.g. time, power consumption while system is performing computations e.g. cryptographic operations. In certain systems timing information can lead to entire secret key, though it seems that timing information can give very little information but it has been found that with proper study of timing sequence entire secret key can be found.

Along with this power consumption can also lead to the entire secret key, well equipped labs have the equipment that can measure the changes in the power consumption with about 1% accuracy and are very inexpensive. To overcome timing attacks one may add random timing delays to various operations, in similar manner we can overcome power consumption attacks by adding random noise or by proper shielding of the equipment but it leads to increased cost of the equipment.

4.3. Software Attacks

Software attacks are very common in embedded systems capable of downloading application from internet or have some means of communication to interact with external world. As compared to physical and side channel attacks, software attacks are very cheap and does not require any big infrastructures thus making it an immediate challenge for embedded system design. These attacks could further be classified in three categories 1. Virus attacks 2. Buffer Overflow and 3. Exploiting Software Vulnerabilities.

4.3.1. Virus, Worm and Trojan

These attacks are executed through malicious agents like virus, worm, Trojan. Following table 1 lists some viruses and respective embedded operation system.

Table 1 Viruses and Respective Embedded Operation System

<i>Operating system</i>	<i>Creator</i>	<i>Known viruses (including variants)</i>	<i>First virus appearance</i>
BlackBerry OS	Research In Motion	1	August 2006
Embedded Linux	GNU Project, Linus Torvalds and al.		
Mac OS X	Apple Inc.	0	-
Palm OS	PalmSource, Inc.	4	September 2000
Symbian OS	Symbian Ltd.	83	June 2004
Windows Mobile	Microsoft	2	July 2004

4.3.2. Vulnerability Exploitation

These agents exploit weaknesses in end-system architecture [7, 8, 9, 10]. "A vulnerability allows the attacker to gain direct access to the end-system, while an exposure is an entry point that an attacker may indirectly exploit to gain access" [6]. Following table lists latest list of vulnerability published by CERT.

Table 2 . Vulnerability List Published By CERT [10]

Vulnerability ID	Description
VU#298521	SonicWall NetExtender NELaunchCtrl ActiveX control stack buffer overflow
VU#446897	CUPS buffer overflow vulnerability
VU#180345	Microsoft Kodak Image Viewer code execution vulnerability
VU#342793	RSA Keon cross-site scripting vulnerabilities
VU#871673	RealPlayer playlist name stack buffer overflow
VU#559977	Mozilla products vulnerable to memory corruption in the browser engine
VU#755513	Mozilla products vulnerable to memory corruption in the JavaScript engine
VU#349217	Mozilla XUL web applications may hide the titlebar
VU#230505	Cisco IOS LPD buffer overflow vulnerability
VU#179281	Electronic Arts SnoopyCtrl ActiveX control and plugin stack buffer overflows

4.3.3. Buffer Overflow

The buffer overflow is a common flaw in OS and Application software. It's the inability of the buffer to store the information, which results from adding more information to a buffer than it was designed to hold. An attacker may exploit this vulnerability to take over a system". This problem arises when buffers used with poor boundary checks. Buffer bounds may be violated due to incorrect loop bounds, format string attacks, etc. Buffer overflow effects can include overwriting stack memory, heaps, and function pointers. Intruders use buffer overflow to overwrite program addresses stored nearby. "This may allow the attacker to transfer control to malicious code, which when executed can have undesirable effects. A good high-level introduction to the challenges involved in writing secure code can be found in [10]". In figure 3, the left-hand chart shows, for

Each year, the total number of CERT-reported vulnerabilities and the number that can be blamed primarily on buffer overruns whereas right-hand chart graphs the percentage of CERT-reported vulnerabilities that were due to buffer overruns for each year (taken from [15]).

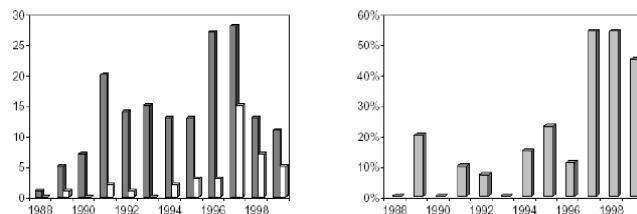


Figure 3. Frequency of buffer overrun vulnerabilities, derived from a classification of CERT advisories[16].

5. Countermeasure for software attacks

When designing countermeasures against software attacks, confidentiality and integrity of data and code is ensured, all security loopholes that make system vulnerable are removed. “A common feature of these countermeasures involves regulating the accesses of various software components (operating system, downloaded code, etc.) to different portions of the system (registers, memory regions, security co-processors, etc.) during different stages of execution (boot process, normal execution, interrupt mode, etc.), through a combination of hardware and software changes. Since an effective countermeasure must allow the system to provide guarantees about the security of the system starting from the powered-on state, most measures define notions of trust or *trust boundaries* (also referred to as *security perimeters*) across the various hardware and software resources. This allows the system to detect infringements of trust boundaries (such as illegal accesses to memory regions) and enforce recovery mechanisms (such as zeroing processor registers and memory regions). Thus, a trust boundary provides a natural and convenient foundation for the system to make judicious decisions about its security (or compromise, thereof)” [6].

5.1. Hardware Support

At hardware level security commonly implemented by using secure co-processor module [17, 18, 19, 22] this co-processor processes sensitive information. Information that needs to be send out of the co-processor is encrypted. Many embedded system also have secure memory areas. These secure memory areas are accessible to trusted system components only.

5.2. Secure Bootstrapping

Integrity check could be implemented at boot process level. After power is switched on, system should only be able to access the next layer if all integrity check found intact. This is done by comparing the securely saved value with hashed value of boot process component.

5.3. Operating System (OS) Enhancements

Secure Operating Systems provide features like process isolation, process attestation and secure storage. Secure storage is ensured by using of cryptographic file systems (CFSs) [20,21].

6. Security Pyramid

Embedded systems have three main components, Hardware, Software and communication mechanism / communication protocols. It is important that all three components should have built in security mechanism. Following figure shows suggested security pyramid for embedded systems.

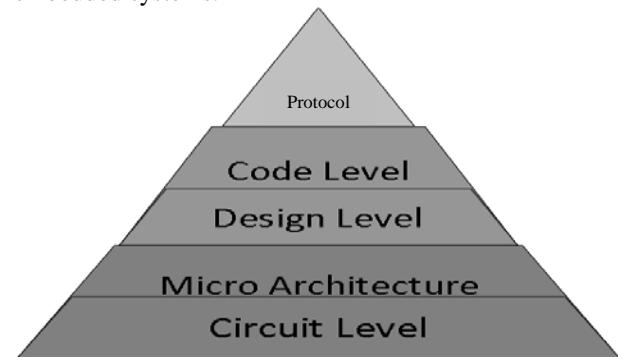


Figure 4. Security Pyramid for embedded systems

6.1. Communication / Protocol Level

Various security protocols have been developed to address security at different layers of network protocols, the most popular being IPSec, a network layer protocol, and TLS/SSL which works on transport layer [13,14]. Security protocols use algorithms (asymmetric or public-key ciphers, symmetric or private-key ciphers, hashing functions, etc.) to make communication secure.



Figure 5. A simple embedded system under protocol attack [16]

6.2. Software Level

To ensure that the software for embedded system meets security requirements, it is important that security should be implemented on various levels.

6.2.1. Code and Algorithm Level

Static analysis tool could be used to scan code to uncover common vulnerabilities. These tools can detect bugs at code level so that they could be fixed [14].

6.2.2. Design and Architecture Level

System must be coherent and present a unified security architecture that takes into account security principles (such as the principle of least privilege) [14].

6.2.3. Requirements Level

Security requirement should cover functional security.

6.2.4. Hardware Level Security:

At hardware level, security should be implemented on Micro-Architecture Level and at Circuit Level.

6.2.5. Micro Architecture Level

Incorporating security at hardware design of the modules (the processors and coprocessors) which is specified at the architecture level

6.2.6. Circuit Level

Implementing security at this level means use of techniques at transistor level and package-level to prevent various physical-layer attacks.

7. Future Work

Active research is going on to develop secure platforms for embedded systems. One approach is trusted computing, a project of Trusted Computing Group. The embedded processor community has also presented some design alternatives to achieve security e.g. ARM has developed new security architecture [23]. The security is achieved by portioning all the software and hardware resources so that they exist in one of two worlds - the secure world for the security subsystem, and the Normal world for everything else. This Architecture is deployed at both the levels i.e. the hardware level and the software level.

At hardware level they have designed the extended bus; the most significant feature of this bus is extended control signals that will restrict non-secure masters to access secure slaves. ARM has developed new architecture by using its "ARM Trust Zone Technology"

MIPS has also developed a secure processor core that includes secure memory management, protection against side-channel attacks, and uses an architecture that provides fast software cryptography using the SmartMIPS extensions to the MIPS32 architecture..

8. Conclusion

In this survey paper, we have examined security requirements and challenges in development of embedded systems. We have also examined how an embedded system can be attacked and their counter measures are also examined. We believe that implementing security pyramid discussed in this article will enable developers and

designers of embedded system to develop more secure system.

9. References

- [1]. Srivaths Ravi , Paul Kocher , Ruby Lee , Gary McGraw , Anand Raghunathan, Security as a new dimension in embedded system design, Proceedings of the 41st annual conference on Design automation, June 07-11, 2004, San Diego, CA, USA
- [2]. McCoy, J. A. An embedded system for safe, secure and reliable execution of high consequence software. In HASE 2004: The 5th IEEE International Symposium on High Assurance Systems Engineering.
- [3]. Online encyclopedia available at www.wikipedia.com
- [4]. Summers, Rita C. Secure Computing: Threats and Safeguards. New York: McGraw-Hill, 1997
- [5]. LAPRIE, J.C. Dependable computing and fault tolerance: concepts and terminology. In Proceedings of the 15th IEEE Symposium on Fault Tolerant Computing Systems (FTCS-15, June 1985). IEEE Computer Society Press, Los Alamitos, CA, 2-11.
- [6]. Srivaths Ravi , Anand Raghunathan , Srinath Chakradhar, Tamper Resistance Mechanisms for Secure, Embedded Systems, Proceedings of the 17th International Conference on VLSI Design, p.605, January 05-09, 2004
- [7]. Common Vulnerabilities and Exposures. Available at cve.mitre.org
- [8]. Latest Virus Threats. Symantec Corporation. Available at <http://www.symantec.com/avcenter/vinfodb.html>
- [9]. Virus Information. Computer Security Resource Center, National Institute of Standards and Technology. Available at <http://csrc.nist.gov/virus/>
- [10]. Vulnerability notes database. CERT coordination center Available at <http://www.kb.cert.org/vuls/>
- [11]. M. Howard and D. LeBlanc. Writing Secure Code. Microsoft Press, 2001.
- [12]. Buffer over flow techniques in computer viruses. Symantec white paper. Available at <http://securityresponse.symantec.com/avcenter/whitepapers.html>
- [13]. W. Stallings, Cryptography and Network Security: Principles and Practice. Prentice Hall, 1998.
- [14]. B. Schneier, Applied Cryptography: Protocols, Algorithms and Source Code in C. John Wiley and Sons, 1996.
- [15]. John Viega, Gary McGraw , Building Secure Software: How to Avoid Security Problems the Right Way, Addison-Wesley Professional, 2001
- [16]. Lawrence P. Ricci , Larry B. Mcginness , White Paper Embedded System Security Designing Secure Systems with Windows CE.
- [17]. B. Yee, Using Secure Co-processors. PhD thesis, Carnegie Mellon University, 1994.
- [18]. Secure Co-processing. IBM Inc. Available at <http://www.research.ibm.com/scop/>
- [19]. The IBM PCI Cryptographic Coprocessor. IBM Inc. Available at <http://www-3.ibm.com/security/cryptocards/>
- [20]. M. Blaze, "A Cryptographic File System for UNIX," in Proc. ACM Conf. on Computer and Communications Security, pp. 9-16, Nov. 1993.
- [21]. T. Garfinkel, M. Rosenblum, and D. Boneh, "Flexible OS Support and Applications for Trusted Computing," in Proc. 9th Wkshp Hot Topics in Operating Systems, May 2003.
- [22]. E. Goh, H. Shacham, N. Modadugu, and D. Boneh, "SiRiUS: Securing Remote Untrusted Storage," in Proc. ISOC Network and Distributed Systems Security (NDSS) Symp., pp. 131-145, 2003.
- [23]. Jaggar, D.; , "Arm Architecture And Systems," Micro, IEEE , vol.17, no.4, pp.9-11, Jul/Aug 1997

Implementing VoIP over Fatima Jinnah Women University

Ammara Tahir¹, Tahira Mahboob², Malik Sikandar Hayat khyal³

¹ Software Engineering Department, Fatima Jinnah Women University
Rawalpindi, Pakistan

² Software Engineering Department, Fatima Jinnah Women University
Rawalpindi, Pakistan

³ Software Engineering Department, Fatima Jinnah Women University
Rawalpindi, Pakistan

Abstract

Working people employee the use of technology in such very effective natural ways that permit them to do what they want: they can communicate with anyone as per need, in the time and space that suits them the best. Easily accessible and user-friendly, collaboration tools allow them to explore, share, engage, and connect with people and content in meaningful ways that help them learn. Traditional telephony carriers use circuit switching for carrying voice traffic. Circuit switching was designed for voice from the outset; hence it carries voice in an efficient manner. However it is an expensive solution. Nowadays people want to talk much more on phone, but they also want to communicate in a myriad of other ways – through e-mail, instant messaging, video, the World Wide Web, etc. Circuit switching is not suitable for this new world of multimedia communication. Therefore, in this paper a solution is proposed to implement scalable VOIP system over Fatima Jinnah University that will lead to better Quality of service and facilitates enhanced communication.

Keywords: Private branch exchange, Real time protocol, Session Initiation Protocol, Voice over Internet Protocol, Internet protocol, Public switched telephone networks, pulse code modulation.

1. Introduction

In this world of fast growing technological developments and the multimedia communications, need for fast paced, integrated services are required. Whereby the Internet protocol provides a healthy environment for integrated services combining voice and data. VoIP is a well suited solution for current networks such as LAN based corporate networks that replaces relatively expensive and circuit switched based PSTN network. This

research paper deals with implementing VoIP over Fatima Jinnah Women University Network. The idea of using a IP-based PBX is quite useful in such a case. Firstly, this system integrates the corporate telephone system with the corporate computer network, removing the need for two separate networks. The PBX itself becomes just another server or group of servers in the corporate LAN, which helps to facilitate voice/data integration.

The proposed system has been developed by implementing a soft PBX which acts as the server which performs call-routing functions, replacing the traditional legacy PBX or key system. This PBX would allow a number of attached soft phones to make calls to one another and to connect to other telephone services. The basic software would include many features available in proprietary PBX systems: voice mail, conference calling, interactive voice response, and automatic call distribution, just to name a few. The proposed system also helps the user in building the dial plan for the network.

The scope of this research project is to facilitate faculty/users and enable new forms of communication and engagement in the classroom, permitting extensions and variations of the informal interactions already occurring in classrooms and hallways, and creating new frontiers for collaboration across geographic boundaries.

2. Background:

Since 1990's phone has been a preferred and convenient way to communicate. In the developed world the phone network has grown and now it is almost everywhere. As we have experienced that internet is surely a public network and therefore connection to it is very cheap, merely costing a local call at the most which is totally unlike phone network.

The very first way for communication through VoIP was by running software on a PC. The first internet phone

software was released by Vocaltec, Inc. in 1995 [1], this software was designed to be able to run on a 33MHz 486 PC. The use of soft phones on the computer systems is one option for transferring voice over the IP network. On the other hand using phone adapters [2] is another way to enable VoIP services.

VoIP is a technology used to transmit voice conversation over a network using the Internet Protocol [3]. VoIP has nowadays become the hottest telecommunication research topic and as one of them it has many advantages over the traditional Public Switched Telephone Networks (PSTNs), which are discussed:

- Voice/Data Integration and Advanced Services [4]
- Lower Equipment Cost
- Lower Bandwidth Requirements
- Widespread availability of IP.

The protocols and services used in VoIP are detailed in Table 1.

Table 1: VoIP Protocols/Services

Application Layer	Audio RTP, RTCP, (SIP, H.323..)
Transport Layer	UDP
Network Layer	IP
Physical Layer	Ethernet, SDH,...

2.1 RTP

The RTP is an essential streaming protocol. It is evidently essential for real time applications that are designed for synchronization of traffic streams compensating for delay variations and de-sequencing. However not ensure on-time delivery or traffic signals or address the issue of QoS, relevant to guaranteed bandwidth availability for specific applications. RTP is generally used in conjunction with UDP, but it can surely make use of any packet-based lower layer protocol [5].

2.2 RTCP

The RTCP is a companion protocol of RTP that is used to transmit periodic control packets on the periodic transmission of control packets to all participants in the session, using the same distribution mechanism as the data packets [7]. RTCP provides QoS feedback and session information. The packets of RTCP carry information related to the state of end points and the statistical information regarding the current RTP stream such as delay times, packet loss and jitter [5]. Both RTP and RTCP run on the top of User Datagram Protocol (UDP) that provide better real-time responsiveness and lower overhead. After compression and digitization of a sample

of analogue voice signal, RTP and RTCP information can be packed inside an IP packet with an UDP header to accomplish the requests of transmitting a voice packet.

2.3 SIP

The SIP is a signaling protocol which is simple and light weighted, it is most commonly implemented on top of the User Datagram Protocol (UDP), but it can also be implemented on top of the Transmission Control Protocol (TCP) [8].

3. Related Work:

[8] Regards VoIP as appropriate technology for roaming academics and local deployment for telephony services. [9] Claims of the high merits of SIP protocol being used for implementing VOIP that has lead Nokia to earn more business. [10] Research survey carried on 280 companies revealing the cost optimization he driving force for deploying (VoIP Ezilon.com, 2010). [11] discusses the comparative analysis of POTS with VoIP where voice converted to digitized form(binary 0s and 1s) and transferred as packets. [12] The project deployed on a high bandwidth LAN so therefore the efficiencies of the protocols used in VoIP networks are thus not particularly an issue: the range and flexibility of providing services was more important with the use of high quality codec.

4. Research Goal:

The aim of the research paper was to develop an efficient, cost effective and bandwidth optimized soft PBX solution that implements VoIP aimed to facilitate dynamic academics as well as local deployment.

5. Implementation:

The proposed system is implemented as a soft PBX that acts as a server responsible for performing routing functionalities replacing the traditional legacy PBX or key system. The system has been developed in fast growing cutting edge technology of .Net and C# framework. Operating system is windows 2000 or windows XP. The client/server architecture is followed with Soft PBX as server as in Fig 2 and the soft phones acting as clients as in Fig 3. G.729 codec together with SIP and RTP/RTCP protocol is implemented as shown in Fig 1. G.729 codec has been chosen due to variety of factors such as lower bit rates that works efficiently well in congestion conditions. Also adaptive jitter buffers have been implemented to improve performance when compared with fixed jitter buffers. Furthermore, with Forward Error correction

receiver recovers from packet loss without retransmissions but it increases the delay.

Session Initiation protocol has been implemented. Unnecessary ports/services in SIP and H.323 gateways have been disabled to decrease the possibility of unauthorized access and remote code execution. Packet loss concealment technique has been implemented in order to improve VoIP performance.

Real Time protocol has been implemented to ensure a proper delivery of packets with real time characteristics. Specifically RTCP is being implemented to provide reception quality feedback of RTP data distribution to all the participants/clients in a session.

```

namespace WaveLib
{
public class WaveStream : Stream, IDisposable
{
    private Stream m_Stream;
    private long m_DataPos;
    private long m_Length;
    private WaveFormat m_Format;

    public WaveFormat Format
    {
        get { return m_Format; }
    }

    private string ReadChunk(BinaryReader reader)
    {
        byte[] ch = new byte[4];
        reader.Read(ch, 0, ch.Length);
        return System.Text.Encoding.ASCII.GetString(ch);
    }

    private void ReadHeader()
    {
        BinaryReader Reader = new BinaryReader(m_Stream);
        if (ReadChunk(Reader) != "RIFF")
            throw new Exception("Invalid file format");
        Reader.ReadInt32();
        // File length minus first 8 bytes of RIFF description,
        // we don't use it
        if (ReadChunk(Reader) != "WAVE")
            throw new Exception("Invalid file format");
        if (ReadChunk(Reader) != "fmt ")
            throw new Exception("Invalid file format");
        int len = Reader.ReadInt32();
        if (len < 16) // bad format chunk length
            throw new Exception("Invalid file format");
        if (m_Stream != null)
            m_Stream.Close();
        GC.SuppressFinalize(this);
    }

    public override bool CanRead
    {
        get { return true; }
    }

    public override bool CanSeek
    {
        get { return true; }
    }

    public override long Length
    {
        get { return m_Length; }
    }

    public override long Position
    {
        get { return m_DataPos; }
        set { m_DataPos = value; }
    }

    public override void SetLength(long value)
    {
        m_Length = value;
    }

    protected void Dispose(bool disposing)
    {
        if (disposing)
            m_Stream?.Close();
    }

    public void Dispose()
    {
        Dispose(true);
        GC.SuppressFinalize(this);
    }

    private void InitializeFormat()
    {
        m_Format = new WaveFormat(22050, 16, 2); // initialize
                                                // to any format
        m_Format.wFormatTag = Reader.ReadInt16();
        m_Format.nChannels = Reader.ReadInt16();
        m_Format.nSamplesPerSec = Reader.ReadInt32();
        m_Format.nAvgBytesPerSec = Reader.ReadInt32();
        m_Format.nBlockAlign = Reader.ReadInt16();
        m_Format.wBitsPerSample = Reader.ReadInt16();
        // advance in the stream to skip the wave format block
        len -= 16; // minimum format size
        while (len > 0)
        {
            Reader.ReadByte();
            len--;
        }
        // assume the data chunk is aligned
        while (m_Stream.Position < m_Stream.Length &&
            ReadChunk(Reader) != "data");

        if (m_Stream.Position >= m_Stream.Length)
            throw new Exception("Invalid file format");
    }

    private void ReadData()
    {
        if (m_Stream.Position >= m_Stream.Length)
            throw new Exception("Invalid file format");
        m_Length = Reader.ReadInt32();
        m_DataPos = m_Stream.Position;
    }
}

```

```

}

public override bool CanWrite
{
    get { return false; }
}

public override long Length
{
    get { return m_Length; }
}

public override long Position
{
    { return m_Stream.Position - m_DataPos; }
    set { Seek(value, SeekOrigin.Begin); }
}

public override void Close()
{
    Dispose();
}

public override void Flush()
{
}

public override void SetLength(long len)
{
    throw new InvalidOperationException();
}

public override long Seek(long pos, SeekOrigin o)
{
    switch(o)
    {
        case SeekOrigin.Begin:
            m_Stream.Position = pos + m_DataPos;
            break;
        case SeekOrigin.Current:
            m_Stream.Seek(pos, SeekOrigin.Current);
            break;
        case SeekOrigin.End:
            m_Stream.Position = m_DataPos + m_Length - pos;
            break;
    }
    return this.Position;
}

public override int Read(byte[] buf, int ofs, int count)
{
    int toread = (int)Math.Min(count, m_Length - Position);
    return m_Stream.Read(buf, ofs, toread);
}

public override void Write(byte[] buf, int ofs, int count)
{
    throw new InvalidOperationException();
}
}

```

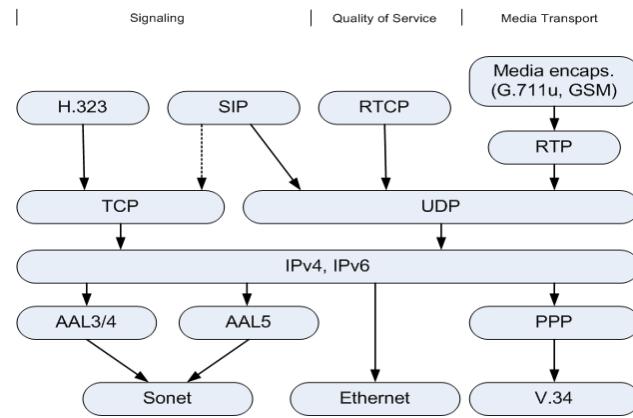


Fig 1: VoIP Architecture

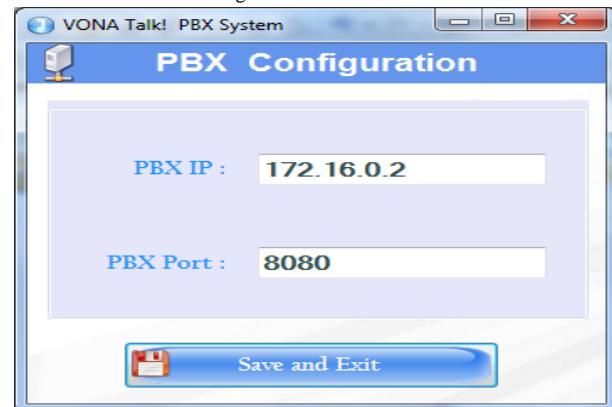


Fig 2: PBX Server



Fig 3: Soft phone Client

5. Conclusion & Future Work:

With the faster pace of development and growing demands of the users for higher quality of voice data transfer at lower bandwidths and better performance over faster growing networks. Integrated services and communications on-the-fly are the demands of industrial, commercial, telecommunications, information technology and the educational sector. With data and voice services being provided with committed quality of service which is

the main goal of this research paper. The proposed system has been tested in environment of the campus at Fatima Jinnah Women University and works effectively, fulfilling the needs of the current user base (lecturers/staff/students). However, with networks becoming larger in size and services being offered, the compression techniques being improved at faster pace the existing system can be further improved / refined to meet the user needs.

References

- [1] Dvorak, John C. Pirillo, Chris. Taylor, Wendy. (October 2003). Online! The Book. Prentice Hall PTR
- [2] Linksys. Feature-rich telephone service through your Internet connection. Cisco Systems Inc 2005.
<http://www.linksys.com/products/product.asp?grid=33&scid=38&prid=651>
- [3] Walter J. Goralski and Matthew C. Kolon, IP telephony. ISBN 007135221X, McGraw-Hill, 1st edition, 2000
- [4] J. Bellamy, Digital Telephony, 2nd ed., Wiley, 1991.
- [5] <http://www9.org/w9cdrom/349/349.html>
- [6] The RTCP gateway: scaling real-time control bandwidth for wireless networks , K. Brown, Computer Communications 23 (2000) 1470–1483
- [7] CHONG, H., MATTHEWS, H. 2004. Comparative Analysis of Traditional Telephone and Voice-Over-Internet Protocol (VoIP) Systems. IEEE international symposium on Electronics and the Environment, 10th–13th May 2004.
- [8] http://www.cisco.com/en/US/tech/tk652/tk701/technologies_configuration_guide_chapter09186a00800eadf9.html
- [9] VOIP—AN INSIGHT INTO A PROGRESSING TECHNOLOGY, David Tarrant, Researcher in Intelligence, Agents and Multimedia School of Electronics and Computer Science University of Southampton
- [10] White Paper: Advantages of SIP for VoIP by Nokia in October 2003
- [11] Background Research in perspective of InformationWeek Research
- [12] VoIP as a communications tool in South African business. By Lydia Uys (Department of Accounting, Stellenbosch University, Private Bag X1, Matieland, 7602, South Africa.) Accepted on 5 February, 2009
- [13] VoiceXML dialog system of the multimodal IP-Telephony—the application for voice ordering service,M. Stai [2005].

Ammara Tahir is a graduate Software Engineer from Fatima Jinnah Women University, Rawalpindi, Pakistan (2011). Her research interests are networks and communications.

Tahira Mahboob graduated from University of Engineering And Technology, Lahore Pakistan in 2007. She is registered with Pakistan Engineering Council and is currently enrolled in Masters degree program of Computer Engineering at Center for Advanced Studies in Engineering (CASE), UET Taxila (2010). She has been associated with telecom sector and now serving as a Lecturer at Fatima Jinnah Women University. She has supervised thesis/ projects in the field of communication, networks, mobile automation and voice recognition at Bachelor's and Master Degree programs. Her research interests are computer/mobile communications, networks, information security and mobile automation.

Dr. Malik Sikandar Hayat Khiyal is Chairperson Department of Computer Sciences and Software Engineering at Fatima Jinnah Women University, Pakistan. He received his M.Sc degree from Quaid-e-Azam University, Islamabad. He got first position in the faculty of Natural Science of the University. He was awarded the merit scholarship for Ph.D. He received his Ph.D. degree from UMIST, Manchester, U.K. He developed software of underground flow and advanced fluid dynamic techniques. His areas of interest are Numerical Analysis, Modeling and Simulation, Discrete structure, Data structure, Analysis of Algorithm, Theory of Automata and Theory of Computation. He has more than hundred research publications in National and International Journals and Conference proceedings.

Analysis of Image Denoising using Wavelet Coefficient and Adaptive Subband Thresholding Technique

S.Kalavathy¹ and R.M.Suresh²

¹ Research Scholar,
Dr.M.G.R Educational and Research Institute University, Maduravoyal, India
&
Department of Mathematics, R. M. D. Engineering College,
Kavaraipettai, TamilNadu, India

² Department of Computer Science and Engineering, R. M. D. Engineering College,
Kavaraipettai, TamilNadu, India

Abstract

Image denoising is a common procedure in digital image processing aiming at the removal of noise which may corrupt an image during its acquisition or transmission while sustaining its quality. A statistical model is proposed depending on the magnitude of wavelet coefficients and noise variance for each coefficient is estimated based on the subband it belongs to using Maximum Likelihood (ML) estimator or a Maximum a Posterior (MAP) estimator. An adaptive thresholding is proposed which is applied to each subband coefficient except the low pass or approximation subband. This is done by fixing the optimum thresholding value depending on the decomposition level. The proposed method describes a new method for suppression of noise in image by fusing the wavelet denoising technique with optimized thresholding function to which a multiplying factor (α) is included to make the threshold value dependent on decomposition level. Due to this, the proposed technique yields significantly superior image quality by preserving the edges, producing a better PSNR value. The efficiency is proved on comparing with Bayes shrink (BS), Modified Bayes Shrink (MBS) and Normal Shrink (NS) for different noise level.

Keywords: *Image denoising, decomposition level, wavelet coefficients, optimum thresholding*

1. Introduction

Estimating an image that is corrupted by additive noise has been of interest to many researchers for practical as well as theoretical reason. The problem is to recover the original image from the noisy data by sustaining the possible important image features. Various statistical wavelet models for images have been proposed. A simple and popular model is independent and identically distributed (iid) Generalized Gaussian Distribution (GGD) model for wavelet coefficients [1, 2]. This model has been successfully used in image denoising and restoration. In recent years, there has been a fair amount of research on filtering using wavelet coefficients thresholding because wavelets provide an appropriate basis for separating noisy

signal from the image signal. The problem of wavelet based denoising can be expressed as an estimation of clean coefficients from noisy data with Bayesian estimation techniques. If the Maximum a Posterior (MAP) estimator is used for this problem, the solution requires a prior knowledge about the distribution of wavelet coefficients. Based on the distribution type, the corresponding estimator (shrinkage function) is obtained. The classical soft threshold shrinkage function can also be obtained by a Laplacian probability density function (pdf) [3]. Many researchers have proposed the bivariate pdfs for modeling the interscale dependency [3, 4, 5, 6]. Usually these pdfs improve the denoising results they may lead to complicated algorithms. Intrascale dependency states that pdfs using spatial local parameters are able to capture the statistical properties of wavelets [7, 8]. Mihcak [8] proposes a Gaussian pdf with local variance for denoising and earns impressive results with his simple algorithm. In this paper, we use a Laplacian pdf with local variance to model the heavy-tailed property and interscale and intrascale dependencies of wavelet coefficients. This pdf is univariate we estimate the local variance of each coefficients using its spatial adjacent and its parent's spatial adjacent to incorporate both inter and intrascale dependencies in this estimation. Denoising is commonly done by wavelet shrinkage irrespective to the type of DWT applied.

Wavelet shrinkage is a method of removing noise from images by shrinking the empirical wavelet coefficients in the wavelet domain and it is a non linear image denoising procedure to remove the noise. A common shrinkage approach is thresholding [9, 10] which sets the wavelet coefficients with "small" magnitudes to zero while retains shrinking in magnitude for the remaining ones. Originally, Donoho and Johnstone proposed the use of a universal threshold uniformly throughout the entire wavelet decomposition tree which was found to be more efficient [11, 12, 13, 14, 15]. Although thresholding with a uniform threshold per subband is attractive due to its simplicity, the

performance is limited and the denoising quality is often not satisfactory. Thus wavelet shrinkage methods using separate threshold in each subband have been developed over recent years. Some methods of selecting thresholds that are adaptive to different spatial characteristics have been recently proposed and investigated [16, 17, 18]. In general, adaptive approaches have found to be more effective than their global counterparts.

In this paper a new model is proposed based on wavelet coefficients and noise variance is estimated for each coefficient depending on the subband using Maximum Likelihood (ML) estimator. A multiplying factor (α) is

included in the optimum threshold formula to make the threshold value dependent on decomposition level. After computing threshold, apply soft thresholding to each noisy coefficient. By inverting the multi scale decomposition, the resultant quality image with less blurring and preserving more detail information is reconstructed.

2. Wavelet Coefficient Model

Considering the advantages and limitations of the statistical model a new model is proposed based on the wavelet coefficients. Wavelet coefficients with large magnitudes are representatives of edges or some textures. While those with small magnitude are associated with smooth regions such as the background.

In this smooth region the signal variance for every sub band are estimated by a ML estimator. This method presents effective results but their spatial adaptivity is not well suited near object edges where the variance field is not smoothly varied. To overcome this the coefficient in each subband except the first fine scale is partitioned into two classes based on the magnitudes of their parents, namely significant class and insignificant class in the corresponding region. The significant class represents high activity regions and insignificant class corresponds to smooth region. The sizes of the two classes are controlled by the significance threshold T. If the magnitude of the parent is larger than T then the coefficient is included in significant class otherwise it is included in insignificant class.

The two classes have different statics. The histogram of the coefficient in insignificant class is highly concentrated around zero while that of significant class is more spread out. Hence the coefficients in significant class are modeled as independent identically distributed (iid) Laplacian with zero mean. For the coefficient in insignificant model which corresponds to homogeneous regions, the usage of

intrascale model in Estimation Quantization [EQ] coder is appropriate [19]. It provides a good fit for the first order statics of wavelet coefficients and well models the nonstationary nature of low-activity regions.

2.1 Statistical Model

The observation model is expressed as follows $Y = X + V$, where Y is the wavelet transform of the degraded image, X is the wavelet transform of the original image, V denotes the wavelet transform of the noise components following the Gaussian distribution $N(0, \sigma_v^2)$. Since X and V are mutually independent, the variances of Y , X and V are given by

$$\sigma_Y^2 = \sigma_X^2 + \sigma_V^2 \quad (1)$$

3. Estimation of Noise Variance

The following steps are used to evaluate the variance estimate for each wavelet coefficients depending on the subband.

Step 1: The noise variance σ_v^2 can be accurately estimated from the first decomposition level diagonal subband HH_1 , by the robust and accurate median estimator [20].

$$\sigma_v^2 = \left(\frac{\text{median}|y(V)|}{0.6745} \right)^2 \quad (2)$$

Where $y(V)$ represents the coefficients HH_1 subband.

Step 2: The coarse subbands are not processed because the coarse subband has very high SNR. These coefficients are considered reliable.

Step 3: For each of the three subbands (horizontal, vertical and diagonal orientations) coefficients within the subband are modeled as identically independently distributed with zero mean and variance $\sigma_{x,j}^2$ (where j indicates the subband). The variance estimate is computed from the noisy coefficients in subband j as

$$\sigma_{x,j}^2 = \max\{0, \text{var}\{\bar{y}_i, i \in \text{subband } j\} - \sigma_v^2\} \quad (3)$$

Using MAP, estimation of \bar{x} is obtained by applying a soft threshold λ as given in equation (4) to each noisy coefficient.

$$\lambda = \sqrt{2\sigma_v^2 / \sigma_{x,j}^2}, j \in \text{subband} \quad (4)$$

Step 4: In each of the other high sub bands, coefficients are assigned either to significant or insignificant classes

depending on the magnitude of their estimated parent relative to the significance threshold T , where

$$T = \sigma \sqrt{2 \log N^2} \quad (5)$$

- (i) Coefficients in significant class are modeled as iid Laplacian with zero mean and their variance $\sigma_{x,insig}^2$ is estimated from the noisy coefficients as mentioned in step 3. Again the MAP estimator is a simple soft thresholding scheme where its threshold value is adjusted to the signal variance.
- (ii) Coefficients in insignificant class which has small magnitude representing smooth areas, $\sigma_{x,insig}^2$ is estimated using ML estimator in order to have an estimate for a local neighborhood σ_x^2 where variance is assumed to be constant. The estimate of the class coefficient variance is

$$\sigma_{x,insig}^2 = \frac{1}{M} \left(\sum_{V=1}^M y^2(V) - \sigma_V^2 \right) \quad (6)$$

where M represents the number of wavelet coefficients residing in local neighborhood N . Considering the coefficients belonging to a insignificant class inside the window are used by excluding the one which belong to significant class, the MAP estimator is given by

$$\bar{x} = \frac{\sigma_{x,insig}^2}{\sigma_{x,insig}^2 + \sigma_V^2} \bar{y}_i \quad (7)$$

Thus the coefficient of estimates corresponding to the high subband are obtained by repeating the above steps from parent to child subband, starting from the coarse scale and terminating in the highest subband.

4. Optimum Value Threshold and Proposed Technique

Wavelet thresholding [21, 22, 23] is a signal estimation technique that exploits the capabilities of wavelet transform for signal denoising. It removes noise by killing coefficients that are insignificant relative to some threshold and turns out to be simple and effective which depends heavily on the choice of a thresholding parameter. The choice of this threshold determines the efficacy of denoising to a great extent.

4.1 Threshold Selection

Finding an optimum value thresholding is not an easy task. A small threshold may yield a result close to the input, but

the result may still be noisy. A large threshold on the other hand, produces a signal with a large number of zero coefficients. This leads to a smooth signal. Paying too much attention to smoothness destroys details and it may cause blur and artifacts in image processing.

Soft thresholding method is used to analyze the performance of denoising system for different levels of DWT decomposition, as it results in better denoising performance than other denoising methods. Also it leads to less severe distortion of the object of interest than other thresholding methods [24]. Several approaches have been suggested for setting the threshold for each band of the wavelet decomposition. A common approach is to compute the sample variance σ^2 of the coefficients in each band and set the threshold to any multiple of standard deviation σ for that band [25]. Thus, to implement a soft threshold of the DWT coefficients for a particular wavelet band, the coefficients of that band should be thresholded as shown in Fig. 1(a). The soft thresholding is generally represented by,

$$d_{ik}^{soft} = \begin{cases} sign(d_{ik})(|d_{ik}| - \lambda^*) & if d_{ik} > \lambda^* \\ 0 & if d_{ik} \leq \lambda^* \end{cases} \quad (8)$$

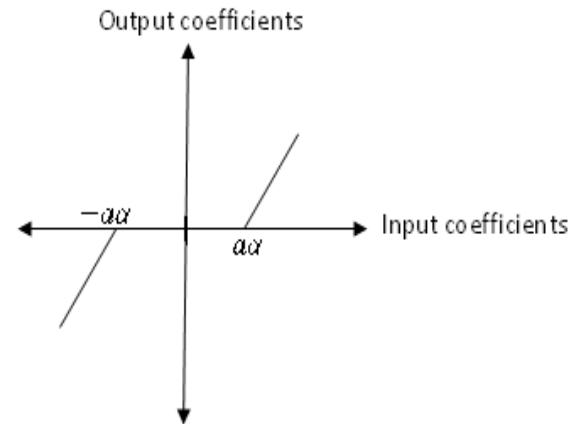


Fig.1 Soft threshold Characteristics with $\lambda = a\alpha$

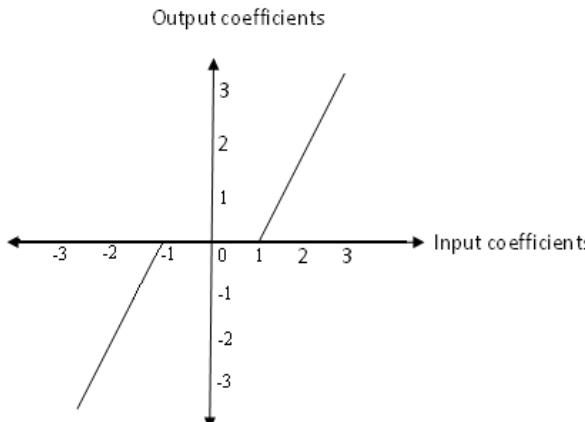


Fig.2 Soft threshold Characteristics with $\lambda = 1$

4.2 Optimum Value Threshold

An adaptive thresholding is proposed by fixing the optimum thresholding value depending on the decomposition level. At every decomposition level, four frequency subbands are obtained namely LL, LH, HL, HH. The next level should be applied to the low frequency subband LL only. This process is continued until a prespecified level (level 2) is reached. In wavelet domain, as the level of subbands increases its coefficients becomes smoother. That is, subband HL_2 is smoother than the corresponding subband in the first level (HL_1) and so the threshold value of HL_2 should be smaller than that of HL_1 .

In the wavelet decomposition, the magnitude of the coefficient varies depending on the decomposition level. Therefore, if all levels are processed with one threshold value, the processed image may be overly smoothed so that sufficient information preservation is not possible and the image gets blurred. To overcome this problem and to obtain a significantly superior quality image, the multiplying factor α is included in the threshold formula to

get better PSNR value by preserving edges where

$$\alpha = 2^{L-K} \sqrt{\log M} \quad (9)$$

L is the number of wavelet decomposition level, K is the level at which the subband is available, M is the total number of wavelet coefficients. Using this multiplication factor α the optimum threshold formula for the proposed technique is given by

$$\lambda^* = \alpha \lambda \quad (10)$$

Where λ, σ are calculated using equations (4) and (9) respectively.

4.3 Proposed threshold algorithm

The proposed block diagram of wavelet based image denoising system is shown in Fig. 2. Wavelet Based Denoising method relies on the fact that noise commonly manifests itself as fine-grained structure in the image and DWT provides a scale based decomposition. Thus, most of the noise tends to be represented by wavelet coefficient at the finer scales. Discarding these coefficients would result in a natural filtering of the noise on the basis of scale. As the coefficients at such scales also tend to be primary carriers of edge information, the DWT noisy coefficients can be made zero if their values are below its optimum threshold value. On the other hand, the edge relating coefficients are usually above the threshold. The inverse DWT of the thresholded coefficients is the denoised image.

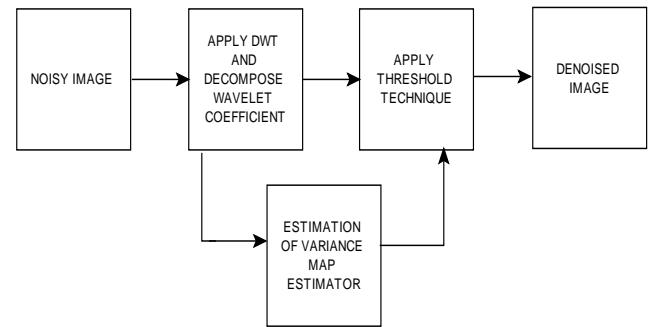


Fig. 3 Block diagram of the proposed method

Algorithm:

The complete algorithm of the proposed wavelet based denoising technique is explained in the following steps:

Input: Noisy image

Output: Denoised image

Step 1: Perform Multiscale decomposition of the image corrupted by Gaussian noise using wavelet transform.

Step 2: Estimate the noise variance σ_v^2 using equation (2) for each scale and compute the scale parameter.

Step 3: For each of the three subbands variance estimate is computed from the noisy coefficient in subband j using equation (3).

Step 4: In each of the other high subbands the estimates of the class coefficient variance are estimated using equation (3) and (6).

Step 5: Calculate threshold value using optimum value threshold formula as given in equation (10) after finding the multiplying factor σ for each subband using the relation given in (9)

After computing threshold for each subband except the low pass or approximation subband, apply soft thresholding to each wavelet coefficient using threshold

given in equation (8), by substituting the threshold value obtained in Step 5.

Step6: Invert the multiple decomposition to reconstruct the denoised image.

5. Experiment and Results

The above algorithm has been applied on several natural gray scale test images like Lena Barbara and pepper at different Gaussian Noise level (standard deviation $\sigma = 0.001, 0.002, 0.003, 0.004$). Here we used Daubechies (Db4), the least asymmetric compactly supported wavelet at two levels of decomposition. Performance of noise reduction algorithm is measured using quantitative performance measure such as Peak Signal to Noise Ratio (PSNR) as given in Table-1 and in terms of visual quality of images, as shown in fig-4. To evaluate the performance of the proposed method, it is compared with the Baye's shrink, Modified Baye's shrink and Normal shrink using PSNR which is defined as

$$PSNR = 20 \log_{10} \left(\frac{255^2}{MSE} \right) dB \quad (11)$$

Where MSE denotes the Mean Square Error between the original and denoised image given by

$$MSE = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (X_{ij} - Y_{ij})^2 \quad (12)$$

Where M, N are width and height of image. Y –Noisy image X –original image.

5.1 Statistical Analysis

Consider the PSNR value obtained for Barbara Image of different noise level using our proposed technique and Modified Bayes Shrink method as two samples X_1 and X_2 of sizes n_1 and n_2 ($n_1 = n_2 = 10$) respectively. The significant differences between these two samples were tested using Student's t-test: Two samples assuming equal variance.

By using the test statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} (\sim t(n_1+n_2-2)) \text{ d.f (degrees of freedom)}$$

for the above sample it is seen that the calculated value of t is greater than the critical value of t at 5% level of significance for 18 d.f. Due to this the null hypothesis $H_0: \mu_1 = \mu_2$ is rejected and $H_0: \mu_1 > \mu_2$ (1 tailed test) is accepted. By this significant test the PSNR value obtained for Barbara image of different noise level by our proposed method is significant than by using modified Bayes Shrink method. A similar test is also done

by considering the other sample as Bayes Shrink Method, Normal Shrink Method. In these cases also it is observed that the proposed method yields a significant value. Hence Adaptive Subband Threshold Technique out performs the other wavelet method by possessing high PSNR value. This method finds application in denoising images those are corrupted during transmission which is normally random in nature.

Table- 1 Comparison of PSNR of different wavelet shrinkage method for different images corrupted by Gaussian noise.

Image	Noise level	Noisy	Baye's Shrink	Normal shrink	Modified Baye's Shrink	Proposed method
Barbara	0.001	69.08	73.10	73.15	73.05	73.01
	0.002	62.17	66.14	66.53	66.20	67.17
	0.003	58.13	61.80	62.24	61.88	63.06
	0.004	55.28	58.69	59.12	58.79	59.99
Lena	0.001	69.07	75.11	75.63	75.19	76.37
	0.002	62.15	67.35	67.99	67.46	69.27
	0.003	58.10	62.63	63.23	62.73	64.51
	0.004	55.24	59.28	59.84	59.38	61.04
House	0.001	69.01	75.12	75.79	75.29	76.77
	0.002	62.09	67.13	67.86	67.34	69.14
	0.003	58.04	62.37	63.05	62.57	64.45
	0.004	55.18	59.02	59.64	59.21	60.75

From Table -1 it is observed that the proposed thresholding technique out performs the other denoising methods by possessing high PSNR value

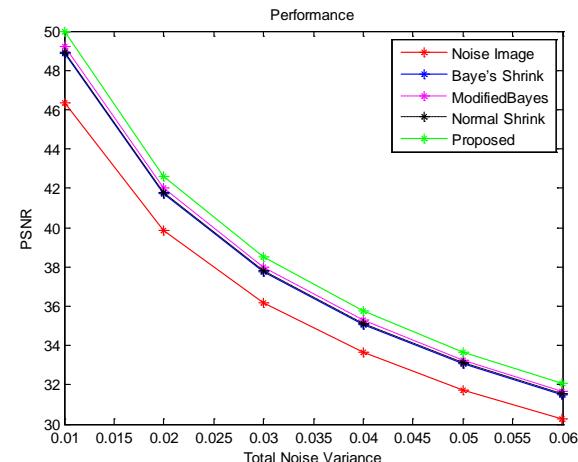


Fig. 3: Comparison of PSNR Value of different denoising methods for Barbara image

From Fig. 3 it is observed that Baye's shrink performs little denoising in high activity sub regions to preserve the sharpness of the edges but completely denoise the flat subparts of the image. Normal shrink preserve edges better than noise removal method using Baye's shrink. Modified Baye's shrink yields a better results for denoising and also adopts thresholding strategy by preserving edges better than Baye's and Normal shrink. The proposed thresholding algorithm gives better performance than other spatial domain filter like Baye's shrink, Normal shrink and Modified Baye's shrink by giving a better PSNR value. Further it out performs the performance of the above mentioned thresholding algorithm by preserving the edges as well as removing the noise, due to the advantages of using the multiplying factor α included in the optimum value threshold formula and subband thresholding.

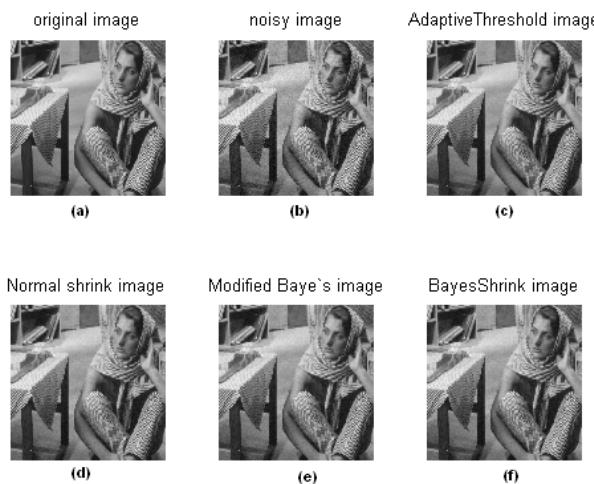


Fig. 5: Quality performance of different denoising methods for Barbara image (noise standard deviation $\sigma_v = 0.001$) (a) Original image (b) Noisy Image (PSNR=69.07) (c) Adaptive Threshold (PSNR=73.01) (d) Normal shrink (PSNR=73.15) (e) Modified Baye's shrink (PSNR=73.05) (f) Baye's Shrink (PSNR=73.02)

6. Conclusion

The proposed threshold estimation method is based on the analysis of statistical parameters like standard deviation, variance of the sub band coefficients using ML or MAP estimator which is more sub bands adaptive. Since the decomposition level dependent is included as a multiplying factor α in the optimum value threshold formula along with sub band variance estimation, the proposed technique yields significantly superior image quality by preserving edges and a better PSNR value. It is also observed that the

images corrupted with less noise densities, single level of decomposition is sufficient. While for images corrupted with higher noise density second level of decomposition is required irrespective of the images.

References

- [1] S.G.Mallat, "A Theory for Multiresolution Signal Decomposition: The wavelet Representation," IEEE trans. on PAMI, vol 11, July 1989, pp. 674-693.
- [2] P.Moulin and J.Liu, "Analysis of Multiresolution Image Denoising Schemes using Generalized-Gaussian and Complexity Priors," IEEE Trans. on Info. Theory. Vol 45, April 1999, pp 909-919.
- [3] L. Sendur, I. Selesnick, "Bivariate Shrinkage Functions for Wavelet-Based Denoising Exploiting Interscale Dependency," IEEE. Tran. Signal Processing, Vol. 50, No. 11, 2002, pp. 2744–2756.
- [4] V. Strela, J. Portilla, and E. Simoncelli, "Image Denoising using a Local Gaussian Scale Mixture Model in the Wavelet Domain," in Proc. SPIE 45th Annu. Meet., 2000.
- [5] H. Rabbani, M. Vafadust, I. Selesnick, S.Gazor "Image Denoising Based on a Mixture of Bivariate Laplacian Models in Complex Wavelet Domain," in Proc. IEEE International Workshop on Multimedia Signal Processing, Victoria, BC, Canada, 2006.
- [6] A. Achim and E. E. Kuruoglu. "Image Denoising using Bivariate Alpha-Stable Distributions in the Complex Wavelet Domain," IEEE Signal Processing Letters, 12, 2005, pp. 17-20.
- [7] L. Sendur and I. W. Selesnick, "Bivariate Shrinkage Functions for Wavelet-Based Denoising Exploiting Interscale Dependency," IEEE Trans. Signal Processing, Vol. 50, No. 11, Nov. 2002, pp. 2744–2756.
- [8] M. K. Mihailescu, I. Kozintsev, K. Ramchandran, P. Moulin, "Low Complexity Image Denoising Based on Statistical Modeling of Wavelet Coefficients," IEEE Signal Processing Letters, Vol. 6, Dec. 1999, pp. 300–303.
- [9] D. L. Donoho, I. M. Johnstone, "Ideal Spatial Adaptation by Wavelet Shrinkage," Biometrika, Vol. 81, No. 3, 1994.
- [10] D. L. Donoho, "De-noising by Soft-Thresholding," IEEE, Trans. Inform. Theory, Vol. 41, May 1995, pp. 613–627.
- [11] D.L.Donoho, "Denoising and Softthresholding" IEEE. Transaction information theory, Vol. 41, 1995, pp. 613 – 627.
- [12] D.L.Donoho, "Nonlinear Wavelet Methods for Recovery of Signals, Densities, and Spectra from Indirect and Noisy Data," in proc. of symposia in applied mathematics, Vol. 00, 1993, pp 173-205.
- [13] S.Zhong, and V.Cherkassy, "Image Denoising using wavelet thresholding and model selection," proc.IEEE.Int.Conf.on image processing.[ICIP].Vancouver.bc,Sept.2000.
- [14] D.L.Donoho, and I.M.Johnstone, "Adaptive to Unknown Smoothness via Wavelet Shrinkage," Journal of American statistical Assoc., Vol.90, No.90, 1995, pp.1200-1224.

- [15] D.L.Donoho, and I.M.Johnstone, "Ideal Spatial Adaptation via Wavelet Shrinkage," *Biometrika*, Vol.81, 1994, pp.425-455.
- [16] S. Grace Chang, "Adaptive Wavelet Thresholding for Image Denoising and Compression," *IEEE transactions on Image processing*, Vol.9, No.9 september.2000.
- [17] S.G.Chang, B.Yu, and Martin Vetterli, "Spatially Adaptive Wavelet Thresholding with Context Modeling for Image Denoising," *IEEE.trans.on image proc.*,VOL.9,no.9,PP.1522-1531,2000.
- [18] S.G.Chang, B.YU, and Martin Vetterli, "Bridging Compression to Wavelet Thresholding as a Denoising Method," in *proc. conf. information sciences systems*, Baltimore, MD, PP.568-573, 1997.
- [19] S.LoPresto K.ramchandran, and M.T.Orchard, "Image Coding Based on Mixture Modeling of Wavelet Coefficients and a Fast Estimation-Quantization Frame-Work", in *Data Compression Conference 97,(Snow-bird Utah)*,pp.221-230,1997.
- [20] Iman Elyasi, and Sadegh Zarmehi "Elimination Noise by Adaptive Wavelet Threshold" *World Academy of Science, Engineering and Technology* 56, 2009, pp.462-466.
- [21] David L.Donoho and Iain M.Johnstone, , "Adapting to Unknown Smoothness via Wavelet Shrinkage", *Journal of the American Statistical Association*, Vol.90, No432, 1995, pp.1200-1224.
- [22] Andrew Bruce, David Donoho and Hung-YeGao, "Wavelet Analysis", *IEEE Spectrum*, 1996, pp.27-35.
- [23] D.L.Donoho, 1995, "De-noising by Soft-Thresholding," *IEEE Trans.on Information Theory*, vol.41, No.3, 1996, pp.613-627.<http://wwwstat.stanford.edu/~donoho/Reports/1992/denoiser.elease3.ps>.
- [24] Arivazhagan. S., Deivalakshmi. S. and Kannan. K., Technical Report on Multi-resolution Algorithms for Image Denoising and Edge Enhancement for CT images using Discrete Wavelet Transform, 2006.
- [25] Raghuvir M.Rao and Ajit S.Bobadikar, "Wavelet Transforms: Introduction to Theory and Applications", Addison Wesley Longman Inc., 1998, pp.183-189.

Strategical Modelling with Virtual Competition for Analyzing Behavior of Malicious Node in Mobile Adhoc Network to Prevent Decamping

Anil G.N^a, Dr. A. Venugopal Reddy^b

^aAssistant Professor, Department of CSE, BMS Institute of Technology, Bangalore, India,

^bProfessor & Principal, University College of Engineering, Osmania University, Hyderabad, India,

Abstract: The proposed system highlights a novel approach of identifying dynamic misbehaviour of the malicious node in mobile adhoc network. The main aim of the proposed work is to implement a new technique in the attack scenario of distributed network of MANET in order to analyse the policies and patterns of attack originated from malicious node and their tendency to decamp to another new logical region from the old one after attack which is accomplished using Perfect Bayesian Equilibrium. The architecture is also designed in most challenging scenario as when the mobile node will decamp to new logical region it will erase its previous actions in old logical region, for which it will be most difficult task to identify the malicious node. Such situation also will give rise to false positive rate of detection by the regular nodes. Therefore a trust model is designed which is always updated by the regular node. The proposed simulation identifies an instance of decamping with estimation of improbabilities as well as various policies owned by the malicious node.

Keywords: Node Misbehaviour, Mobile Adhoc Network, Non-Cooperation, Routing Misbehaviour

I. INTRODUCTION

From the last decade there has been an extensive research [1] for the cause of secure routing in mobile adhoc network (MANET). Due to the inherent characteristics of dynamic topology, excessive energy consumption, difficult in taking decision for appropriate routing protocol and no certificate authorization in mobile adhoc network, it has pose a huge challenge in field of security of routing in MANET. Various prior researches has focused on different security protocols designed like reputation system, virtual currency etc. has been implemented to understand the cause of misbehaviour of mobile nodes. The MANET consists of regular as well as malicious node whose objective is to increase the

destruction of valuable resources in the MANET. The main focus of the proposed system will be:

- How to distinguish between the regular and malicious node in MANET environment based on their dynamic behaviour which is difficult to predict?
- What exactly happens when the any malicious node attempt to drop the packet, which is quite possible even by using reputation system used extensively in prior research work?
- In case of multiple-logical region of simulation, what if the malicious node generates an attack in one logical region and itself arrives in one new logical region by erasing the past identification of the malicious node?
- As malicious node will attempt to behave exactly like a regular node in distributed environment of MANET for gaining the trust, it will eventually violate the price-based system proposed in the many prior research work [2].
- Can the appropriate value of the probability of malicious node and regular node can be estimated in run-time?
- Can the uncertainty or the trust value can be estimated? Can the number of detected co-operation or attacks can be estimated accurately in run-time?

If the above mentioned parameters can be defined and computed, there is a possibility of creation of a robust and secure model for better analyzation of nodes strategy (Regular / Malicious) in the MANET system. The rest of this paper is organized as follows. We discuss Problem Statement in Section II. The related work is discussed in Section-III. Proposed System will be discussed in Section-IV. Methodology for the research work is elaborated in Section V. Implementation is described in Section-VI and performance analysis followed in Section-VII. Finally Conclusion is discussed in Section-VIII.

II. PROBLEM STATEMENT

The fundamental issue with frequently used routing protocols is that they rely all mobile nodes of network and based on the assumption that nodes will behave or cooperate properly but there might be a condition where some nodes are not behaving properly. Majority of the adhoc network routing protocols becomes inefficient and shows reduced performance while mitigating with big number of misbehaving nodes. Such set of misbehaving nodes support the flow of route discovery traffic but interrupt the data flow, causing the routing protocol to restart the route-discovery process or to select an alternative route if one is available.

The absence of an infrastructure and consequently the absence of authorization facilities impede the usual practice of establishing a line of defense-distinguishing nodes as trusted or non-trusted. Freely roaming nodes form transient associations with their neighbors: they join and leave sub-domains independently with and without notice. In such vulnerable situation, the compromised mobile node will definitely cause potential Byzantium failures in the routing protocols in MANET. In such type of Byzantine failure, a set of the mobile nodes could be attacked in a specific procedure that erroneous and malicious actions cannot be even identified.

The malicious node present has the policy of decamping to another logical region from their present one in order to circumvent of being trapped. The worst situation is when the malicious node decamps to a new logical region; they will tend to erase the previous history of their actions which will defiantly results in increase in false positive rates in updates in MANET. However, this supplementary approach does not confirm that malicious nodes will incessantly compromise other nodes present in the network and will attempt to escape since decamping is also connected with a cost (e.g., the power depleted to shift to the newly chosen destination).

III. RELATED WORK:

Recently, numerous approaches have been proposed to deal with the node non-cooperation problem in wireless networks. They generally can be classified into two main categories: reputation systems and price-based systems. We use a monitoring and reputation system [3] as the basic setting for regular nodes. Many related works also use reputation systems [4]–[6] and a game theory model [7] to

analyze the problem. Some recent works have studied the incentives for malicious nodes and modeled their behavior more rationally. In [8], Liu et al. present a general incentive-based method to model the attackers' intents, objectives, and strategies. In [9], Theodorakopoulos and Baras further study the payoff of the malicious nodes and identify the influence of the network topology. However, the good nodes' behavior in [10] is simple, and it fails to consider the possibility that an attacker might choose different attack frequencies toward different opponents.

The security problem and the misbehaviour problem of wireless networks including MANETs have been studied by many researchers e.g. [10], [11], [12], [13]. Various techniques have been proposed to prevent selfishness in MANETs. These schemes can be broadly classified into two categories: credit-based schemes and reputation based schemes. The basic idea of credit-based schemes is to provide incentives for nodes to faithfully perform networking functions. Nodes get paid for providing services to other nodes. When they request other nodes to help them for packet forwarding, they use the same payment system to pay for such services.

Sanjeev Rana [14] has created a mechanism with the help of which it prevents various replay attacks and also activate the neighboring nodes to control the behavior of its neighbors to thwart active attacks. Rakesh Kumar e.t. al [15] has implemented a prototype of key management service by using genetic algorithm. Rajib Das e.t. al [16] has proposed a solution against black hole attack and has illustrated the effect of black hole attack on network performance. However, the results cannot be considered as optimal. Kannan e.t. al [17] has published an extensive survey on various attacks, and their respective countermeasures with respect to vulnerability in routing protocols. Aishwarya Sagar [18] has proposed an approach based on reputation system that deals with routing misbehaviour and consists of identification and separation of misbehaving nodes. Hariharan e.t. al [19] has proposed a new technique termed as recommendation based on identification of routes with misbehaved nodes. Usman e.t. al [20] have analyzed the effects of different types of jammers using Conservation of Flow (CoF), which has been useful for detecting other attacks, in the wired networks. Abbas [21] have categorized reputation based schemes based on monitoring approaches: active and passive based acknowledgments. Finally, the authors have discussed their pros and cons as well as some other important identity related issues.

IV. PROPOSED SYSTEM

Although there considerable amount of work done in security of the routing protocols on mobile adhoc network, but the proposed system gives higher contrast result compared to all major previous work by considering the probabilistic approach of identifying the routing behaviour of the malicious node. One of the most significant criteria considered for the proposed system is analyzing and distinguishing the behaviour of either regular node or the malicious node. The proposed model will be composed for multiple logical regions where the mobile nodes will be distributed and will be addressed as an interactive virtual competition between regular node and malicious node as Perfect Bayesian Equilibrium.

Virtual Competition represents the implications of respective roles of individual actions for regular nodes and malicious nodes. In this proposed approach of virtual competition, the mobile nodes will scrutinize the results of each specific communication occurring. In the experiment, each mobile node will design a trust factor towards its neighboring nodes and update their trust information in accordance to the neighbor's actions as the virtual competition evolves. The best responses of the both regular and malicious node are governed by the threats about specific events from opposite mobile nodes which are completely dependent on their current trust level. The regular node configures a reputation threshold and decides the trust level and its threshold. Whereas the malicious nodes will also estimates the risk, which is computated by the feasibility that a regular node will decide to update other mobile nodes under that condition.

Depending on the threat level and expected decamping cost, the malicious node makes a decision on decamping to other logical region. The contributions of this proposed system are as follows: 1) We formulate a Perfect Bayesian Equilibrium to study the behavior policy of regular and malicious nodes in mobile adhoc network with respect to its routing; 2) we propose decision rules for regular nodes to update and malicious nodes to escape; 3) we study the equilibrium strategy profiles for both parties based on the trust and expected payoff and expose the association between nodes' superlative response and the cost and gain of each individual policy; and 4) we will present several countermeasures to restrict the decamping policy. The main intention of the proposed system will be to understand the condition of decamping to other logical region by malicious nodes, because the

behaviour of the malicious node will be programmed in such a way that whenever they will attempt to decamp to another logical region, it will almost erase the routing history of the previous logical region where it was prior residing.

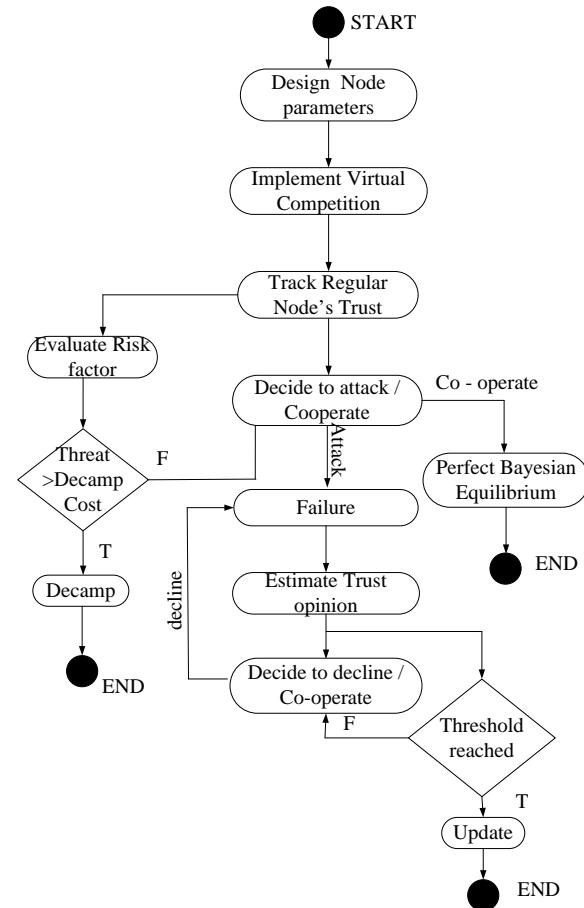


Fig 1. Flow of the Proposed Model

The flow diagram of the proposed system is shown in Fig 1. Prior researches in this field has not succeeded to consider the feasibility that an intruder might select different threat frequencies toward different opponents whereas the proposed project work considers more "Smart" malicious nodes, making the regular and malicious nodes' competition in this proposed model more realistic. This is the reason of deploying the proposed algorithm.

V. METHODOLOGY

The proposed contribution for project is to represent the regular / malicious node with virtual competition as a multi phase scheme to find the optimal policy of regular and malicious nodes for computing the

general decision process of regular and malicious mobile nodes. The neighboring monitoring policy will assist the regular node to receive the feedback from neighboring mobile node at that instant and computes the trust and adequacy of the proof towards the opposite mobile node depending on the quantity of the identifying cooperation and number of identifying attacks on routing. The proposed system also observes a threshold schema to choose whether to update other mobile nodes in the logical region or not. If not the regular node chooses to assist with the probability which is estimated depending on the trust level. Not only this, the malicious node also estimates the threat of being caught in its existing location, so it follows its protocol to decide whether it should decamp to another logical region or not. If not, the malicious mobile node will choose to attack. The prime issue in this decision process is the decision rules for both regular and malicious nodes and the event profiles shown by the probability that the regular node cooperates and the probability that the malicious node attacks. The proposed system analyzes the mobile adhoc network to identify the optimal decision protocols and events by deploying the framework which chooses to achieve perfect Bayesian Equilibrium.

The algorithms deployed are discussed below.

Algorithm: Mobility Model

Objective: The main objective of this algorithm is to estimate all the parameters (nodes, velocity of node, length and width etc) and it also assigns the new positions for the movement of the nodes from its old position.

Input: x-node, y-node, speed, length and width

Output: The program estimates all the parameters responsible for the mobility model of MANET.

Steps:

- 1 Initialize x-node, y-node, speed, length and width
- 2 Create a function for calculating parameters
- 3 Calculate first random position (r_1) for the node to travel
- 4 Calculate second random position (r_2) for the node to travel
- 5 Estimate new position of x-node (x-new)
- 6 Estimate new position of y-node (y-new)
- 7 If ($x\text{-new} < 0 \text{ || } x\text{-new} > \text{length}$)
- 8 Assign ($x\text{-node}-r_1$) to x-new
- 9 End
- 10 If ($y\text{-new} < 0 \text{ || } y\text{-new} > \text{width}$)
- 11 Assign ($y\text{-node}-r_2$) to y-new
- 12 End

Algorithm: Estimating Probability of Malicious Node

Objective: The main objective of the algorithm is to estimate the probability of malicious node in the MANET environment.

Input: N_c and N_a

Output: Calculation of the probability of malicious node.

Steps:

- 1 Initialize number of detected cooperation (N_c)
- 2 Initialize numbers of detected attacks (N_a)
- 3 Create a function for estimating probability of malicious node.
- 4 Initialize probabilities that the node is a malicious node (P_m)
- 5 Apply Formula:

$$P_m = N_a / (N_c + N_a)$$

Algorithm: Estimating Trust Factor

Objective: The main objective of the algorithm is to estimate the trust factor by considering number of detection cooperation, attacks, and improbability in the vulnerable environment of MANET

Input: N_c , N_a , I_o

Output: The program estimate the trust factor

Steps:

- 1 Initialize number of detected cooperation (N_c)
- 2 Initialize number of detected attacks or declines (N_a)
- 3 Initialize improbabilities in the opinion (I_o)
- 4 Create a function for calculating the trust system
- 5 Assign ($N_c/(N_c + N_a)$) to t_{v1}
- 6 Assign ($1-I_o$) to t_{v2}
- 7 Estimate trust (t) in the opinion by applying formula

$$t = t_{v1} + t_{v2}$$

Algorithm: Calculating improbability of opinion

Objective: The main objective of the algorithm is to estimate the improbability

Input: N_c , N_a , I_o

Output: The program gives the estimation of uncertainty.

Steps:

- 1 Initialize number of detected cooperation (N_c)
- 2 Initialize number of detected attacks (N_a)
- 3 Create a function for estimating improbability
- 4 Estimate improbability in the opinion (I_o) by formula:

$$I_o = 12 \times N_c \times N_a / \{(N_c + N_a)^2 \times (N_c + N_a + 1)\};$$

VI IMPLEMENTATION

The proposed system has considered a wide range of policies of intrusion in dynamically generated mobile adhoc network in Matlab platform. The regular node is considered to follow its respective neighboring transmitted message by neighbor monitoring. A simulation framework of 1000 x 1000 m is designed with 200 mobile randomly positioned with a transmission range of 300 meters. The entire simulation area is designed with 9 logical area. A mobility model is designed and any two nodes in the same cluster are considered as neighboring nodes.

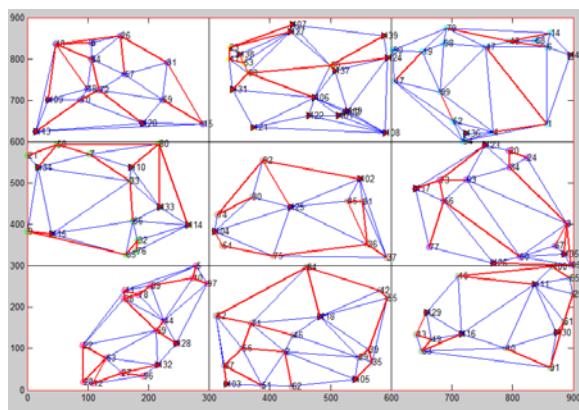


Fig 2. Nine logical region with 200 mobile nodes randomly distributed.

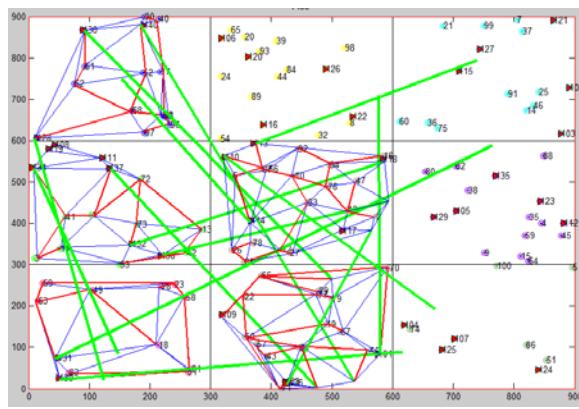


Fig 3. Nine logical region with 200 mobile nodes with an attempt of decamping.

Fig 2 represents an instance of simulation where 200 mobile nodes are randomly distributed in nine logical region. The red colored line shows the communication between two mobile nodes in the same logical region. The blue colored line represents probable communication link between two mobile nodes in the same logical region. In figure 3. The

green colored line represents the process of decamping from region of attack to a new logical region.

VII. PERFORMANCE ANALYSIS

The performance analysis is done by checking the outcomes of various phases of the virtual competition to understand the abnormal behaviour of the mobile nodes. Figure 4 shows the graphical representation of the normal node movement with respect to the simulation time. A polynomial trend line is used to understand the graphical representation.

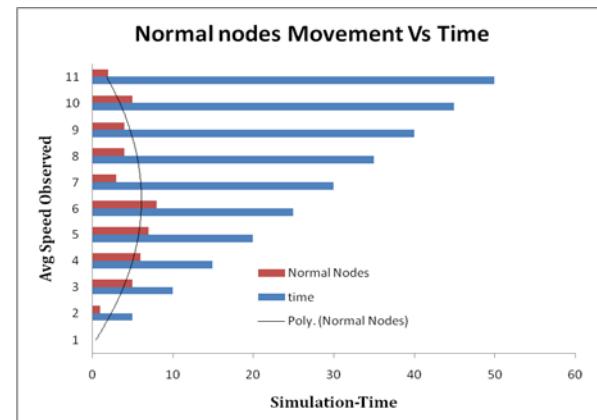


Fig 4. Analysis of Normal node movement Vs Time

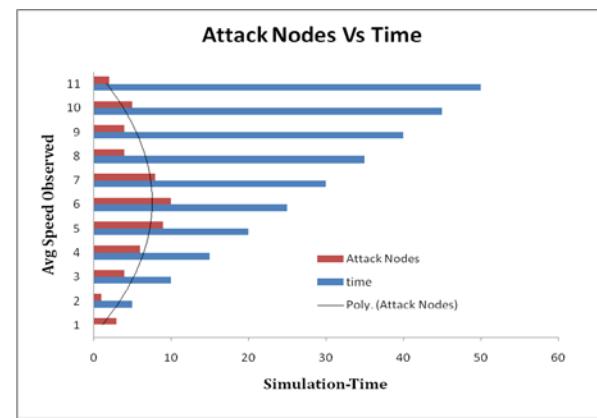


Fig 5. Analysis of Attack node movement Vs Time

In Fig-4 and Fig 5, trend line is almost similar, but a slight deviation on average speed at 5, 6, 7 seconds. Which represents that almost all the normal node and malicious node poses the same behavior with respect to mobility in mobile adhoc network topology, which will provoke the application very difficult situation to distinguish between the normal and attack nodes.

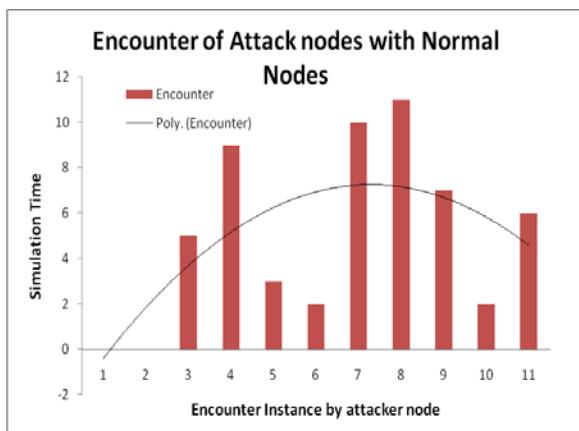


Fig 6. Analysis of Encounter of attack nodes with Normal node

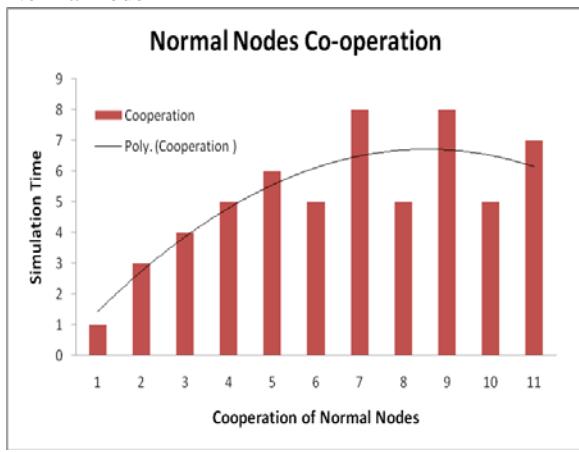


Fig 7. Analysis of Normal Nodes Co-operation

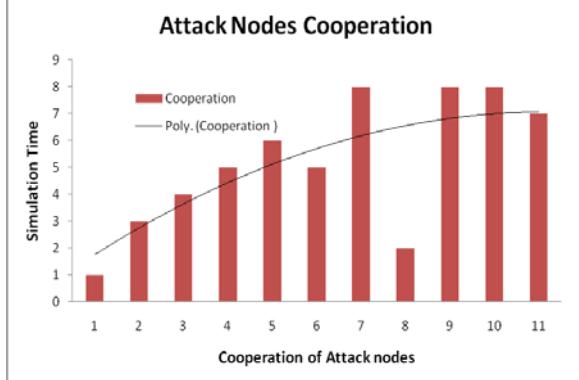


Fig 8. Analysis of Attack Nodes Cooperation

Fig-6 represents encounter of attack nodes with regular nodes. The graph represents very uneven variation proving difficult to predict the encountering strategy of attack nodes in the MANET environment. Fig-7 and Fig-8 represents cooperation instances of attack nodes and regular nodes. Comparison proves that malicious nodes try to duplicate the similar behavior of the regular node. The observation is carried out in 60 second interval. In case of continuing for another cycle of observation, it can be

observed that attack nodes have higher tendency of cooperation at the end of each simulation, which indirectly proves the detrimental strategy of infection by the malicious nodes.

The simulation results also highlight the various cooperation probabilities of the nodes in the MANET as shown in Figure 9.

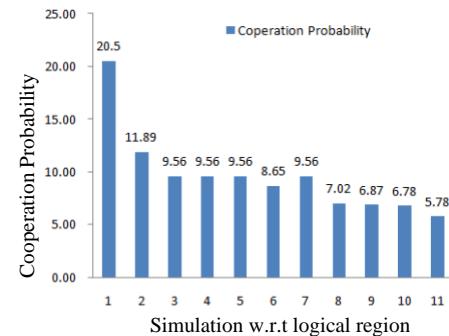


Fig 9. Graphical Representation of Cooperation Probability by nodes [regular / selfish].

This also highlights that when the simulation has been conducted in various stages with respect to every individual clusters defined, the co-operation factor decrease down which is an indication of advent of either selfish or malicious node. Once the attack has been created by any malicious node in a specific cluster, when it attempts to jump to another cluster, the proposed methodology also makes sure that any such attempt information will be instantly updated to all other clusters in the neighborhood for providing security against intrusion from such malicious nodes. With every communication the belief system and the trust opinion is gathered for estimating this information.

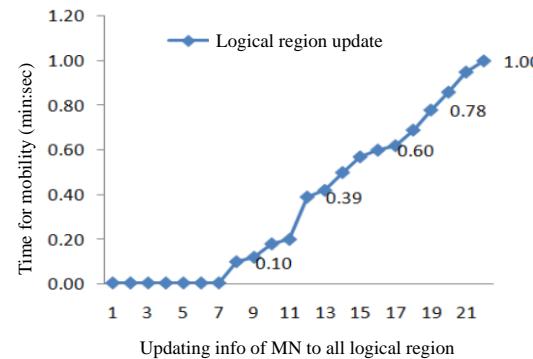


Fig 10. Cluster updates during attack by malicious nodes

Figure 10 represents that with every attempt of attack, the countermeasures provided to update the

attack information to the other cluster increase, which shows the robustness of the proposed methodology.

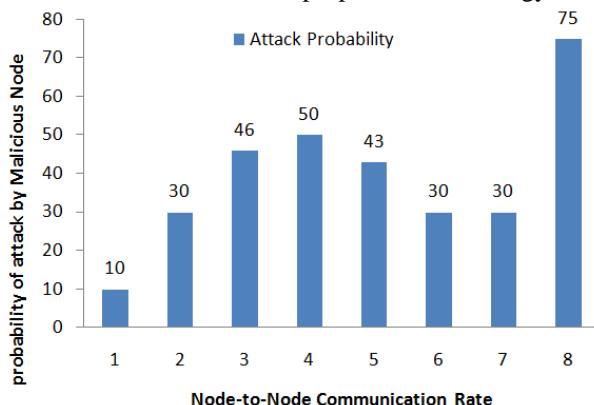


Fig 11 Attack probability by malicious node

One of the main attempts of the proposed system is to understand the misbehaviour of the nodes in the manet environment. In real-time scenario, it is almost difficult task to estimate time of the attack probability, which pose a great threat to the existing cluster as well as neighborhood clusters. So the figure 11 represents that with every node-to-node communication, there is a peak seen representing highest attack probability. This also represents that it is highly possible to understand the attack strategy if the simulation parameters are kept on changes based on multi-stages according to game theory

VII CONCLUSION

In this paper, we have discussed about a novel technique of analyzing the behaviour of the malicious node in the MANET. A framework is designed as virtual competition which is mapped with expected policies to be adopted by regular node to update and malicious node to attack using Perfect Bayesian Equilibrium. The simulation shows better accuracy of catching the malicious nodes, their behavior at every cycle, and their policy adopted to decamp to a new logical region.

ACKNOWLEDGEMENT

I'm thankful to the R&D division of CBK Infotech India (P) Ltd, Bangalore to conduct my experiment for validating simulation prototype.. I'm also thankful to Dr. G. Narsimha, Assistant Professor, JNTUH College of Engineering, Jagtial, Karimnagar, Hyderabad for his valuable guidance and inputs.

REFERENCE:

- [1] Sunita Sahu & Shishir K. Shandilya, "A comprehensive survey on intrusion detection in MANET", *International Journal of Information Technology and Knowledge Management*, July-December 2010, Volume 2, No. 2, pp. 305-310
- [2] Azadeh Omrani and Mehran S. Fallah, "Stimulating Cooperation in MANETs Using Game Theory", Proceedings of the *World Congress on Engineering 2007* Vol II, WCE 2007,
- [3] S. Buchegger and J. Boudec, "A robust reputation system for P2P and mobile ad-hoc networks," in Proc. 2nd Workshop Econ. Peer-to-Peer Syst., 2004, pp. 403–410.
- [4] F. Li and J. Wu, "Mobility reduces uncertainty in MANETs," in Proc. IEEE INFOCOM, 2007, pp. 1946–1954.
- [5] A. Josang, R. Ismail, and C. Boyd, "A survey of trust and reputation systems for online service provision," *Decis. Support Syst.*, vol. 43, no. 2, pp. 618–644, Mar. 2007.
- [6] F. Li, A. Srinivasan, M. Lu, and J. Wu, "Uncertainty mitigation for utility-oriented routing in MANETs," in Proc. IEEE GLOBECOM, 2007, pp. 427–431.
- [7] F. Li and J. Wu, "Hit and run: A Bayesian game between malicious and regular nodes in mobile networks," in Proc. IEEE SECON, 2008, pp. 432–440.
- [8] P. Liu, W. Zang, and M. Yu, "Incentive-based modeling and inference of attacker intent, objectives and strategies," *ACM Trans. Inf. Syst. Secur.*, vol. 8, no. 1, pp. 78–118, Feb. 2005.
- [9] G. Theodoropoulos and J. Baras, "Malicious users in unstructured networks," in Proc. IEEE INFOCOM, 2007, pp. 884–891.
- [10] Kejun Liu, Jing Deng, Pramod K. Varshney, and Kashyap Balakrishnan, "An Acknowledgment-Based Approach for the Detection of Routing Misbehaviour in MANETs", *IEEE Transactions on mobile computing*, Vol. 6, NO. 5, May 2007.
- [11] Dhanalakshmi, Dr.M.Rajaram , "A Reliable and Secure Framework for Detection and Isolation of Malicious Nodes in MANET", *IJCSNS International Journal of Computer Science and Network Security*, VOL.8 No.10, Oct2008
- [12] Zan Kai Chong, Moh Lim Sim, Hong Tat Ewe, and Su Wei Tan , " Separation of Detection Authorities (SDA) Approach for Misbehavior Detection in Wireless Ad Hoc Network",PIERS Online, VOL. 4, NO. 8, 2008.
- [13] S. Buchegger and J.-Y. Le Boudec, "Performance Analysis of the CONFIDANT Protocol: Cooperation of Nodes, Fairness in Dynamic Ad-Hoc Networks," Proc. MobiHoc, June 2002.

- [14] Sanjeev Rana, Manpreet Singh, "Performance Analysis of Malicious Node Aware Routing for MANET using Two-Hop Authentication", International Journal of Computer Applications (0975 – 8887), Volume 25– No.3, July 2011
- [15] Rakesh Kumar, Piush Verma, Yaduvir Singh, "Design and Development of a Secured Routing Scheme for Mobile Adhoc Network", International Journal of Computer Applications (0975 – 8887) Volume 13– No.2, January 2011
- [16] Rajib Das, Bipul Syam Purkayastha, Prodipto Das, "Security Measures for Black Hole Attack in MANET: An Approach", International Journal of Engineering Science and Technology (IJEST), 2011
- [17] S. Kannan, T. Maragatham, S.Karthik, V.P. Arunachalam, "A study of Attacks, Attack Detection, and Prevention Methods in Proactive and Reactive Routing Protocols", International Business Management, 2011
- [18] Aishwarya Sagar, Anand Ukey, Meenu Chawla, "Detection of Packet Dropping Attack Using Improved Acknowledgement Based Scheme in MANET", IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 4, No 1, July 2010
- [19] Sowmiya Hariharan, Jothi Precia, Suriyakala.C.D, Prayla Shyry, "A Novel Approach for Detection of Routes with Misbehaving Nodes in Manets", International J. of Recent Trends in Engineering and Technology, Vol. 3, No. 2, May 2010
- [20] Usman Yaseen, Ali Zahir, Faraz Ahsan, and Sajjad Mohsin, "Estimating the Effects of Jammers via Conservation of Flow in Wireless AdHoc Networks", International Journal for Advances in Computer Science, Volume 1, Issue 1, 2010
- [21] Sohail Abbas, Madjid Merabti, and David Llewellyn-Jones, "A Survey of Reputation Based Schemes for MANET", The 11th Annual Conference on The Convergence of Telecommunications, Networking & Broadcasting (PGNet 2010), Liverpool, UK, 21-22 June 2010

Authentication of DICOM Medical Images using Multiple fragile watermarking Techniques in Wavelet Transform Domain

A.Kannammal^{1*}, Dr.S.Subha Rani²

¹ Assistant Professor

Department of Electronics and Communication Engineering
PSG College of Technology, Peelamedu, Coimbatore-641004
Tamil Nadu,India.

Professor and Head

Department of Electronics and Communication Engineering
PSG College of Technology, Peelamedu, Coimbatore-641004
Tamil Nadu,India.

ABSTRACT

A multiple, fragile image authentication scheme is proposed for DICOM images using discrete wavelet transform. This scheme addresses critical health information management issues, including source authentication, data authentication and transfer of patient diagnosis details. The robustness of the method is enhanced through a form of hybrid coding, which includes repetitive embedding of BCH encoded watermarks. The watermarked images were tested with common attacks to evaluate the behaviour of the algorithm. Conclusions were drawn based on the algorithms performance when the images were subjected to various attacks. The algorithm was also tested by changing the wavelet function used in the algorithm and the results show that the Haar wavelet is the most suitable wavelet. The experimental results on different medical images demonstrate the efficiency and transparency of the watermarking scheme, which fulfils the strict requirements concerning the acceptable alterations of medical images.

Key Words- DICOM images, multiple watermarking, discrete wavelet transform.TAF.

1. Introduction

Digital watermarking has been used to increase medical image security, confidentiality and integrity. Watermarked medical images should not differ perceptually from their original counterparts. Digital data or images can be easily manipulated without leaving any trace of modification. Image authentication is the process of giving a legal validity to the data. By authentication technique, content tampering can be detected and we can indicate the true origin of the data [1]. Recent innovations in information and communication technologies have led to a new era in healthcare delivery and medical data management [2]. New challenges have arisen as

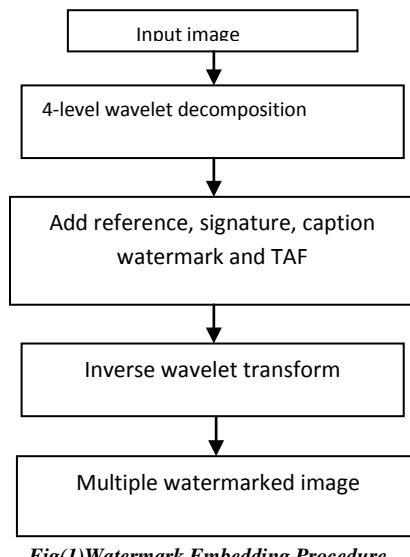
a result of easier access and distribution of digital data, especially regarding security of sensitive medical information. The research community seeks complementary solutions to confront these challenges and to effectively deal with a range of substantial healthcare information management issues. Image authentication can be done using watermarking or encryption schemes.

The work is to watermark medical images [3] in an irreversible way and to use the watermarked data for authentication. The medical images used will be in DICOM (Digital Imaging and Communications in Medicine) format. Multiple watermarks [4] will be embedded in the image for source authentication, data authentication and transfer of patient diagnosis details. The watermarked images will be tested with common attacks to evaluate the behaviour of the algorithm. Conclusions will be drawn based on the algorithms performance when the images are subjected to various attacks. The performance of the algorithm will be measured using Peak Signal to Noise Ratio (PSNR) and a Tamper Assessment Factor (TAF). A good image authentication technique should satisfy the following criteria: It should be invertible and should support all data formats. It should be robust against incidental image processing operations, but it must detect malicious tampering. The test images used are 256X256 size medical (DICOM) images.

2. Proposed algorithm

2. 1 .Watermark Embedding Algorithm:

In this method multiple watermarks such as, the reference watermark, TAF[6] of the watermarked image, patient diagnosis information and the physician's signature are embedded into wavelet transform domain.. The reference watermark is a fixed bit pattern that is used for tamper proofing or image authentication. The TAF value and patient diagnosis information act as caption watermark which gives additional information about the image. The physician signature is used for source authentication. All the data to be watermarked are represented in ASCII values. These values are BCH coded and embedded in many places to give additional robustness for the watermarking scheme. The steps involved in embedding the watermark are shown in Fig.1.



Initialization:

- Given $f(m,n)$ – image to be watermarked.
- Given $w(m,n)$ – watermark information to be embedded
 - The watermark information consists of the patient information, physician's signature and the TAF value of the watermarked image. All the watermark information is represented using ASCII values and BCH encoded.
 - These watermarks are improving the robustness and also used to identify the tampered images

Step 1: Decomposing the image into Wavelet Coefficients

Perform the 4 level discrete Haar wavelet transform [7,8] on the image $f(m,n)$ to produce 12 detail coefficient images $f_{k,l}(m,n)$ where $k = h, v, d$ (horizontal, vertical or diagonal detail coefficient) and $l = 1, 2, 3, 4$ is the particular detail coefficient resolution level, and a gross approximation at the lowest resolution level $f_{a,4}(m,n)$. That is,

$$\{f_{k,l}(m,n)\} = \text{DWT}_{\text{Haar}}[f(m,n)] \quad (1)$$

Step 2: Embedding watermark

The multiple watermark such as patient information(PI),physician's signature(PS),Tamper assessment Factor(TAF).The reference watermark embedding locations in a four level wavelet decomposed image is shown in Fig.(5).All the data to be watermarked are represented in ASCII values. These values are BCH coded and embedded in many places to give additional robustness for the watermarking scheme. Perform quantization of the wavelet coefficients in places where watermark is to be embedded using the quantization function

$$Q(f) = \begin{cases} 0, & \text{if } r\Delta \leq f < (r+1)\Delta \text{ for } r = 0, \pm 2, \pm 4, \dots \\ 1, & \text{if } r\Delta \leq f < (r+1)\Delta \text{ for } r = \pm 1, \pm 3, \pm 5, \dots \end{cases} \quad (2)$$

where Δ - quantization parameter (positive real number)

- If $Q(f_{k,l}(m,n)) = w(i)$, then no change in the coefficient is required.
- Otherwise, change $f_{k,l}(m,n)$ such that we force $Q(f_{k,l}(m,n)) = w(i)$ using the following assignment:

$$z_{k,l}(m,n) = \begin{cases} f_{k,l}(m,n) + \Delta & \text{if } f_{k,l}(m,n) = 0 \\ f_{k,l}(m,n) + 2\Delta & \text{if } -\Delta < f_{k,l}(m,n) < 0 \\ f_{k,l}(m,n) - 2\Delta & \text{if } 0 < f_{k,l}(m,n) < \Delta \\ f_{k,l}(m,n) - \Delta & \text{if } f_{k,l}(m,n) \geq \Delta \end{cases} \quad (3)$$

Step 3: Getting the Watermarked image

Perform the 4th level inverse discrete Haar wavelet transform on the transformed wavelet coefficients $\{z_{k,l}(m,n)\}$ to produce the watermarked image $z(m,n)$. That is,

$$z(m,n) = IDWT_{Haar} [\{z_{k,l}(m,n)\} - (4)]$$

for $k = h, v, d, a$ and $l = 1, 2, 3, 4$.

3. Testing for various attacks

The water marking procedure was done for twelve radiological images with different anatomy. The effect of various commonly occurring attacks [9] like Modification, Rotation and cropping, Brightness and Contrast adjustment, noises,[10] was observed. The various attacks for which the authentication effects were observed are listed below:

1. **Modification:** In modification, small area is removed from the image completely. The effect of this attack is tested at the center, top left, top right, bottom left and bottom right of the image.
 2. **Rotation and Cropping:** The rotation operator performs a geometric transform which maps the position (x_1, y_1) of a picture element in an input image onto a position (x_2, y_2) in an output image by rotating it through user-specified angle θ about an origin o. In most implementations, output locations (x_2, y_2) which are outside the boundary of the image are ignored. The image was rotated for the angles 15° , 30° , 45° and 90° . The cropped corners after rotation were lost in the process.
 3. **Brightness adjustment:** The brightness attack was performed using the formula,
- $$f(x, y) = g(x, y) + \beta \quad \text{---(5)}$$
- The various values added for testing are $\beta = 35, 600, 2500, 10000$.
4. **Contrast adjustment:** The contrast attack was performed using the formula,
- $$f(x, y) = g(x, y)^\gamma \quad \text{---(6)}$$

The various exponents used were,
 $\gamma = 0.5, 0.8, 1.2, 3.0$ when $\gamma < 1$, the image contrast decreases otherwise it increases.

5. **Horizontal flip and Vertical flip**
 Here the image is flipped about its horizontal axis by interchanging the rows above and below the horizontal axis. Here the image is flipped about its vertical axis by interchanging the columns to the left and right of the vertical axis
6. **Blurring:**
 Here a frequency domain Gaussian Low

Pass Filter with standard deviation of 0.2, 0.5, 0.8 and 2 were used to blur the edges of the image.

7. Sharpening

The sharpening of the image was carried out using Laplacian high pass filter with standard deviation of 0, 0.3, 0.7 and 1. The edges of the sharpened image were enhanced during the process.

8. Averaging filter

Averaging filter was implemented using two dimensional spatial masks of sizes: 3×3 , 5×5 , 7×7 and 9×9 . When size of the mask is increased the blurring caused also increases.

9. Addition of Salt & Pepper Noise

Salt and pepper noise is also known as impulse noise or shot noise. Salt and pepper noise with variance of 0.01, 0.05, 0.1 and 0.5 were added to the image

10. Addition of Gaussian Noise

Gaussian Noise is the most common form of noise encountered in most of the communication channels. The Gaussian noise with variance of 0.01, 0.05, 0.1 and 1 were added to the image.

11. High Boost filtering

In high boost filtering a blurred version of the image is subtracted from the image to obtain a sharpened image. The weight given to the center pixel in the mask is varied using an alpha factor which is given the values 0.2, 0.5, 0.8 and 1.

4. Quality measure of proposed algorithm:

Any algorithm should be evaluated on basis of certain quality measures. Here we have used two quality measures: Peak Signal to Noise Ratio (PSNR) and Tamper Assessment Factor (TAF).

4.1. Peak Signal to Noise Ratio (PSNR)

PSNR is the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. It is most easily defined via the mean squared error (MSE) for which two $M \times N$ images f and z where one of the images is considered a noisy approximation of the other is defined as:

$$MSE = \frac{1}{M \times N} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (f(x, y) - z(x, y))^2 \quad \text{---(7)}$$

The PSNR is defined as:

$$PSNR = 10 \times \log_{10} \frac{MaxBits^2}{MSE} dB \quad \text{---(8)}$$

Here, *MaxBits* is the maximum possible pixel value of the image. When samples are represented using *B* bits per sample, *MaxBits* is $2^B - 1$.

4.2. Tamper Assessment Factor (TAF):

It gives the difference between the actual embedded watermark and the reconstructed watermark. It is given by the expression

$$TAF(w, \tilde{w}) = \frac{1}{M \times N} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} w(x, y) \oplus \tilde{w}(x, y) \quad \text{---(9)}$$

where *M, N* - dimensions of the image

w - embedded watermark

\tilde{w} - reconstructed watermark

\oplus - denotes exor operation

The value of TAF ranges between zero and one.

5. Results and Discussion about various attacks and wavelet transforms:

The water marking procedure was done for different radiological images and the overall results were tabulated in Table.2. Figure(2) shows the original Image. Figure(3) shows the four level wavelet decomposed image. Figure(5) shows the location of four different watermarks (TAF, physical signature, patient information, reference watermark) in four level wavelet coefficients. After applying all the watermarks in the wavelet coefficients the inverse wavelet transform are applied, then the multiple watermarked image was obtained is shown in Figure (4). Then the watermarked image is analysed for different attacks and the results are shown in Table.1

5.1. Different Attacks

1. *Modification attack*: The embedded information was retrieved in most of the cases and the average TAF was at least 19% more than the actual TAF of the watermarked image.
2. *Rotation and Cropping attack*: The embedded information was lost completely when the image is subjected to Brightness attack
3. *Brightness attack*: The embedded information was retrieved when Beta is between 100 and 4000. TAF was at least

70% more than the actual TAF of the watermarked image.

4. *Contrast attack*: The embedded information was lost completely when the image is subjected to Contrast attack.
5. *Flip*: The flipped image when seen by a person may be mistaken to be the actual radiological image that was taken. But on retrieving the embedded watermark we can observe the effect of flipping very clearly.
6. *Blurring*: The reference watermark is modified completely even though Blurring of the image is very less and the image cannot be differentiated from the actual image by the human eye.
7. *Sharpening*: The Sharpened image may appear to be the actual image but on seeing the recovered reference watermark we can say that the image has been subjected to modification.
8. *Averaging*: The reference watermark is modified completely even though Blurring of the image is very less and the averaged image is difficult to differentiate from the original image by the human eye.
9. *Salt and pepper Noise*: The embedded reference watermark is lost completely when the image is corrupted by Salt & Pepper Noise.
10. *Gaussian Noise*: The embedded reference watermark is lost completely when the image is corrupted by Gaussian Noise.
11. *High boost filtering*: The Laplacian sharpened image can be easily differentiated by the human eye. The drastic increase in TAF and decreased PSNR further confirm that the image has undergone some modification.

5.2 .Performance of the algorithm for various wavelet functions

The embedded TAF value was retrieved in most of the cases is shown in Table.3.. The TAF was retrieved for all cases when using Haar, Symlet and Integer wavelet. But the patient information and the physician signature were retrieved only when using the Haar wavelet in the algorithm. Thought the PSNR of the image watermarked using Integer wavelet transform is more than the PSNR of watermarked image using Haar wavelet, the TAF is considerably high. So from the results we can say that the Haar wavelet is best suitable for algorithm used for watermarking the DICOM images.



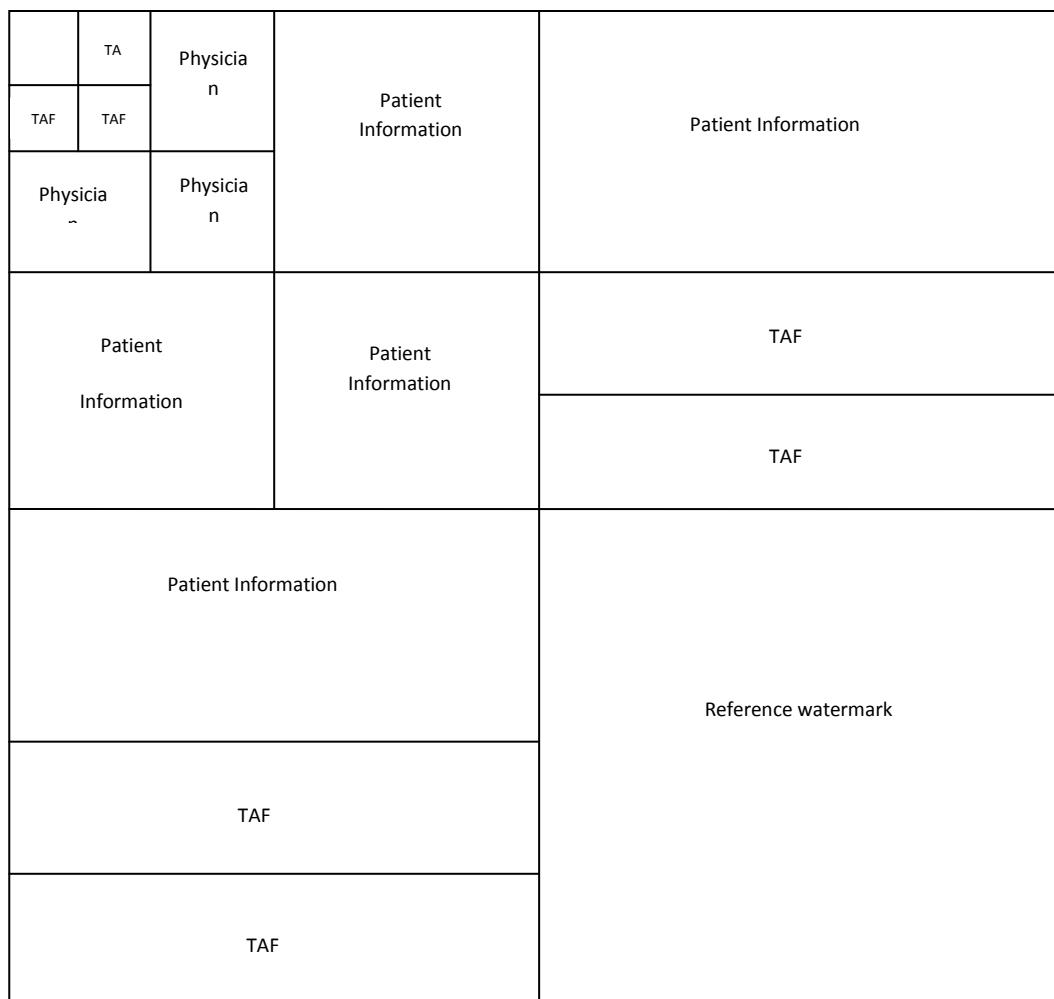
Fig(2) Original Image



Fig(3) Four level Wavelet decomposed image

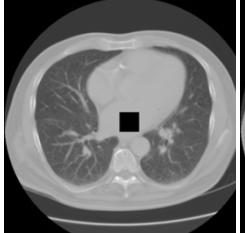
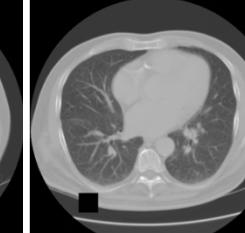
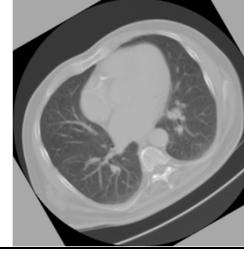
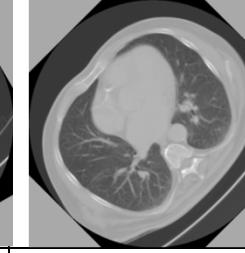
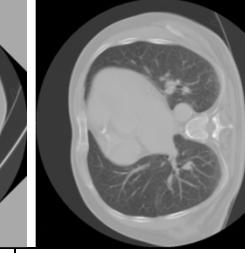
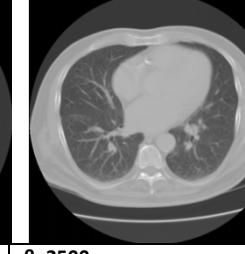
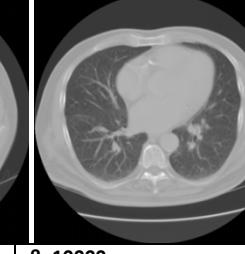
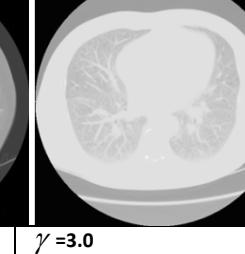


Fig (4) Multiple watermarked image



Fig(5) Watermark embedding locations in a 4-level wavelet decomposed image

Table.1.Output of Different attacks outputs

Modification attacks				
				
At center	At top left	At top right	At bottom left	At bottom Right
Rotation and cropping				
	$\theta=15$	$\theta=30$	$\theta=45$	$\theta=90$
Brightness				
	$\beta=35$	$\beta=600$	$\beta=2500$	$\beta=10000$
Contrast				
	$\gamma =0.5$	$\gamma =0.8$	$\gamma =1.2$	$\gamma =3.0$
Flip		Horizontal		Vertical

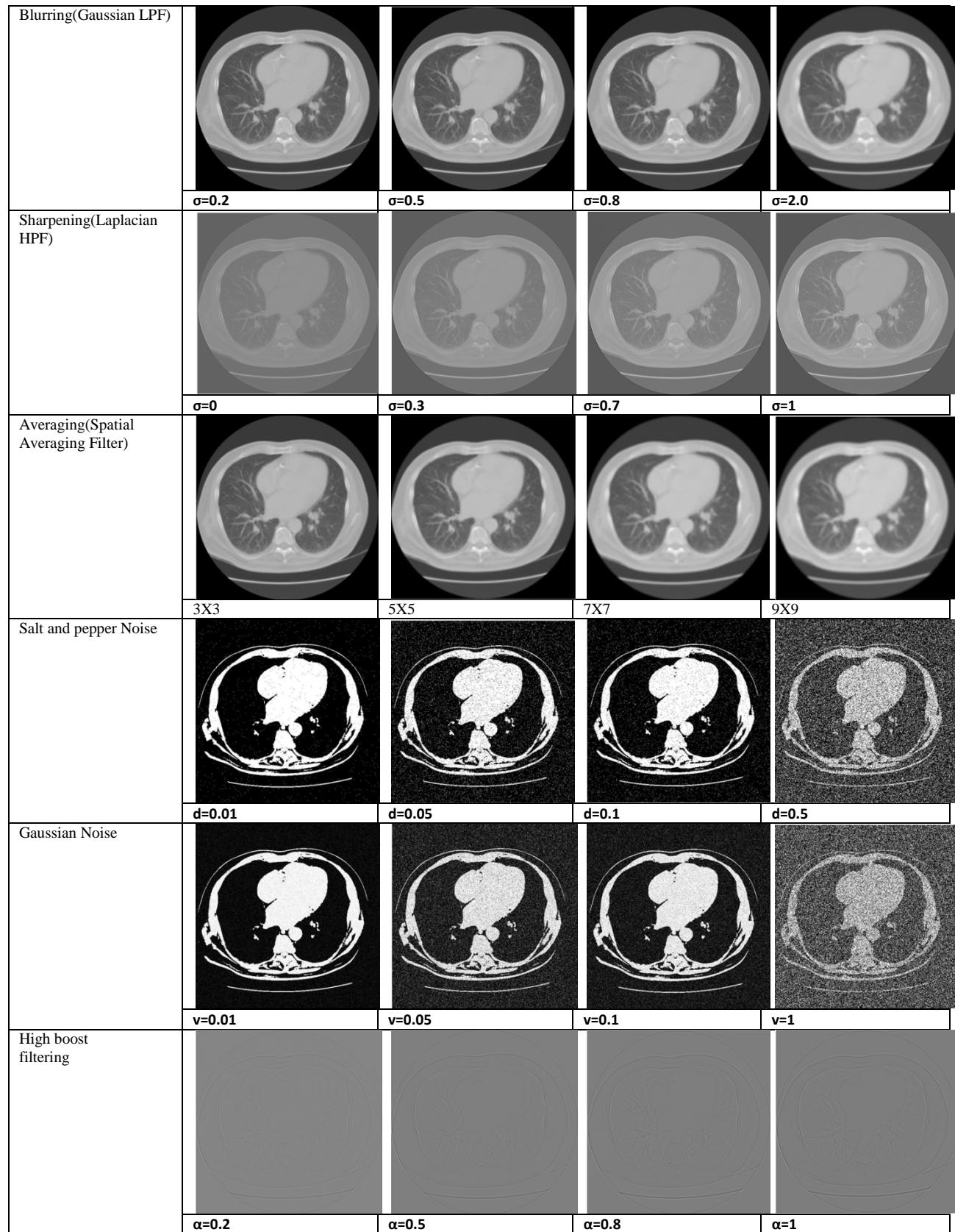


Table2.Anaylsis of Different attacks and different watermarks

Attacks	Types	Average TAF	Average PSNR	Patient info	Physician sign	Embedded TAF
Modification	At center	0.0084	55.58894167	Retrieved	Retrieved	Retrieved
	At top left	0.0084	59.18993333	Retrieved	Not Retrieved	Retrieved
	At top right	0.0083	59.30213	Retrieved	Retrieved	Retrieved
	At bottom left	0.0083	58.20526	Retrieved	Retrieved	Retrieved
	At bottom right	0.0083	58.54638	Retrieved	Retrieved	Retrieved
Rotation and Cropping	$\theta=15$	0.1406	42.49909	Error	Error	Error
	$\theta=30$	0.1406	40.86071	Error	Error	Error
	$\theta=45$	0.1405	40.28192	Error	Error	Error
	$\theta=90$	0.1407	43.19115	Error	Error	Error
Brightness	$\beta=35$	0.0134	65.4288	Not Retrieved	Not Retrieved	Retrieved
	$\beta=600$	0.0142	40.76633	Retrieved	Retrieved	Retrieved
	$\beta=2500$	0.009	28.37064	Retrieved	Retrieved	Retrieved
	$\beta=10000$	0.0241	16.32948	Not Retrieved	Not Retrieved	Not Retrieved
Contrast	$\gamma=0.5$	0.14	39.35022	Error	Error	Error
	$\gamma=0.8$	0.1398	41.35969	Error	Error	Error
	$\gamma=1.2$	0.1229	30.87605	Error	Error	Error
	$\gamma=3.0$	0.1245	86.40327	Error	Error	Error
Flip	Horizontal	0.086392	44.66864	Error	Error	Error
	Vertical	0.068108	48.08443	Error	Error	Error
Blurring(Gaussian LPF)	$\sigma=0.2$	0.14	39.35022	Error	Error	Error
	$\sigma=0.5$	0.1398	41.35969	Error	Error	Error
	$\sigma=0.8$	0.1229	30.87605	Error	Error	Error
	$\sigma=2.0$	0.1245	86.40327	Error	Error	Error
Sharpening(Laplaci an HPF)	$\sigma=0$	0.1478	53.81487	Error	Error	Error
	$\sigma=0.3$	0.10985	55.48749	Error	Error	Error
	$\sigma=0.7$	0.149167	56.78416	Error	Error	Error
	$\sigma=1$	0.115308	57.33012	Error	Error	Error
Averaging (Spatial averaging filtering)	hsizen=3	0.207967	65.837	Error	Error	Error
	hsizen=5	0.213008	61.77748	Error	Error	Error
	hsizen=7	0.210958	59.28639	Error	Error	Error
	hsizen=9	0.21455	57.57739	Error	Error	Error
Salt and pepper noise	d=0.01	0.2203	39.22999	Error	Error	Error
	d=0.05	0.2203	39.22978	Error	Error	Error
	d=0.1	0.2203	39.22954	Error	Error	Error
	d=0.5	0.2203	39.22757	Error	Error	Error
Gaussain Noise	v=0.01	0.2203	39.22964	Error	Error	Error
	v=0.05	0.2203	39.22915	Error	Error	Error
	v=0.1	0.2203	39.2288	Error	Error	Error
	v=1	0.2203	39.2269	Error	Error	Error
High Boost filtering	$\alpha=0.2$	0.111525	38.86525	Error	Error	Error
	$\alpha=0.5$	0.137925	38.91058	Error	Error	Error
	$\alpha=0.8$	0.139192	38.93458	Error	Error	Error
	$\alpha=1.0$	0.149333	38.94469	Error	Error	Error

Table .3.Performance of the algorithm for various wavelet functions

Wavelet function	Average TAF	Average PSNR	Embedded TAF	Patient info	Physician sign
Haar	0.0075	87.44158	Retrieved all	Retrieved all	Retrieved all
Biortogonal	0.0379	85.57798	Retrieved	Error	Error
Coiflet	0.0507	86.32846	Retrieved	Error	Error
Symlet	0.0354	86.5734	Retrieved	Error	Error
Meyer	0.0776	86.89085	Retrieved	Error	Error
Integer	0.0116	90.47334	Retrieved	Error	Error

6. Conclusion

Almost all the images responded in a similar way to the various attacks. The embedded watermarks were retrieved when the image is subjected to Modification and Brightness attack, within certain limits. But the watermarks were completely destroyed in Rotation and Cropping and Contrast Attacks. In the Modification and Brightness attack the embedded TAF value was found to be significantly higher than the retrieved TAF value from the tampered image. So we can conclude that a image is tampered when the reconstructed TAF is greater than the actual embedded TAF, as in Modification and Brightness attack or when the TAF value is not at all recovered from the image, as in Contrast, Rotation and Cropping attack.

References

- [1]José Alberto Martínez Villanueva, Claudia Feregrino Uribe and Jezabel Z. Guzmán Zavaleta, “Watermarking algorithms analysis on radiological images,” *International Conference on Electrical Engineering, Computing Science and Automatic Control*, pp. 298-303, 2008.
- [2]A. Giakoumaki, S. Pavlopoulos, D. Koutsouris “Secure and efficient health data management through multiple watermarking on medical images” *Med Bio Eng Comput* pp. 619–631, 2006.
- [3]M. Kutter, F. Hartung, “Introduction to watermarking techniques. In: Katzenbeisser S, Petitcolas FAP (eds) Information hiding techniques for steganography and digital watermarking,” Artech House, Norwood pp 97–120, 2000
- [4]C.S. Woo, J. Du, and B. Pham “Multiple Watermark Method for Privacy Control and Tamper Detection in Medical Images” *WDIC2005 pages* pp. 59-64, Australia, February, 2005.
- [5]J. Nayak, P. Subbanna Bhat, R. Acharya U and Niranjan UC “Simultaneous storage of medical images in the spatial and frequency domain: A comparative study” *BioMedical Engineering OnLine*, pp. 3-17, 2004.

[6]D. Kundur, D. Hatzinakos, “Digital watermarking for telltale tamper proofing and authentication,” *Proc IEEE Vol. 87*, pp. 1167–1180, 1999.

[7]A. Giakoumaki, S. Pavlopoulos, D. Koutsouris “A medical image watermarking scheme based on wavelet transform” *IEEE EMBS Annual International Conference, Cancun, Mexico* pp. 17-21, September, 2003

[8]P. Meerwald, A. Uhl, “A survey of wavelet-domain watermarking algorithms”. *Proc of the SPIE security and watermarking of multimedia contents*, Vol. 4314, San Jose, pp. 505–516, 2001.

[9]D. Kundur, D. Hatzinakos, “Diversity and attack characterization for improved robust watermarking”, *IEEE Trans Signal Process Vol.49*, pp. 2383–2396, October, 2001

[10]W. Puech, JM. Rodrigues, “A new crypto-watermarking method for medical images safe transfer,” *Proc of the 12th European signal processing conference, Vienna, Austria*, pp. 1481–1484, 2004.

Mrs.A.Kannammal registered Ph.D. in Engineering from Anna University coimbatore in the year 2008, M.E. degree from Thiagarajar college of Engineering , Madurai,Tamilnadu,India in 2004. B. E degree in Electronics and communication engineering, in 2002. In 2004, she joined the Deptt. of ECE,Thiagarajar college of Engineering as Lecturer. She joined PSG College of Technology, India as a Lecturer in the Deptt. of Electronics and Communication in 2007. She is presently working as Assistant Professor in the Deptt. of Electronics and Communication Engineering at PSG College of Technology ,Coimbatore,Tamilnadu, India. Her interests include, Medical Image Processing and Soft computing Techniques. She has contributed 6 technical papers in various journals and conference.

Dr.S.Subha Rani received Ph.D. in Engineering from Bharathiar University in the year 2000. M.E. degree from Anna University, in 1987. B. E degree in Electronics and communication engineering, in 1984. In 1988, she joined the Deptt. of Electronics and communication engineering as Lecturer. She is presently working as Professor and Head in the Deptt. of Electronics and Communication Engineering at PSG College of Technology, Tamil Nadu, India. Her interests include, Optimal Control, Wireless systems.MEMS and Soft computing Techniques. She has contributed more than 67 technical papers in various journals and conference. She is a life member of IEEE and IETE, she has completed 04 sponsored projects.

E-Business: Application of software and technology in selected Ethiopian Banks: Issues and challenges

Bhaskar Reddy Muvva Vijay¹

Tewdros Sisay Asefa²

¹ Department of Computer Science and Information Systems, Ethiopian Institute of Technology, Mekelle University
Mekelle, Tigray Region, Ethiopia

² Department of Computer Science and Information Systems, Ethiopian Institute of Technology, Mekelle University
Mekelle, Tigray Region, Ethiopia

Abstract

The application of software and technology is inevitable in the present competitive banking industry. Ethiopian banking system is one of the most underdeveloped compared to the rest of the world. In Ethiopia cash is still the most dominant medium of exchange and Electronic-banking is not well known, let alone used for transacting banking issues. The article tries to examine specific issues and challenges in Ethiopian banking system. The present study further highlights various selecting issues and service quality systems practiced in selected banks. Survey method is used to collect data from practicing managers of banking industry. It further analyzes the present obstacles and hindrances in improving and facilitating the present banking services with application of customize software and technology.

Keywords: Application software, Information and communication Technology, Security Issues, Electronic-banking system.

1. Introduction

Ethiopia is the only African country to have without a meaningful colonial past. Still it is one of the poorest countries in the world with an estimated per capita income of just \$203 (IMF 2007 cited by the Financial Standards Foundation). The country's economy is heavily dependent on agriculture (45% according to the CIA World Fact book) but the sector has been buffeted for decades by drought, poor cultivation practices, and a border war with Eritrea. Under Ethiopia's constitution, the state owns all land, providing 99-year leases. This system continues to hamper growth in the industrial sector as entrepreneurs are unable to use land as collateral for loans. While measures have been taken to stimulate private sector enterprise (e.g. simplifying red tape, clarifying the rules that govern business activities and shortening the time required in

obtaining required licenses), the government still maintains a major role in the economy. Indeed, several sectors of the economy are reserved solely for the government. These including banking, insurance, broadcasting, high capacity air transportation, motels, saw mills, movie theatres, travel agencies, domestic baked good, raw coffee export, retail and wholesale trade, brokerage services and shipping. The Ethiopian banking history as whole has a network of 521 branches at the end of fiscal period ended in June 2010. Which is lowest compared to the size of the country (1.1million square km) and number of population (78 million) and this shows that the number of population being served by a single branch stood at around 149,172. Internet is the driving force in the 21st century. Due to pervasive and steady growth of information and communication technology, the world banking industry is entering into new phenomena of unprecedented form of competition supported by modern information and communication infrastructure. E-commerce has become a buzzword for companies over a couple of years with increased awareness about the use of computers and internet. Information and communication technology applications are paramount concerns to the banks in today's business environment and internet has become a major platform for all financial, banking and commercial transactions. Statistics shows Ethiopia is lagging behind in the adaption of E-commerce. The software used in the Ethiopian banking industry is almost same as the rest of the African nations like South Africa, Egypt, and Tunisia. The internet infrastructure only in their major cities, Due to lack of internet facility it is very difficult to engage E-commerce activities in the Sub-urban and rural areas. Now a days some of the banks providing best services to their customers with advanced software and technology, but they are suffering with security Issues. Ethiopian banking system is comparatively good, but the Electronic banking

system is still not at in use. Thus this study is conducted with the following objectives.

- To describe the Historical development of Ethiopian banking system.
- To describe and differentiate between different Software's used in Ethiopian banks.
- To analyze the status of E-banking in Ethiopia.
- To investigate the main challenges and opportunities for software's and E-banking.
- To recommend appropriate actions to be taken to improve software's and E-banking in Ethiopia.

2. Review of Literature

Building software is hard. Building E-business applications even harder. Current E-business represents a hybrid of distributed, client-server, concurrent and networked systems. All these systems require the solving hard problems. Since E-business lays on the front line of modern business, it also requires dealing with scalability, high-availability and fail-over. To perform these operations banking industry need good software. According to Astley (2001) "software executing on distributed systems represents a unique synthesis of application code and code for managing requirements such as heterogeneity, scalability, security and availability". Jensen (2003), feel that most of the developing countries in Africa lacking with Internet infrastructure, so it's difficult to improve E-banking and Mobile banking. According to (ITU4-2006), the ICT itself to commit fraud and other cyber crimes. According to Gradchew Worku (2010), the Electronic banking in Ethiopia is facing lot of challenges due to lack of software, awareness, fear of risk and lack of trained persons in the key organizations. According to Dragos A. Manolescu and Adrian E. Kunzle (2011), they are giving a first step towards the direction of Patterns for E-business application development, will make developers aware of hard problems that they need to deal with and show them ways to solve them. According to Financial security authority (2004), it is very difficult to maintain electronic transactions without proper software. According to Magembe S and Shemi A (2002), Adoption of E-commerce in the developing countries tends to agree with the theory of planned behavior, but attitude seems to weigh more than subjective manner and perceived behavioral control. According to R.K.Mishra and J.Kiranmai (2009), presents case study approach has been used to compare various banks for rendering different internet banking services to its customers. According to Ole hanseth (2002), propose a design theory that tackles dynamic complexity in the design for Information Infrastructures (IIs) defined as a shared, open, heterogeneous and evolving socio-technical system of Information Technology (IT) capabilities.

According to UNCTAD (2004), the recent trends and developments in the area of ICT, e-commerce and economic development, including some aspects of ongoing international discussions on matters such as Internet governance. Specific discussions on e-commerce and ICT in developing countries focus on selected topics such as the use of digital and Internet technologies in the creative industries, in particular in the music industry, and their application to higher online learning. As e-commerce is not the sole domain of private business, the report also looks at government e-commerce applications in e-procurement. Finally, the report looks at the legal issues and challenges of data privacy and its role as a trust-building mechanism for information society development.

3. Historical Background of Ethiopian Banking System

Banking in Ethiopia began in 1905 with the Bank of Abyssinia, a private company controlled by the Bank of Egypt. In 1931 it was liquidated and replaced by the Bank of Ethiopia which was the bank of issue until the Italian invasion of 1936. During the Italian occupation, Bank of Italy banknotes formed the legal tender. Under the subsequent British occupation, Ethiopia was briefly a part of the East Africa Currency Board. In 1943, the State Bank of Ethiopia was established, with 2 departments performing the separate functions of an issuing bank and a commercial bank. In 1963, these functions were formally separated and the National Bank of Ethiopia (the central and issuing bank) and the Commercial Bank of Ethiopia were formed. In the period to 1974, several other financial institutions emerged including the state-owned:

1. The Agricultural and Industrial Development Bank (established largely to finance state owned Enterprises)
2. The Savings and Mortgage Corporation of Ethiopia
3. The Imperial Savings and Home Ownership Public Association (which provided savings and Loan services)

Major Private commercial institutions, many of which were foreign owned, included

1. The Addis Ababa Bank
2. The Banco di Napoli
3. The Banco di Roma

4. Current Conditions

The financial system of Ethiopia is very underdeveloped. There is no stock exchange and of the eleven banks that exist. There is a centralized bank called National bank of Ethiopia, three are state owned banks dominate the sector

and the remaining seven are private banks. The state owned banks are

1. Commercial bank of Ethiopia.
2. Development bank of Ethiopia.
3. Construction and business bank.

The private banks are

1. Dashen bank.
2. Awash International bank.
3. Bank of Abyssinia.
4. Wegagen bank.
5. United bank.
6. Cooperative bank of Oromia.
7. Nib International bank.

There are no foreign banks in the country, and the system remains isolated from the effects of globalization while policy-makers fear that liberalization will lead to loss of control over the economy. The government controls interest rates and sets them below the high inflation rate. Corruption, though strictly sanctioned, remains a concern. The National Bank of Ethiopia is the country's central bank. The state-owned Commercial Bank of Ethiopia is the largest bank in Ethiopia and controls 2/3 of the assets of the entire banking system. Kiyota (2007) describes the Ethiopian banking sector as 88% concentrated versus 59% for Kenya, 67% for Tanzania, 63% for Uganda, and 81% for sub-Saharan Africa as a whole.

Table 1: Value of Ethiopian Bank Assets

	1998	1999	2000	2001	2002	2003	2004	2005	2006	1998 % of Total	2006 % of Total
State-owned banks	19,732	18,938	23,437	25,035	25,674	27,897	31,113	35,000	37,668	94%	70%
Commercial Bank of Ethiopia	17,503	17,434	19,828	21,489	22,146	24,200	27,975	33,163	35,849	83%	66%
Development Bank of Ethiopia	2,229	2,502	2,615	2,578	2,549	2,554	4,091	n.a.	n.a.	31%	n.a.
Construction and Business Bank	n.a.	n.a.	974	968	958	942	1,057	1,832	1,797	n.a.	3%
Private banks	1,354	2,040	3,157	4,016	5,234	5,968	9,093	12,253	16,443	6%	30%
Dashen Bank	511	674	865	1,100	1,486	1,594	2,677	3,420	4,546	2%	8%
Awash International Bank	452	536	759	907	1,112	1,401	1,770	2,226	2,954	2%	5%
Bank of Abyssinia	206	388	718	891	1,142	1,333	1,585	2,057	2,834	1%	5%
Wegagen Bank	185	366	514	583	645	889	1,110	1,616	2,259	1%	4%
United Bank	n.a.	76	143	214	314	469	674	1,073	1,599	n.a.	3%
Cooperative Bank of Oromia	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	129	224	n.a.	0%	
Nib International Bank	n.a.	n.a.	158	336	534	885	1,247	1,732	2,027	n.a.	4%
Total	21,084	23,976	25,570	29,073	30,907	34,065	42,206	47,254	54,089	100%	100%

Source: Reproduced from Kiyota (2007) which gives sources as Annual Report of Individual Banks

4.1 Stock Market

There is no stock exchange exist in the country for trading.

4.2 Other Types of Finance/Financial Markets

The other type of finance market is Micro finance. The formal microfinance industry began in Ethiopia in 1994/1995 with the government's the Licensing and Supervision of Microfinance Institution Proclamation designed to encourage Microfinance Institutions (MFIs) to extend credit to both the rural and urban poor of the

country. By 2005, there were 23 MFIs with almost 1 million clients. Since the government prohibits any foreign national from providing banking services in Ethiopia, MFIs in the country must be established as share companies with capital wholly owned by Ethiopian Nationals or by organizations wholly owned and registered under the laws with a head office in Ethiopia. This has led to lack of transparency in the sector since much of 12 the initial capital comes from foreign donors who must enlist "nominal" shareholders to act as fronts. (MFIs are licensed under the central bank). Gobezie (2005) notes, these shareholders are precluded from selling or transferring their shares and "voluntarily forsake" their claim on dividends, if any, declared by the MFI. Such shareholders do not have a real stake in the organization and would be unlikely to lend it support at a time of financial crisis. Interest rates charged on loans are not fixed, but a minimum interest rate of 3% to depositors is required by law, which sometimes discourages mobilization in hard-to-reach areas (where administrative costs added to the cost of capital make investment too expensive). Such high transaction costs mean that most MFIs operate in urban or semi-urban areas, leaving the rural poor underserved. On the other hand, MFIs are exempt from Income and Sales Tax on their profits. Other than the formally-licensed MFIs, there are NGOs informally involved in the delivery of microfinance. Their practices include subsidized interest rates, charity and lax delinquency penalties, which Gobezie notes may undermine the health of the microfinance industry as a whole.

Table 2: Lending and Deposits at Ethiopian Banks (ETB mm)

	2001/02	2002/03	2003/04	2004/05	2005/06
Lending					
Central government	21,791	23,393	27,624	26,042	34,245
Other sectors	8,513	10,275	13,743	8,337	13,220
Nonfinancial public enterprises @	13,278	13,118	13,881	17,705	21,025
Financial public enterprises *	1,099	785	1,572	2,447	2,525
Cooperatives	505	449	404	302	275
Private sector	317	314	336	817	1,477
% of Total lending					
Central government	39.1	43.9	49.8	32	38.6
Other sectors	60.9	56.1	50.2	68	61.4
Nonfinancial public enterprises @	5	3.4	5.7	9.4	7.4
Financial public enterprises *	2.3	1.9	1.5	1.2	0.8
Cooperatives	1.5	1.3	1.2	3.1	4.3
Private sector	52.1	49.5	41.9	54.3	46.9
Deposits					
Demand deposits	24,298	27,095	31,313	37,090	44,380
Public enterprises	12,124	13,396	15,621	18,060	21,587
Cooperatives	4,122	3,637	3,735	4,062	4,234
Private sector	3,773	4,392	5,186	6,341	7,813
Central government	1,244	2,215	3,108	4,245	5,984
Other +	2,592	2,721	2,991	2,624	2,631
Savings deposits	11,071	12,529	14,447	17,403	20,688
Public enterprises	37	52	15	42	52
Cooperatives	238	268	378	895	603
Private sector	10,788	12,199	14,039	16,429	19,890
Other +	9	10	15	36	143
Time deposits	1,102	1,170	1,245	1,627	2,105
Public enterprises	119	113	138	165	210
Cooperatives	28	33	52	47	38
Private sector	504	556	616	794	888
Central government	8	8	8	8	8
Other +	444	459	432	613	960
% of Total deposits					
Demand deposits	49.9	49.4	49.9	48.7	48.6
Savings deposits	45.6	46.2	46.1	46.9	46.6
Time deposits	4.5	4.3	4	4.4	4.7

5. Review of Software Technologies, Trends and Services Offered by Ethiopian Banking Systems

Ethiopia is one of the fastest developing countries in Africa. The Ethiopian banking system is still under developing. We will review software technologies and services offered to customers by major banks in Ethiopia.

5.1 Dashen Bank

Dashen bank is one of the private banks in Ethiopia, and it was established in September 20, 1995 according to commercial code of Ethiopia. The website of the company the mission of the bank says “to provide efficient and customer focused domestic and international banking services, overcoming the challenges for excellence through the application of appropriate technology”. The bank has 59 share holders, quarter million customers, 34 area banks, 1051 number of employees and it stood 2nd on the market share among other banks in the country. It is the first bank providing ATM cards to customers in Ethiopia.

➤ Services offered:

Currently, the bank renders four major services in all of its branches namely.

- Credit Facility
- Saving Scheme
- International Banking
- Fund Transfer.
- MasterCard and Visa cards

➤ Credit card facility:

The bank provides a credit facility to its customers in different forms depending on their need and the nature of their business they are to invest on. Some of the credit lines offered include; overdraft facilities, term loans, letter of credit facilities, merchandise loans and personal loans.

➤ Saving Scheme:

The other service the bank renders is deposit services including demand deposit, savings deposit, youth savings deposit and time/fixed deposit.

➤ International banking:

The bank also renders international banking services providing services like; opening letters of credit for importers, handling of incoming LCs for exporters, purchase of outward bills purchasing and selling of foreign currency denominated notes, receiving and transferring foreign currency payment by swift and handling incoming and outgoing international letters of guarantee.

➤ Fund transfer:

Furthermore, the bank is currently offering fund transfer. The bank provides both domestic fund transfer all over the

country and international fund transfer, rendered in cooperation with Western Union.

➤ Master card and Visa cards:

Moreover, the bank is providing the customers MasterCard and Visa cards so that they can use it internationally. Dashen Bank has established account maintenance relationship with thirteen correspondent Banks. Overall banking relationship in SWIFT has expanded to 109 banks in 55 cities and 44 countries.

➤ Technology used in the bank:

The bank uses various kinds of communication and computing technologies to carry out its day to day activities. The communication technologies range from telephone and fax to dial up internet connections and very fast high speed broadband network. Now the bank is under transition of changing the old system with the new one for all branches. The change includes replacement of old hardware with modern and competent one. All branches in the capital started to work using the new system and other branches especially branches in rural areas are under process (training users, preparing the necessary hardware and equipment...). Wide Area Network (WAN) is another attractive feature of the Bank's technology; customers having a deposit account in one of the area banks can access their accounts from any other area bank in the country. Micro Banker is used as major bank software (UNIX sco version) starting from its establishment 9 years before. The system is decentralized in a sense the servers are distributed to the branches to support the service provision. Those branches are connected by WAN through dial up connection. The software supports different banking activities like customer registration, transaction processing, generating reports to different departments and so on. Micro banker works only in a UNIX environment obligating workstations to use only command lines and prohibiting them to change between applications while using the software. As the need for control and efficiency increases, the bank decides to change the software to new banking software called Flex cube. This system enables the bank to have competitive advantage and more control over the operations. "This system will not only allow the Bank improve its competitive edge but also make a difference by ensuring further capacity as well as providing assured centralized control and centralized database"(Bank President). The new software has a lot more facilities and modules that the previous software does not have. Though currently, the bank uses only some of the functionalities that the banking software provides, it can use more functionalities in the future as more and more services are introduced in the bank. The bank also uses different applications for performing different tasks. For instance, Microsoft Excel and Access for payroll and human resource database respectively.

➤ **Features of FlexCube Software:**

FlexCube software provides an Information Infrastructure. Based on this any system has to qualify the following criteria.

- **Sharable:** in a sense it has to be common resource to all users (of a community)
- **Evolving:** it should exhibit growth in terms of number of users, technological improvement, diversification of services
- **Open:** It should not have limit or boundary to number of users , or groups
- **Heterogeneous:** The different users because it has no limit to group or number of users, the different technological progress and the diversification of service it supports makes it heterogeneous
- **Standardized:** means there must be agreement between different users about the arrangement of the infrastructure in order to be used by all users.

The software is going to provide to their customers better features compared to the previous software. These includes

➤ **Change in services**

The bank is introducing new services to the user community. Services like ATM, MasterCard and Visa cards are introduced to the user community this year.

➤ **Change in technology**

The bank has changed its system from the previous computer system called Micro banker which does not have a centralized database to a new one called FlexCube which can enable centralized database processing as well as distributed one. This means all the branch banks will have their own database of users, which as well be propagated to the central server of the bank. This helps the recovery of data easier than that of the previous system. If the servers at the branches crash the data can be reclaimed from the central database since the data from the branches is synchronized with the central server. The changes have not only affected the software part, but also the hardware that the previous system used to use. This includes the change in computers and network facilities. The workstations at the service end have changed to meet the requirements of the FlexCube client software. To this end, high speed computers with high memory and hard disk capacity have been introduced. Regarding the network, it has been changing over the time from dial-up to broadband technology. As a result the components have changed from

wired network solution to wireless one using the UBR technology which can increase the bank's bandwidth. This has enabled the branches in the regional towns where the infrastructure is not at the level of the one in the capital city benefit from the new improved network.

5.2 Wegagen Bank

The bank, which was established in 1997 with a paid-up capital of 30 million birr, the highest initial capital at the time, has now reached 633.2 million birr. The press statement of the bank also indicated that during the budget year Wegagen has mobilized 3.9 billion birr deposits through its 51 branches and lend 2.5 billion birr to its customers. Its total asset has also increased to 5.7 billion birr (around 346 million US dollars at the current exchange rate) with 1,859 employees. Wegagen Bank has now 2,140 shareholders. Recently, it announced "Agar Visa Card" for their customers for effective banking transactions.

5.3 Commercial Bank of Ethiopia

The leading bank in Ethiopia, established in 1942. Pioneer to introduce modern banking to the country. It has 352 branches stretched across the country. A leading African bank with assets of Birr 73.3 billion as on June 30th 2010. Plays a catalytic role in the economic progress & development of the country. Currently CBE has about 2 million Account holders. It has strong correspondent relationship more than 50 renowned foreign banks and a SWIFT bilateral arrangement with 500 others. CBE combines a wide capital base with more than 9,700 talented and committed employees. Pioneer to introduce Western Union Money Transfer Services in Ethiopia. CBE has reliable and long-standing relationships with many internationally acclaimed Banks throughout the world.

Software used:

Commercial Bank of Ethiopia (CBE), the country's largest bank with more than two million customers, has selected Temenos (SIX: TEMN), the market leading provider of banking software. Founded in 1993 and listed on the Swiss Stock Exchange (SIX: TEMN), Temenos Group AG is a global provider of banking software systems in the Retail, Corporate & Correspondent, Universal, Private, Islamic and Microfinance & Community banking markets. Headquartered in Geneva with 56 offices worldwide, Temenos serves over 1000 customers in more than 120 countries. Temenos' software products provide advanced technology and rich functionality, incorporating best practice processes that leverage Temenos' experience in over 600 implementations around the globe. Temenos' advanced and automated implementation approach,

provided by its strong Client Services organization, ensures efficient and low-risk core banking platform migrations. They announced that to transform its entire operational infrastructure across more than 200 branches. CBE's operational improvement plans will modernize its services to its growing client base in Ethiopia as part of its vision to become a World Class bank in the region. CBE is a government owned bank which offers a range of retail banking services, which has already opened a subsidiary company in Southern Sudan with plans to extend its services to other East African countries. "Temenos' software delivers all the functionality and efficiencies we need to match a world class commercial bank. Its dedicated investment in product development to continually meet the needs of its clients and their customers was a very compelling proposition for us. The lack of flexibility and homogeneity in our current systems impacts product innovation, service delivery, risk management and cost control. Being able to replace all systems with T24 will deliver a single view of the business and customer and therefore provide a better understanding of our customers to improve services, enable us to effectively monitor and manage risk and lower our total cost of ownership. We will have an enormous wealth of functionality and flexibility to develop a wide range of new products, enabling us to strengthen our presence in the retail space, as well as penetrate new markets in the face of increased competition. By using this software CBE is going to provide to their customers better services like E-banking, Mobile banking etc. Temenos' software was benchmarked against solutions from Oracle FS and Infosys and proved to be the most functionally rich platform to deliver end to end operational support and broaden CBE's service offering, such as mobile banking, to enter new markets. Temenos and CBE will jointly implement the software in a phased roll out, with the first stage expected to go live in 2011.

5.4 National bank of Ethiopia

The National Bank of Ethiopia was established in 1963 by proclamation 206 of 1963 and began operation in January 1964. Prior to this proclamation, the Bank used to carry out dual activities, i.e. commercial banking and central banking. The proclamation raised the Bank's capital to Ethiopian dollars 10.0 million and granted broad administrative autonomy and juridical personality. Following the proclamation the National Bank of Ethiopia was entrusted with the following responsibilities.

- To regulate the supply, availability and cost of money and credit.
- To manage and administrate the countries international reserves.

- To license and supervise banks and hold commercial banks reserves and lend money to them
- To supervise loans of commercial banks and regulates interest rates.
- To issue paper money and coins.
- To act an agent of the government.
- To fix and control the foreign exchange rates.

However, monetary and banking proclamation No. 99 of 1976 came into force on September 1976 to shape the Bank's role according to the socialist economic principle that the country adopted. Hence the Bank was allowed to participate actively in national planning, specifically financial planning, in cooperation with the concerned state organs. The Bank's supervisory area was also increased to include other financial institutions such as insurance institutions, credit cooperatives and investment-oriented banks. Moreover the proclamation introduced the new Ethiopian currency called 'birr' in place of the former Ethiopia Dollar that ceased to be legal tender thereafter. The proclamation revised the Bank's relationship with Government. It initially raised the legal limits of outstanding government domestic borrowing to 25% of the actual ordinary revenue of the government during the proceeding three budget years as against the proclamation 206/1963, which set it to be 15%. This proclamation was in force till the new proclamation issued in 1994 to reorganize the Bank according to the market-based economic policy so that it could foster monetary stability, a sound financial system and such other credit and exchange conditions as are conducive to the balanced growth of the economy of the country. Accordingly the following are some of the powers and duties vested in the Bank by proclamation 83/1994.

➤ Features of Advanced technologies:

Ethiopia uses the advanced software's like Temenos and FlexCube in their banking system, leads tremendous changes and provides more facilities to their customers. These include E-banking and Mobile banking (M-banking).

6. E-banking and Mobile banking

6.1 E-banking

All E-payment methods share a number of common characteristics. These are: independence, interoperability, portability, security, anonymity, divisibility, ease of use and transaction fees. Independence refers to the ability of E-commerce methods to operate without installing specialized software. Interoperability and portability refers to the ability of forms of E-commerce to interlink with

other enterprise applications and systems. Security is an important consideration that encompasses the safety of transfer and the chance of the transfer being intercepted. E-cards offer a number of benefits to the issuing banks and customers of the bank including:

- Dramatically reduce printing, mailing and financial handling costs associated with processing transaction.
- Enhance payment security by minimizing theft or loss.
- Reduce undeliverable payments via electronic delivery to the card account.
- Prevent fraud through automated controls.
- Increase customer satisfaction.
- Improve operational efficiency and profitability of the issuing banks.

6.2 Mobile banking (M-banking)

Mobile banking is a subset of e-banking in which customers access a range of banking products like savings accounts and credit instruments, via electronic channels. M banking requires the customer to hold a deposit account to and from which payments or transfers may be made. M banking reduces the transactions costs of payments because there is an electronically accessible store of value.

6.3 Challenges

In most regulatory regimes, creating account-based stores of value is regarded as banking-related business. The question of who may hold the deposit balance turns out to be a crucial issue affecting the development of mobile banking. Technology and the regulation, in turn, affect the effectiveness of m-banking. The issue relevant to many financial systems particularly in developing societies, soundness, might have concerned many. In other words, M-banking could jeopardize financial stability and prevent economic and financial growth. However, despite the natural conservatism of the banking industry, M banking innovation has proceeded to become a rapidly-growing tool across developing countries.

6.4 Benefits

A major benefit of M banking is drawing in the “unbanked” who generally can’t afford the cost of formal banking services and infrastructure. There is the potential to bank people outside the realm of traditional financial services and the mobile phone is a pervasive device that has fewer barriers to entry than most technologies and has

penetrated some of the poorest economies due to the overwhelming demand for any form of telecommunications. The evolution of the system necessarily started out as a simple transaction to purchase airtime, strictly to make calls. Very soon, people in rural areas in just about every sub-Saharan African country were purchasing prepaid airtime from local vendors in cities and selling it on to merchants in rural locales, who in turn either rented the use of mobile phones to rural dwellers or sold the airtime on to them at a profit.

7. Issues and Challenges

Banking sector in Ethiopia faces numerous challenges to adopt advanced technologies as well as E-banking applications and seize the opportunities presented by ICT applications in general. Key challenges in technology and E-banking applications are:

7.1 New Technologies

Banking sector in Ethiopia it is still under developing. The banks try to use advanced technologies and attract customers. The new software's requires high configured hardware devices (computers) to maintain centralized servers and individual PC's and also to use these software's employee's needs more skills. Ethiopia is lagging with proper trained persons. Software is failed or error should be occurred, they are unable to rectify it.

7.2 Internet and Telecommunication

Lack of infrastructure for internet and telecommunications impede smooth development and improvements in E-banking and M-banking in Ethiopia. 80% of population they are living in the rural areas of the country, where majority of small and medium business are concentrated, have no internet facilities and thus are unable to engage banking services like E-banking and M-banking. Still today, the whole country is totally depending on only one network called Ethiopian Tele Communication. The government is not allowed any other private telecommunication network with in the country.

7.3 Lack of Suitable Legal and Regulatory Framework for Banking

Ethiopian current laws do not accommodate electronic contracts and signatures. Ethiopia has not yet enacted legislation that deals with E-commerce concerns including enforceability of the validity of the electronic contracts, digital signatures and intellectual copyright and restrict the use of encryption technologies. ICT in Ethiopia is now collaborating with Europe and other countries in the world and developing regulatory frame work for banking sector.

7.4 Political Instabilities in Neighboring Countries

Political and economic instabilities in Somalia, Sudan and Eritrea are threatening traits that do not provide a very conducive environment for to improve banking sector. Political instabilities inevitably disturb smooth operations of business and free flow of goods and services.

7.5 High rates of Illiteracy

Illiteracy we have to consider it as major problem. Banking sector provide proper features like ATM machines, E-banking and M-banking to their customers, they in a position unable to taste the sweetness of these facilities. Most of the customers to the bank, they don't know how to read and write.

7.6 Integration of Different Financial Networks

There is 100% absence of financial networks that links different banks. Banks are not yet automated. Most of the transactions currently taking place use credit and debit cards supplied by visa and master cards. For conducting E-business, the use of credit or debit cards is mandatory, thus requiring the need for specialized systems which are not currently available.

7.7 Frequent Power Interruptions

Lack of reliable power supply is a key challenge for government. Due to this, industrialization is very difficult. This will impact the banking sector, and also difficult for smooth running E-business in Ethiopia.

7.8 Resistance to changes in Technology among customers and staff

- i) Lack of awareness on the benefits of new technologies
- ii) Fear of risk
- iii) Lack of trained personnel in the key organizations.
- iv) Tendency to continue with existing infrastructure.

People and government may resist the new technologies

7.9 Cyber Security Issues

Cyber security is not permitted only to Ethiopia, is a global challenge that requires global and multi-dimensional response with respect to policy, socio-economic, legal and technological aspects. The customers are expecting from

banks well secured transactions. Banks introduces new technologies like E-banking and M-banking in the market, attention should be drawn to the prevention of cyber crimes.

8. Conclusions

Certainly the banking industry in Ethiopia is underdeveloped and therefore there is an immediate need is required to embark on capacity building arrangements and modernize the banking system by employing the state of art technology being used anywhere in the world. Day by day improvements in import and export business, international trading and relationships, the current banking system is little bit short of providing efficient and dependable services. Therefore all operating banks in Ethiopia should recognize the needs and introduce the new software and technologies like electronic banking, mobile banking, international banking services and more ATM machines. The government has to implement new policies and give free hands to their banks and then only it is possible to overcome the challenges. Finally I conclude to Develop a comprehensive regulatory and legal framework for e-commerce and e-payment, Raise public awareness on the use of ICT, e-commerce, and e-Payment, Provide incentives for financial institutions to invest rigorously on ICT and use of e-commerce, and e-Payment, Encourage the current efforts to develop and expand ICT infrastructure.

9. References

- [1] Jensen, "Resource wealth and political regimes in Africa", 2003.
- [2] Gardachew Worku, "Electronic-Banking in Ethiopia-practices, opportunities and challenges", journal of Internet banking and commerce, vol. 15. no.2, August 2010.
- [3] Magembe S and Shemi A P, "challenges and opportunities for adopting electronic commerce in a developing countries", The Botswana perspective, IAABD Conference, 2002.
- [4] UNCTAD, "electronic commerce and development report", New York and Geneva, 2004.
- [5] Ole Hanseth, "Complexity and information infrastructure", University of Oslo, 2002.
- [6] Astlet, "customizable middleware for modular distributed software", CACM, volume 44, number 5, 2001.

- [7] ITU4D “Cyber Security Issues”, ITU, 2006.
- [8] R.K.Mishra and J. Kiranmai “E-banking: A case study of India”, the ICfai University journal of public administration, vol. 5, no. 1, 2009.
- [9] Dragos A. Manolescu and Adrian E. Kunzle “Several patterns for eBusiness Applications”, Applied reasoning Systems Corporation, 2001.
- [10] Hanseth, O. “The economics of standards. In from control flow to drift”, Oxford University press, 2000.
- [11] FSA United Kingdom “Financial Security Authority”, FSA, 2004.
- [12] Kiyota “Static and Dynamic Consequences of a Korus FTA”, 25th Anniversary, Korea Economic Institute, 2007.
- [13] Getanesh Gobezie “Sustainable rural finance: Prospects, Challenges and Implications”, International NGO Journal Vol. 4 (2), pp. 012-026, February 2009.

First Author

Bhaskar Reddy Muvva Vijay B.E, M.TECH, (PhD)

I have completed my Masters in computer science and Engineering in the year of 2008 and Bachelors in Electronics and Communication Engineering in 2004. Presently pursuing doctorate from Dravidian University, Kuppam, A.P, and India. Currently I am working as a lecturer in Department of Computer Science and Information Systems, Ethiopian institute of technology, Mekelle University, Mekelle, Ethiopia. I have six years of academic experience in teaching various IT modules and published two books.

Second Author

Tewdros Sisay Asefa, B.Sc, M.Sc

I have completed both my Masters and my Bachelors in computer science, in University of Pune, India. Currently I am working as a lecturer and programming and software Engineering chair head in Department of Computing, Ethiopian institute of technology, Mekelle University, Mekelle, Ethiopia. I have four years of academic experience in teaching various IT courses and modules.

Design and Implementation of Memory-less Forbidden Transition Free Crosstalk Avoidance CODECs for On-Chip Buses

J.Venkateswara Rao¹ and P.Sudhakara Rao²

¹ Electronics & Communication Engineering Department, Vignan Institute of Technology & Science, Deshmukhi(V), Nalgonda Dist., Andhra Pradesh-508284, India

² Electronics & Communication Engineering Department, Vignan Institute of Technology & Science, Deshmukhi(V), Nalgonda Dist., Andhra Pradesh-508284, India

Abstract

Recently, reducing crosstalk noise delay is an important issue in VLSI design. As circuit geometries become smaller, wire interconnections become closer together and taller, thus increasing the cross-coupling capacitance between nets. At the same time, parasitic capacitance to the substrate becomes less as interconnections become narrower, and cell delays are reduced as transistors become smaller. In this work, we present a CODEC design for the forbidden transition free crosstalk avoidance CODEC. Our mapping and coding scheme is based on the Binary number system. In this paper, we investigate and propose a bus forbidden transition free CODECs for reducing bus delay and our experimental results show that the proposed CODEC complexity is orders of magnitude better compared to the existing techniques. Compared to the best existing approaches, we achieved a 3 times faster design and improvement in logic complexity.

Keywords: Crosstalk, Crosstalk Avoidance Codes, Forbidden Transition Free, Encoding, System on Chip, Parasitic, Coupling Capacitance, Deep-submicron.

1. Introduction

With shrinking device sizes, increasing chip complexity and faster clock speeds, wire delay is becoming increasingly significant [11, 12]. The propagation delay through long cross-chip buses is proving to be a limiting factor in the speed of some designs, and this trend is only expected to get worse. It has been shown that the delay through a long bus is strongly dependent on the coupling capacitance between the wires. In particular, the crosstalk effect when adjacent wires simultaneously transition in opposite directions is particularly detrimental to the delay. When the cross-coupling capacitance is comparable to or exceeds the loading capacitance on the wires, the delay of such a transition may be twice or more than that of a wire transitioning next to a steady signal. This delay penalty is commonly referred to as the capacitive crosstalk delay. The capacitive crosstalk delay strongly depends on the transition activities of the adjacent signals, hence the crosstalk type. Type-4 and type-3 crosstalk have the worst

delay characteristics, followed by type-2 and then type-1. A few techniques involving selective skewing of bus data signals [13], transistor sizing [14], and repeater sizing [15] to reduce capacitive crosstalk induced delay have been proposed. Encoding is one of the more effective ways to reduce capacitive crosstalk delays. Here we present encoding techniques that focus on reducing crosstalk delay. The rest of the paper is organized as follows. Section 2 explains about Crosstalk Classification, Section 3 discusses about forbidden transition free crosstalk avoidance codes (FTF-CAC). In Section 4, we discuss about circuit implementation and experimental results. In Section 5 we compare the results of experiments that we have performed to quantify the CODEC performance. We conclude the paper in Section 6.

2. Crosstalk Classification

Figure 1 illustrates a simplified on-chip bus model with crosstalk. In the figure, C_L denotes the load capacitance, which includes the receiver gate capacitance and also the parasitic wire-to-substrate parasitic capacitance. C_I is the inter-wire coupling capacitance between adjacent signal lines of the bus. In practice, this bus structure is typically modelled as a distributed RC network, which includes the non-zero resistance of the wire as well.

The on-chip bus crosstalk is classified into five types [3, 4] as shown in Table 1.

Table 1: Transition pattern crosstalk classification Margin specifications

Crosstalk class	C_{eff}	Sample transition patterns
0C		000→111
1C	$C_L(1+\lambda)$	011→000
2C	$C_L(1+2\lambda)$	010→000
3C	$C_L(1+3\lambda)$	010→100
4C	$C_L(1+4\lambda)$	010→101

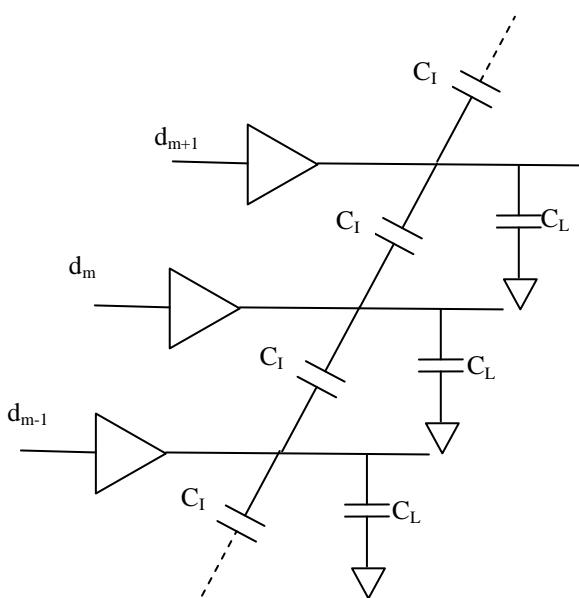


Fig. 1. On-chip bus model with crosstalk

This classification is based on the effective capacitance, in the j^{th} line in a bus as, $C_{\text{eff},j}$

$$\begin{aligned} C_{\text{eff},j} &= C_L [1 + \lambda ((1 - \delta_{j,j-1}) + (1 - \delta_{j,j+1}))] \\ &= C_L + C_{\text{lw},j} + C_{\text{rw},j} \end{aligned} \quad (1)$$

It separates $C_{\text{eff},j}$ into three components: the intrinsic capacitance C_L , the crosstalk capacitance to the wire on the left side, $C_{\text{lw},j} = \lambda (1 - \delta_{j,j-1}) C_L$, and the capacitance to the wire on the right side, $C_{\text{rw},j} = \lambda (1 - \delta_{j,j+1}) C_L$. It is easy to see that $C_{\text{lw},j}, C_{\text{rw},j} \in \{0, 1C_L, 2C_L\}$.

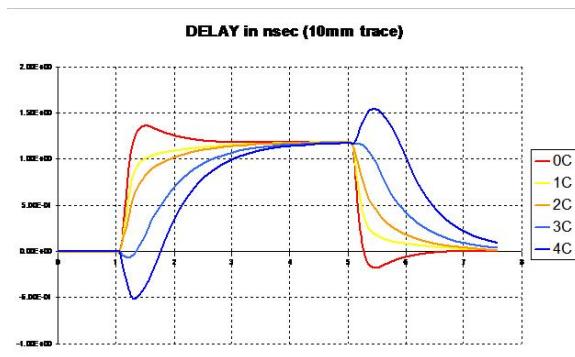


Fig. 2. Delay impact of different sequences confirmed by SPICE simulations-0.1μm CMOS process.

3. Forbidden Transition Free Crosstalk Avoidance Codes

The forbidden transition free Crosstalk Avoidance Codes (FTF-CAC) are an efficient 3C-free memory-less codes. It was first proposed by Victor and Keutzer in 2001 [7], and we will use the FTF-CAC to derive an efficient CODEC. The basic idea of the FTF code is to prohibit two adjacent bits from transitioning in opposite directions, i.e., the forbidden transitions $01 \rightarrow 10$ or $10 \rightarrow 01$ are not allowed. This guarantees that $\delta_{j,j+1} \geq 0$, therefore $(2 - \delta_{j,j-1} - \delta_{j,j+1}) \leq 2$ and the bus satisfies $\max_j (C_{\text{eff},j}) = 2$ from equation 1. Hence the transition is 3C-free. An FTF code is a set of codewords such that transitions among these codewords do not produce forbidden transitions on any two adjacent bits.

A forbidden transition is defined as the simultaneous transition (in opposite directions) on two adjacent bits, i.e., $01 \rightarrow 10$ or $10 \rightarrow 01$. We first observe that to guarantee forbidden transition freedom on the boundary $d_j d_{j+1}$ between any two codewords in an FTF-CAC, the 01 and 10 patterns cannot coexist in the same set of codewords. This can be easily confirmed by examining the transitions among codes in $\{00, 01, \text{ and } 11\}$, or $\{00, 10, 11\}$. If we eliminate 01 or 10 from all the boundaries in the codewords in a set of codewords R, we can guarantee that R is forbidden transition free. Therefore, once again, the problem of eliminating forbidden transitions is transformed into a problem of eliminating specific patterns.

Theorem-1 The largest sets of codewords satisfying the forbidden transition free condition is the set of codewords that can transition to a class 1 codeword (defined as a codeword with alternating 0 and 1 bits) without generating forbidden transitions [7].

For any given size bus, there are two class 1 codewords: “1010...” and “0101...”. From Theorem-1, we can see that there exist two different sets with the same maximum cardinality. In one set (set A), in all codewords, the 01 pattern is eliminated from $d_{2j+1} d_{2j}$ boundaries and the 10 pattern is eliminated from $d_{2j} d_{2j-1}$ boundaries. In the second set (set B), the 10 pattern is eliminated from $d_{2j+1} d_{2j}$ boundaries and the 01 pattern eliminated from $d_{2j} d_{2j-1}$ boundaries. Table-2 lists all set-A FTF codewords for 2, 3, 4 and 5-bit busses. Take the 5-bit bus as an example, we can see that 10 is not present in $d_2 d_1$ and $d_4 d_3$, and 01 is not present in the boundaries $d_3 d_2$ and $d_5 d_4$. If one of these two sets is known, the other set can be produced by simply complementing all the codewords in the first set. There are multiple methods to produce all the n-bit codewords in the set that satisfy Theorem-1.

- Start with a complete set of 2^n vectors and remove codewords that do not satisfy the boundary constraints.

- Start with a set consisting of a single class 1 codeword and grow the FTF-CAC codewords by adding compatible codewords to the set.

Start from a small FTF set (say 2-bit FTF codes) and inductively append bits to the codewords in the set until the codeword length reaches n -bits. Clearly, the first (pruning) method is impractical when n is large, since a complete set of codewords have $2n$ entries and searching through $n - 1$ boundaries requires $O(2(n-1)n)$ searches. Both the second and the third methods listed above actually “grow” the FTF codewords instead of “pruning”, and therefore require less computation. The method of “growing” codewords by appending bits to codewords in an existing set is given in Algorithm-1.

Algorithm-1 is the pseudo code for generating the FTF codewords.

Table 2: FTF-CAC codewords for 2, 3, 4 and 5-bit busses

2-bit	3-bits	4-bits	5-bit
00	000	0000	00000
01	001	0001	00001
11	100	0100	00100
	101	0101	00101
	111	0111	00111
		1100	10000
		1101	10001
		1111	

Algorithm-1 FTF codeword generation

```

 $S_2 = \{00, 01, 11\}$ 
for  $m > 2$  do
    if  $m$  is odd then
        for  $\forall V_{m-1} \in S_{m-1}$  do
            add  $1 \cdot V_{m-1}$  to  $S_m$ ;
            if  $d_{m-1} = 0$  then
                add  $0 \cdot V_{m-1}$  to  $S_m$ ;
            end if
        end for
    else
        for  $\forall V_{m-1} \in S_{m-1}$  do
            add  $0 \cdot V_{m-1}$  to  $S_m$ ;
            if  $d_{m-1} = 1$  then
                add  $1 \cdot V_{m-1}$  to  $S_m$ ;
            end if
        end for
    end if
end for

```

The inductive codeword generation method given in Algorithm-1 can be used to derive the cardinality of the FTF codes. We first define,

Definition-1

$T_{t(m)}$: number of m -bit FTF vector,

$T_{tl(m)}$: number of m -bit FTF vector with the MSB being 1,

$T_{t0(m)}$: number of m -bit FTF vector with the MSB being 0.

The following relationship can be derived from Algorithm-1:

$$T_{t(m)} = T_{t0(m)} + T_{tl(m)} \quad (2)$$

$$T_{t(2m)} = T_{tl(2m-1)} + T_{t(2m-1)} \quad (3)$$

$$T_{t(2m-1)} = T_{t0(2m-2)} + T_{t(2m-2)} \quad (4)$$

and with some simple manipulation of Eqs. (2), (3) and (4), we get

$$T_{t(2m)} = T_{t(2m-1)} + T_{t(2m-2)}, \quad (5)$$

$$T_{t(2m+1)} = T_{t(2m)} + T_{t(2m-1)}. \quad (6)$$

and they can be combined into a single recursive equation

$$T_{t(m)} = T_{t(m-1)} + T_{t(m-2)} \quad (7)$$

Given the initial condition of $T_{t(2)} = 3 = f_4$ and $T_{t(3)} = 5 = f_5$, we get

$$T_{t(m)} = f_{m+2} \quad (8)$$

Compare Eq. 8 with the maximum cardinality of FPF codes, we find that the FTF codes have slightly lower cardinality because $2f_{m+1} > f_{m+2} = f_{m+1} + f_m$ for all $m > 0$. However, the asymptotic overhead percentage of the FTF code is still $\sim 44\%$, the same as the FPF code cardinality. Therefore for large busses, the coding gain for the FTF-CAC is the same as the coding gain for FPF codes $G_{FTF} \approx 39\%$.

4. Circuit Implementation and Experimental Results

The coded busses in the simulation were 6-bits wide. A 4-to-6-bit encoder and a 6-to-4-bit decoder logic were manually implemented using an arbitrary mapping of data words to codewords. Our simulations using Synopsys Design Compiler show that the maximum delay of both the encoder and the decoder was 8.21 ns.

Table-3: 6 bit FTF code words for 4 bit data words

4 bit data word	6 bit code word
$(d_3d_2d_1d_0)$	$(c_5c_4c_3c_2c_1c_0)$
0000	000000
0001	000001
0010	000101
0011	010001
0100	110100
0101	110101
0110	011101
0111	011111
1000	110000
1001	000100
1010	010000
1011	110111
1100	110001
1101	111101
1110	111100
1111	111111

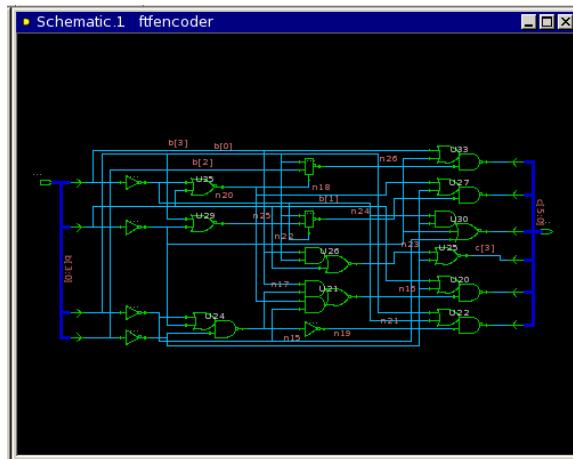


Fig. 3. Gate level schematic for FTF encoder

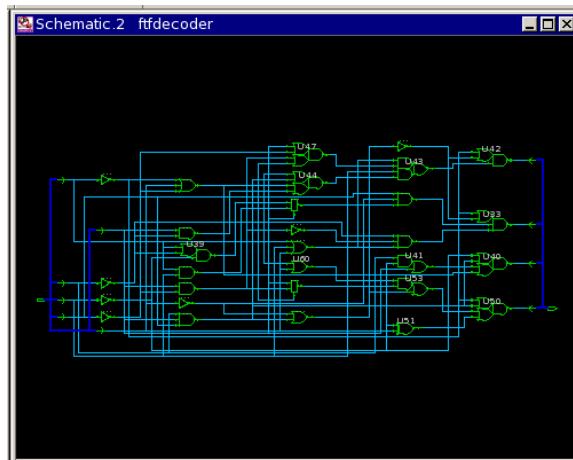


Fig. 4. Gate level schematic for FTF decoder

Table 4: Timing Reports for FTF Encoder

<i>Point</i>	<i>Incr. Delay</i>	<i>Path Delay</i>
input external delay	0.00	0.00 r
b[3] (in)	0.00	0.00 r
U36/Z (IV)	0.53	0.53 f
U35/Z (NR2)	1.88	2.41 r
U34/Z (MUX21L)	0.65	3.06 f
U33/Z (AO7)	0.66	3.72 r
c[0] (out)	0.00	3.72 r
Data arrival time		3.72

The above timing report shows the data arrival time for the FTF Encoder from input side to output side. Prime Time reports the worst delay path from input to output. The b [3] is the input port of the encoder and the name in the bracket is the reference name for that port. Incr. path is the incremental path delay. The delay from input port, b [3] to the output of not gate, (U36/Z) is 0.53 ns. The delay from the output of not gate, (U36/Z) to the output of next

nor gate, (U35/Z) is 1.88 ns. The delay from the output of nor gate, (U35/Z) to the output of next mux, (U34/Z) is 0.65 ns. The delay from the output of mux, (U34/Z) to the output of next AOI gate, (U33/Z) is 0.66 ns. The output of AOI gate, (U33/Z) is the output port of encoder, c(0). So, the total delay from input port, b [3] to output port c [0] is 3.72 ns. The f in the third column indicates a transition from 1 to 0 and r indicates a 0 to 1 transition.

Table 5: Timing Reports for FTF Decoder

<i>Point</i>	<i>Incr. Delay</i>	<i>Path Delay</i>
input external delay	0.00	0.00 r
c[5] (in)	0.00	0.00 r
U62/Z (IV)	0.47	0.47 f
U46/Z (OR3)	1.44	1.92 f
U44/Z (AO4)	1.32	3.24 r
U43/Z (AO2)	0.60	3.83 f
U42/Z (AO7)	0.66	4.49 r
b[1] (out)	0.00	4.49 r
Data arrival time		4.49

The above timing report shows the data arrival time for the FTF Decoder from input side to output side. So, the total delay from input port, c [5] to output port b [1] is 4.49 ns.

5. Comparison to Other Techniques

This paper is based on the concepts proposed in [1]. The encoder and decoder presented in fig.3 and fig. 4 has a data arrival time of 3.72 ns and 4.49 ns. The encoder and decoder proposed in [1] have a data arrival time of 16.38ns and 7.63 ns. Thus the total delay for the FTF CODEC is 8.21 ns compared to encoder proposed in [1] have a total delay of 24.01 ns. So, we can say that our CODEC is around 3 times faster than the CODEC proposed in [1]. When the data arrival time is small, it creates a more positive slack.

6. Conclusions

The 1C-free bus does not require CODECs, or we can say that these CODEC designs are trivial (repeating the input bit by N times). On the other hand, 3C-free and 2C-free codes need CODECs. For these codes to be used in practice, efficient CODEC designs are necessary. In the case of crosstalk avoidance codes, the complexity and speed of the CODEC are both critical for overall bus performance. Also, the power consumption of the CODECs should be factored in when the overall power consumption is evaluated. In this paper, we present efficient CODEC design technique for the memory less CACs. The advantages of these CODECs are their low complexity, high speed as well as minimum area overhead.

Our CODEC is around 3 times faster than existing efficient CODECs.

Acknowledgments

We grateful to Synopsys Tools providers and the management of Vignan Institute of Technology & Science for providing the Back end Tools for carrying out the research.

References

- [1] Duane .C, Cordero. V and Khatri. S. P. "Efficient On-Chip Crosstalk Avoidance CODEC Design", IEEE Transactions on VLSI Systems, April 2009, pp 551 – 560.
- [2] Sotiriadis .P and Chandrakasan .A, "Low power bus coding techniques considering inter-wire capacitance". Proc. of IEEE-CICC, 2000, pp 507-510.
- [3] Chem. J, Huang .J, Aldredge .L, Li .P, and Huang .P, "Multilevel metal capacitance models for CAD symbol design synthesis systems". In IEEE Electron Device Letter, 1992, pp 32–34.
- [4] Arora .N, Raol .K, Shcumann .R, and Richardson .L, "Modeling and extraction of interconnect capacitance for multilayer VLSI circuits". In IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 1996, 15(1) pp 58–67.
- [5] Sridhar .S.R, Ahmed .A, and Shanbhag .N. R., "Area and Energy-Efficient Crosstalk Avoidance Codes for On-Chip busses", Proc. of ICCD, 2004, pp 12-17.
- [6] Duan .C, Tirumala .A and Khatri .S.P., "Analysis and Avoidance of Cross-talk in On-chip Bus", Hot Interconnects, 2001, pp 133-138.
- [7] Bret Victor and Keutzer .K, "Bus Encoding to Prevent Crosstalk Delay", ICCAD, 2001, pp 57-63.
- [8] Duan .C and Khatri S. P. "Exploiting Crosstalk to Speed up On-chip busses", Design, Automation and Test in Europe Conference and Exhibition, 2004, pp 778-783.
- [9] Duan .C, Gulati .K and Khatri .S. P, "Memory-based Cross-talk Canceling CODECs for On-chip busses", ISCAS, 2006, pp 4-9.
- [10] Ma .J and He .L,"Formulae and applications of interconnect estimation considering shield insertion and net ordering", ICCAD, 2001, pp 327-332.
- [11] Davis J. A, al. et, "Interconnect limits on giga-scale integration (GSI) in the 21st century" in *Proceedings of the IEEE*, 2001, 89, pp. 305–324.
- [12] Bohr .M. T, "Interconnect scaling—The real limiter to high performance ULSI" in *Proceedings of IEEE Electron Devices Meeting*, 1995, pp. 241–244.
- [13] Hirose .K and Yasuura .H, "A bus delay reduction technique considering crosstalk" in *Proceedings of Design, Automation and Test in Europe*, 2000, pp. 441–445.
- [14] Xiao .T and Sadowska .M, "Crosstalk reduction by transistor sizing" in Proceedings of Asia and South Pacific Design Automation Conference (ASPDAC), 1999, pp. 137–140.
- [15] Li .D, Pua .A, Srivastava .P and Ko .U, "A repeater optimization methodology for deep submicron, high-performance processors" in IEEE International Conference

on Computer Design: VLSI in Computers and Processors (ICCD), Austin, TX, 1997, pp. 726–731.

First Author J.Venkateswara Rao received B.Tech Degree from, VRSEC, Nagarjuna University, Guntur, in 2000. M.Tech. Degree from NITW, Warangal, India in 2003, pursuing Ph.D in the Department of Electronics and Communication Engineering, JNT University, Hyderabad, INDIA. Currently he is working as Associate Professor in the Department of Electronics and communication Engineering, VITS, Hyderabad, INDIA, His research interest includes on chip crosstalk noise reduction in VLSI Circuits. He published 4 papers on "on-chip crosstalk noise reduction in VLSI circuits" in international journals and international conferences.

Second Author Dr.P.Sudhakara Rao Completed Ph.D., Information and Communication Engineering from Anna University, India, **Masters** in Electronics and Communication Engineering from Anna University, India, **Bachelors** in Electronics and Communication Engineering from Mysore University, India. Worked as deputy **Director**, "Central Electronics Engineering Research Institute centre, India" for over 25 years, 2 years as Vice-president, "Sieger Spintech Equipments Ltd., India", established an electronic department for the development of electronic systems for nearly 2 years. Presently working with Vignan Institute of Technology and Science, Nalgonda District, AP, INDIA as DEAN R&D, HOD ECE. He has one patent and 55 technical publications/ conference papers to his credit. Conducted many international and national conferences as chairman.

Extending XACML to support Credential Based Hybrid Access Control

Nirmal Dagdee¹ and Ruchi Vijaywargiya²

¹ S D Bansal College Of Technology,
Indore, MP, INDIA

² Department of Computer Science and Engineering, S D Bansal College Of Technology,
Indore, MP, INDIA

Abstract

Various research efforts are in progress to enforce credential based access control using XACML standard. The current standard of XACML supports attribute based access control [4,5,9,19]. While XACML accepts certified attributes through digital certificates, it does not support credential based access control in which the access conditions are defined not only in terms of credential attributes but also in terms of types of credentials. Credential based hybrid access control[7,11,14,20,21] has been proposed for systems having diversified access control requirements. The use of various types of credentials in access control policy specification provides easy and immediate access to unknown user in open access environment. Fine grained access control in closed administrative domain is achieved using Identity Credential and the attributes associated with the credentials. In this paper, we propose extensions to the XACML standard that support credential-based hybrid access control. The XACML access policy language has been extended to define access policy in terms of heterogeneous credentials. Each credential is uniquely identified by associating a *category* and *type* with it. The access policy contains various conditions over credentials and the attributes associated with the credentials. Enhancement to XACML framework has also been proposed so that credential based hybrid access policies can be evaluated and enforced.

Keywords: XACML, credential, access control.

1. Introduction

With the development of digital certificate technology, credential based approaches are becoming popular for enforcing security and access control on shared resources. In credential based access control systems[1,2,3,4,6,8,9,12], the access decision depends on the properties possessed by the user and can prove by submitting one or more credentials. The user submits a set of credentials along with the data access request. Credential based hybrid access control approach [7,11,14,20,21] has been proposed for the environment

having varied access control requirements. It provides easy and immediate access to unknown user in open access environment along with fine grained access control in closed administrative domain. The access policy is defined in terms of various types of credentials. Various types of credentials along with attributes associated with credentials, data sources and the environment affect the access decision.

XACML is an OASIS standard that provides a formal representation model of access control policies and mechanisms for their evaluation[15,16,19]. It defines a XML-based language for specifying and exchanging access control policies over the web. The XACML framework describes a modular architecture for the evaluation of policies and a communication protocol for message interchange between various modules. XACML includes standard extension points for the definition of new functions and data types that provide a great potential for enhancing the language for customized needs[13,15,16]. Although XACML enjoys a large adoption in industry due to its simplicity and powerful extension mechanism, it has few limitations[15,16,18,19]. The current XACML standard enforces attribute-based access control[4,5,9,19] in which the authorizations applicable to a user depends on the attributes that the user presents and the conditions that the user satisfies. While the XACML permits that the attributes can be presented by means of certificates and can accept certified attributes but this is limited however to conditions on attributes associated with the credentials[15,16,18,19]. The real support for credential based access control requires expressing the access control conditions not only on values of attributes associated with the credentials but also on the credentials themselves. Intuitively, XACML supports attribute-based access control but does not really support credential-based access control[19]. In this paper, we propose extensions to XACML framework so that it can support credential based hybrid access control.

The rest of this paper is organized as follows. Section 2 highlights the credential based hybrid access control and the XACML standard framework has been discussed in section 3. The proposed extension to XACML framework has been described in section 4. Few illustrative usecases from the medical domain are discussed in section 5. In section 6, we have summarized the salient advantages of the proposed work.

2. Credential based hybrid access control

Various credential based access control approaches [1,2,3,4,6,8,9,12,16,17] have been proposed that make use of digital credentials for authentication and authorization. For systems having varied access control requirements, credential based hybrid access control[7,11,14,20,21] has been proposed. This approach not only enables immediate and open access to shared information by competent users but also provides fine grained access control on domain confined data. The access control policy is defined in terms of heterogeneous credentials. Use of Identity credentials enable fine grained user specific accesses to known users. Attribute certificates are found suitable in open access environment where the resource provider and the resource requester are not known to each other and the access decision depends on the properties possessed by the user and can prove by submitting one or more attribute credentials. Recently, a new type of credential called as Standard credential[11, 14, 22] has also been proposed. Standard credentials are general purpose credential that grants a status to the bearer of the credential. A Credential *type* is associated with each credential and has a fixed set of attributes contained in it. The type of the credential identifies the attribute contained in the credential. Use of standard credentials provides easy and immediate access to resource in open access environment. Granular access control can be achieved by defining the conditions on attributes associated with the credentials, the resources and the environment. The authorization certificate overrides the existing access policies and grants immediate access to the desired resource. The main components of credential based hybrid access control policy are as follows:

1. Credentials: The credentials are the basic elements of the access control policy. The credentials are considered as a bundle of attributes. Access control policy is defined in terms of various credentials like identity credential, attribute credential, standard credential and the authorization credential. Every credential is uniquely represented in the policy and the policy is defined as a set of conditions over heterogeneous credentials.
2. Attributes: In credential based hybrid access control, the access policy is defined not only in terms of credentials but

also in terms of various attributes associated with the credentials, the data source and the environment. The credential attributes are similar to any other attribute with a difference that the credential attributes are always associated with some credential.

3. Conditions: The credential based access policy is defined as a set of various conditions that must be satisfied in order to gain access on the resource. The conditions can be of two types: the credential condition and the attribute condition. The credential condition specifies a credential or a set of credentials that the user must submit. The attribute condition contains various attributes associated with the credential, user, resource or environment as operands. The conditions are evaluated and if found satisfied, access is granted else denied.

3. XACML Framework

3.1 XACML Policy Specifications

XACML[13,15,16] defines an XML-based access control policy language for specifying the access control policy. It provides a processing model for evaluating these policies on the basis of the given XACML access request. The access request is specified by means of attributes like which subject wants to perform which action on which resource. The XACML Request Context is embedded in the *<Request>* element. The *<Request>* element contains *<Subject>*, *<Resource>* and *<Action>* elements. The *<Subject>* element contains the various attributes associated with the user, the *<Resource>* indicates the attributes of the requested resource and the *<Action>* specifies the action which the user is requesting on the resource.

A XACML policy contains one *<Policy>* or *<PolicySet>* as root element, which is a container for other Policy or PolicySet elements. Element *<Policy>* consists of a *<Target>*, a set of *<Rule>*, an optional set of *<Obligation>*, and a rule combining algorithm. A *<Target>* element includes conditions on Subject, Resource, Action, and Environment. If a request satisfies the conditions specified in the Target, the corresponding policy applies to the request. A *<Rule>* corresponds to a positive (permit) or a negative (deny) authorization depending on its effect. It includes an element *<Target>* and an element *<Condition>* specifying further restrictions on subject, resource and action. Each condition can be defined through element *<Apply>* with attribute FunctionID denoting the XACML predicate (e.g., string-equal, integer-less-than, etc) and with appropriate sub-elements denoting both the attribute against which the condition is evaluated and the

comparison value. The rule's effect is then returned whenever the condition evaluates to true. The <Obligation> element specifies an action that has to be performed in conjunction with the enforcement of the authorization decision. Each element <Policy> has attribute RuleCombiningAlgID specifying how to combine the decisions of different rules to obtain a final decision of the policy evaluation (e.g., deny overrides, permit overrides, first applicable, only one applicable, etc). According to the selected combining algorithm, the authorization decision can be permit, deny, not applicable (i.e., no applicable policies or rules can be found), or indeterminate (i.e., some information is missing for the completion of the evaluation process).

3.2 XACML Architecture

XACML standard[13,15,16] defines a modular architecture for the evaluation of policies and a communication protocol for message interchange between various modules. The architecture includes different functional components that interact to take an access control decision. The access request is received by the Policy Enforcement Point (PEP) that is responsible for the enforcement of the access decision. The Policy Decision Point (PDP) is the component that evaluates the access policy and produces the access decision by retrieving the applicable policies from the Policy Administration Point (PAP). The PAP provides functionalities to administer access control policies. The Context Handler(CH) provides the interface between the PDP and PEP. It buffers the attributes that were given to the PEP along with the request and provides them to the PDP on demand. In addition to what is included in the access request, for decision process the PDP may also need additional information related to subject, resource or environment. The context handler provides this information via the Policy Information Point (PIP) that acts as a source of attribute values. PIP retrieves the attribute values by interacting with the subject, resource and environment modules.

Data flow in the standard XACML model can be summarized as follows:

1. The requester sends an access request to the PEP module.
2. The PEP module sends the access request to the Context Handler that translates the original request into a canonical format, called XACML Request Context, and sends it to the PDP.
3. The PDP extracts the applicable policies by means of the PAP module and retrieves the required attributes and the resource from the Context Handler. The Context Handler relies on the PIP module to acquire the attribute values associated with the subject,

resource and the environment. For this purpose, the PIP interacts with the Subject, Resource and the Environment modules.

4. The PDP evaluates the policies and returns the XACML Response Context, together with an optional set of obligations, to the Context Handler. The Context Handler translates the XACML response context to the native format of the PEP and returns it to the PEP. If some information is missing, the PDP cannot take the decision and returns an error (Indeterminate response).
5. The PEP fulfills the obligations and, if permitted, it gives access to the requester. Otherwise, the access is denied.

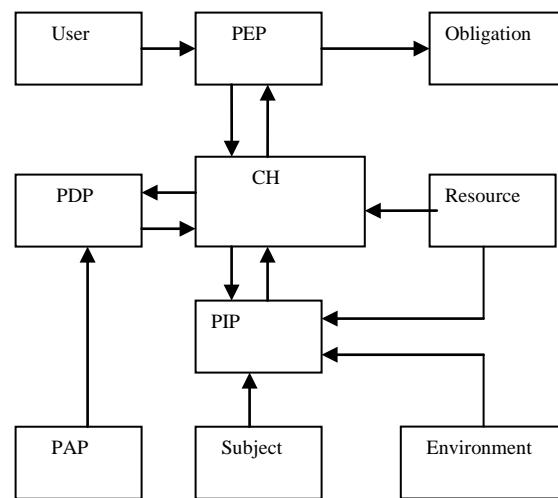


Fig 1: XACML Architecture

3.3 XACML and Credential based Hybrid Access Control

Although the XACML has been established as a solution for controlling access in a flexible way, it suffers from some limitations in supporting credential based access control. The current XACML standard enforces attribute-based access control[4,5,9,19] in which the access policy is defined in terms of various attributes. Here, the authorizations applicable to a user depend on the attributes that the user presents and the conditions that the user satisfies. While XACML permits that the attributes can be presented by means of certificates, the credential based access control requires the support for expressing the access control conditions not only on values of attributes contained in the credentials but also in terms of various types of credentials[10,13,15,16,18,19,20].

Several research efforts are in progress to extend the functionality of XACML to support credential based access control[10,13,15,16,18,19]. Representation of credential attributes in policy specification has been proposed in [16,19] but is limited to the evaluation of attributes like issuer, time and date only. Credential-based access control needs support for expressing the conditions on the types of the credentials in addition to the various attributes of the credentials. Need has been realized to uniquely identify each credential in the policy. The Standard credentials[11,14,22] have been proposed that have unique credential types. The credential type determines the attributes contained in the credential. The various credential types and the attributes contained in each credential is published and this facilitates the policy designers and the access requester in specifying and understanding the policy respectively. Recently, the similar concept of credential type has also been proposed in [10,13,16,19]. Credentials are realized as a bundle of attributes having a specific type. However, these works are more focused on addressing the problem of preserving the privacy of the various credential attributes. In [15], the author has pointed the need of having a common ontology of credential types and their attributes and to share this ontology between the policy makers and the user. Some recent proposals have investigated credential-based access control with anonymous credentials. In [23], an access control language has been introduced that is explicitly targeted to anonymous credentials. A card based access control system[10] has been proposed in which the card is similar to the credential. Rather than submitting various cards, depending on the policy, the user creates a claim from various cards. The claim contains a statement about the attributes of one or more of the cards along with their evidence. A card based access control requirements language (CARL) for technology-independent credential-based access control[10] has been introduced but the language follows a proprietary syntax, which makes deployment in real-world access control scenarios difficult. In [18,19], the author has focused on extending XACML to support credential based access control in web based scenarios. However their focus is on developing a mechanism for communicating the access policy to the access requester. In our work, we propose extensions to the XACML standard to enforce credential based hybrid access control. The extended XACML framework is suitable for providing open as well as closed access and thus handles diversified access control requirements.

4. Proposed Extension to XACML

In our work, we propose extension to XACML framework that supports credential based hybrid access control. Our

proposal is divided into two parts. First we extend the access control policy specification language to support heterogeneous credentials, credential attributes and the conditions having credentials and attributes as operands. Second, we propose the functional enhancements in some of the architectural modules of XACML. The enhancements are designed in such a way that there is minimal impact on the policy language and the evaluation mechanism.

Following are the typical requirements of the hybrid credential based access control that entails enhancements in the XACML framework:

1. A Credential is the basic element of the access control policy. Access control policy is defined in terms of various categories of credentials like Identity credential, Attribute credential, Standard credential and Authorization credential. This is represented using *CredentialCategory* that specifies the category to which the credential belongs. Every standard credentials[11, 14, 22] has a unique type associated with it. Thus in addition to *CredentialCategory*, the standard credential also has a unique type identifier called as *CredentialType*. The credentials are uniquely represented in the policy using *CredentialCategory* and the *CredentialType*.
2. Every credential should be uniquely represented in the policy. A *CredentialId* is assigned to each credential that is being referred in the policy. The *CredentialId* is unique within the namespace of the policy.
3. As per the access policy, the user may have to submit one or more credentials. The access policy should support the specification of logical combination of credentials. For example, an access policy may expect the user to either submit the Standard Credential of Senior Citizen or an Attribute certificate issued by CityCorporation containing Date of birth as attribute.
4. The credential attributes are considered similar to other attributes. However, within the policy, the credential attributes are always referred along with the associated credential. The access policy should be able to define condition on credential attributes in addition to other types of attributes.
5. A standard credential is issued by the Standard Credential Authority and is easily verifiable using CIA hierarchy. Therefore the standard credential may not have issuer associated with it while being referred in the policy specification. Whereas the identity, attribute and authorization credentials can be issued by anyone and thus the policy should specify the issuer from whom these credentials can be accepted.

4.1 Extension to XACML Policy Specification

To support credential based hybrid access control, the extensions to XACML policy specification language have been proposed. The extensions are defined in such a way that they do not alter the semantics of existing elements or attributes. The proposed extensions are as follows:

4.1.1 Representation of Credential

To represent credential in the access policy, we propose a new element `<Credential>` that corresponds to the definition of a credential and its attributes. To uniquely identify a credential with in the policy, `<Credential>` element has an attribute called as `CredentialId` whose value is the identifier by which the credential will be referred in the rest of the policy. `CredentialId` is unique within the namespace of the policy. The various attributes of the credentials are defined using `<CredentialMatch>` and `<CredentialAttributeDesignator>` elements that are similar to `<SubjectMatch>` and `<SubjectAttributeDesignator>`. For example the Standard Credential of Doctor can be represented as follows:

```
<Credential CredentialId = "SC1">
    <CredentialMatch MatchId= "string-equal">
        <CredentialAttributeDesignator
            DataType= "string"
            AttributeId= "CredentialCategory"/>
        <AttributeValue> Standard </AttributeValue>
    </CredentialMatch>
    <CredentialMatch MatchId= "string-equal">
        <CredentialAttributeDesignator
            DataType= "string"
            AttributeId= "CredentialType"/>
        <AttributeValue>Doctor </AttributeValue>
    </CredentialMatch>
</Credential>
```

4.1.2 Logical combination of credentials

To specify logical combination of multiple credentials within the policy, we introduce a new element that corresponds to the specification of combination of credentials in the policy. We define `<CredentialRequirements>` within the `<Target>`. The `<CredentialRequirements>` contains one or more `<Credentials>` elements within it. The `<Credentials>` contain one or more `<Credential>` elements. Credentials expressed within one `<Credentials>` are considered in conjunction, whereas the `<Credential>` in different `<Credentials>` are considered in disjunction. The following code specifies that the user should submit either C1 and C2 or C3 in order to gain access where C1, C2 and C3 are defined as described in section 4.1.1.

```
<Target>
    <CredentialRequirements>
        <Credentials>
            <Credential>
                C1
            </Credential>
            <Credential>
                C2
            </Credential>
        </Credentials>
        <Credentials>
            <Credential>
                C3
            </Credential>
        </Credentials>
    </CredentialRequirements>
</Target>
```

4.1.3 Conditions involving Credential Attributes

In XACML, the operands appearing in conditions are typically the attributes associated with the subject, resource, action or environment. In our work, we extend XACML conditions to support credential attributes as well. Credential attributes can be treated just like any other attribute but the only difference is that they must be associated with a credential. For this purpose, we propose the inclusion of `CredentialId` attribute in the `<CredentialAttributeDesignator>` element. The `CredentialId` contains the value as defined in the `<Credential>` element for the credential in which the attribute being referred is contained. For example, the attribute “specialization” of the standard credential of doctor can be represented as follows:

```
<CredentialAttributeDesignator
    DataType= "string"
    CredentialId = "SC1"
    AttributeId= "specialization"/>
    <AttributeValue>orthopedics </AttributeValue>
```

4.1.4 XACML Request Context

The Context Handler generates the XACML Request Context which is sent to the PDP. The XACML Request Context has been extended so as to specify the various credentials submitted by the user. The standard Request Context contains `<Subject><Resource>` and `<Action>` element within the `<Request>` element. We propose one more element `<Credentials>` within the `<Request>` element. The `<Credentials>` element contains multiple `<Credential>` element one each for every credential submitted by the user. The `<Credential>` element contains details of all the attributes associated with the credential.

```
<Request>
```

```

<Credentials>
  <Credential>
  <\Credential>
  .....
  <Credential>
  <\Credential>
<\Credentials>
... Other Subject, Resource and Action
elements...
<\Request>
    
```

4.2 Proposed XACML Architecture

Functional enhancements have been proposed to the XACML framework so as to make it suitable for enforcing credential based hybrid access control policies. The extensions have been proposed in such a way that the existing functionality of XACML is not altered and the extensions can be used in combination with the existing standard features. The modified XACML architecture is as shown in Figure 2. The greyed modules are modified as follows for the credential based access control.

- PEP: The PEP accepts the data access request from the user and sends it to the Context Handler. It has been modified to accept a credential or a set of credentials from the user. The PEP sends the credentials to the Context Handler.
- Context Handler(CH): The Context Handler is enhanced to have Credential Manager (CM) that verifies the credentials received from the PEP and extracts the values of various attributes associated with the credential. The CH then translates the access request into a canonical format, called XACML Request Context. The modified XACML Request Context is as described in section 4.1.4. The modified XACML Request Context is sent to the PDP.
- PAP: The PAP has been modified so as to administer the heterogeneous credentials based policies.
- PDP: The functionality of PDP has been extended to evaluate the credential based hybrid access control policies.

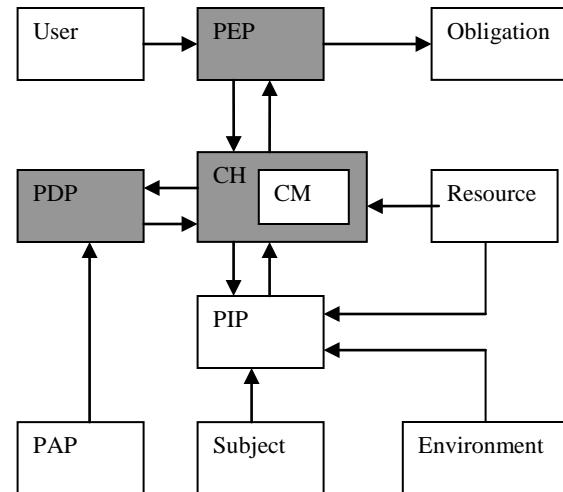


Fig 2: The Extended XACML Architecture

The data flow in the modified XACML model can be summarized as follows:

1. The requester sends an access request along with the set of credentials to the PEP.
2. The PEP accepts the credentials along with the request and sends them to the Context Handler.
3. The Context handler receives the credentials and the request from the PEP. The Credential Manager (CM) verifies the received credentials and extracts the values of various attributes associated with the credentials. The CH then translates the access request into a canonical format called XACML Request Context. The XACML Request Context is sent to the PDP.
4. The PDP identifies the applicable policies by means of the PAP module and retrieves the required attributes from the Context Handler. The context handler relies on the PIP module to retrieve the various attribute values.
5. The PDP evaluates the access policies. The PDP returns the XACML Response Context together with an optional set of obligations to the Context Handler. If some information is missing, the PDP cannot take a decision and returns an error (Indeterminate response).
6. The Context Handler translates the XACML Response Context to the native format of the PEP and returns it to the PEP.
7. The PEP fulfills the obligations and, if permitted, it gives access to the requester. Otherwise, the access is denied.

5. Illustrative Usecases

In this section, we have discussed few usecases from the medical domain. Some of the typical access control rules that exist in medical domain are as follows:

Rule 1: A Doctor can access details of only those patients that have disease same as his own specialization ie. an orthopedic doctor can access details of orthopedic patients, a gynecologist can access details of gynecology related patients, etc.

Rule 2: A known or registered doctor can access all details of his patients only.

Rule 3: A doctor can access the medical details of any patient.

Rule 4: A medical student who has enrolled in the university before year 2009 can access the medical details of patients between 9 AM and 5 PM.

Rule 5: A person can access all details of a specific patient if he brings an authority letter from the health secretary.

The XACML representation of the first two rules is shown below. The other rules can be defined similarly. The code snippet shown below is just an illustration of the various access control rules.

```
<Rule RuleId= "Rule1" Effect= "Permit">
<Target>
<CredentialRequirements>
<Credentials>
<Credential CredentialId = "SCI">
    <CredentialMatch MatchId= "string-equal">
        <CredentialAttributeDesignator
            DataType= "string"
            AttributeId= "CredentialCategory"/>
        <AttributeValue> Standard </AttributeValue>
    </CredentialMatch>
    <CredentialMatch MatchId= "string-equal">
        <CredentialAttributeDesignator
            DataType= "string"
            AttributeId= "CredentialType"/>
        <AttributeValue>Doctor </AttributeValue>
    </CredentialMatch>
</Credential>
</Credentials>
</CredentialRequirements>
</Target>
<Condition >
    <Apply FunctionId= "string-equal">
        <ResourceAttributeDesignator
            DataType= "string"
            AttributeId= "Disease"/>
        <CredentialAttributeDesignator
            DataType= "string"
```

```
        CredentialId = "SC2"
        AttributeId= "Specialization"/>
    </Apply>
</Condition>
</Rule>

<Rule RuleId= "Rule2" Effect= "Permit">
<Target>
<CredentialRequirements>
<Credentials>
<Credential CredentialId = "IC1">
    <CredentialMatch MatchId= "string-equal">
        <CredentialAttributeDesignator
            DataType= "string"
            Issuer = "XYZ"
            AttributeId= "CredentialCategory"/>
        <AttributeValue> Identity </AttributeValue>
    </CredentialMatch>
    <Credential>
        <AttributeValue> </AttributeValue>
    </Credential>
</Credentials>
</CredentialRequirements>
</Target>
<Condition >
    <Apply FunctionId= "string-equal">
        <CredentialAttributeDesignator
            DataType= "string"
            CredentialId = "IC1"
            AttributeId= "Subject">
        <ResourceAttributeDesignator
            DataType= "string"
            AttributeId= "DoctorId"/>
    </Apply>
</Condition>
</Rule>
```

6. Conclusion

In this paper, we have presented extensions to the standard XACML framework for supporting credential based hybrid access control. The extended XACML policy language supports specification of various types of credentials. Fine grained access control is provided by defining conditions associated with credentials in addition to attributes associated with subject, resource and environment. The extended language does not alter the semantics of the existing elements and attributes of the policy language. The functionality of some of the modules of the XACML architecture has been enhanced to accept various types of credentials and to evaluate and enforce credential based policies defined using extended XACML language. To facilitate the adoption of our approach in existing XACML

code bases, we have designed our extensions for minimal impact on the language and the XACML evaluation mechanism. The proposed extensions are suitable for developing access control systems that support open access as well as fine grained control in closed domain. Access to high priority user by overriding the existing policies is also possible.

References

1. Sudhir Agarwal, Barbara Sprick, Sandra Wortmann, Credential Based Access Control for Semantic Web Services, American Association for artificial Intelligence, 2004
2. Mary R. Thompson, Abdellilah Essiari, Srilekha Mudumbai, Certification based Authorization policy in a PKI Environment, ACM Transactions on Information and System Security (TISSEC), Volume 6 , Issue 4, pages: 566 – 588, 2003
3. William Johnston, Srilekha Mudumbai and Mary Thompson, Authorization and Attribute Certificates for widely Distributed Access Control, IEEE WETICE, 1998
4. Damiani, E., Di Vimercati, S.D.C., Samarati, P., New paradigms for access control in open environments, Proceedings of the Fifth IEEE International Symposium on Signal Processing and Information Technology, 2005
5. Shen Hai Bo, Hong Fan, An attribute based access control model for web services, proceeding of the seventh international conference on Parallel and Distributed Computing, Applications and Technologies, IEEE 2006
6. Mary Thompson, William Johnston, Srilekha Mudumbai, Gary Hoo, Keith Jackson, Abdelilah Essiari, Certificate-Based Access Control For Widely Distributed Resources, ACM, Proceedings of the 8th conference on USENIX Security Symposium - Volume 8, Pages: 17 - 17 , 1999
7. Nirmal Dagdee, Ruchi Vijaywargiya, Credential Based System for realizing Open and Domain confined accesses on Shared Data Sources, International Conf on data management, 2008
8. Ioannis Mavridis, Christos Georgiadis, George Pangalos, Marie Khair, Access Control based on Attribute Certificates for Medical Intranet Applications”, Journal Of Medical Internet Research, Vol3, 2001
9. Sabrina De Capitanidi Vimercati, Pierangela Samarati, New Directions in Access control, www.spdp.dti.unimi.it/papers/nato.pdf, 2002
10. Jan Camenisch, A Card Requirements Language Enabling Privacy-Preserving Access Control, ACM 2010
11. Nirmal Dagdee, Ruchi Vijaywargiya: Credential based hybrid access control methodology for shared Electronic Health Records, IEEE International Conference on Information Management and Engineering, Kuala Lumpur, Malaysia, 2009
12. P. Bonatti and P. Samarati. A unified framework for regulating access and information release on the Web, Journal of Computer Security, 10(3):241–272, 2002
13. Jan Camenisch et al, Credential-Based Access Control Extensions to XACML, www.w3.org/2009/policy-ws/papers/Neven.pdf, 2009
14. Nirmal Dagdee, Ruchi Vijaywargiya, Policy architecture for credential based access control in open access environment, Journal of Information Assurance and Security 6, 039-047, 2011
15. Claudio A. Ardagna et al , Enabling Privacy-Preserving Credential-Based Access Control with XACML and SAML, www.zurich.ibm.com/~new/papers/credprofile.pdf, 2010
16. Claudio A. Ardagna, An XACML-Based Privacy Centered Access Control System”, Conference on Computer and Communications Security”, Proceedings of the first ACM workshop on Information security governance, Chicago, Illinois, USA, pages: 49-58, 2009
17. Pierangela Samarati et al, Enriching Access Control to Support Credential-Based Specifications, subs.emis.de/LNI/Proceedings/Proceedings19/GI-Proceedings.19-10.pdf, 2002
18. Claudio A. Ardagna et al, Extending XACML for Open Web-based Scenarios, <http://www.w3.org/2009/policy-ws/papers/Samarati.pdf>, 2009
19. Claudio A. Ardagna, Expressive and Deployable Access Control in Open Web Service Applications, IEEE Transactions on Services Computing, Vol. 4, No. 2, April-June 2011
20. Sonia Jahid et al, Enhancing Database Access Control with XACML Policy, www.sigsac.org/ccs/CCS2009/pdf/abstract_23.pdf, 2009
21. Mariemma I. Yagüe, Survey on XML-Based Policy Languages for Open Environments, Journal of Information Assurance and Security 1, pages 11-20, 2006
22. Dr. Nirmal Dagdee, Access control methodology for sharing of open and domain confined data using Standard Credentials, International Journal on Computer Science and Engineering Vol.1(3), 148-155, 2009
23. Claudio A. Ardagna, Exploiting cryptography for privacy enhanced access control, JCS, vol. 18, no. 1, 2010

Dr. Nirmal Dagdee has earned his BE, ME and PhD degrees in Computer Engineering. His major research interests are in the fields of Service oriented computing, Data security and Soft Computing. He has authored several research papers that are published in reputed journals and conference proceedings. Presently, he is Director of S. D. Bansal College of Technology, Indore, India. He is a member of IEEE and ACM.

Ruchi Vijaywargiya, a faculty of computer science in S.D. Bansal College of Technology, Indore, India has around 20 years of experience in academics and software industry. She has done BE and ME in Computer Engineering and is pursuing PhD under the supervision of Dr. Dagdee in the field of data security and access control. Her areas of interest are data security, computer networks and object oriented technology. She is a member of IEEE.

New Channel Assignment Method for Access Points in Wireless LANs

Dr. Mohammed Fawzi Al-Hunaity
Al-Balqa' Applied University
Al-Salt 19117, Jordan

Abstract

Wireless LANs topology communicates using radio frequencies. The number of these frequencies is limited and not enough to assign a special frequency for each Access Point, this means that the communication topology should use a "re-use" mechanism, which allows the system to assign the same frequency that assigned previously for another access point. In this paper, we suggest a method to assign frequency channels to access points in a Wireless LANs system. The goal of this paper is to reach the assignment that prevents interference between access points especially neighbor ones. We used Genetic Algorithm to work on this issue, it is considered as an Expert System and one of the main multi point search technique used in computing to solve optimization and search problems.

Keywords: *Wireless LANs, Mobile Communications, Channel Assignment, Optimization, Genetic Algorithm, Soft Computing Techniques.*

1. Introduction

The most important issues on the design and deployment of wireless infrastructure mode are continuous coverage for the geographical area and an optimal channel assignment for access points (APs) while satisfying both the user's traffic demand and quality of services [1].

To satisfy all aims above, designers should have a mechanism that helps using the available narrow frequency spectrum efficiently [2].

When thinking about efficiently frequency use, designer should understand one more fact: All users of mobile and wireless system will use one shared medium: air, this means that users will be close to each other, and they will be able to "hear" or may be interfere each other [3].

WLAN communicates use radio frequencies (RF) or electromagnetic wave to carry a signal over part or the entire communication path similar way in AM/FM radios and FDMA [2].

FDMA, is based on dividing the available frequency range into smaller ranges (sub-frequencies), then use each sub-frequency as a standalone frequency. Each sub-frequency is called Channel [2].

The Federal Communications Commission (FCC) regulates the use of WLAN devices. To organize WLAN standards IEEE 802.11 appeared. IEEE 802.11 is a group of standards developed by the Institute of Electrical and Electronics Engineers Inc. to represent all WLAN computer communication [3]. Overviews of these standards are mention below.

IEEE 802.11 (legacy mode) - The original WLAN standard was released in 1997. It specified two rates 1 or 2 Mbps that operates in the 2.4 GHz frequency hopping spread spectrum (FHSS) and in the 2.4 GHz direct sequence spread spectrum (DSSS) [4].

IEEE 802.11a- It operates in the 5 GHz unlicensed frequency band Orthogonal Frequency Division Multiplexing (OFDM) with a maximum data rate of 54 Mbps. It offers 12 non interference channels, 8 channels for indoor use and 4 for outdoor [4], [5].

IEEE 802.11b- It operates in the 2.4 GHz unlicensed frequency band direct sequence spread spectrum with a maximum data rate of 11 Mbps. It divides band into 11 channels (North America) and 13 channels (Europe), but only 3 of them (channels 1, 6, 11) do not have band overlap [4], [6].

IEEE 802.11g- It operates in the 2.4 GHz unlicensed frequency band Orthogonal Frequency Division Multiplexing with a maximum data rate of 54 Mbps [4], [7].

The feature to use the same amount of information that sent formerly using a narrow band carrier signal spread it out over a higher frequency range is called Spread Spectrum Technology. This feature reduces the chance that will be damaged or jammed [3].

(FHSS)The process of splitting the spectrum into 79 nonoverlapping channels across the 2.402 to 2.480 GHz

frequency range called Frequency Hopping Spread Spectrum. FHSS WLANs support 1 Mbps and 2 Mbps data rate. FHSS device changes or hops frequencies within the RF frequency band [8].

(DSSS) is another WLAN physical method for the original 802.11. DSSS use 22 MHz channels. This model is well known, because it is easy to implement and it has a high data rates. These days most of wireless LAN uses DSSS model in the markets, because it allows sending a large data at a higher data rate than FHSS systems such as in IEEE 802.11b [8].

DSSS uses channels, with a band of frequencies 22 MHz wide, while in the FHSS, it uses 1 MHz carrier frequencies. For more details see the following example; there are two channels: channel 1, operate from 2.401 GHz to 2.423 GHz ($2.412 \text{ GHz} \pm 11 \text{ MHz}$); channel 2 operates from 2.406 to 2.429 GHz ($2.417 \pm 11 \text{ MHz}$)” [3] as shown in figure (1).

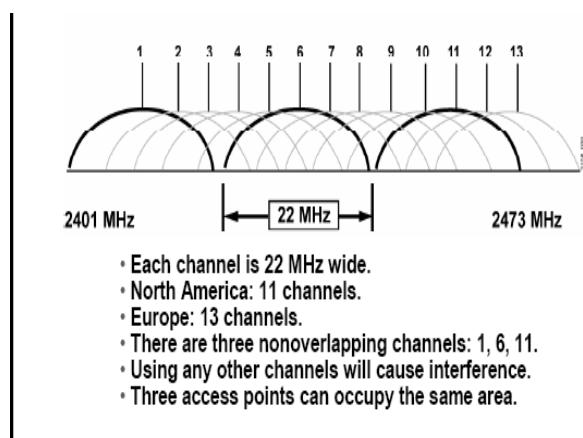


Fig. 1 DSSS channel allocation [4]

WLAN standard IEEE 802.11b and IEEE 802.11g uses three nonoverlapping channels 1, 6, 11 and using any other channels will cause interference, this means that the system should use a "re-use" mechanism. This re-use mechanism allows the system to assign the same channel which assigned previously for another AP if the distance is far enough to prevent interference.

In this paper, a new channel assignment mechanism is introduced. It uses genetic algorithm to specify a channel for each AP that prevents interference between APs especially neighbor ones.

The next section presents a channel assignment related works and section 3 explains genetic algorithm. Section 4 explains simulation method and the parameters used in the simulation; section 5 explains the result of simulation. Finally, section 6 draws the conclusions.

2. Related Work

The work [9] titled “Study on efficient channel assignment method using the genetic algorithm for mobile communication systems” presents a group channel assignment method for the optimization and its simulation results. This method utilizes the genetic algorithm (GA). Prior to the optimization calculation, whole channel cells are divided into subgroups and then, each subgroup is optimized. Using this result, the whole cells are further optimized. This process enables an efficient calculation to optimize the channel assignment. The results of simulation shows that the proposed method gives a shorter computation time compared with the method called as the individual assignment. See figure (2).

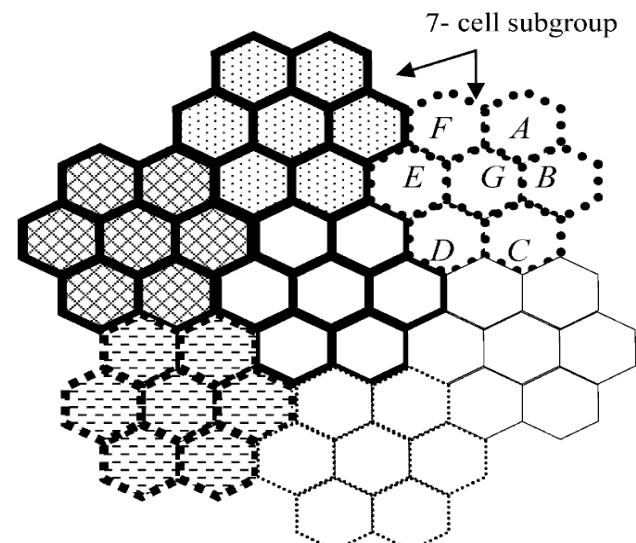


Fig. 2 Frequency Reuse - Channel Pattern [9]

In the study of [10] authors highlights the potential of using genetic algorithms to solve cellular resource allocation problems. Authors consider that cellular resource allocation deals with how and when to allocate radio frequency channels to mobile hosts. The objective of study is to gauge how well a GA-based channel borrower performs when compared to a greedy borrowing heuristic.

In the study of [11] authors stated an optimal access points allocation schema using GA which provided a solutions that allows a wireless user to choose an optimal AP according to its capacity not only to AP's power of signal.

3. Genetic Algorithm

3.1 How GA works

The basic concepts of Genetic algorithms were created by John Holland in the 1960s. After that Holland and his students from the University of Michigan developed these concepts [12], [13].

Genetic Algorithm is a multi point search technique used in computing to solve optimization and search problems by finding approximate. It tries to solve optimization problems by mimicking the behavior of nature. Nature always takes a decision of re-producing or eliminating individuals in each generation according to their fitness with their life conditions [13].

The operation of GA usually starts from a population of randomly generated “individuals” or “phenotypes”, this step called “Initialize chromosomes” and this group of individuals will be called the first generation; each individual represents a possible solution for a given problem. The individuals in this first generation will be represented by a string of bits “genes” which forms a chromosome [13], [14].

Each individual is evaluated and assigned a fitness value according to how good a solution to the problem it is “Evaluate Fitness”. The algorithm knows the fittest individuals by running a function called: Fitness function. This function returns the ratio of fitness for each individual in the generation [14], [15].

The highly fit individuals are stochastically selected from the current population, and modified (crossover and possibly mutated) to form a new population (second generation). The new population is then used in the next iteration of the algorithm. The algorithm continues until the maximum number of generations has been produced, or a satisfactory fitness level has been reached for the population [14], [15].

The cycle described in Figure (3) will be repeated more and more, and at the end of the day, genetic algorithm will find the fittest individual, which will be hopefully the best solution for the problem. See figure-3 which describes the complete cycle of genetic algorithm.

The selection, crossover and mutation are the most important part of genetic algorithm. The performance is influenced mainly by these operators [15].

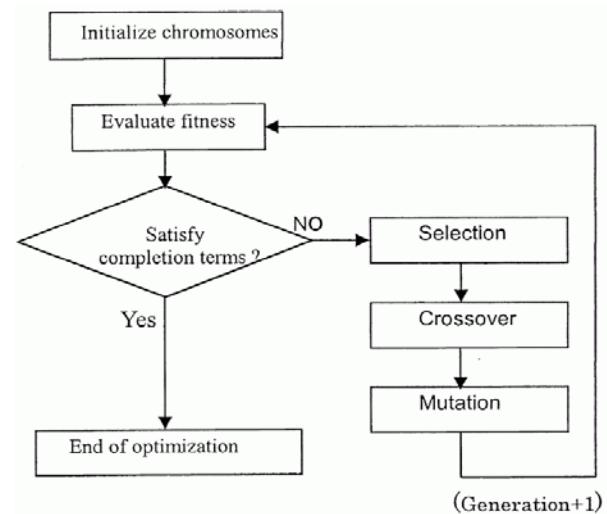


Fig. 3 Complete Cycle of Genetic Algorithm [13]

3.2 Crossover

It is the process of generating a child from two parents by taking a part from one of the parents and replaces it with the corresponding part from the second parent and vice versa [13].

3.3 Mutation

It is a change done on some of the children resulted from the crossover process by flipping the value of one of the bits randomly. The benefit of such operation is to restore the lost genetic values when the population converges too fast [14].

4. Simulation Method

4.1 Environment Circumstances

As we know WLAN standard IEEE 802.11b and IEEE 802.11g uses three nonoverlapping channels 1, 6, 11, so Simulation done supposed that there are three channels (i.e. sub-frequencies): Channel 1, 6 and 11. The goal is to find the best assignment that minimizes the interference. In other words, we have to find an assignment that guarantees that no neighbor APs use the same channel.

4.2 Representation and Encoding of the Chromosome

We used a chromosome which contains number of genes just like the number of APs in the network; each gene was represented by two bit, these two bits represent the channel number that used by the AP. so gene “00” represent channel 1, gene “01” represent channel 6, gene “10” represent channel 11, gene “11”

represent channel **16** which replaced by channel 1 or 6 or 11 according to the optimal solution. For example, if the chromosome is : **00 10 01 00 11 10 11 01 00**, then this means that the Network consists of **9 APs**, where **AP#1** uses channel 1 (which represented 00 in binary), **AP#2** uses channel 11 (which represented 10 in binary), **AP#3** uses channel 6 (which represented by 01 in binary) and so on. See Figure (4) below.

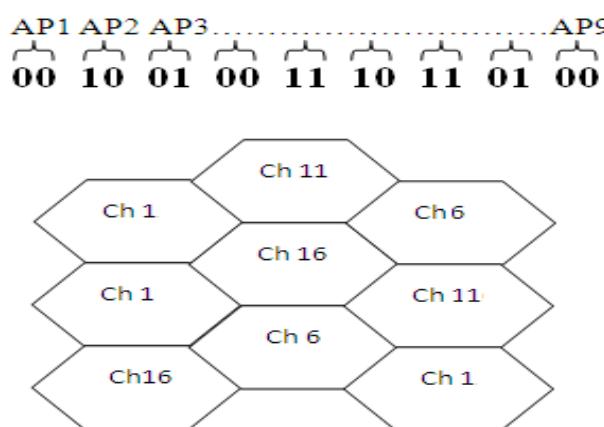


Fig. 4 Chromosome representation

4.3 Fitness Function

The fitness value has a value equals to 0 at the start of GA process. This value will be incremented by 1 for each two neighbor APs that use the same frequency channel. Chromosome with the minimum fitness value means that it is the optimal solution for the APs Allocation Channel Assignment problem.

- Fitness of each individual in a chromosome is calculated by using this Fitness Function:
 1. Fitness value is 0 by default.
 2. - The function starts checking the “gene” one by one.
 3. - Fitness value will be increased by 1 for each two neighbor APs “genes” use the same frequency channel.

4.4 Selection

Selection is the process of choosing structures for the next generation from the structures in the current generation [14]. In the simulation the Selection algorithm used is the "Roulette Wheel Selection",

which allocate to each individual a portion of the wheel proportional to the individual's fitness.

4.5 Simulation Parameters

Table 1 below shows the parameters that used in the simulations.

Table 1: Simulation Parameters

Total APs Number	16 , 36, 64
Crossover Rate	20% , 40% , 60%
Mutation Rate	0.1% , 0.7% , 2%
Selection Method	Roulette Wheel Selection
Population Size	100

5. Simulation Results

After running the experiments using the parameters defined previously, we started drawing the relation between the generations and the fitness value. Here are the results in details:

5.1 Results by Crossover rate

Figures (5), (6) and (7) show the relation between the fitness value and generations when crossover rate value was 20%, 40% and 60%. Mutation rate was 0.1% for each experiment. Number of APs in experiments represented in Figure (5) was 16 APs; in Figure (6) was 36 APs; while it was 64 APs in experiment represented in Figure (7).

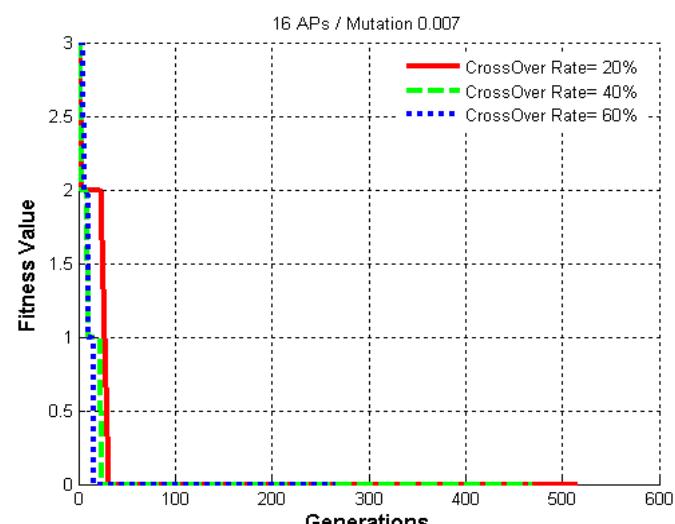


Fig. 5 Change by the Crossover rate in 16 APs

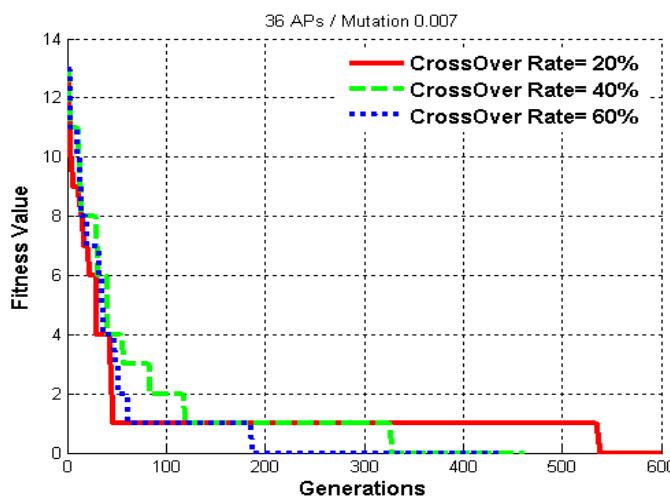


Fig.6 Change by the Crossover rate in 36 APs

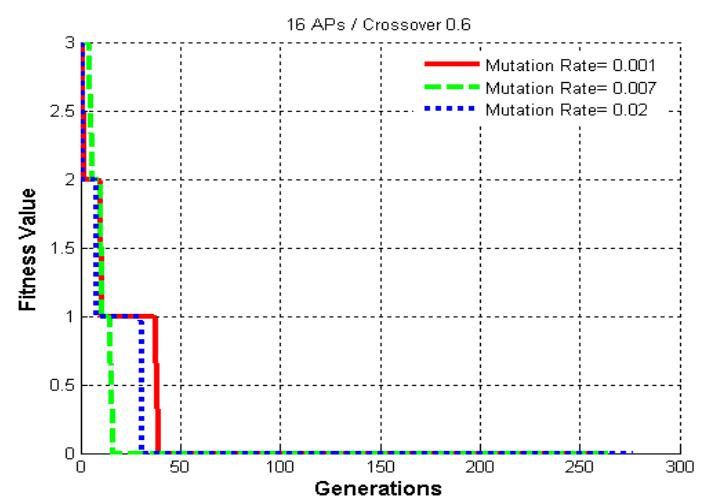


Fig.8 Change by the Mutation rate in 16 APs

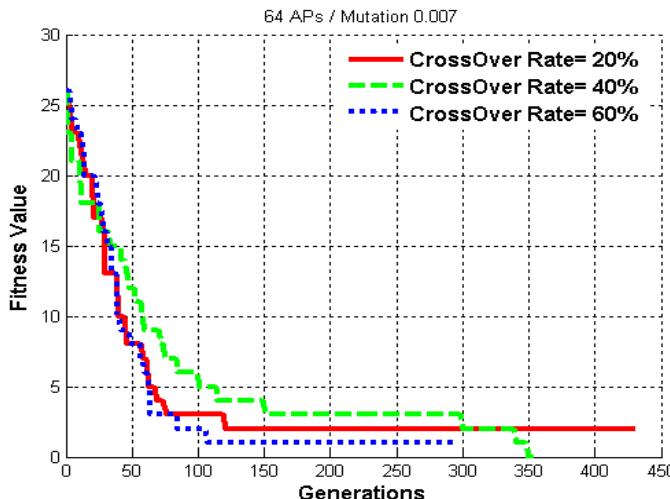


Fig. 7 Change by the Crossover rate in 64 APs

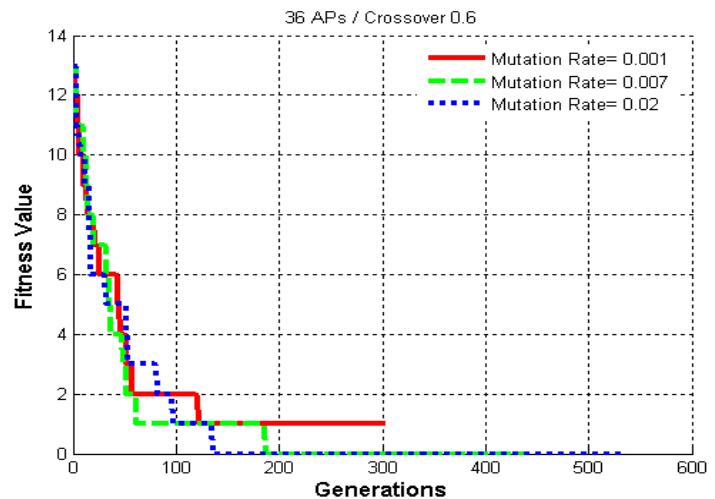


Fig. 9 Change by the Mutation rate in 36 APs

5.2 Results by Mutation rate

Figures (8), (9) and (10) show the relation between the fitness value and generations when mutation rate value was 0.1%, 0.7% and 2%. Crossover rate was 60%. Number of APs in experiments represented in Figure (8) was 16 APs, in Figure (9) was 36 APs, while it was 64 APs in experiment represented in Figure (10).

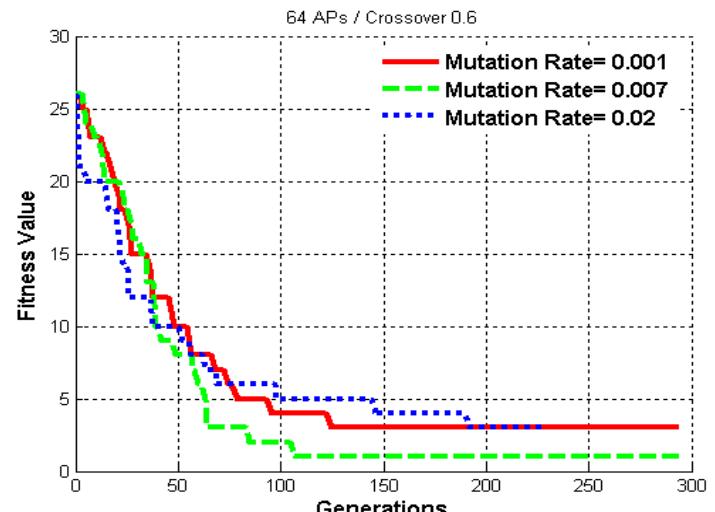


Fig. 10 Change by the Mutation rate in 64 APs

6. Conclusions

This paper introduced a frequency assignment method using genetic algorithm. This method differs from the method traditionally used in mobile systems (channel pattern).

This method worked well for small WLAN systems that maximum contain number of APs around 49 and introduces a multi and rapid solutions, but the performance of this method was decrease when using it with bigger WLAN systems.

From our point of view, this happened because the tool we used will face a tremendously large number of solutions that can be generated when the chromosome is long. This made the job of the tool more hard.

From the above figures we could note that the optimal Crossover Rate was between 40% and 60%, and the optimal Mutation Rate was around 0.7%.

References

- [1] R. Akl and S. Park. "Optimal Access Point selection and Traffic Allocation in IEEE802.11 Networks", proceeding of 9th World Multiconference on Systemic, Cybernetics and Informatics (WMSCI 2005): Communication and Network Systems, Technologies and Applications, paper no.S464ID, July 2005.
- [2] Upamanyu Madhow, "Fundamentals of Digital Communication", Cambridge University Press, 2008.
- [3] T. Carpenter and A. Barrett, "CWNA Certified Wireless Network Administrator Official Study Guide", McGraw-Hill Osborne Media; 4 edition, 2006.
- [4] The free encyclopedia, (Inc 2009). WWW.Wikipedia.com. Wikipedia ® is a registered trademark of the Wikipedia Foundation.
- [5] 802.11b working group, "Wireless LAN medium access control (MAC) and physical layer (PHY) specification: Higher-speed physical layer extension in the 2.4 GHz band"1999, IEEE standard.
- [6] 802.11a working group, "Wireless LAN medium access control (MAC) and physical layer (PHY) specification: Higher-speed physical layer in the 5 GHz band"1999, IEEE standard.
- [7] 802.11g working group, "Wireless LAN medium access control (MAC) and physical layer (PHY) specification: Further higher data rate extension in the 2.4 band"2003, IEEE standard.
- [8] P. Roshan and J. Leary, "802.11 Wireless LAN Fundamentals," Cisco Press publishing, 2004.
- [9] Junichi Yoshino, Isao Ohtomo, "Study on efficient channel assignment method using the genetic algorithm for mobile communication systems", Springer-Verlag, 2004.
- [10] Albert Y. Zomaya, Senior Member, IEEE, and Michael Wright. "Observations on Using Genetic-Algorithms for Channel Allocation in Mobile Computing" . IEEE Transactions on Parallel and Distributed Systems, Volume. 13, No. 9, Sep. 2002
- [11] Jong-Hyouk Lee, Byung-Jin, Hyo-Keun Bang, and Tai-Myoung Chung. "An Optimal Access Points Allocation Scheme based on Genetic Algorithm", Internet Management Technology Laboratory, 2008.
- [12] J.H. Holland, "Adaptation in Natural and Artificial Systems" University of Michigan Press, 1975.
- [13] Michael Liven, "The Evolution of Understanding: A Genetic Algorithm Model of the Evolution of Communication" Genetics Dept. Harvard Medical School, 2000.
- [14] D.E. Goldberg, "Genetic Algorithms in Search, Optimization and Machine Learning," Addison Wesley, Reading, MA, 1989.
- [15] Mitchell Melanie, "An Introduction to Genetic Algorithms", MIT Press, 1999.



Author biography He got his B.Sc., M.Sc. (Avionics Engineering) and PhD (Telecommunications Engineering and IT) degrees from International University of Civil Aviation, Kyiv-Ukraine republic in 1994, 1996 and 1998 respectively. He was an Assistant Professor at Communications Eng. Dept. and Dean Assistant, Faculty of Engineering in Al-Ahlyyah Amman University (1998-2002), Amman-Jordan. He was a Head of IT Dept. and currently he is an Associate professor at CIS Dept., Faculty of IT in Al-Balqa' Applied University since 2002, Al-Salt-Jordan. He is a Member of editorial and steering committees for some international journals and conferences. His research interests include: Networking, Wireless Telecommunications and Networks and Information Security. He has published several (over 20 papers) international journals and Conferences papers in the above area, and one book. He is a supervisor of many under graduated and Master graduated projects and master thesis in the above fields. He is a Member of IAENG organization.

User Navigation Pattern Discovery using Fast Adaptive Neuro-Fuzzy Inference System

J Vellingiri¹, Dr. S.Chenthur Pandian²

¹Assistant Professor, Department of CSE
Kongunadu College of Engineering and Technology
Thottiam, Trichy (Dt) - 621 215, TamilNadu, India.

²Principal, Dr. Mahalingam College of Engineering and Technology
Pollachi-642 003, TamilNadu, India.

Abstract

World Wide Web is a huge repository of web pages and links. It provides abundance information for the Internet users. The growth of web is incredible as it can be seen in present days. Users' accesses are recorded in web logs. From the user's perspective, it is very difficult to extract useful knowledge from the huge amount of information and secondly, it is also difficult to extract for the users to access relevant information efficiently. From business point of view, the webmasters and administrators find it difficult to organize the contents of the websites to cater to the needs of the users. Both the problems can be solved if the web navigation behavior of a user can be understood. One way to extract such information is to use Web Usage Mining. Web Usage Mining consists of preprocessing, pattern discovery and pattern analysis. Preprocessing is the step which transforms the raw log file into a form that is more suitable for mining. Four steps are used in preprocessing, they are, data cleaning, user identification, session identification and formatting the result to suit the clustering algorithm. The clustering technique used in this paper is Fuzzy-Possibilistic C-Means clustering technique. In pattern analysis, this paper uses Fast Adaptive Neuro-Fuzzy Inference System (FANFIS). The experimental results suggest that the proposed technique for web log mining results in better prediction of user behaviors when compared to the existing web usage mining techniques.

Keyword: Raw log file, Preprocessing, Fuzzy-Possibilistic C-Means, Web Usage Mining, Adaptive Neuro Fuzzy Interference System (ANFIS).

1. INTRODUCTION

World Wide Web (WWW) [9, 10], in the current era of information explosion, has become the large source of online data, which includes text, graphics, videos, sound, etc. WWW is global data medium where users read, write and communicate through computers associated with the Internet. It has become the default technology for sharing innovative schemes and content exchange. The impact of the Internet on everyday life is tremendous and it has changed the way of doing business, providing and receiving education, organization management, etc. The manner of information collection and sharing has changed with the advancement of hardware and communication software. The number of people using WWW is around 20.1 per cent in Asia alone and more than 234 million websites and 126 blogs [16] which shows the enormous growth and development of the WWW.

Thus, it can be concluded that the Web is a diverse, dynamic and unstructured data repository, which provides vast amount of useful information and also illustrates the complexity of handling huge amount of information from the different viewpoints, users, Web service providers, business analysts. The effective search tools are very much necessary to identify meaningful and useful information precisely.

The growth has inspired the web service providers to forecast the user's web usage behaviors so that, they can

- (i) Personalize the information provided to them
- (ii) Make the websites more user friendly
- (iii) Reduce the traffic load
- (iv) Create or modify their website to suit different group of people.

The application of machine learning approaches to web-based information for learning or extracting data is broadly called as Web mining [14]. The approaches in web mining concentrates on offering solutions to content provider, web designer and programmers to enhance their website and also to the web users with navigation assistance tools. It is a segment of data mining in which knowledge is obtained from WWW. This approach comprises of preprocessing, pattern discovery and pattern analysis. Pattern discovery is carried out through Fuzzy Possibilistic C-Means clustering approach. In pattern analysis, this paper uses Fast Adaptive Neuro-Fuzzy Inference System.

2. RELATED WORKS

This section reviews some proposals made in web personalization [5, 6], system and business intelligence through the use of web usage mining.

Web mining [7] has been significantly applied to several areas like research trend identification, robot detection and filtering to split human and non-human nature, user profiling, fraud and threat investigation and discovering web communities. Several tools and techniques have been proposed which examine web log data to offer knowledge that will facilitate web administrators in building effective websites. Chi et al. [1] presented a Web Ecology and Evolution Visualization (WEEV) tool to recognize the association between Web content, Web structure and Web

Usage over a period of time. Web Usage Mining [19] is the most sought after tool in the Internet community where data from online web is converted to meaningful knowledge. The knowledge [12] thus discovered can be used in web personalization, general system improvement, improve business intelligence, site modification and discover usage characteristics. Identifying usage characteristics from navigational patterns has been one of the vital areas of research.

Mobasher et al. [2] presented a potential approach for identifying user profiles depending on association rules discovery and usage based clustering integrating with present position of an on-going activity to carry out real time personalization [15]. This approach according to web usage mining provides Personalized Site Maps that are focused to the interest of each individual visitor.

Shahabi and Kashani [3] illustrated a whole structure for web-usage mining to fulfill the difficult necessities of web-personalization applications [8]. The author proposed a distributed user-tracking technique for correct, scalable, and inherent collection of the usage data [11] and presented a Feature-Matrices (FM) approach, to identify and interpret the access patterns of the users. A new similarity evaluation depending on FM was developed for correct categorization of partial navigation prototypes in real time. This proposed approach is well suited for both synthetic and real data for anonymous and efficient web personalization. A study of the popular approaches and tools presented for web personalization was provided by Eirinaki and Vazirgiannis [4]. The personalization agent employs the user model data along with the past identified sequential patterns and uses a collection of personalization [13] rules to provide the personalization works or operations like memorization of personal information, user salutation, recommendation of links connected to what users in the same category past selected or links that the same user generally views, objects separation by giving various features of each object.

3. METHODOLOGY

Data cleaning is one of the preprocessing techniques which are used to eliminate inappropriate records that are not essential for mining. Data cleaning comprises of the following steps:

- Removal of records of graphics, videos and the format information.
- Removal of records with the failed HTTP status code.
- Robots cleaning.

The unwanted and unrelated data are removed using the above steps. Knowledge extraction is the next step after the preprocessing phase. The main objective of this step is to find out the user behavior and the navigation patterns. For this purpose clustering algorithm is used. Clustering plays a significant role in data analysis and understanding the behavior of users in the websites. It combines the data into classes or clusters with the intention that the data objects

inside a cluster have huge similarity in relationship to one another, but are very dissimilar to those data objects in other clusters. In this paper, Fuzzy Possibilistic C-Means Algorithm is used to find out to extract the user behavior.

The Fuzzy C-Means (FCM) [17, 18] can be regarded as the fuzzified form of the k-means algorithm. It is an approach of clustering which permits one piece of data to fit in to two or more clusters. This method is commonly used in pattern recognition. The algorithm is an iterative clustering approach that generates an optimal c partition by reducing the weighted within group sum of squared error objective function.

FPCM is an integration of both possibilistic C-Means (PCM) and Fuzzy C-Means (FCM) that are supposed to circumvent a variety of difficulties of PCM and FCM. FPCM completely ignores the noise sensitivity deficiency of FCM, overcomes the coincident clusters problem of PCM. To predict the user behavior existing approaches have been used FCM and PCM. But the existing approaches are inadequate because of its sensitivity towards noise. Thus with the help of FPCM, noise is reduced, provides more accuracy and thus provides better result in predicting the user behavior.

In the online phase, when a new request appears at the server, the wanted URL and the session to which the user belongs are determined, the primary knowledge base is restructured, and a list of implication is suggested to the demanded page. After the clustering is performed, the output will be a set of clusters $np' = \langle np_1, np_2, \dots, np_n \rangle$ where $np_i = \langle P_1, P_2, \dots, P_k \rangle$ where k represents the set of web pages identified as user navigation patterns and $1 \leq i \leq n$. Sequence $W' = \langle P_1, P_2, \dots, P_m \rangle$ represents a current active session and m indicates size of active session window.

The web pages present in the active session are sorted and after this the prediction list is determined by means of classification. After this process, for building the prediction list, the technique used is Fast Adaptive Neuro-Fuzzy Inference System algorithm.

The ANFIS [20] is a framework of adaptive technique to assist learning and adaptation. This kind of framework formulates the ANFIS modeling highly organized and not as much of dependent on specialist involvement. To illustrate the ANFIS architecture, two fuzzy if-then rules according to first order Sugeno model are considered:

Rule 1: If (x is A₁) and (y is B₁) then (f₁ = p₁x + q₁y + r₁)

Rule 2: If (x is A₂) and (y is B₂) then (f₂ = p₂x + q₂y + r₂)

where x and y are represents the inputs, A_i and B_i indicating the fuzzy sets, f_i indicates the outputs within the fuzzy region indicated by the fuzzy rule, p_i, q_i and r_i shows the design parameters that are determined while performing training procedure.

The ANFIS architecture to execute these two rules is represented in figure 1, in which a circle shows a fixed node and a square shows an adaptive node.

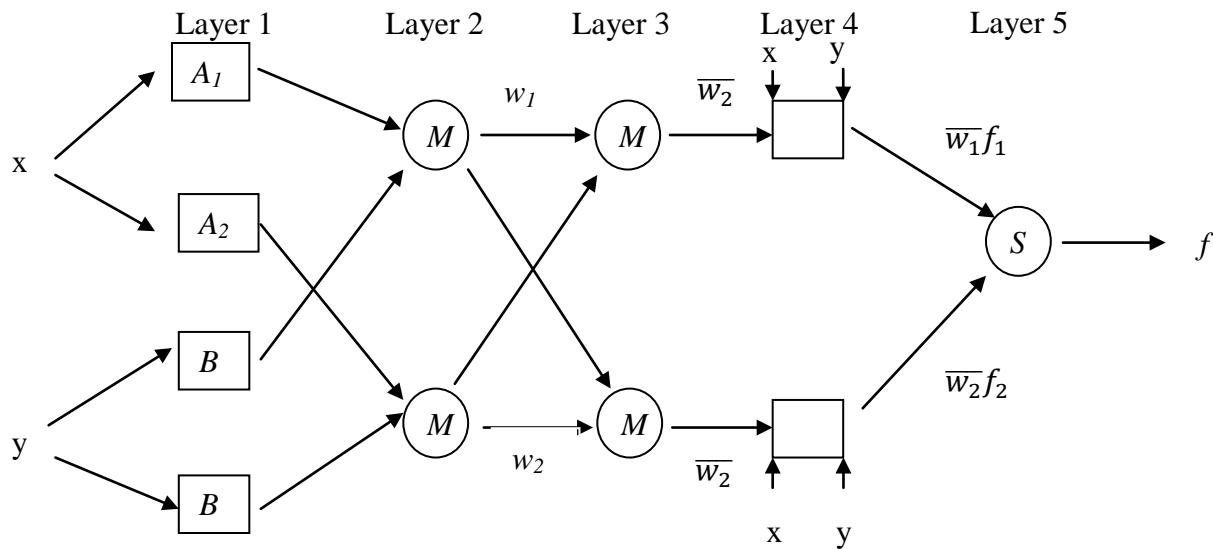


Figure 1: ANFIS Architecture

All the nodes in the initial layer are adaptive. The outcomes from these adaptive nodes are fuzzy membership grade of the inputs that are indicated by:

$$O_i^1 = \mu_{A_i}(x) \quad i = 1, 2 \quad (1)$$

$$O_i^1 = \mu_{B_{i-2}}(y) \quad i = 3, 4 \quad (2)$$

where $\mu_{A_i}(x), \mu_{B_{i-2}}(y)$ can allow any fuzzy membership function. For instance, if the bell shaped membership function is utilized, $\mu_{A_i}(x)$ is given by:

$$\mu_{A_i}(x) = \frac{1}{1 + \left\{ \left(\frac{(x - c_i)}{a_i} \right) \right\}^{b_i}} \quad (3)$$

where a_i, b_i and c_i is nothing but the parameters of the membership function which controls the bell shaped functions accordingly.

In the second layer, the nodes are fixed. These nodes are named with M, indicating that they perform as a simple multiplier. The outcome of this layer can be given by:

$$O_i^2 = w_i = \mu_{A_i}(x)\mu_{B_i}(y) \quad i = 1, 2 \quad (4)$$

which represents the firing strengths of the rules.

Also, the nodes in third layer are fixed. They are named with N, indicating that they are occupied in a normalization function to the firing strengths from the earlier layer.

The outputs of this layer can be represented as:

$$O_i^3 = \bar{w}_i = \frac{w_i}{w_1 + w_2} \quad i = 1, 2 \quad (5)$$

which represents the normalized firing strengths.

All nodes that are in fourth layer are adaptive. The outcome of the entire node in this layer is just the multiplication of the normalized firing strength with the first order polynomial. As a result, the outcome of this layer is represented by:

$$O_i^4 = \bar{w}_i f_i = \bar{w}_i (p_i x + q_i y + r_i) \quad i = 1, 2 \quad (6)$$

There is only one node named 'S' in the layer 5. This node performs the addition of all the incoming signals. Thus, the overall output of the model is given by:

$$O_i^5 = \sum_{i=1}^2 \bar{w}_i f_i = \frac{\sum_{i=1}^2 w_i f_i}{w_1 + w_2} \quad (7)$$

It can be distinguished that layer 1 and the layer 4 are adaptive layers. Layer 1 composes of three adjustable parameters like a_i, b_i and c_i that is related to the input membership functions.

These parameters are represented as premise parameters. In layer 4, there exists three adjustable parameters namely p_i, q_i and r_i , related to the first order polynomial. These parameters are called consequent parameters.

Learning algorithm of ANFIS

The intention of the learning algorithm is to adjust all the modifiable parameters such as $\{a_i, b_i, c_i\}$ and $\{p_i, q_i, r_i\}$, for the intention of matching the ANFIS output with the training data. If the parameters such as a_i, b_i and c_i of the membership function are unchanging, the outcome of the ANFIS model can be given by:

$$f = \frac{w_1}{w_1 + w_2} f_1 + \frac{w_2}{w_1 + w_2} f_2 \quad (8)$$

Substituting Eq. (5) into Eq. (8) yields:

$$f = \bar{w}_1 f_1 + \bar{w}_2 f_2 \quad (9)$$

Substituting the fuzzy if-then rules into Eq. (15), it becomes:

$$f = \bar{w}_1(p_1 x + q_1 y + r_1) + \bar{w}_2(p_2 x + q_2 y + r_2) \quad (10)$$

After rearrangement, the output can be expressed as:

$$f = (\bar{w}_1 x)p_1 + (\bar{w}_1 y)q_1 + (\bar{w}_1)r_1 + (\bar{w}_2 x)p_2 + (\bar{w}_2 y)q_2 + (\bar{w}_2)r_2 \quad (11)$$

which is nothing but the linear display of the changeable resulting parameters like p_1, q_1, r_1 and p_2, q_2, r_2 . The least squares distance method can be used to identify the optimal values of these parameters easily. If the basis parameters are not changeable, the search space turns to be larger and directs to allowing for large time for convergence. A hybrid technique combining the least squares distance measure and the gradient descent method is used for the purpose of solving this drawbacks. The hybrid algorithm contains a forward pass and a backward pass. The least squares method that functions as a forward pass is used for the purpose of determining the outcome parameters with no alterations in the premise parameters. Once the optimal consequent parameters are identified, the backward pass starts directly. The gradient descent method that functions as a backward pass is used to fine-tune the premise parameters equivalent to the fuzzy sets in the input domain.

The outcome of the ANFIS is determined by utilizing the outcome parameters gathered in the forward pass. The output inaccuracy is used to modify the principle parameters with the help of standard back propagation technique. It has been accepted that this hybrid method is very proficient in training the ANFIS. Learning can be fast up in ANFIS using Modified Levenberg-Marquardt algorithm.

Modified Levenberg-Marquardt algorithm

A Modified Levenberg-Marquardt algorithm is used for training the neural network. Considering performance index is $F(w) = e^T e$ using the Newton method we have as:

$$W_{K+1} = W_K - A_K^{-1} \cdot g_K \quad (12)$$

$$A_K = \nabla^2 F(w)|_{w=w_k} \quad (13)$$

$$g_k = \nabla F(w)|_{w=w_k} \quad (14)$$

$$[\nabla F(w)]_j = \frac{\partial F(w)}{\partial w_j} = 2 \sum_{i=1}^N e_i(w) \cdot \frac{\partial e_i(w)}{\partial w_j} \quad (15)$$

The gradient can write as:

$$\nabla F(x) = 2 J^T e(w) \quad (16)$$

Where

$$J(w) = \begin{bmatrix} \frac{\partial e_{11}}{\partial w_1} & \frac{\partial e_{11}}{\partial w_2} & \cdots & \frac{\partial e_{11}}{\partial w_N} \\ \frac{\partial e_{21}}{\partial w_1} & \frac{\partial e_{21}}{\partial w_2} & \cdots & \frac{\partial e_{21}}{\partial w_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial e_{KP}}{\partial w_1} & \frac{\partial e_{KP}}{\partial w_2} & \cdots & \frac{\partial e_{KP}}{\partial w_N} \end{bmatrix} \quad (17)$$

$J(w)$

is called the Jacobian matrix.

Then, Hessian matrix is identified. The k, j elements of Hessian matrix provides:

$$[\nabla^2 F(w)]_{k,j} = \frac{\partial^2 F(w)}{\partial w_k \partial w_j} = 2 \sum_{i=1}^N \left\{ \frac{\partial e_i(w)}{\partial w_k} \frac{\partial e_i(w)}{\partial w_j} + e_i(w) \cdot \frac{\partial^2 e_i(w)}{\partial w_k \partial w_j} \right\} \quad (18)$$

The Hessian matrix can then be expressed as follows

$$\nabla^2 F(w) = 2 J^T (W) \cdot J(W) + S(W) \quad (19)$$

$$S(w) = \sum_{i=1}^N e_i(w) \cdot \nabla^2 e_i(w) \quad (20)$$

If $S(w)$ is small assumed, the Hessian matrix can be approximated as

$$\nabla^2 F(w) \approx 2 J^T (w) J(w) \quad (21)$$

$$\begin{aligned} W_{k+1} &= W_k - [2 J^T (w_k) \cdot J(w_k)]^{-1} 2 J^T (w_k) e(w_k) \\ &\approx W_k - [J^T (w_k) \cdot J(w_k)]^{-1} J^T (w_k) e(w_k) \end{aligned} \quad (22)$$

The main benefit of Gauss-Newton is that it does not need computation of second derivatives.

There is a problem the Gauss-Newton method is the matrix $H = J^T J$ may not be invertible. This can be overcome by using the following modification.

Hessian matrix can be written as

$$G = H + \mu I \quad (23)$$

Suppose that the eigenvalues and eigenvectors of H are $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ and $\{z_1, z_2, \dots, z_n\}$. Then:

$$\begin{aligned} G z_i &= [H + \mu I] z_i \\ &= H z_i + \mu z_i \\ &= \lambda_i z_i + \mu z_i \\ &= (\lambda_i + \mu) z_i \end{aligned} \quad (24)$$

Thus, the eigenvectors of G are the similar as the eigenvectors of H , and the Eigen values of G are $(\lambda_i + \mu)$. The matrix G is positive definite by rising μ until $(\lambda_i + \mu) > 0$ for all i thus the matrix will be invertible.

This leads to Levenberg-Marquardt algorithm:

$$w_{k+1} = w_k - [J^T (w_k) J(w_k) + \mu I]^{-1} J^T (w_k) e(w_k) \quad (25)$$

$$\Delta w_k = [J^T (w_k) J(w_k) + \mu I]^{-1} J^T (w_k) e(w_k) \quad (26)$$

The learning parameter, μ denotes illustrator of steps of actual output movement to preferred output. In the standard Levenberg-Marquardt technique, μ denotes a constant number. This paper modifies LM technique using μ as

$$\mu = 0.01 e^T e \quad (27)$$

Where e denotes a $k \times 1$ matrix thus $e^T e$ is a 1×1 therefore $[J^T J + \mu I]$ is invertible.

Hence, if actual output is out of range than preferred output or likewise, errors are huge so, it converges to preferred output with large steps. Similarly, when quantity of error is very less

then, actual output approaches to preferred output with soft steps. Thus, error oscillation greatly minimizes.

4. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed FANFIS technique, numerous experiments have been conducted on the data from the UCI dataset repository (<http://www.ics.uci.edu>) that comprises of several logs from msnbc.com for the month of September 1998.

Every sequence related to the page views of a user throughout that 24 hour time period. Each event in the sequence corresponds to a users request for a page.

TABLE I: Sample Sequence

Sequence	Order of user page visits
1	1 1
2	2
3	3 2 2 4 2 2 2 3 3
4	5
5	1
6	6
7	1 1
8	6
9	6 7 7 7 6 6 8 8 8 8
10	6 9 4 4 4 10 3 10 5 10 4 4 4
11	1 1 1 1 1 1 1 1
12	12 12
13	1 1

Consider there are 17 page categories “Home Page”, “News”, “Technology”, “Cinema”, “Opinion”, “Politics”, “Miscellaneous”, “Weather”, “Health”, “Living”, “Business Opportunities”, “Sports”, “Summary”, “bbs” (bulletin board service), “Travel”, “Books”, and “Photos”.

Each category is related in sequential order with an integer starting from “1”. For example, “Home Page” is numbered as 1, “News” with 2, and “Technology” with 3. Each row below “% Sequences.” describes the hits in order of a single user. For instance, the first user hits “Home Page” twice, and the second user may hits “News” once.

The first similarity upper approximation at weighting threshold (m) value 0.5 is given by

$$\begin{aligned}
 R(T1) &= \{T1, T5, T7, T11, T13\}, \\
 R(T2) &= \{T2\}, \\
 R(T3) &= \{T3\}, \\
 R(T4) &= \{T4\}, \\
 R(T5) &= \{T1, T5, T11, T13\}, \\
 R(T6) &= \{T6, T8\}, \\
 R(T7) &= \{T1, T7, T11, T13\}, \\
 R(T8) &= \{T6, T8\}, \\
 R(T9) &= \{T9\}, \\
 R(T10) &= \{T10\} \\
 R(T11) &= \{T1, T5, T7, T11, T13\}, \\
 R(T12) &= \{T12\}, \\
 R(T13) &= \{T1, T5, T7, T11, T13\}
 \end{aligned}$$

The similarity upper approximations for

$$\begin{aligned}
 S1 &= \{T1, T5, T7, T11, T13\}, \\
 S2 &= \{T2\}, \\
 S3 &= \{T3\}, \\
 S4 &= \{T4\}, \\
 S5 &= \{T1, T5, T7, T11, T13\}, \\
 S6 &= \{T6, T8\}, \\
 S7 &= \{T1, T5, T7, T11, T13\}, \\
 S8 &= \{T6, T8\}, \\
 S9 &= \{T9\}, \\
 S10 &= \{T10\}, \\
 S11 &= \{T1, T5, T7, T11, T13\}, \\
 S12 &= \{T12\}, \\
 S13 &= \{T1, T5, T7, T11, T13\}.
 \end{aligned}$$

This indicates that user visiting the hyper links in T1 may also visit the hyperlinks in T5 and then T7, T11 and hyper links in T13. Also one who visits the hyper links in T6 possibly will also visit the hyper links in T8.

The proposed prediction accuracy is compared by means of using the sample sequence as 1-6-2, 3-7-1-10, 2-1-9-5-6, 9-2-5-7-1 and 1-3-4-2.

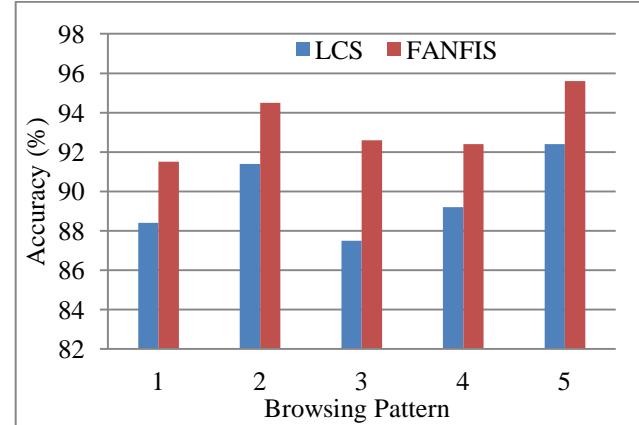


Figure 2: Accuracy of Web Page Prediction

Figure 2 represents the web page prediction accuracy by using various patterns suggested by the users. From the figure, it can be observed that the proposed technique results in better accuracy of prediction when compared to the existing technique.

5. CONCLUSION

Web mining is the extraction of interesting and useful knowledge and implicit information from artifacts or activity related to the WWW. Web servers record and accumulate data about user interactions whenever requests for resources are received. Analyzing the Web access logs can help understand the user behavior and the web structure. An important knowledge that can be obtained from web log files is the user’s navigation pattern. The navigation pattern knowledge can be used to help users from getting lost in the cyberspace by predicting their future request. For clustering the user behavior, this paper uses FPCM clustering technique. From the clustered results, several important statistics like number

of visits made to a webpage, the most popular web page among different users, common navigation pattern between users can be identified. Classification is performed in this paper using Fast Adaptive Neuro Fuzzy Interference System. Experimental results show that the proposed web usage mining technique results in better accuracy of predicting the web pages.

REFERENCES

- [1] Ed Chi, James Pitkow, Jock Mackinlay, Peter Pirolli, Rich Gossweiler, and Stuart Card, (1998). Visualizing the evolution of web ecologies. In CHI 98, pages 400-407, Los Angeles, CA.
- [2] Mobasher, B., Cooley, R. and Srivastava, J. (2000) Automatic Personalization Based on Web Usage Mining, Communications of the ACM, Vol. 43 , Issue 8, Pp: 142 - 151
- [3] Cyrus Shahabi, Farnoush Banaei-kashani, (2003), Efficient and anonymous web-usage mining for web personalization, Information Processing and Management
- [4] Magdalini Eirinaki, Michalis Vazirgiannis, (2003),Web mining for web personalization, ACM Transactions on Internet Technology (TOIT), Vol. 3, No. 1.
- [5] K. Devipriya and B. Kalpana. (2010) Users' Navigation Pattern Discovery using Ant Based Clustering and LCS Classification, Journal of Global Research in Computer Science, Pp. 1-5, Vo. 1, No. 1.
- [6] Jalali, M., Mustapha, N., Mamat, A., Md Sulaiman, N. (2008) OPWUMP An architecture for online predicting in WUM-based personalization system, 13th International CSI Computer Science, Springer Verlag, Vol. 6, Pp. 838-841.
- [7] Yan, W.T., Jacobsen, M., Garcia-Molina, H. and Umeshwar, (1996) From user access patterns to dynamic hypertext linking, Computer Networks and ISDN Systems, Proceedings of the Fifth International World Wide Web Conference, Elsevier, Computer Networks and ISDN Systems, Vol. 28, Issues 7-11, Pp. 1007-1014
- [8] Eui-Hong Han, Daniel Boley, Maria L. Gini, Robert Gross, Kyle Hastings, George Karypis, Vipin Kumar, Bamshad Mobasher, Jerome Moore, (1998), WebACE: a Web agent for document categorization and exploration, Pp. 408-415.
- [9] Joachims, T., Freitag, D. and Mitchell, T. (1997) Webwatcher: A tour guide for the world wide web. In The 15th International Conference on Artificial Intelligence, Nagoya, Japan.
- [10] Ngu, D.S.W. and Wu, X. (1997) Sitehelper: A localized agent that helps incremental exploration of the world wide web. In 6th International World Wide Web Conference, Santa Clara, CA.
- [11] Lieberman, H.(1995) Letizia: An agent that assists web browsing. In Proc. of the 1995 International Joint Conference on Artificial Intelligence, Montreal, Canada.
- [12] Mobasher, B. and Moore, J. (1998) Webace: a web agent for document categorization and exploration, Proc. of the 2nd International Conference on Autonomous Agents, ACM Press, Pp. 408—415.
- [13] Yan, T. W., Jacobsen, M., Garcia-Molina, H. and Umeshwar, D. (1996) From user access patterns to dynamic hypertext linking. Fifth International World Wide Web Conference, Pp. 52-54.
- [14] Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P.(1994) From data mining to knowledge discovery: An overview, Proc. ACM KDD, Pp. 23-39.
- [15] Ivancsy, R. and Kovacs, F. (2006) Clustering techniques utilized in web usage mining, Proceedings of the 5th WSEAS International Conference on Artificial Intelligence, World Scientific and Engineering Academy and Society (WSEAS), Knowledge Engineering and Data Bases, Pp. 237-242.
- [16] [\[www.thisblogrules.com/2010/07/facts-about-the-internet-infographic.html\]](http://www.thisblogrules.com/2010/07/facts-about-the-internet-infographic.html).
- [17] Lin Zhu, Fu-Lai Chung and Shitong Wang "Generalized Fuzzy C-Means Clustering Algorithm with Improved Fuzzy Partitions," IEEE Transactions on Systems, 2009.
- [18] Cheul Hwang, Frank Chung-Hoon Rhee, "Uncertain Fuzzy Clustering: Interval Type-2 Fuzzy Approach to C-Means," IEEE Transactions on Fuzzy Systems, 2007.
- [19] Chu-Hui Lee and Yu-Hsiang Fu, "Web Usage Mining Based on Clustering of Browsing Features", Eighth International Conference on Intelligent Systems Design and Applications, Vol. 1, Pp. 281-286, 2008.
- [20] K.AnandaKumar and Dr.M.Punithavalli, "Efficient Cancer Classification using Fast Adaptive Neuro-Fuzzy Inference System (FANFIS) based on Statistical Techniques", (IJACSA) International Journal of Advanced Computer Science and Applications, Special Issue on Artificial Intelligence, Pp. 132-137, 2011.



J VELLINGIRI doing research in Anna University of Technology, Coimbatore. He received M.E(CSE) degree from K.S.Rangasamy College of Technology which was affiliated under Anna University, Chennai. He worked as Lecturer in M.Kumarasamy college of Engineering and Technology, Karur and Selvam College of Technology, Namakkal from 2006 – 2010, and now he is with Kongunadu College of Engineering and Technology,Thottiam(Tk),Trichy(Dt), Tamilnadu, India as Assistant Professor in Department of Computer Science and Engineering.



Dr. S. CHENTHUR PANDIAN had obtained his Ph.D. degree (highly commendable) in Fuzzy Applications for Power Systems from Periyar University, Salem in 2006. Under his guidance 3 scholars are completed their Ph.D programme and he is presently guiding 13 research scholars for their Ph.D. programme. He has published 25 research articles in various National and International Journals and participated 41 Conferences in national and international level.

Approach to a conceptual model of indexation in a weak ontology described in RDFS

Traoré Issa¹, Souleymane Oumtanaga¹, Babri Michel¹ and Claude Lischou²

¹ Laboratory for Informatics and Telecommunications Research (LARIT), INPHB,
08 BP 475 Abidjan 08 (225), Cote d'Ivoire.

²University Cheikh Anta Diop (UCAD), BP 5353 Dakar-Fann-Senegal, Dakar (205), Senegal.

Abstract

In this paper, we describe a model of conceptual indexation adapted to the RDFS structured resources. The process of the structured IR (Information Research) in XML corpus, based on the balancing of a document terms improves the IR. However, silence in the resources return remains a concern. Operating a weak ontology described by RDFS can improve the percentage of relevant return of documents. The use of semantic relationship between pairs of synonymous modeled concept by graphs enables to take into account the meanings of the terms in a RDFS corpus. This model allows us to consider some types of structural information by exploiting the semantic properties, particularly the synonyms and abbreviations. What considerably improves the performance of the SIR (Systems of Information Research). Experiments on a collection of documents were used to evaluate our model.

Keywords: *structured IR, SIR, ontology, RDFS, BM25, synonyms.*

1. Introduction

The explosion of data volume and the improvement of the storage capacity of databases were not supported by the development of analytical tools and the IR to exploit this mass of information. The achievement of intelligent research systems has become a matter of urgency or even a necessity. Research methods based on the key-words are not accurate enough for the description of the text content. Given the increasing number of documents on the web, several models of SIR are created. Indeed, the main issue of the SIR is to find the ten or thousand odd relevant documents among millions of documents. This utmost makes this task even more difficult. There are some classical models of information research such as the Boolean model [18], the matching score [21], the vector model [28], [5], the probabilistic model [6], [22], and so on. These models are based on the IR of structured or non-structured documents. The main drawback of these models

is that they generate a significant silence [1], [27]. Indeed, two words with different syntaxes may have the same meaning. For instance: "father" and "dad" or "web 3.0" and "semantic Web". A user seeking information on a "semantic Web" will be unlucky to be informed of the "Web 3.0".

The structure and the content processing are the main issues of the structured information research. It is essential to obtain highly relevant scores in a graduated scale. In the purpose of increasing the research models, various strategies are implemented so as to be grafted. These strategies use various sources of evidence: semantic relationship defined in the thesaurus , classes and contextual use of the concepts, research results, relevant judgments of the users, element of the information theory, heuristics [3], [16], and [14].

The semantic relations provides a great advantage in IR, it allows exploiting the meaning of the concepts in the ontology, these latter being constructed with structured languages such as the XML (Extensible Markup Language). Several works have been done in IR with the format XML documents. The well-known are the works of TREC [1], [4] and [12], INEX [27], [13], then project WorldNet [7], [29] and MESH [31], they are ontologies to facilitate IR [26] and [33], by using the similarity distance between concepts. However, XML documents cannot efficiently operate on the meaning of terms.

Concerning the approach presented in this paper, the consideration of semantic in a weak ontology is exploited. These resources are described in the RDFS language (Resource Description Framework Schema). RDF is a markup language which has the advantage of describing objects. It was standardized in 1999 (by a W3C recommendation) [10]. It enables human beings to read multiple metadata schemes as well as their analysis by machines. It uses XML to express a structure allowing the metadata communities to define the real semantics. The RDFS is an extendable language knowledge representation

[11]. The RDFS resource is characterized by a content (of the text) and a structure (tags). However such documents cannot be efficiently exploited by the classical methods of IR. Indeed these latter treat a document only in terms of its content (key words), whereas XML and RDF are used to add structural constraints (tags). From time to time this requires to adapt the classical methods of IR or introduce new mechanisms in order to well exploit the meaning of the terms of the document.

This paper is organized as follows: in addition to the introduction, the first part develops the problematic and gives a state of the art XML documents in IR. The first contribution of this paper is the suggestion of a formal operational, semantics in a light ontology described in RDFS, presented in the second section. Then, considering explicitly the semantic of the terms of documents, we calculate the relevance score of a term paper and the document containing it. In part 4, we evaluate our model by an experiment on its corpus.

2. State of the art.

2.1. problematic

The basic question is: how must the information of the structured XLM or RDFS documents be indexed in order to respond efficiently to a request?

By focusing on the structure, we can identify some specific issues related to XML. It doesn't consider the semantics in terms of the documents [27]. Indeed, following a request from a user, only documents containing the query terms will be ranked and then returned. The problem is then: how to define a formalism which will reduce the silence by returning the documents containing the demanded requests and their synonyms? Can we combine the silence reduction and the improvement of the documents returned by the system? It's all these questions we will try to answer in this article.

2.2. Some research work

The research of a document on the web is based on the key words. In fact, the relevance of a web resource for a given query depends on one hand, on the weight of the query words, and on the frequency of the occurrence of words on the other hand.

Generally speaking, the web resources are not structured, the most existing used methods in IR are: Boolean method, vector method, Page Rank method, BM25, and so on.

The IR process role is to establish a correspondence between the relevant information sought by the user, usually represented through a query, and the set of

available documents. It is based on two essential steps: the indexing phase and the research phase.

The indexing phase focuses on the documents and requests analysis in order to create a representation of their textual content which should be usable by SIR. Each document (and research) is then associated with a descriptor, represented by a set of the extracted indexing terms.

The research phase aims at connecting the indexing of the documents with the user's request. It is based on a precise formality defined by a model of IR. The documents presented as a result to the user, and considered as the more relevant, are the ones which indexing terms are the closest to those of the request.

Many works have been done in the finding structured or semi-structured resources [3], [17]. XML documents are used in the background (words) and forms (tags). These principles have already been used separately in research INEX [14] and [11] and in other research work using the balancing of the tags [8]. Therefore, the XML research works done have sought not only to identify some more concise information units but also to use these markers to detect the information in a more appropriate response to a need of information. As a result, the relevance of a resource (Fig.1) not only depends on the frequency with which the request terms appear, but also on the weight of the tags which mark these words in the document.

Generally speaking, any SIR has two basic purposes: find all the relevant documents, and reject all the irrelevant ones. These purposes are evaluated by the measures of recall and precision.

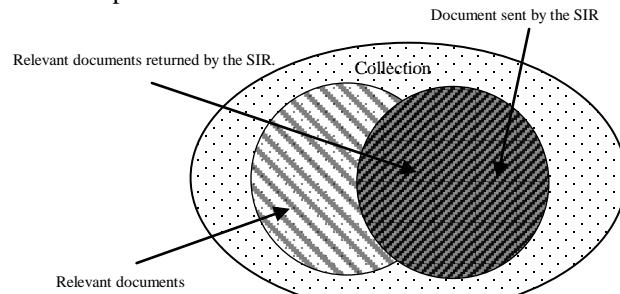


Fig. 1 Precision and Recall

In works [15] and [12], the tags judged important as well as their weights are selected manually. For instance in work [12], the weight of the "title" tag is set to 2 and the one of the "abstract" tag to 1.5. Another method in [9], consists in automatically taking these weights. For instance [12] and [2] use the genetic algorithm, whereas [32] uses techniques based on simulated annealing. In the case of a set of structured documents, the weight of resources is obtained by combining the weight of words and tags.

Let N_r be the number of documents returned by the system for a given query, N_p , the number of relevant

documents in the collection for this query and N_{pr} , the number of relevant documents returned by the system. Precision and recall are given by the following formulas: Precision measures the proportion of the relevant documents restored by the system. It is expressed by:

$$\text{Precision} = \frac{N_{pr}}{N_r} \quad (1)$$

The recall measures the proportion of relevant documents restored by the system related to the set of relevant documents in the basic information. It is expressed by:

$$\text{Recall} = \frac{N_{pr}}{N_p} \quad (2)$$

By these formulas, silence and noise rates are given by the following formulas:

$$\text{Silence} = 1 - \text{Recall} \quad \text{and} \quad \text{Noise} = 1 - \text{precision} \quad (3)$$

In the case of an ideal system, the precision rate is equal to the recall rate that is all the relevant documents are selected.

In the model we present, the structure and the meaning of the documents are operated at two levels:

- Logical structure and layout: the tags are used to determine the granularity of the index, so the granularity of the elements the system will likely return. The weight of each of the tags is expressed by learning.
- Formalization of the semantics: the terms with the same meaning are linked by a relation R. they are measured weighted in accordance with the R relation.

At the query step, the probability for a resource to be relevant is estimated by combining the weights of the words it contains with the weight of the tags which label them.

In the next section, we present a light ontology described in RDFS, then we will find the formality enabling to consider the meaning of the terms by the SIR.

3. The ontology built with lightweight RDFS

3.1. The RDFS language

Most of the standardized languages by the W3C as parts of the semantic web are WML dialects; this is the case of RDF and RDFS we are going to deal with in this article. RDFS provides some basic elements for defining

ontologies and vocabularies for structuring the RDF resources [10]. The language SPARQL (SPARQL Protocol and RDF Query language) is a query language for the RDF. Just like SQL for relational databases or XQuery for XML documents, this language is used to extract information from the RDF documents.

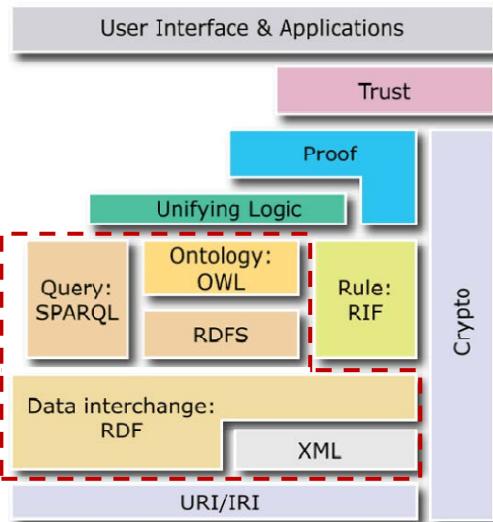


Fig. 2 Architecture of the semantic web

More specifically, the semantic web is based on an architecture called the Semantic Web Stack (see Fig.2). It essentially deals with an illustration of the languages hierarchy (XML, RDF, RDFS, OWL, and so on) used in the semantic web. It's important to notice that all these technologies have been standardized in the last decades by the W3C consortium to permit the passage across the semantic web. We can also notice that from the bottom of the stack down to this ontology, all the documents have been accepted and standardized.

The ontology construction requires a consensus in order to avoid the lexical disambiguation, due to hypernyms and polysemes. The light ontology constructed in RDFS is a graph in which inferences can be exploited.

RDFS descriptions are presented as oriented graph (see Fig.3 and Fig.4). There are various notations of RDFS graphs. The simplest one is the N-Triples, according to which the RDFS graph is represented by a collection of triplets [11] with the following abbreviated form:

ex : Patient rdf : SubClassOf ex : Person

This graph means that: “a patient” is described as a “sub class” of “persons”.

3.2. The light weight ontology in RDFS semantics

RDFS extends RDF language to describe more precisely the resources used to label the graph. For that, it provides a

mechanism to specify the classes whose resources will be instances like properties. The use of this set of theories to describe these models has two interests: the generic of set notion based on mathematical basis and its universality that is the semantic domain is for experts in this field.

RDFS resource can also be translated into a formula of the positive logic (without negation), conjunctive, and existential of the first order (without functional symbols), which models are the same with those defined by the direct semantic in the theory of models (see [19]). However, the extension to RDFS only provides primitive mechanism for specifying these classes.

In our approach, RDFS is enough to warrant the construction of a lightweight ontology describing some synonymous concepts. In this work, the literals will have a particular importance since they contain words whose scores are used by the SIR.

Let's consider the two RDFS below graphs representing the resources describing some diseases in the medical field.

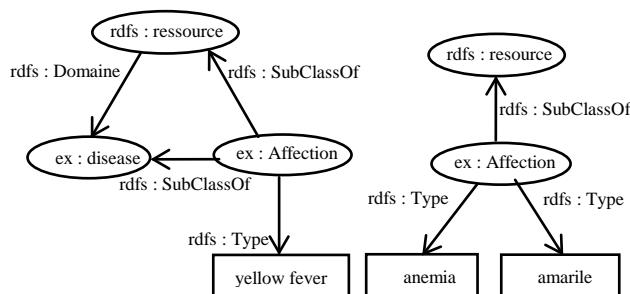


Fig.3 Resource 1

Fig.4 Resource 2

The abbreviated N-triple entry of the two graphs, an inference on the concepts see [29] and [30]. So the two examples bellow:

ex: Affection rdf:type "Yellow Fever"
 ex: Affection rdf:type "Amarile"

Let's suppose that these two RDFS graphs in are from the same ontology of a medical research. At the request of a user, who seeks documentation of Yellow fever, document n° 1 will be returned because it has a literal containing the term "Yellow fever". Document 2 will not be returned because it no longer contains the term "Yellow fever". Our approach consists in reducing the silence during the return of the relevant documents by the SIR. Knowing that " Amarile" stands for "Yellow fever" the following equivalence can be written:

IF rdf:hasSynonym rdf:type rdf:SymmetricProperty
AND ex:Amarile rdfs:hasSynonym ex:yellow_fever
THEN ex:yellow_fever rdfs:hasSynonym ex:Amarile

Thus, according to this algorithm, it is possible to infer the

terms of a document. Although the vocabulary with which to build ontology in RDFS is limited, it is possible to infer the concepts [19] and [20], and formally justify that two concepts are equivalent. The query language SPARQL is the RDFS whose syntax is:

```

SELECT ? Subject ? Object
WHERE {
    ? subject rdfs: subClassOf? object.
}
    
```

4. The database of synonyms

4.1. Properties of the relation of synonymy.

The database of synonyms will be composed of a set of couple of words. These couples will be composed of words and their synonyms or their abbreviations such as: (Web_3.0, Web semantic),

(Amarile, Yellow fever) or even (PC, Personal Computer), and so on. In some situations, a term may have various synonyms, so the exploitation of the synonym relationship noted *syn* () will solve this problem. In our approach we use the pure synonymy defined by:

Two lexical units are in a pure synonym relation if any occurrence of one can be replaced by the occurrence of the other in any environment without changing significantly the meaning of the statement in which it is. Let S be the set of synonym terms, $\forall A, B \in S$ $syn(A, B)$ means that the term B is the synonym of term A . The synonym relation thus defined checks if:

reflexivity : $syn(A, A) \quad \forall A \in S$;

symetry : $syn(A, B) \Rightarrow syn(B, A) \quad \forall A, B \in S$;

transitivity: $syn(A, B), syn(B, C) \Rightarrow syn(A, C) \quad \forall A, B, C \in S$.

So the relation of pure synonym is a relation of equivalence. It allows the construction of classes of lexical units taking into account all the word which are yonyms. These synonyms will be grouped in the database of synonyms.

The synonym relation enables to select in the database of synonym all the synonyms of a term. When the system receives an event (query), it searches simultaneously the "key" terms in the database of synonym. As soon as a term is found, then the other element of the couple is considered as the synonym of the term query. Thus, the synonym terms enrich the query in order to take into account the semantics.

The particularity of our approach is that it takes into account the various synonyms of the query key terms, so the database of synonyms will have a great importance. For

the enrichment, two basic methods are considered:

- First, it is manually created by experts in this domain, preferably those that will create the knowledge base (see Fig.5).
- Then a learning method that allows the automatic addition of new synonyms for user input. This can offer synonyms for certain query terms, which will be subject to registered and subsequently used by the SIR.

It is important to have a minimum consensus on the definitions of the field in order to avoid problems of polysemy and hypernyms. Since the performance of the SIR depends on the one hand, on the response time of the system and also the quality of answers that is to say, the relevance of the documents that will be returned.

Definition 1: Let I and J be two sets of the finite integers, $\rho_i = \bigcup_{i \in I} < s_i, P_i, o_i >$ is a collection of triplets of a resource, and $O = \bigcup_{k \in J} \rho_{k_i}$, a set of all the resources describing an ontology of documents in RDFS field.

Remark 1: The s_i objects are subjects or literals. In our approach we look for terms in the literals. As a result, the literals play an important role in our approach to SIR.

There, $\rho_1 = <s_1, P_1, o_1 = \{4t_1, 2t_2, t_3, t_4\}>$, a triplet of a resource which object is the literal with terms. The literal shows that the resource contains four times term t_1 , twice term t_2 , and once terms t_3 and t_4 . Without losing general information, we will simplify the writing of the resources by the following:

$$\nabla_1 = \{4t_1, 2t_2, t_3, t_4\}$$

To show that document ∇_1 contains four times term t_1 , twice term t_2 , and once the terms t_3 and t_4 .

Definition 2: Let $I \in \{1, 2, \dots, n\}$ and $J \in \{1, 2, \dots, m\}$.

$\rho_j = \bigcup_{j \in J} < s_j, P_j, o_j >$ be a document in the format of a collection of RDFS triplet. We call representative of ρ_j the set ∇_j of all the terms of the documents.

4.2. Approach of the system model of information research.

The search of relevant information which responds to the need of a user consists in corresponding the representation of the information contained in a document collection with those of the user's need. In the figure below (Fig.5), we present an architecture of our SIR model in the form of the "U" process. This process manipulates some documents

collection, a database of synonyms and some user's needs for search from a collection of documents, those which best meet a user's need.

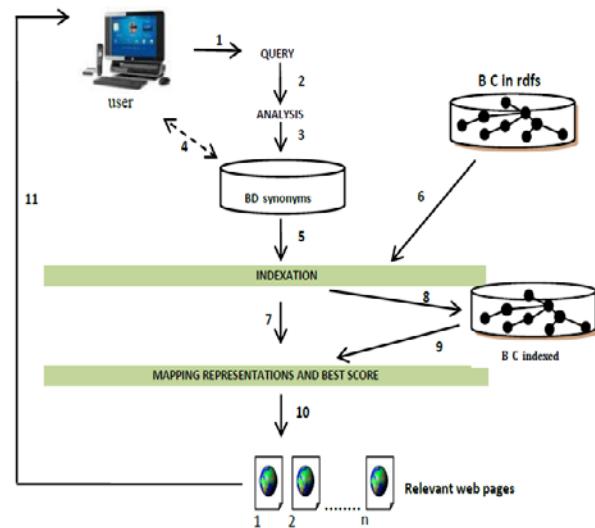


Fig.5. Conceptual architecture of the SIR.

The process of SIR is shown in Fig.5: (1) a user formulates his information need as a query, (2) this query is analyzed and then put in a RDFS format (triplet $<s, P, o>$). (3) Then the obtained description enables to look for synonyms in a database of synonyms, which enriches the query with the terms of the same meaning. (4) The enrichment of the synonym base by experts in the field or by users. (5) After identifying the synonyms, the indexing is done. At the same time or previously, (6) the RDFS collection documents in the knowledge base is also indexed. Thinks to the index (indexed collection and query), (8) the system can build the representations, then (7 and 9) mapping of the representation of the query with the representations of the document in the collection. (10) The calculation of the documents scores being done, (11) it returns a list of documents considered by the IR engine as relevant compared with the user request.

It is important to notice that in our model, the SIR proceeds to the removal of the tool-words (stop words). It deals with suppression the words of the common language which do not contain more semantic information (example: "a", "the", "of").

In the following paragraph, an evaluation of our approach allows checking the influence and advantage of the consideration of the meaning of documents.

4.3. A probabilistic approach of the resources representation.

The relevance of document ∇_j relatively to a query Q depends on the weight of the terms which appear in the document and in the request. We note w_{ji} the weight of term t_i in the document ∇_j . Let suppose that the weight of the t_i term in Q equals to 1 (the t_i term appears once in the query).

We define as X_j a vector of random variables and, $x_j = (x_{j1}, \dots, x_{ji}, \dots, x_{jn})$ a realization of this vector X_j , with $x_{ji} = 1$ (resp .0). If the term: t_i appears (resp. doesn't appear) in the document ∇_j .

Given these notations, let's consider π as the relevance of X_j based on the weights of the terms and U the universe of representative terms collection. π is given by the score :

$$\pi(x_j) = \sum_{t_i \in U \cap Q} x_{ji} \times w_{ji} \quad (4)$$

The BM25 weighting scheme (BM stands for “Best Match”) which has been developed by [22], and then BM25 OKAPI in [28] and [23], is a weighting scheme based on the probabilistic model. They use the probability distribution of 2-Poisson.

Several versions of this formula have been suggested, and BM25 is a compromise on the nature and the length of the documents and the queries. This formula has been set on the corpus of the TREC competition. The apparent complexity of the formula (compared to the vector model), and its high efficiency (this formula is currently one of the most successful) well shows the advantages of the probabilistic model. Below is the BM25 formula which enables to calculate the weight w_{ji} of a term t_i in a document ∇_j .

$\forall k_1, b \in \mathbb{R}$

$$w_{ji} = \frac{tf_{ji}(k_1 + 1)}{tf_{ji} + k_1 \times ((1 - b) + b \times ndl)} \times \log \frac{(N - dl_i + 0,5)}{(dl_i + 0,5)} \quad (5)$$

With:

tf_{ji} : the frequency of t_i in the document ∇_j .

- N : the number of documents in the collection
- dl_i : the number of documents with the term t_i
- $ndl = \frac{dl_i}{avg - dl_i}$: the ratio between the size of the document and the avg (average) size of the elements (in terms number)
- k_1 and b : the classical parameters of BM25

Parameter k_1 permits to adjust the overloading of tf_{ji} , and parameter b is used to adjust the focus on ndl , that is the importance of standardization of the items size. We will take $k_1=2$ and $b=0,75$.

Let's notice that we will choose k_1 and b so as BM25 will be a non-linear function compared to the terms frequency as stipulated by [9].

The weight of the t_i term of the request, called w_{Qi} takes into account the appearance number of term t_i in the request tf_{Qi} , and a parameter k_2 , of value equals to 8 (found in experiments).

$$w_{Qi} = \frac{tf_{Qi} \times (k_2 + 1)}{k_2 \times tf_{Qi}} \quad (6)$$

Let $\nabla_1 = \{4t_1, 2t_2, t_3, t_4\}$ be the representation of a document ∇_1 and $Q = <s ; P; \{t_1, t_2\}\rangle$ a user request. We consider $\nabla_Q = \{t_1, t_2\}$ as a representative of the request.

In our approach, we also consider the meaning of the query terms. Let t_i be a term representing ∇_Q t_{ik}^* the t_k term representing ∇_j as the synonym of t_i . The SIR will look for the concepts t_i and $t_{2,4}^*$ ($syn(t_i, t_{ik}^*)$) in the different documents. Indexing will be done with the BM25 formula with a change in the calculation of the tf_{ji} frequencies of the t_i term in a representative ∇_j .

$$tf_{ji} = \frac{n_i + \sum_{t_k \in \nabla_j} n_{ik}^*}{|\nabla_j|} \quad (7)$$

With:

- n_i : the number of terms in the representative ∇_j ;
- n_{ik}^* : The number of synonyms of the t_i term in the representative ∇_j ;
- $|\nabla_j|$: The sum of all the terms of the representative ∇_j .

So, tf_{ji} is the arithmetical average of the number of words having the same meaning in the resource ∇_j . We can deduce from it the following remark:

Remark 2: Let's focus on n_{ik}^* , the number of synonyms of the term t_i in the representative ∇_j .

- When $n_{ik}^* = 0$, we have the formula for BM25.

- When $n_{ik}^* \neq 0$, the representatives with t_i and t_k as terms will have their scores improved.
- The representatives with t_k term and not the t_i term will no longer have their nil score.

5. Evaluation of our approach

5.1. Score concepts

To evaluate our approach, we'll consider the universe representatives collection terms:

$U = \{t_1, t_2, t_3, t_4, t_5, t_6\}$. In this universe it is assumed that t_2 and t_4 have the same meaning doing . Learning will allow the system to identify the literal meaning like the same. The query representative $\nabla Q = \{t_1, t_2\}$.

Let the following representatives be ones of a corpus of 12 documents in RDFS:

$$\begin{aligned}\nabla_1 &= \{t_2, 5t_3, 7t_{2,4}^*, 2t_5\}, \nabla_2 = \{5t_1, 3t_2, 7t_5, 2t_6\}, \\ \nabla_3 &= \{t_1, t_2, 3t_{2,4}^*, 2t_5, 5t_6\}, \nabla_4 = \{2t_3, 8t_{2,4}^*, 2t_5, 8t_6\}, \\ \nabla_5 &= \{10t_2, t_3, 2t_5\}, \nabla_6 = \{2t_1, 7t_3, t_{2,4}^*, t_6\}, \\ \nabla_7 &= \{5t_1, 3t_3, 2t_5\}, \nabla_8 = \{t_1, t_2, 4t_5\}, \\ \nabla_9 &= \{t_1, t_2, t_3, t_{2,4}^*, t_5, t_6\}, \nabla_{10} = \{t_1, 4t_5, t_6\}, \\ \nabla_{11} &= \{3t_1, 5t_{2,4}^*\} \text{ et } \nabla_{12} = \{2t_2, 7t_{2,4}^*, 3t_6\}.\end{aligned}$$

The calculation of the representative scores terms lead to this document, we have the graphs below for the query Q . This score gives the first performance of the various documents by the BM25 method. From Fig.6, the relevant documents are respectively: ∇_2 , ∇_9 , ∇_8 , and ∇_3 .

In the same figure, the synonyms of query terms Q are taken into account. Indeed, terms t_2 and t_4 have the same meaning with the document containing them, they are calculated accordingly. We obtain the following graph below:

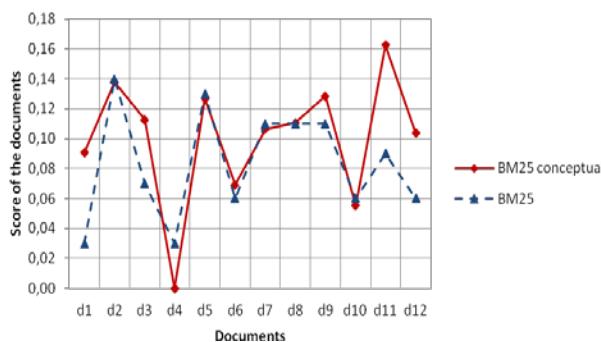


Fig.6 Score of the documents according to BM25 method and our approach.

The observation of this curve suggests that some documents with nil scores now have good scores. The relevant documents are in the following order: ∇_{11} , ∇_2 , ∇_9 , ∇_3 , ∇_8 et ∇_6 .

The consideration of the meaning of the terms increases the number of relevant documents and reduce the silence.

5.2. Probability of relevance of a concept

The SIR evaluation is a main issue on which the IR community has worked a lot. INEX (Initiative of the evaluation of XML retrieval), TREC (Text Retrieval Conference) and CLEF (Cross Language Evaluation Forum) designed for flat documents (see [4], [12], [24] and [25]). These evaluation campaigns of the SIR enabled a comparison as strict as possible of the systems of information research in the XML collections-oriented documents.

The return of the documents as an answer to a user need is based on the "Probability Ranking Principle (PRP)", stimulating a better performance of the system when the documents are returned in a decreasing number of their relevance.

In our approach, we will use the evaluation methods advised by the INEX, in which measures based on the recall-precision curves are widely used.

The probability for a document ∇ to be relevant for a given query Q , for measuring the performance of the SIR. For this purpose, we have two events:

- R , the event, the document is relevant for Q ;
- NR , the event, the document is not relevant for the query Q .

The recall also measures the probability for document ∇_k to be selected, knowing that it is relevant:

$$\text{Recall } 1 = P(t_{ik}=1|R) \quad (8)$$

The precision also measures the probability for a document ∇_k to be pertinent knowing that it is selected:

$$\text{Precision} = P(R| t_{ik}=1) \quad (9)$$

In practice, we must estimate the probability p_{ik} and q_{ik} , $i \in \{1, \dots, n\}$, $k \in \{1, \dots, m\}$ where:

- p_{ik} : the probability for t_i to appear in the document ∇_k , knowing that a relevant document.
- q_{ik} : the probability for t_i to appear in the document ∇_k , knowing that a non relevant document.

To make the evaluation of the ∇_k documents possible, p_{ik} and q_{ik} must be estimated on a set of pre-definite queries.

Given the set R (resp. NR) which has the relevant documents, and (resp. irrelevant), (see Table 1) the contingency can be constructed for each t_i term of the document we have . Given the consideration of the synonyms, we have the equalities below:

We write:

$\mu_{ik} = \sum_{i \in I} n_{ik}^*$ the sum of the various synonyms of t_i in the document ∇_k .

$$p_{ik} = \frac{r_{ik}}{R} \quad \text{and} \quad 1 - p_{ik} = \frac{R - r_{ik}}{R} \quad (10)$$

$$q_{ik} = \frac{n_i + \mu_{ik} - r_{ki}}{N - R} \quad (11)$$

$$\text{and} \quad 1 - q_{ik} = \frac{N - R - n_i - \mu_{ik} + r_{ik}}{N - R} \quad (12)$$

Hence the following table.

Table 1. Under the contingency documents

	Relevant (R)	Non relevant (NR)	Total
$t_i \in D_k$	r_{ik}	$n_i + \mu_{ik} - r_{ik}$	$n_i + \mu_{ik}$
$t_i \notin D_k$	$R - r_{ik}$	$N - n_i - \mu_{ik} + R - r_{ik}$	$N - n_i + \mu_{ik}$
Total	R	$N - R$	N

Table 1 enables to construct the recall/precision curve. The recall measures the capability of the system to find all the relevant documents and the precision measures its ability to find only relevant documents. The measure of the silence evaluation is a complementary notion to the recall, it is defined by;

$$S = 1 - \text{Recall} \quad \text{and} \quad S = 1 - \frac{r_{ik}}{R} \quad (13)$$

- Where S is the silence.

This measure is in the interval $[0,1]$. It also evolves toward to 0 (for r_{ik} evolves toward to R) for the best SIR. To examine efficiently the results, we calculate the pair of measures (recall rate, precision rate) for each returned document. Table 1 illustrates some calculations of precision and recall for the first ten documents returned by a system for a different query, for which the collection contains 4 relevant documents (in first BM25model) and 5 relevant documents (in our approach, BM25 conceptual). The associated recall-precision graphs are drawn on Fig.7. As we can notice it in Table 1, several precision values can correspond to the same point of each recall. In order to have curves which are easy to read, we generally present the interpolated precision calculated at each point of the

recall (which means that each relevant document is returned).

The perfect system will only find the relevant documents with a precision and a recall of 100%. In practice, the precision and recall measures increase inversely, which means that the interpolated curve according to the recall decreases. The higher the curve is, more performing is the system. We consider the 10 first referred documents.

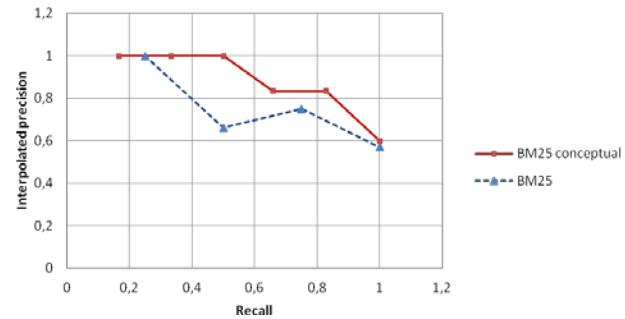


Fig.7 Curve of Recall / Precision

By comparing the rates of precision and recall of the suggested collection, we can notice that by integrating the meaning of the terms in our indexing system, we have better performances compared to the best results obtained by using BM25. In fact, the integration of semantics into the indexing system has shown a real influence on the performances of a SIR.

6. Conclusion

The web is considered as the first source of information all over the world, and the search for relevant information is viewed as one of the new needs of the information society. The interest in consulting this media is linked to the efficiency of the information research engines.

In this article, we have presented a new approach of the consideration of the RDFS structure for IR. We not only consider the structure so as to define the types of RDFS elements indexed by the system, but also the semantic of the field concepts. In fact, a consideration of the semantic permits to diminish the silence in the system while increasing the recall. It allows improving the precision of the SIR through the addition of a data base of the synonym terms.

During the questioning phase, the calculation of term relevance for a query is a combination of the weight of the required terms with their synonyms in the resource.

The main contribution of this work consists in modeling the capability of the system to identify the synonym terms and highlighting them, in accordance with the reviewed probability BM25 model. In doing so, the regulation of the score of the documents is done almost automatically.

The enrichment of the query enables to understand the research well and have results coming from the research of the synonym terms.

The second contribution of this work is the experience of the BM25 models in the context of the IR conceptual described in the RDFS language.

First of all, we have evaluated our model with a corpus of 12 documents. Our experiment consists in comparing the classical BM25 model and the BM25 adapted to the conceptual model.

Our approach gives the best results compared to the recall and the precision. Therefore it is more performing in a particular context and it will be interesting to take great advantages of the documents it offers.

In addition, the positive results of our SIR model pave the way for interesting prospects as the presentation of the result to a user is concerned. Our model brings a partial solution to the problems arisen into 2.1. This model could be improved by optimizing b and k parameters b and k_i of the BM25 formula of our conceptual approach, or to use the weight of the markup and the weight of the subjects, predicates in triplets $< s, P, o >$ of RDFS. Then, it will be interesting to automate the construction process of the database of synonyms entirely.

The research system based on a conceptual representation of documents and queries are promising since they go from the symbol level ("characters chain ") to the conceptual one. What will enable them to raise (or slash) the morphological constraint of the synonymy and the polysemy which are known since a long time in IR as generative of silence and noise.

References

- [1] G. Amati, E. Ambrosi, M. Bianchi, C. Gaibisso, and G. Gambosi. FUB, IASI-CNR and university of Tor Vergata at trec 2007 blog track. In Proc. TREC, volume Special Publication 500-274. 2007.
- [2] Boyan J., Freitag D., Joachims T.. A Machine Learning Architecture for Optimizing Web Search Engines. AAAI Workshop on Internet-Based Info. Systems, 1996.
- [3] Ciro Cattuto, Vittorio Loreto, and Luciano Pietronero. Semiotic dynamics and collaborative tagging. Proceedings of the National Academy of Sciences (PNAS), 104, 1461–1464, 2007.
- [4] C. Clarke, N. Craswell, and I. Soboro_. Preliminary report on the TREC 2009 Web track. 2009.
- [5] E. Fox. Extending the Boolean and Vector Space Models of Information Retrieval with P-Norm Queries and Multiple Concept Types. PhD thesis, Cornell University, University Microfilms, Ann Arbor, Michigan, 1983.
- [6] Fuhr, N.. Probabilistic models in information retrieval. The Computer Journal 35, 3, 243–255. 1992.
- [7] George A. Miller. WordNet: a lexical database for English. Commun. ACM, 38:39–41, November 1995.
- [8] Géry M., Largeron C., Thollard F.. Integrating structure in the probabilistic model for Information Retrieval, Web Intelligenc. p. 763-769, 2008.
- [9] Géry M., Largeron C., Thollard F., UJM at INEX 2008 : pre-impacting of tags weights , Proc. of INitiative for the Evaluation of XML Retrieval (INEX), Dagstuhl. 2009.
- [10] <http://www.w3.org/TR/rdf-schema/Langage> de description de vocabulaire RDF 1.0 : RDF Schema. Recommandation du W3C du 10 février 2004
- [11] <http://www.w3.org/TR/rdf-mt/RDF> semantic; Recommandation du W3C du 10 février 2004;
- [12] Kim Y.-H., Kim S., Eom J.-H., Zhang B.-T..SCAI Experiments on TREC-9, Text Retrieval Conference (TREC-9), p. 392-399, 2000.
- [13] J. Kim, X. Xue, and W. Croft. A probabilistic retrieval model for semistructured data. Advances in Information Retrieval, pages 228–239, 2009.
- [14] M. Koolen, R. Kaptein, and J. Kamps. Focused search in books and Wikipedia. Categories, links and relevance feedback. In S. eva, J. Kamps, and A. Trotman, editors, Focused Retrieval and Evaluation: 8th International Workshop of the Initiative for the Evaluation of XML Retrieval (INEX 2009), volume 6203 of LNCS, pages 273–291, 2010.
- [15] Lalmas M.. XML Information Retrieval, Encyclopedia of Library and Information Sciences, J. Bates and M.N. Maack (Eds), 2009b.
- [16] H. Li, T. Y. Liu, and C. Zhai, Learning to rank for information retrieval, SIGIR Forum, vol. 42, no. 2, pp. 76–79, LR4IR 2008.
- [17] Louis, A., Nenkova, A.. Performance confidence estimation for automatic summarization. In: EACL, The Association for Computer Linguistics 541–548. 2009.
- [18] Noreault, T., M. Koll, and M. McGill. Automatic Ranked Output from Boolean Searches in SIRE. J. American Society for Information Science, 28(6), 333-39. 1977.
- [19] Pascal Hitzler, Marcus Krötzsch, Sebastian Rudolph. Foundations of Semantic Web Technologies. Chapman & hall, CRC / 2009.
- [20] Pascal Hitzler . Markus Krötzsch, Sebastian Rudolph, York Sure. Semantic Web, Grundlagen, Erste Auflage springer.com; Springer-Verlag Berlin Heidelberg ISBN 978-3-540-33993-9. 2008.
- [21] S. Ribaric, I.Fratric. A Matching-Score Normalization Technique for Multimodal Biometric Systems. Proc. 3rd COST 275 Workshop: Biometrics on the Internet, Hatfield, UK, 27-28 pp. 55-58. October 2005.
- [22] Robertson S., and Walker, S. Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval. In Proc. of the International ACM-SIGIR Conference, pp. 232–241. 1994.
- [23] Robertson S., Walker, S., and Beaulieu, M.. Experimentation as a way of life : Okapi at trec. In Information Processing and Management, 36, 95–108.808,809–840. 2000.
- [24] Robertson S. and H. Zaragoza. The Probabilistic Relevance Framework: BM25 and Beyond. Foundations

- and Trends in Information Retrieval Vol. 3, No. 4 (2009) 333–389 ; DOI: 10.1561/1500000019. 2009
- [25] G. Salton. A comparison between manual and automatic indexing. American Documentation, 20(1):61–71. 1969.
- [26] Schenkel, R., Suchanek, F.M., Kasneci, G.: Yawn. A semantically annotated wikipedia xml corpus. In Kemper, A., Schöning, H., Rose, T., Jarke, M., Seidl, T., Quix, C., Brochhaus, C., eds.: BTW. Volume 103 of LNI., p. 277–291. GI 2007.
- [27] Shlomo Geva, Jaap Kamps, Andrew Trotman. INEX 2010 Workshop Pre-proceeding. Published by: IR Publications, Amsterdam. ISBN 978-90-814485-2-9.; December 6–10, Woodlands of Marburg, Ipswich, Queensland, Australia <http://www.inex.otago.ac.nz/>. 2009.
- [28] R. Song, M. J. Taylor, J. R. Wen, H. W. Hon, and Y. Yu. Viewing term proximity from a different perspective. Advances in Information Retrieval (ECIR 2008), Springer LNCS 4956, pp. 346–357, 2008.
- [29] Sun, K.-T., Huang, Y.-M., & Liu, M.-C. A WordNet-Based Near-Synonyms and Similar-Looking Word Learning System. Educational Technology & Society, 14 (1), 121–134. 2011.
- [30] Stevenson M., Guo Y., Gaizauskas R., Martinez D.. Knowledge sources for word sense disambiguation of biomedical text. BioNLP'08, p. 80-87. 2008.
- [31] Trieschnigg D., Pezik P., Lee V., de Jong F., Kraaij W., Rebholz-Schuhmann D., MeSH Up : effective MeSH text classification for improved document retrieval , Bioinformatics, vol. 25, n° 11, p. 1412-1418, June, 2009.
- [32] Trotman A.. Choosing document structure weights. Processing and Management, vol. 41, n° 2, p. 243-264, 2005.
- [33] Zargayouna, H. et Salotti, S. Mesure de similarité sémantique pour l'indexation de documents semistructurés. In IC'2004 : 15e journées francophones d'Ingénierie des connaissances. Cité 1 fois, p. 11.2004.

Traoré Issa, is a student preparing a Ph.D, at the university Cheick Anta Diop (UCAD) of Dakar in Senegal. He is also engineer in Systems, Networks and telecommunication. This research focuses on semantic Web, data mining and web service discovery. He made a communication on this work during a

workshop named CARI 2010. He has taken up web mining as his desired interest area of research.

Oumtanaga Souleymane, is a Professor in Data processing and telecommunication. He teaches in Inp-Hb and is the director of LARIT Laboratory, he is making a lot of research in telecommunication and web mining.

Babri Michel, is a doctor in data processing and carries out research in the field of the data bases. He teaches data processing and the telecommunication in Inp-HB.

Claude Lichou, is a Professor in data processing. He also teaches at the university Cheick Anta Diop (UCAD) of Dakar in Senegal. He supervises the students in Thesis of doctorate. He is my teacher adviser for my thesis, he is making in many fields as follows: mathematics, computing and telecommunication. He is in charge of the technological computing doctorate school of UCAD in Dakar.

A Novel Approach to Fast Image Filtering Algorithm of Infrared Images based on Intro Sort Algorithm

Kapil Kumar Gupta¹, M. Rizwan Beg², Jitendra Kumar Niranjan³

¹ Department of Computer Science & Engg., Integral University,
Lucknow, Uttar Pradesh, 226001, India

² Department of Computer Science & Engg., Integral University,
Lucknow, Uttar Pradesh, 226001, India

³ Department of Computer Science & Engg, IMS Engineering College
Ghaziabad, Uttar Pradesh 201009, India

Abstract

In this study we investigate the fast image filtering algorithm based on Intro sort algorithm and fast noise reduction of infrared images. Main feature of the proposed approach is that no prior knowledge of noise required. It is developed based on Stefan-Boltzmann law and the Fourier law. We also investigate the fast noise reduction approach that has advantage of less computation load. In addition, it can retain edges, details, text information even if the size of the window increases. Intro sort algorithm begins with Quick sort and switches to heap sort when the recursion depth exceeds a level based on the number of elements being sorted. This approach has the advantage of fast noise reduction by reducing the comparison time. It also significantly speed up the noise reduction process and can apply to real-time image processing. This approach will extend the Infrared images applications for medicine and video conferencing.

Keywords: *Image filtering, Intro Sort, infrared Images, Noise reduction, Digital Image Processing.*

1. Introduction

In Infrared images, impulse noise detection and removal is an important process as the images are corrupted by those noise because of transmission and acquisition. The main aim of the noise removal is to suppress the noise when preserving the edge information. Images and videos belong to the most important information carriers in today's world (e.g., traffic observations, surveillance systems, autonomous navigation, etc.). However, the images are likely to be corrupted by noise due to bad acquisition, transmission or recording. Such degradation negatively influences the performance of many image processing techniques and a preprocessing module to filter the images is often required.

The sensitive spectrum of an IR camera is about 3–5 μ m and 8–14 μ m. So, the IR images are robust under a wide-range of lighting conditions. However, the low signal-to-noise (S/N) ratio [1,2] is the inherent limitation of IR images that affect their quality and hinder their deployment. The low S/N ratio results in low signal and high noise that degrades the quality of IR images. This is significant for un-cooled IR camera, even though the un-cooled IR camera is much cheaper than the cooled one and more prevalently used to capture IR images in recent years. The high noise is caused by the IR sensors and read-out circuits of IR cameras, and the low IR signal detected by IR sensors is due to the bad atmospheric weather's degrading the IR signal radiating from objects. To enhance image quality and improve the adoption of IR-based applications, image preprocessing is necessary. Improvement in noise reduction is the crucial task of IR image preprocessing.

In this paper, a Fast Image Filtering approach to Infrared images is developed that is based on *Infrared imaging mechanism* to detect noise and median-based to remove noise with low computation load. It is performed without any prior knowledge about the IR image noise is necessary and any parameters must be preset. This property is quite different from some state-of-the-art noise reduction methods [5–8] which performance relies on one or more external heuristically preset parameter.

2. Previous Research

The standard median filter (MF) [3] has been prevalently used in for noise reduction of image preprocessing. However, there are two inherent limitations of the MF. The first is high computation load. The second is that it

removes the thin lines and small objects of interest and blurs the details even at low noise densities, while the size of the filter window increasing. The later makes it worse when the objects of interest are with few pixels in IR images. The MF will consider the few pixel objects as noise and remove them. The weighted median filter and the center-weight median filter (CWMF) [3,4] which are modified the MF to alleviate the inherent limitations of MF at the expense of reduced noise removal performance. In addition, there are many methods [5–20] combine the MF with impulse detection have been proposed to remedy the MF's limitations. Their performances inherently rely on the performance of the impulse detector. Mean-based filters [21–24] are the alternative approach to remedy the limitations of MF. These filter usually exhibit good filter performance at the cost of increased computational complexity. In recent years, a number of literatures [25–28] proposed the impulse noise reduction based on fuzzy technologies. However all these literatures mentioned previously considered on visual images.

3. Fast noise reduction approach

There is a main difference between visual image and infrared image imaging mechanism. The visual sensor receives the visual light reflected from object's surface to image visual image. It needs an external light source to offer a sufficient light power. This approach is called active imaging mechanism. An object reflects the light power based on the texture, color, roughness and other factors of its surface. These factors induce the reflected light power irregularly. Median filter filters the impulse noise from visual images based on an assumption that signal pixels have high correspondence with their neighbor pixels inside a small area. This assumption is reasonable but not theoretical. Because of this, the images processed by median filter will possibly lose some thin lines, textures or details. On the other hand, the IR sensor receives the infrared emitted from objects themselves to image IR images. It does not need external light sources to offer the IR to illuminate objects. This approach is called passive imaging mechanism. The temperature distribute on object's surface monotonically. Thus, object's surface also emits infrared power to IR sensor monotonically. The gray-level of signal pixels on the same object surface will vary monotonically while noise pixels cannot do it. The proposed algorithm developed based on the infrared image imaging mechanism theoretically. The details are described in following.

3.1 Stephen-Boltzmann law

The imaging mechanism is derived based on the Stefan-Boltzmann law [29] (heat radiation law) and the Fourier

law (heat conduction law). The heat radiation law is shown in Eq. (1):

$$PW = \epsilon \sigma T^4 \quad (1)$$

where PW is the radiant emittance (W/cm^2), ϵ is the emissivity, σ is the Stefan-Boltzmann constant ($\approx 5.6705 \times 10^{-12} \text{ W/cm}^2 \text{ K}^4$), T is the temperature (K) of the object surface. The intensity of IR radiation emitted by the objects in the range of 3–14 μm [30] is dependent on the emissivity of the object surface, the surface temperature, the air molecules, the humidity of the air, and the distance between the IR camera and the objects. The IR transmission spectrum for the atmosphere is about 3–5 μm and 8–14 μm [31], which means that the radiant emittance of the IR spectrum at 3–5 μm and 8–14 μm has only minimum attenuation in the atmosphere. In Eq. (1), σ is a constant and ϵ is also a constant for objects. The temperature T of the object surface is the only variable that dominates the PW that significantly affects the thermal image contrast and quality.

3.2 Fourier law (heat conduction law)

the gray-level of pixels has a positive relationship with the temperature T of the object surface. Moreover, heat conduction affects the temperature distribution on the skin surface. Based on the Fourier law (heat conduction law (2)), the temperature gradient will vary over the surface of an object, and the direction of the temperature gradient changes slowly from high to low.

$$Q_1 = -kA (dT/dl), \quad (2)$$

where Q_1 is the rate of heat flow through area A in the positive l direction, and the constant k is the thermal conductivity of the material. The heat conduction law means that the rate of heat flow is proportional to the area and the temperature gradient in a given direction. Thus, the temperature T of the object varies monotonically on a surface.

3.3 Noise detection

In order to explain how to perform the noise detection, we must define some parameters of image pixels. x and y represent the horizontal and vertical coordinates of a pixel, respectively. $p(x, y)$ is the pixel with coordinates x and y. $g(x, y)$ is the gray-level of the pixel $p(x, y)$. S_{xy} is the set that includes $g(x, y)$ and its neighbor pixels. For example, a 3×3 window, $S_{xy} = \{g(x-1, y-1), g(x, y-1), g(x+1, y-1), g(x-1, y), g(x, y), g(x+1, y), g(x-1, y+1), g(x, y+1), g(x+1, y+1)\}$, g_m is the median gray-level of the S_{xy} , $g(x, y)$ represents the gray-level of the central position pixel that will be processed, (s, t) denotes the coordinates of the pixels belonging to S_{xy} , and $g(s, t)$ represents the gray-level of the pixels belonging to S_{xy} .

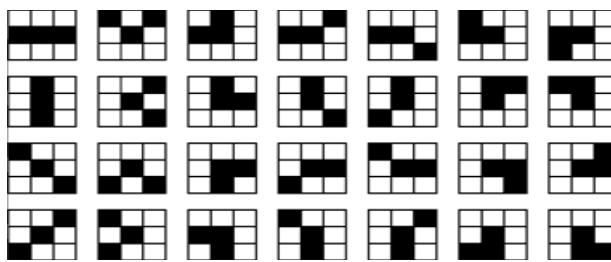


Fig. 1. Comparisons of central pixel to its two neighbors pixels in 3×3 Window

we propose the noise detection algorithm to find noisy pixels inside the filter window in IR images based on two steps. First, the noise detection method employs Eq. (3a) and (3b) to find the maximum and minimum gray-level inside the filter window. Next, we check the gray-level $g(x, y)$ with g_{\min} and g_{\max} to consider whether the pixel $p(x, y)$ is noise or signal. This process can be performed in the following steps.

Step 1. Determine the maximum and minimum gray-level inside the filter window.

$$g_{\max} = \operatorname{Arg} \max_{(s,t) \in S_{xy}} \{g(s,t)\} \quad (3a)$$

$$g_{\min} = \operatorname{Arg} \min_{(s,t) \in S_{xy}} \{g(s,t)\} \quad (3b)$$

Step 2. Based on the IR imaging mechanism, check $g(x, y)$ with gray-level g_{\min} and g_{\max} to determine whether the pixel $p(x, y)$ is noise or signal.

If $g(x, y) = g_{\min}$ or g_{\max} then $p(x, y)$ is a noisy pixel
 Otherwise $p(x, y)$ is considered as a signal pixel .

3.4 Noise removal based on Intro sort algorithm

In order to reduce noise, a pixel is considered as a noisy pixel which has to be removed from the IR image. According to the property of IR imaging mechanism, the pixel with median gray-level inside the window is adopted replacing the noisy pixels. In the proposed approach, the procedure to find out median gray-level is performed by the sort algorithm with a low computation complexity. In addition, it only processes the noisy pixels, but not the signal pixels.

Sorting is the main computation load of the noise removal. So, in order to speed up the noise removal, reducing the computation load is critical. This paper adopts a sort algorithm with low computation load. Sort algorithms are normally divided into two groups called internal sort and external sort. The former is suitable to small databases, and the latter is usually used on large databases. Hence, the number of pixels inside the filter window is small that may be 3×3 , 5×5 , 7×7 , and so on. Based on the

properties of the sort algorithm and the number of pixels sorted, we adopt the internal sort algorithm to sort the pixel gray-levels inside a filter window. The internal sort algorithm includes Bubble sort, Insertion sort, Selection sort, Shell sort, Quick sort, Heap sort, Radix sort, and so on. [32,33]. From an analysis of the properties of the versatile sort algorithms, a Intro sort with suitable parameters should run faster even than Quick sort or Heap sort. The complexity of Intro sort is described as Eq. (4).

$$\text{Complexity of Intro sort} = O(n \log n) \quad (4)$$

Where n is the number of data.

It begins with Quick sort and switches to Heap sort when the recursion depth exceeds a level based on the number of elements being sorted. It is the best of both worlds, with a worst-case $O(n \log n)$ runtime and practical performance comparable to Quick sort on typical data sets. Since both algorithms it uses are comparison sorts, it is a comparison sort too. Then the Intro sort algorithm is performed on bits to select the pixel with the median gray-level g_m inside a filter window. The selected pixel with gray-level g_m is utilized to replace the noisy pixel in the central position inside the filter window and the noise removal is accomplished. The Fast noise reduction approach does not act like the Median Filter in sorting each pixel in a whole Infrared image; it only processes that on noisy pixels. In general, there are far fewer noisy pixels than signal pixels. The FNR approach has to spend extra computation load to determine the maximum and minimum gray-level in the noise detection procedure.

(3a)

4. Experimental Results

4.1 Platform for evaluation

The platform utilized to evaluate the proposed approach includes a dual core CPU, the Intel Core 2 E6600 with clock rate 2.4 GHz and memory 1 Gbytes DDR2 667. The display card has GPU GeForce 7600 GT of NVIDIA Inc. and 256 MB memory. The program to simulate the approach was developed by Matlab. One noisy life-time IR images are used as test samples to assess the effect on the proposed approach. Their details are described in the following section.

4.2 Test samples of the IR image

In order to validate the proposed approach, one gray-scale noisy life-time IR images are collected as test sample, as shown in Fig. 2. There are 364×244 pixels in the images, respectively, and each pixel is represented by 8-bits in gray-scale. Fig. 2 shows people walking on the street on a rainy day. The people (objects) are small compared with the street scenery rendered as background.



Fig. 2 An Infrared image taken of some people walking on the street on a rainy day.

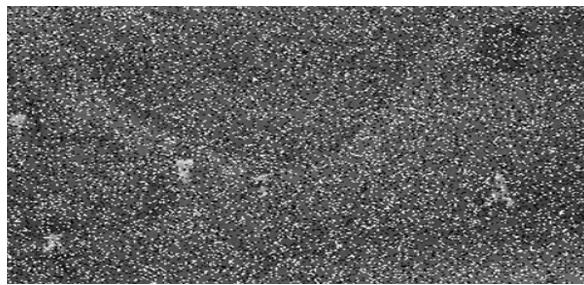


Fig. 3 Shows Fig. 2 with impulse noise processed by FNR iteratively with 20% impulse noise.

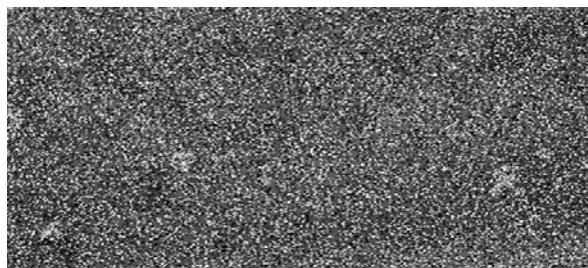


Fig. 4 Shows Fig. 2 with impulse noise processed by FNR iteratively with 30% impulse noise.



Fig. 5 The result of Fig. 3. iteratively processed by Fast noise reduction approach two times.



Fig. 6 The result of Fig. 4. iteratively processed by Fast noise reduction approach two times.



Fig. 7 The result of Fig. 3. iteratively processed by MF two times.,

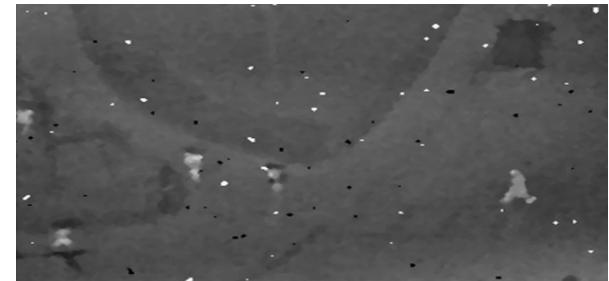


Fig. 8 The result of Fig.4. iteratively processed by MF two times.,

In order for the small objects of interest in these IR images to be observed easily, quickly, and accurately, they have to be preprocessed using a noise reduction.

The performances of FNR approach are address in this paragraph. Fig. 3 & 4 show the Fig. 2 with impulse noise 20%, 30%. Fig. 5 & 6 exhibit the result of performing FNR to iteratively filter two times. the impulse noise is filtered out and preserve the edges, textures and detail information simultaneously Fig.7 & 8 iteratively processed by MF two times they show the more noises are not filtered out and more edges, textures and detail information are lost. In addition, the proposed FNR approach can preserve the edge and texture information regardless of increases in filter size while removing noise. When the filter size increases, the image processed by the MF is blurred, and the edges and texture information are

lost. The experimental results of performing the MF and FNR Table 1 illustrates the different performances between utilizing the MF and FNR approach on Fig. 3 and 4 one time. The MF processes each pixel in a whole image so it has to perform a sorting algorithm 88,816 times for Fig. 3 and spend 24 ms. On the other hand, the FNR approach performs a sorting algorithm for noisy pixels only, performing the sorting algorithm 24,587 times for Fig. 3 and spending 15 ms, which includes finding the g_{\max} and g_{\min} 88,816 times. Hence, the computation load is reduced 37.5% by the proposed FNR approach. The FNR reduces the computation load 25%.

A typical noise measure used is peak signal-to-noise ratio (PSNR) [3], defined as.

$$\text{PSNR} = 10 \log_{10}(\text{MAX}^2 / \text{MSE}) \quad (5)$$

where MSE is the mean square error between the original and processed image and Max is the maximum gray scale of pixels, e.g., 255 for 8 bits. We use the PSNR to assess the noise reduction performance of the FNR approach and MF.

Table 1 : The differences in performance between utilizing the MF and the FNR approach on Fig. 3&4.

Image Size (Pixels) 364×244 (88, 816)			
Noise Image		Fig.3(20% noise)	Fig.4(30% noise)
Noisy image	PSNR(db)	34.11	32.26
Processed by MF	Sort times	88,816	88,816
	Process time(ms)	24	24
	PSNR(db)	40.23	38.79
Processed by FNR approach	Time of finding g_{\max} & g_{\min}	88,816	88,816
	Sort times	24,587	30,832
	Process Time(ms)	15	18
	PSNR(db)	43.45	41.39
Process time improved by FNR (%)		37.50%	25%
PSNR improved by FNR (dB)		3.22	2.60

5. Conclusions

In this paper, we present an effective FNR approach to reduce the noise of IR images. Based on the results shown in Figs. 5,6,7 and 8 and in Tables 1 the FNR approach possesses three main advantages. The first is that FNR approach utilizes noise detection based on IR imaging mechanism to identify noisy pixels in IR images, and median-based noise removal performed by Intro sorting with bits decomposition effectively decreases the

computation load for performing noise reduction. It can be applied to real-time video processing by software. The second advantage of FNR approach is that remedies the shortcomings of MF while increasing the filter window size. Finally, no prior knowledge about the IR images necessary and no parameter must be manually preset to perform the proposed approach.

Experimental results demonstrate that the proposed approach can improve the performance of noise reduction and enhance quality IR images. In the applications of IR images, it is a considerable challenge to provide a removal on noise but not on the edges, text information and small objects of interest. In order to conquer the challenge, we propose an FNR approach consisting of the noise detection and noise removal. Experimental results demonstrate that the proposed approach can meet this challenge.

Acknowledgments

This work is supported by National Natural Science Foundation , resources from Prof. Dr. M. Rizwan Beg, Head of department of computer science and engineering, Integral University, Lucknow. They would also thank the anonymous reviewers for their significant and constructive critiques and suggestions, which substantially improved the quality of this paper.

References

- [1] A. Dawoud, M.S. Alam, A. Bal, C. Loo, Target tracking in infrared imagery using weighted composite reference function-based decision fusion, IEEE Trans. Image Process. 15 (2) (2006) 404–410.
- [2] A. Bal, M.S. Alam, Automatic target tracking in FLIR image sequences using intensity variation function and template modeling, IEEE Trans. Instrum. Meas. 54 (5) (2005) 846–1852.
- [3] A.I. Bovik, Handbook of Image & Video Processing., second ed., Elsevier Academic Press, 2005.
- [4] T.C. Lin, A new adaptive center weighted median filter for suppressing impulsive noise in images, Inf. Sciences 177 (4) (2007) 1073–1087.
- [5] V. Crnojevic, V. Senk, Z. Trpovski, Advanced impulse detection based on pixelwise MAD, IEEE Signal Process. Lett. 11 (7) (2004) 589–592.
- [6] V.V. Khryashchev, I.V. Apalkov, A.L. Priorov, P.S. Zovanyarev, Image denoising using adaptive switching median filter, In: Proc. IEEE Int. Conf. Image Process. (ICIP 2005) 1 (2005) 117–120.
- [7] T. Chen, H.R. Wu, Adaptive impulse detection using center-weighted median filters, IEEE Signal Process. Lett. 8 (1) (2001) 1–3.
- [8] T. Chen, H.R. Wu, Space variant median filters for the restoration of impulse noise corrupted images, IEEE Trans. Circ. Syst. II 48 (8) (2001) 784–789.
- [9] R.H. Chan, C. Hu, M. Nikolova, An iterative procedure for removing randomvalued impulse noise, IEEE Signal Process. Lett. 11 (12) (2004) 921–924.

- [10] I. Aizenberg, C. Butakoff, D. Paliy, Impulsive noise removal using threshold Boolean filtering based on the impulse detecting functions, *IEEE Signal Process. Lett.* 12 (1) (2005) 63–66.
- [11] S. Zhang, M.A. Karim, A new impulse detector for switching median filters, *IEEE Signal Process. Lett.* 9 (11) (2002) 360–363.
- [12] G. Pok, Y. Liu, A.S. Nair, Selective removal of impulse noise based on homogeneity level information, *IEEE Trans. Image Process.* 12 (1) (2003) 85–92.
- [13] E. Bes_dok, M.E. Yüksel, Impulsive noise rejection from images with JarqueBerra test based median filter, *Int. J. Electron. Commun.* 59 (2) (2005) 105–109.
- [14] R. Garnett, T. Huegerich, C. Chui, W. He, A universal noise removal algorithm with an impulse detector, *IEEE Trans. Image Process.* 14 (11) (2005) 1747–1754.
- [15] J.Y. Chang, J.L. Chen, Classifier-augmented median filters for image restoration, *IEEE Trans. Instrum. Meas.* 53 (2) (2004) 351–356.
- [16] S.Q. Yuan, Y.H. Tan, Impulse noise removal by a global-local noise detector and adaptive median filter, *Signal Process.* 86 (8) (2006) 2123–2128.
- [17] B. Smolka, A. Chydzinski, Fast detection and impulsive noise removal in color images, *Real-Time Image* 11 (4) (2005) 389–402.
- [18] H.-L. Eng, K.-K. Ma, Noise adaptive soft-switching median filter, *IEEE Trans. Image Process.* 10 (2) (2001) 242–251.
- [19] M.E. Yüksel, E. Bes_dok, A simple neuro-fuzzy impulse detector for efficient blur reduction of impulse noise removal operators for digital images, *IEEE Trans. Fuzzy Syst.* 12 (6) (2004) 854–865.
- [20] S. Schulte, M. Nachtegael, V. DeWitte, D. Van derWeken, E.E. Kerre, A fuzzy impulse noise detection and reduction method, *IEEE Trans. Image Process.* 15 (5) (2006) 1153–1162.
- [21] S.K. Mitra, G.L. Sicuranza, J.D. Gibson, *Nonlinear Image Processing (Communications, Networking and Multimedia)*, Academic, Orlando, FL, 2001.
- [22] E. Abreu, Signal-dependent rank-ordered mean (SD-ROM) filter, in: S.K. Mitra, G.L. Sicuranza, J.D. Gibson (Eds.), *Nonlinear Image Processing (Communications, Networking and Multimedia)*, Academic, Orlando, FL, 2001. pp. 111–133.
- [23] D. S. Zhang, D. J. Kouri, Varying weight trimmed mean filter for the restoration of impulse noise corrupted images, In: Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP 2005), vol. 4, pp. 137–140.
- [24] W. Luo, An efficient detail-preserving approach for removing impulse noise in images, *IEEE Signal Process. Lett.* 13 (7) (2006) 413–416.
- [25] S. Schulte, V. De Witte, E.E. Kerre, A fuzzy noise reduction method for color images, *IEEE Trans. Image Process.* 16 (5) (2007) 1425–1436.
- [26] A. Toprak, I. Güler, Impulse noise reduction in medical images with the use of switch mode fuzzy adaptive median filter, *Digital Signal Process.* 17 (4) (2007) 711–723.
- [27] S. Morillas, V. Gregori, G. Peris-Fajarnés, A. Sapena, Local self-adaptive fuzzy filter for impulsive noise removal in color images, *Signal Process.* 88 (2) (2008) 390–398.
- [28] M. Tülin Yıldırım, Alper Bas_türk, M. Emin Yüksel, Impulse noise removal from digital images by a detail-preserving filter based on type-2, *IEEE Trans. Fuzzy Logic Fuzzy Syst.* 16 (4) (2008) 920–928.
- [29] J.M. Lloyd, *Thermal Image System*, Plenum Press, 1975.
- [30] E.F. Godik, Y.V. Guljaev, Functional imaging of the human body, *IEEE Eng. Med. Biol.* 10 (4) (1991) 21–29.
- [31] D.C. Harris, *Infrared Window And Dome Materials*, The Int. Society Opt. Eng., SPIE, Bellingham, Washington, 1992. July.
- [32] A. Drozdek, *Data Structures and Algorithms in C++*, second ed., Thomson learning, USA, 2001.
- [33] M.T. Goodrich, R. Tamassia, D.M. Mount, *Data Structures and Algorithms in C++*, John Wiley & Sons, Inc., USA, 2004.
- [34] W.L. Lee, K.C. Fan, Design of optimal stack filter and under MAE criterion, *IEEE Trans. Signal Process.* 47 (December) (1999) 3345–3355.
- [35] Rafael C. Gonzalez, Richard E. Woods, *Digital Image Processing*, second ed., Prentice Hall, 2002.
- [36] Tukey J. M. *Exploratory Data Analysis*[M]. Reading , Massachusetts: Addison, Wesley, 1971. 55-105.
- [37] Castleman Kenneth R. *Digital Image Processing*[M]. Beijing: Electronics Industry Press, 2002. 204–206.
- [38] Xin Wang. Adaptive Multistage Median Filter[J]. *IEEE Transactions on signal processing*,1992, 40(4):1015-1017.
- [39] Xiahua Yang and Peng Seng Toh. Adaptive Fuzzy Multilevel Median Filter[J]. *IEEE Transactions on image processing*,1995,4(5):680-682.
- [40] H. Hwang and R.A.Haddad. Adaptive Median Filters: New Algorithms and Results[J]. *IEEE Transactions on image processing*,1995,4(4):244-502.
- [41] Yüksel M. Emin, Besdok Erkan. A simple neuro-fuzzy impulse detector for efficient blur reduction of impulse noise removal operators for digital images[J]. *IEEE Transactions on Fuzzy Systems*, 2004,12 (12) : 854-865.
- [42] Sorin Zoican. Improved median filter for impulse noise removal[J]. TELSIKS Serbia and Motenegro, Ni,2003,10(123):681-684.
- [43] Nelson H C Yung. Novel filter algorithm for removing impulse noise in digital image[J]. SPIE, 1995,250(1):210-220.
- [44] Chih-Lung Lin, Chih-Wei Kuo, Chih-Chin Lai, Ming-Dar Tsai, Yuan-Chang Chang, Hsu-Yung Cheng, A novel approach to fast noise reduction of infrared image, *Infrared Physics & Technology* 54(1), (2011):1–9.



Kapil Kumar Gupta obtained his B. Tech (IT) degree in 2009 from JSS academy of technical education, Noida, Uttar Pradesh, India . He is currently a student of M. Tech (CSE) from Integral University, Lucknow and Asstt. Professor in Department of Information Technology, Goel Institute of technology and management, Lucknow, Uttar Pradesh, India. His main research interest is in the field of Image Processing, Design and analysis of algorithms, etc.. He is a Member of CSTA.



Prof. Dr. M. Rizwan Beg is M. Tech & Ph.D in Computer Sc. & Engg. Presently he is working as Professor & Head Deptt. Of Computer Sc. & Engg. and Information Technology of Integral University Lucknow, Uttar Pradesh, India. He is having more than 16 years of experience which includes around 14 years of teaching experience. His area of expertise is Software Engg., Requirement Engineering, Software Quality, and Software Project Management. He has published more than 40 Research papers in International Journals & Conferences. Presently 8 research scholars are pursuing their Ph.D in his supervision. Dr. Beg is Associate Editor in Chief of Advancement in Computing Technology & is also member of editorial board for various other International & National Journals. He is member of large member of International professional societies & also member of Advisory Board for various institutions in India. He chaired a no. of workshops, seminars & National Conference.



Jitendra Kumar Niranjan has obtained his M.Tech (Information System) degree in 2011 from Delhi Technological University, Delhi, India. He is currently working as Asstt. Professor in IMS Engineering College Ghaziabad. His main research interest is in the field of Image Processing, and Cloud computing.

A New Factorization Method to Factorize RSA Public Key Encryption

B R Ambedkar¹ and S S Bedi²

¹ Department of CS and IT, MJP Rohilkhand University

Bareilly, Uttar Pradesh 243006, India

² Department of CS and IT, MJP Rohilkhand University

Bareilly, Uttar Pradesh 243006, India

Abstract

The security of public key encryption such as RSA scheme relied on the integer factoring problem. The security of RSA algorithm is based on positive integer N , because each transmitting node generates pair of keys such as public and private. Encryption and decryption of any message depends on N . Where, N is the product of two prime numbers and pair of key generation is dependent on these prime numbers. The factorization of N is very intricate. In this paper a New Factorization method is proposed to obtain the factor of positive integer N . The proposed work focuses on factorization of all trivial and nontrivial integer numbers and requires fewer steps for factorization process of RSA modulus N . The New Factorization method is based on Pollard rho factorization method. Experimental results shown that factorization speed is fast as compare existing methods.

Keywords: Attacks, Pollard rho method, Public Key Cryptography, RSA Algorithm.

1. Introduction

Public key cryptography is one of the mathematical applications that are valuable in sending information via insecure channel. RSA algorithm is a public key encryption algorithm. RSA has become most popular cryptosystem in the world because of its simplicity. According to number theory, it is easy to finds two big prime number, but the factorization of the product of two big prime numbers is very difficult. The difficulty of computing the roots N , where N is the product of two large unknown primes, it is widely believed to be secure for large enough N . Since RSA can also be broken by factoring N , the security of RSA is often based on the integer factorization problem [1]. The integer factorization problem is a well-known topic of research within both academia and industry. It consists of finding the prime factors for any given large modulus. Currently, the best factoring algorithm is the general number field sieve or

GNFS for short. On December 12, 2009 a small group of scientists used the previously mentioned approach to factor a RSA-768 bit modulus, that is, a composite number with 232 decimal digits. Their achievement required more than two years of collaborative work and used many hundreds of computing machines. Hence, factoring large primes is a laborious and complex task [2]. A method for factoring algorithm (specially designed) for semi primes based on new mathematical ideas. Since this method is relatively simple and scalable, it can be suitable for parallel processing [2]. A new algorithm which attacks the RSA scheme. But the main condition of this algorithm is that to break RSA modulus firstly we should have public key and modulus N . On the basis of this public key (e, N) proposed algorithm disclose the private key [3]. An attacker has access to the public key e and N and the attacker wants the private key d . To get d , N needs to be factored (which will yield p and q , which can then be used to calculate d). Factoring n is the best known attack against RSA to date. (Attacking RSA by trying to deduce $(p-1)(q-1)$ is no easier than factoring N , and executing an exhaustive search for values of d is harder than factoring N .) Some of the algorithms used for factoring are as follows [12]: Trial division oldest and least efficient Exponential running time. Try all the prime numbers less than \sqrt{N} . Quadratic Sieve (QS): The fastest algorithm for numbers smaller than 110 digits. Multiple Polynomial Quadratic Sieve (MPQS): Faster version of QS. Double Large Prime Variation of the MPQS Faster still. Number Field Sieve (NFS) Currently the fastest algorithm known for numbers larger than 110 digits. Was used to factor the ninth Fermat number. These algorithms [12] represent the state of the art in warfare against large composite numbers against RSA. We can identify many approaches to attacking RSA mathematically, factor N into two prime factors. A lot of algorithm has been proposed regarding factorization, the Pollard rho algorithm [7], and the Pollard $(p-1)$ algorithm

[8], Brent's method [9], are probabilistic, and may not finish, for small values of N , Trial division algorithm [12] and Fermat method can finish [11]. In this paper we are proposing New Factorization (NF) method which is based on Pollard rho Factorization (PRF) method [7]. By using NF method we can factorize quickly all integer number. It is very suitable for factorization against RSA algorithms, because that have a product of two prime number, so there is no more than two factors, NF method factorize all numbers with minimum elapsed time shown in Fig. 1. MATLAB environment is used for various analyses.

2. Attacks

The security of the RSA cryptosystem is not perfect. An attacker can resort to different approaches to attack RSA algorithm. Among them, Brute force, Mathematical attacks, Timing attacks and Chosen Cipher text attacks are described in following:

2.1 Brute Force Attacks on RSA

Brute force attacks, involves trying all possible public and private keys. The attacker tries all possible combinations to guess the private key. RSA with short secret key is proven insecure against brute force attack [13]. This attack can be easily circumvented by choosing large key. However, the larger key will make the encryption and decryption process little slow as it will require greater computations in key generation as well as in encryption/decryption algorithm. The key length used in the encryption determines the practical feasibility of performing a brute-force attack, with longer keys exponentially more difficult to crack than shorter ones. Brute-force attacks can be made less effective by obfuscating the data to be encoded, something that makes it more difficult for an attacker to recognize when he/she has cracked the code. One of the measures of the strength of an encryption system is how long it would theoretically take an attacker to mount a successful brute-force attack against it.

2.2 Mathematical Attacks on RSA

Mathematical attacks focus on attacking the underlying structure of RSA function. The first intuitive attack is the attempt to factor the modulus N . There are several approaches, all equivalent to in effect to factoring the product of two primes N . Our objective is to provide RSA attacks that decrypts message by factoring N .

2.3 Timing Attack

In RSA, the attacker can exploit the timing variation of the modular exponentiation implementations and able to

determine d by calculating the time it takes to compute $C^d \pmod{N}$ for a given cipher text C [13].

2.4 Factorization Attack

The security of RAS is based on the idea that the modulus is so large that is infeasible to factor it in reasonable time. Bob selects P and Q and calculate $N=P \times Q$. Although N is public, P and Q are secret. If Eve can factor N and obtain P and Q , Eve then can calculate $d = e^{-1} \pmod{\phi(N)}$ because e is public. The private exponent d is the trapdoor that Eve uses to decrypt any encrypted message.

The factorization attack is a extremely giant dispute for security of RSA algorithm. Some existing factorization algorithms can be generating public and private key of RSA algorithm, by factorization of modulus N . But they are taking huge time for factorization of N , in case of P and Q very large. We are focusing on factorization speed and proposing new factorization method to enhance the speed of factorization. Related works for factorization of modulus N are following.

3. Related Work

The RSA Factoring Challenge was started in March 1991 by RSA Data Security to keep abreast of the state of the art in factoring. Since its inception, well over a thousand numbers have been factored, with the factories returning valuable information on the methods they used to complete the factorizations. The Factoring Challenge provides one of the largest test-beds for factoring implementations and provides one of the largest collections of factoring results from many different experts worldwide. In short, this vast pool of information gives us an excellent opportunity to compare the effectiveness of different factoring techniques as they are implemented and used in practice. Since the security of the RSA public-key cryptosystem relies on the inability to factor large numbers of a special type, the cryptographic significance of these results is self-evident. Some factorization methods are explored in following paragraphs:

J. Carlos et al [2] proposed a method for factoring algorithm is: $\gcd[N, (k \cdot \check{N}) + \Delta]$ and $\gcd[N, (k \cdot \check{N}) - \Delta]$ result in nontrivial factors of N for different values of Δ where \check{N} is the reverse of N and k is a positive integer ranging from one to infinity.

Sattar J Aboud [3] was introducing a method that breaking the RSA scheme based on the knowing public key (e, N). This method will work efficiently if the decryption key d is small. It possible therefore, to factor the modulus N .

L. Scripcariu, M.D. Frunze [4] was commencing some weak points of RSA algorithm and proposed a method for secure public key. The main concept of this paper is for encryption (e, N) change every time public key for encryption.

In 1978, RSA developed a public key cryptosystem that is based on the difficulty of integer factoring. The RSA public key encryption scheme is the first example of a provable secure public key encryption scheme against chosen message chosen attacks [5]. The RSA scheme is as follows [6]:

Key generation algorithm, to generate the keys entity A must do the following:

1. Randomly and secretly select two large prime numbers p and q .
2. Compute the $N = p * q$.
3. Compute $(\phi)(N) = (p-1)(q-1)$.
4. Select random integer e , $1 < e < N$, where $\gcd(e, \phi) = 1$.
5. Compute the secret exponent d , $1 < d < \phi$, such that $e * d \equiv 1 \pmod{\phi}$.
6. The public key is (N, e) and the private key is (N, d) . Keep all the values d, p, q and ϕ secret.

Public key encryption algorithm, entity A encrypt a message m for entity B which entity decrypt.

Encryption: Entity B should do the following:

Obtain entity A's public key (N, e) , represent the message M as an integer in the interval $[0 \dots N-1]$. Compute $C = M^e \pmod{N}$. Send the encrypted message c to entity A.

Decryption: To recover the message m from the cipher text c . Entity A do the following: Obtain the cipher text c from entity B. Recover the message $M = C^d \pmod{N}$.

Fermat Factorization [11] was discovered by mathematician Pierre de Fermat in the 1600s. Fermat factorization rewrites a composite number N as the difference of squares as follows: $N = x^2 - y^2$, this difference of squares leads immediately to the factorization of N : $N = (x+y)(x-y)$. Assume that s and t are nontrivial odd factors of N such that $st = N$ and $s \leq t$. We can find x and y such that $s = (x - y)$ and $t = (x + y)$

Example 1:

Let $N=95$ {Product of any two prime numbers.}

Decimal digits: 2 Bits: 7

Let factors := P,Q

Compute X := 10

Compute Y := 2.236 (is not integer number)

(Go to step 4)

```

X := 11
Y := 5.09 (is not integer number)
(Go to step 4)
X := 12
Y := 7 (is an integer number)
P := 5
Q := 19
End

```

Example 2:

Let $N=2320869986411928544793$

Decimal digits: 22 Bits: 73

Let factors := P,Q

Compute X := 48175408524

Compute Y := 219488.97 (is not integer number)

(Go to step 4)

X := X + 1, ..., X + 17073029192103

X := 17121204600627

Y := 17121136822844 (is an integer number)

P := 67777783

Q := 34242341423471

End

4. Pollard Rho Factorization

Pollard rho Factorization [7] method is a probabilistic method for factoring a composite number N by iterating a polynomial modulo N . The method was published by J.M. Pollard in 1975. Suppose we construct the sequence: $X_0 \equiv 2 \pmod{N}$, $X_{n+1} \equiv X_n^2 + 1 \pmod{N}$. This sequence will eventually become periodic. It can be shown that the length of the cycle is less than or equal to N by a proof by contradiction: assume that the length L of the cycle is greater than N , however we have only N distinct x_n values in our cycle of length $L > N$, so there must exist two x_n values are congruent, and these can be identified as the starting points of a cycle with length less than or equal to N . Probabilistic arguments show that the expected time for this sequence (\pmod{N}) to fall into a cycle and expected length of the cycle are both proportional to N , for almost all N [8]. Other initial values and iterative functions often have similar behavior under iteration, but the function $f(N) = (X_n^2 + 1)$ has been found to work well in practice for factorization. Assume that s and t are nontrivial factors of N such that $st = N$ and $s \leq t$. Now suppose that we have found nonnegative integers i, j with $i < j$ such that $X_i \equiv X_j \pmod{s}$ but $X_i \neq X_j \pmod{N}$. Since $s | (X_i - X_j)$, and $s | N$, we have that $s | \gcd(X_i - X_j, N)$. By assumption $s \geq 2$, thus $\gcd(X_i - X_j, N) \geq 2$. By definition we know that $\gcd(X_i - X_j, N) | N$. However, we have that $N / (X_i - X_j)$, and thus that $\gcd(X_i - X_j, N) | N$. So we have that $N | \gcd(X_i - X_j, N)$.

N , $\gcd(X_i - X_j, N) > 1$, and $\gcd(X_i - X_j, N)/N$. Therefore $\gcd(X_i - X_j, N)$ is a nontrivial factor of N .

Example 3:

Consider the Pollard rho algorithm for $N = 21 = 7 \times 3$.

The sequence of X_n values generated by the algorithm is:

$X_0 = 2, X_1 = 5, X_2 = 5, \dots, \text{for } n \geq 1,$

If $n \geq 1, X_{2n} - X_n = 0$.

The algorithm at each step for $n = 1, 2, \dots$

Compute $\gcd(X_{2n} - X_n, N) = \gcd(0, 21) = 21$

We are showing this example, the Pollard rho algorithm is a probabilistic, and may not finish, even for small values of N .

5. New Factorization Method

The new factorization method (NF) based on Pollard rho method and square root of N . The NF method is focusing on Mathematical and Factorization Attacks on RSA. Shown in e.g. 3, the Pollard rho factorization method may not finish small value of N . We are using RSA algorithm for public key cryptography, it is asymmetric encryption technique, by using this technique we cannot encrypt and decrypt message by same key. The generation of public and private key is dependent on the factorization of N . By using Pollard rho factorization method, we can't factorize all integer numbers N see e.g. 3. By using NF method we factorize all integer numbers N . The NF method is an independent of periodic sequence and probabilistic method. For factorization of all integer number, we are proposing to Pollard rho factorization method to modifying and removing the concept of periodic sequence. NF method as follows:

1. Let any positive integer is N .
2. Compute the square root of N .
3. Take ceiling function of step two.
4. Decrement by one in square of step three.
5. Compute the greatest Common divisor of step three and step number one.
6. If step five is greater than one.
7. Step five is a factor of step one.
8. Compute other factor of N divided by step seven.
9. Otherwise increment the value of steps three by one and continues step three to eight, till step four is greater than one.

By using these steps we factorize any positive integer as follows:

Example 4:

Let $N=21$

$S=5$

$P=\gcd(24,21)=3>1$

$Q=21/3=7$

P and Q is factor of N .

Example5:

Let $N=999962000357$

Decimal digits=12 Bits= 40

$S=999981$

$P=\gcd(999962000360, 999962000357)=1$

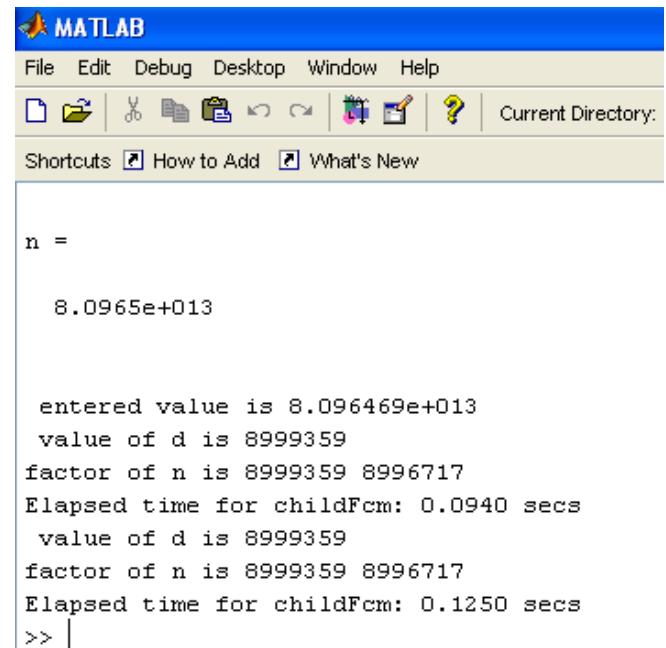
$P=\gcd(999964000323, 999962000357)=999983>1$

$Q=999962000357/999983=999979$

P and Q is a factor of N .

6. Simulation and Results

We are factorized all trivial and nontrivial number shown in e.g. 4. The NF method is an appropriate and essential for factorization of product of any two prime numbers because that have only two factors as shown in simulated result Fig. 1, factors of 14 decimal digit by using NF method elapsed time is 94ms shown in Fig. 1. That domino effect simulated by MATLAB version 7.0 and Intel(R), Core(TM)2 Quad CPU 2.66GHz, 3.24 GB of RAM. As per the results shows factorization becomes more rapidly by NF method.



```

MATLAB
File Edit Debug Desktop Window Help
Current Directory:
Shortcuts How to Add What's New

n =
8.0965e+013

entered value is 8.096469e+013
value of d is 8999359
factor of n is 8999359 8996717
Elapsed time for childFcm: 0.0940 secs
value of d is 8999359
factor of n is 8999359 8996717
Elapsed time for childFcm: 0.1250 secs
>>

```

Fig. 1 Factors of 80964686104403 by using NF method (Elapsed time in sec.= 0.094) simulation shows.

The elapsed time for prime factorization shown in Fig. 2 with respect digits and Fig. 3 with respect bits process elapsed times is awfully tiny as compare to existing method.

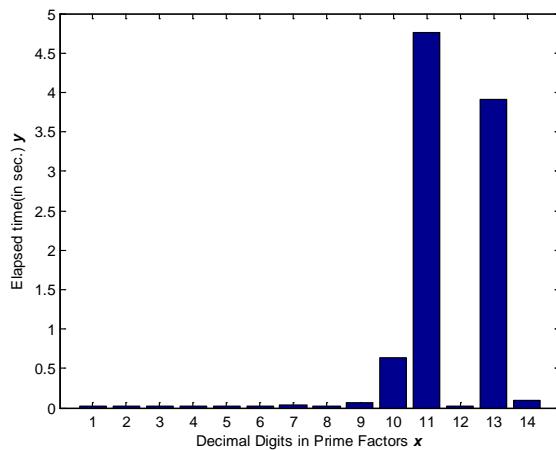


Fig. 2 NF method(Elapsed time vs Decimal digits in prime factors)

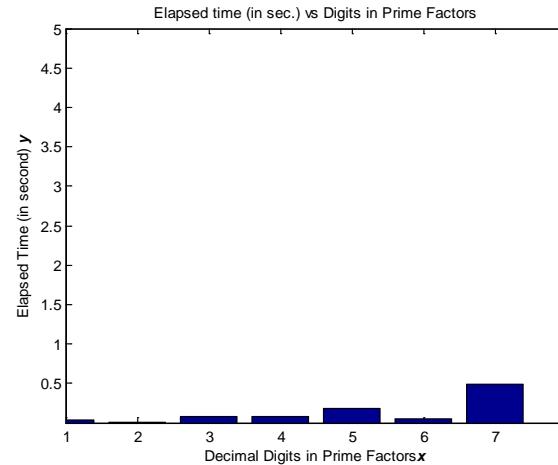


Fig. 4 Pollard rho method (Elapsed time vs Decimal digits in prime factors)

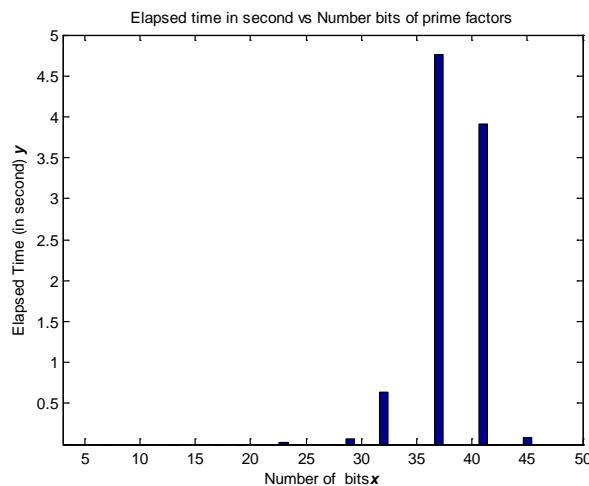


Fig. 3 NF method(Elapsed time vs Total number of Bits of prime factors)

Shown in Fig. 4, a plot of time elapsed and number of digits for given number by using traditional existing Pollard rho, the factorization of 7 digit decimal number elapsed time is 0.5s.

Table 1 shows a comparison of some factorization methods vs NF method. The trial division, Fermat factorization method and NF method always terminate, and upper bounds can be derived for the running times of these algorithms in terms of N , the number to be factored. The Pollard rho algorithm, Brent's method, and the Pollard $p-1$ algorithm are probabilistic, and may not finish, even for small values of N .

Table 1: Comparison of few factorization methods vs MPRF method

Factorization Method	For all numbers	Technique used
Trial	Can factorize	Division based
Fermat	Can factorize	$N=x^2-y^2$
Pollard Rho	Can't factorize	Periodic sequences
Pollard ($p-1$)	Can't factorize	$(a^{k!}-1) \bmod n$
Brent's	Can't factorize	$X_{n+1} = X_n^2 + 2 \bmod n$
NF	Factorize	Square root based

The shown in Fig. 5 comparison between existing algorithms and NF method. The factorization of 7 digit decimal number by NF method the elapsed time is 0.032s. NF method produced all factors of any integer number. The algorithm was executed using MATLAB tool.

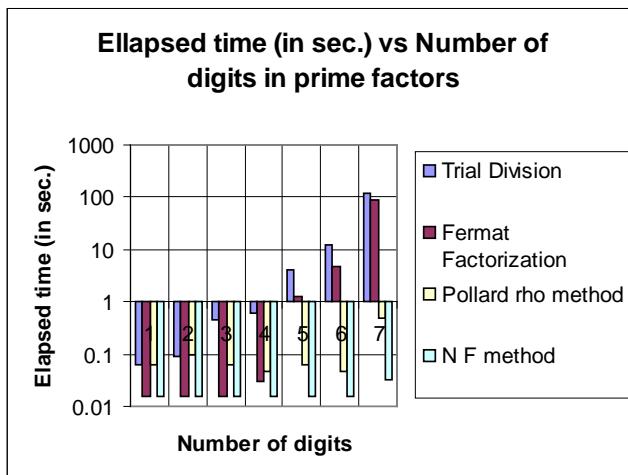


Fig. 5 Number of digits vs. Elapsed time in Prime Factors

7. Conclusions

In this paper New Factorization method is proposed for RSA modulus N factorization. During simulation, Pollard rho factorization method is found not suitable to factor all trivial numbers as shown in example 3. Although NF method factorizes all the integer numbers as described in example 4. The elapsed time with respect to decimal digits in prime factors is found less as compared to considered existing methods. The speed to compute the prime factors is also found more as compare to existing methods. Therefore the proposed method is relatively simple, fast and scalable as compare to existing methods.

References

- [1] Hung-Min Sun, Mu-En Wu, Wei-Chi Ting, and M. Jason Hinek, "Dual RSA and Its Security Analysis", IEEE Transactions on Information Theory, Vol. 53, No. 8, Aug. 2007.
- [2] Joao Carlos Leandro da Silva, "Factoring Semi primes and Possible Implications", IEEE in Israel, 26th Convention, pp. 182-183, Nov. 2010.
- [3] Sattar J Aboud, "An efficient method for attack RSA scheme", ICADIWT Second International Conference, pp 587-591, 4-6 Aug 2009.
- [4] L. Scripcariu, M.D. Frunza, "A New Character Encryption Algorithm", ICMCS 2005, pp. 83 - 86, Sept., 2005.
- [5] Hongwei Si *et al*, "An Improved RSA Signature Algorithm based on Complex Numeric Operation Function", International Conference on Challenges in Environmental

Science and Computer Engineering , vol. 2, pp. 397-400, 2010.

- [6] B. Schneier, Applied cryptography, second edition, NY: John Wiley & Sons, Inc., 1996.
- [7] J. Pollard, "Monte Carlo methods for index computation (mod p)", Math. Comp., Vol. 32, pp.918-924, 1978.
- [8] J. Pollard, "Theorems on factorization and primality testing", Proc. Cambridge Philos.Soc., Vol. 76, pp.521-528, 1974.
- [9] R. P. Brent, "An improved Monte Carlo factorization algorithm", BIT 20 (1980), 176-184. MR 82a:10007, Zbl 439.65001. rpb051 .
- [10] R. Rivest, A. Shamir, L. Adleman, "A Method for Obtaining Digital for Signatures and Public-Key Cryptosystems", Communications of the ACM, vol. 21 (2), pp. 120-126, 1978.
- [11] Bell, E. T., "The Prince of Amateurs: Fermat." New York: Simon and Schuster, pp. 56-72, 1986.
- [12] Ms. Divya Bansal, Ajay Goel, "Interrelation Among RSA Security, Strong Primes, and Factoring"
- [13] M. Wiener, "Cryptanalysis of short rsa secret exponents", IEEE Transactions on Information Theory, 160:553-558, March 1990.

Bhagwant Ram Ambedkar received the B. Tech. Degree in Electronics Engineering from Institute of Engineering and Technology Lucknow, India in 2001 and M. Tech. with specialization in Wireless Communication and Computing from Indian Institute of Information Technology, Allahabad, India in 2004. Presently he is working as Assistant Professor in Department of Computer Science and Information Technology, MJP Rohilkhand University Bareilly, Uttar Pradesh, India.



Dr. Sarabjeet Singh Bedi received the M.E. degree in Computer Science and Engineering from Thapar Institute of Engineering and Technology, Punjab, India in 2002. He has received his Ph.D. from Indian Institute of Information Technology and Management, Gwalior, India. He has teaching and research experience of 16 years. His research interest is Network Management and Security. Currently He is teaching at M.J.P. Rohilkhand University, Bareilly, India. He is involved in various academic and research activities. He is member of selection, advisor, governing and technical committees of various Universities and Technical Institution bodies. Dr. S. S. Bedi received Institute Medal from TIET, Punjab, India, 2002, Rastraya Shikshak Ratan Award from AIBDA, New Delhi in 2002 and Best Paper award at World Congress on Engineering at London, 2008.



A Classical Fuzzy Approach for Software Effort Estimation on Machine Learning Technique

S.Malathi¹ and Dr.S.Sridhar²

¹ Research Scholar, Department of CSE, Sathyabama University
Chennai, Tamilnadu, India

² Research Supervisor, Department of CSE and IT, Sathyabama University
Chennai, Tamilnadu, India

Abstract

Software Cost Estimation with resounding reliability, productivity and development effort is a challenging and onerous task. This has incited the software community to give much needed thrust and delve into extensive research in software effort estimation for evolving sophisticated methods. Estimation by analogy is one of the expedient techniques in software effort estimation field. However, the methodology utilized for the estimation of software effort by analogy is not able to handle the categorical data in an explicit and precise manner. A new approach has been developed in this paper to estimate software effort for projects represented by categorical or numerical data using reasoning by analogy and fuzzy approach. The existing historical datasets, analyzed with fuzzy logic, produce accurate results in comparison to the dataset analyzed with the earlier methodologies.

Keywords: software effort, Analogy, Fuzzy logic, categorical data, datasets.

1. Introduction

The software environment has evolved significantly in the last 30 years. To estimate software development effort, the use of the neural networks has been viewed with skepticism by majority of the cost estimation community. Even though neural networks have exposed their strengths in solving multifarious problems, their limitation of being 'black boxes' has limited their usage as a common practice for cost estimation [4]. Some models carry a few advantageous features of the neuro-fuzzy approach, such as learning capability and excellent interpretability, while maintaining the qualities of the COCOMO model [6].

Estimation by Analogy is simple and flexible, compared to algorithmic models. Analogy technique is applied effectively even for local data which is not supported by algorithmic models [2], [8]. It can be used for both qualitative and quantitative data, reflecting closer types of datasets found in real life. Analogy based estimation has the potential to mitigate the effect of outliers in a historical

data set, since estimation by analogy does not rely on calibrating a single model to suit all the projects. Unfortunately, it is difficult to assess the preliminary estimation as the available information about the historic project data during early stages is not sufficient [11]. The proposed method effectively estimates the software effort using analogy technique with the classical fuzzy approach.

The research paper is organized as follows: Section 2 deals with some of the recent research works related to the proposed technique. Section 3 describes the proposed technique and Section 4 discusses about the experimentation and comparative results with necessary tables and graphs and Section 5 concludes the paper.

2. Related Work

The proposed technique elucidates the effective estimation of effort. Several researchers have carried out researches in the field of effort estimation for the software projects using various techniques [9]. A few of the significant researches have been highlighted here for iris recognition.

Fuzzy logic has been applied to the COCOMO using membership functions such as Symmetrical Triangles and Trapezoidal Membership Function (TMF) to signify the cost drivers. The limitation of the latter function is that a few attributes were assigned the maximum degree of compatibility instead of lower degree. To overcome this drawback, Ch. Satyananda Reddy et al. [7] proposed the usage of Gaussian Membership Function (GMF) for the cost drivers by studying the behaviour of COCOMO cost drivers. COCOMO dataset has been used in the proposed methodology and the experiments envisage the scientific approach and compare the same with the standard version of the COCOMO. It is adduced that the Gaussian function performs better than the trapezoidal function as the latter facilitates a smooth transition in its intervals and the achieved results are closer to the actual effort.

Ahmeda and Muzaffar [6] dealt with the imprecision and uncertainty in the inputs of effort prediction. M.Kazemifard et al. [1] uses a multi agent system for handling the characteristics of the team members in fuzzy system. There are many studies that utilized the fuzzy systems to deal with the ambiguous and linguistic inputs of software cost estimation [3].

Wei Lin Du et al. [5] proposed an approach combining the neuro-fuzzy technique and the SEER-SEM effort estimation algorithm. The continuous rating values and linguistic values are the inputs of the proposed model for avoiding the deviation in estimation among similar projects. The performance of the proposed model has been improved by designing and evaluated with data from published historical projects. The evaluation results indicate that the estimation with the proposed fuzzy model containing analogy reasoning produce better results in comparison with the existing estimated results [4] that uses feature selection algorithm.

3. Proposed Methodology

3.1 Effort Estimation

Fuzzy logic is based on human behaviour and reasoning. It has an affinity with fuzzy set theory and applied in situations where decision making is difficult. A Fuzzy set can be defined as an extension of classical set theory by assigning a value for an individual in the universe between the two boundaries that is represented by a membership function.

$$A = \int_{x} \mu_A(x) / x \quad (1)$$

Where x is an element in X and $\mu_A(x)$ is a membership function. A Fuzzy set is characterized by a membership function that has grades between the interval $[0, 1]$ called grade membership function. There are different types of membership function, namely, triangular, trapezoidal, Gaussian etc.

Fuzzy logic consists of the following three stages:

1. Fuzzification
2. Inference Engine
3. Defuzzification

The Fuzzifier transforms the inputs into a membership value for the linguistic terms. The function of inference engine is to develop the complexity matrix for producing a new linguistic term to determine the productivity rate by using fuzzy rules. A defuzzifier carries out the

Defuzzification process to combine the output into a single label or numerical value as required.

3.2 Fuzzy Analogy

Fuzzification of classical analogy procedure is Fuzzy analogy. It comprises the following procedures, viz., 1) Identification of cases, 2) Retrieval of similar cases and 3) Case adaptation. Each step is the fuzzification of its equivalent classical analogy procedure.

3.2.1 Identification of cases

The goal of this step is the characterization of all software projects by a set of attributes. Selecting attributes, which will describe software projects, is a complex task in the analogy procedure. Indeed, the selection of attributes depends on the objective of the CBR system. In this case, the objective is to estimate the software project effort. Consequently, the attributes must be relevant for the effort estimation task. The objective of the proposed Fuzzy Analogy approach is to deal with categorical data. So, in the identification step, each software project is described by a set of selected attributes which can be measured by numerical or categorical values. These values will be represented by fuzzy sets.

In the case of numerical value x_0 , its fuzzification will be done by the membership function which takes the value of 1 when x is equal to x_0 and 0 otherwise. For categorical values, M attributes are considered and for each attribute M_j , a measure with linguistic values is defined

(A_k^j) . Each linguistic value A_k^j is represented by a fuzzy set with a membership function $(\mu_{A_k^j})$.

$$A_k^j$$

It is preferable that these fuzzy sets satisfy the normal condition. The use of fuzzy sets to represent categorical data, such as 'very low' and 'low', is similar to how humans interpret these values and consequently it allows dealing with imprecision and uncertainty in the case identification step.

3.2.2 Retrieval of Similar Cases

This step is based on the choice of software project similarity measure. In this method, a set of candidate measures for software project similarity has been proposed

for software project similarity. These measures assess the overall similarity of two projects P_1 and P_2 , $d(P_1, P_2)$ by combining all the individual similarities of P_1 and P_2 associated with the various linguistic variables V_j describing the project P_1 and P_2 , $d_{V_j}(P_1, P_2)$. After an

axiomatic validation of some proposed candidate measures for the individual distances $d_{V_j}(P_1, P_2)$, two measures

have been retained [14].

$$d_{V_j}(P_1, P_2) = \begin{cases} \max_k \min(\mu_{A_k^j}(P_1), \mu_{A_k^j}(P_2)) \\ \text{max-min aggregation} \\ \sum_k \mu_{A_k^j}(P_1) \times \mu_{A_k^j}(P_2) \\ \text{sum-product aggregation} \end{cases} \quad (2)$$

Where A_k^j are the fuzzy sets associated with V_j and

$\mu_{A_k^j}$ are the membership functions representing fuzzy

sets A_k^j . Scale factors (SF) are understanding product objectives, flexibility, team coherence, etc., Effort multipliers (EF) are software reliability, database size, reusability, complexity, etc. The imprecision of the cost drivers significantly affects the accuracy of the effort estimates which are derived from effort estimation models. Since the imprecision of software effort drivers cannot be overlooked, a fuzzy model gains advantage in verifying the cost drivers by adopting fuzzy sets.

$$\text{Effort} = A * (\text{SIZE})^{B+0.01*\sum_{i=1}^N SF_i} * \prod_{i=1}^N EM_i \quad (3)$$

Where A and B are constants, SF is the scale factor and EM is effort multipliers. By using the above formula the effort is estimated. The cost drivers are fuzzified using triangular and trapezoidal fuzzy sets for each linguistic value such as very low, low, nominal, high etc. as applicable to each cost driver.

Rules are developed with cost driver in the antecedent part and corresponding effort multiplier in the consequent part. The defuzzified value for each of the effort multiplier is obtained from individual Fuzzy Inference Systems after matching, inference aggregation and subsequent Defuzzification. Total Effort is obtained after multiplying

them together. The high values for the cost drivers lead an effort estimate that is more than three times the initial estimate, whereas low values reduce the estimate to about one third of the original. This highlights the vast differences between different types of projects and the difficulties of transferring experience from one application domain to another

3.2.3 Case adaptation

The objective of this step is to derive an estimate for the new project by using the known effort values of similar projects. We are not convinced in fixing the number of analogies in this step. In our proposed method, all the projects in a dataset are used to derive the new project estimate. Each historical project will contribute, in the calculation of the effort of the new project according to the similarity.

4. Results and Discussion

The datasets used in the study is the Desharnais dataset [15], NASA 93 [12] and COCOMO NASA dataset shown in Table 1 and Table 2.

Table 1: Desharnais project features

Variable	Description	Type
ExpEquip	Experience of Equipment	Continuous
ExpProj Man	Experience of Project Manager	Continuous
Trans	Transactions	Continuous
Raw FP	Raw Function Points	Continuous
Adj. Factor	Technology Adjustment Factor	Continuous
Adj.FP	Adjusted Function Points	Continuous
Dev Env	Development Environment	Categorical
Year Fin	Year Finished	Continuous
Entities	Number of entities	Continuous
Effort	Actual effort	Continuous

Table 2: Nasa93 project features

Variable	Description	Type
Acap	Analysts Capability	Categorical

Pcap	Programmers Capability	Categorical
Aexp	Application Experience	Categorical
Modp	Modern Programming Practices	Categorical
Tool	Use of Software tools	Categorical
Vexp	Virtual Machine Experience	Categorical
Lexp	Language Experience	Categorical
Sced	Schedule Constraint	Categorical
Stor	Main Memory Constraint	Categorical
Data	Database Size	Categorical
Time	Time Constraint for CPU	Categorical
Turn	Turnaround time	Categorical
Virt	Machine Volatility	Categorical
Rely	Required Reliability	Categorical
Cplx	Process Complexity	Categorical
Loc	Line of Code	Continuous
DevEff	Development Effort	Continuous

Table 3, summarizes the number of projects collected under each dataset with the max effort which is compared with the estimated method. From this table, it is inferred that the datasets have a very low max effort when compared to the actual max effort for the proposed method.

Table 3 Comparison of actual max effort
 With estimated max effort

Datasets	No of projects	Actual Max Effort	Estimated Max Effort
Nasa60	60	3240	1594
Desharnais	77	23,940	10,950
Nasa93	93	8211	1290

Figure.1 represents the graphical plot of the datasets versus the max effort of the estimated and the proposed method.

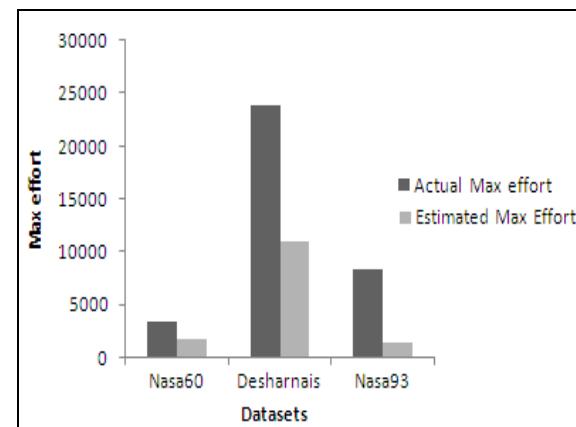


Figure 1: Comparative results of actual and estimated maximum efforts

In this method, 30% of test data sets is taken from NASA60, NASA93 and Desharnais to measure the average effort in comparison with the actual values. From the comparative results given in Table 4, it is predicted that the existing average effort using the selected features is very high compared to the proposed method for each dataset while considering the full-fledged features sets.

Table 4. Comparative results of average effort and actual effort

Datasets	Test set	Actual Avg Effort	Existing Method		Proposed Method	
			No. of Feature	Avg. Effort	No. of Feature	Avg. Effort
Nasa60	16	340.49	3	354.66	16	284.34
Desharnais	22	5119.3	2	4852.17	10	2428.9
Nasa93	26	734.03	4	722.96	16	640.65

Figure 2 represents the graphical plot of the datasets versus the average effort among the existing and the proposed method.

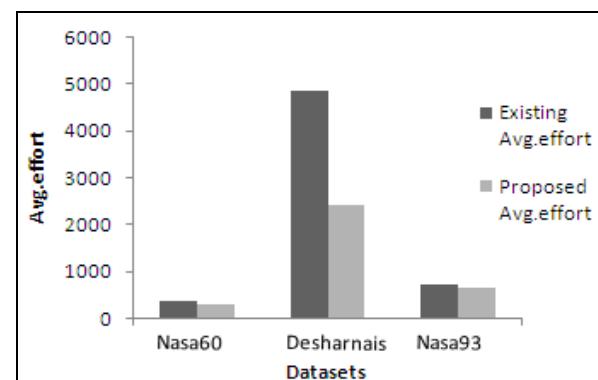


Figure 2: Comparative results of avg. effort in existing and proposed method

Therefore, it is concluded that by taking all the features into consideration for all the datasets, the average effort will be low in the proposed method in comparision to the existing method [4].

5. Conclusions

A new classical approach has been proposed in this paper to estimate the average software project effort. This approach is based on reasoning by analogy, fuzzy logic and linguistic quantifiers, which can be effectively used when the software projects are described by categorical and or numerical data. The new approach improves the classical analogy procedure while using the categorical data. In the fuzzy analogy approach, both categorical and numerical data are represented by fuzzy sets. The advantage of this method is that it can handle the imprecision and the uncertainty quite vividly while describing the software project. From the implementation of the results, it is observed that the proposed method has effectively estimated the average effort for the software project datasets.

References

- [1] M.Kazemifard, A.Zaeri, N.ghasem-ghaee, M.A.Nematbakhsh, F.Mardukhi, "Fuzzy Emotional COCOMO II Software Cost Estimation (FECSCE) using Multi-Agent Systems", Applied Soft Computing, Elsevier, pg.2260-2270, 2011.
- [2] Ekrem Kocaguneli, Tim Menzies, Ayse Bener, Jacky W.Keung, "Exploiting the Essential Assumptions of Analogy-based Effort Estimation", Journal of IEEE Transactions on Software Engineering, Vol.34, No.4, pp. 471-484, July 2010.
- [3] Iman Attarzadeh and Siew Hock Ow , "Improving the Accuracy of Software Cost Estimation Model Based on a new Fuzzy Logic Model", World applied sciences Journal, Vol. 8, No. 2, pp. 177-184, 2010.
- [4] Pichai Jodpimai, Peraphon Sophatsathit and Chidchanok Lursinsap,"Estimating Software Effort with Minimum Features using Neural Functional Approximation", ICCSA.2010.
- [5] Wei Lin Du, Danny Ho and Luiz Fernando Capretz, "Improving Software Effort Estimation Using Neuro-Fuzzy Model with SEER-SEM", Global Journal of Computer Science and Technology, Vol. 10, No. 12, Pp. 52-64, Oct 2010.
- [6] M.A. Ahmeda, Z. Muzaffar, "Handling imprecision and uncertainty in software development effort prediction: a type-2 fuzzy logic based framework", Information and Software Technology 51 (2009) 640–654.
- [7] Ch. Satyananda Reddy and KVSVN Raju, "An Improved Fuzzy Approach for COCOMO's Effort Estimation using Gaussian Membership Function", Journal Of Software, Vol. 4, No. 5, Pp. 452-459, July 2009.
- [8] J.Keung,"Empirical evaluation of analogy-x for software cost estimation", in ESEM '08: Proceedings of the second ACM-IEEE international symposium on Empirical engineering and measurement. New York, NY, USA: ACM, 2008, pp.294-296.
- [9] M.Jorgensen and M.Shepperd,"A systematic review of software development cost estimation studies", IEEE Trans.Softw.Eng., vol.33, no.1, pp.33 -53, 2007.
- [10] Xishi Huang, Danny Ho, Jing Ren and Luiz F. Capretz, "Improving the COCOMO model using a neuro-fuzzy approach", Applied Soft Computing, Vol. 7, Pp.29–40, 2007
- [11] Hasan Al-Sakran, "Software Cost Estimation Model Based on Integration of Multi-agent and Case-Based Reasoning", Journal of Computer Science, Vol. 2, No. 3, Pp. 276-282, 2006
- [12] Sayyad Shirabad, J. and Menzies, T.J. (2005) The PROMISE Repository of Software Engineering Databases. School of Information Technology and Engineering, University of Ottawa, Canada. Available: <http://promise.site.uottawa.ca/SERepository>
- [13] Ali Idri, Taghi M.Khoshgoftaar and Alain Abran, "Can Neural Networks be easily Interpreted in Software Cost Estimation?", 2002 World Congress on computational intelligence, honolulu, Hawaii, pp. 1-8, May 12-17, 2002.
- [14] A.Idri and A. Abran, "Towards A Fuzzy Logic Based Measures For Software Project similarity", In Proc. of the 7th International Symposium on Software Metrics, England, pp.85-96, 2001.
- [15] M.J.Shepperd, C.Schofield and B.Kitchenham, "Estimating Software Project Effort Using Analogies", IEEE Trans. Software Eng., vol. 23, pp. 736-743, 1997.

Intelligent Paging Strategy for Multi-Carrier CDMA System

Sheikh Shanawaz Mostafa¹, Khondker Jahid Reza², Md. Ziaul Amin³ and Mohiuddin Ahmad⁴

¹ Dept. of Biomedical Engineering, Khulna University of Engineering and Technology,
Khulna-9203, Bangladesh.

² Electrical & Electronics Engineering, University Malaysia Pahang,
Pahang, Malaysia.

³ Electronics and Communication Engineering Discipline, Khulna University,
Khulna-9208, Bangladesh

⁴ Dept. of Electrical and Electronic Engineering, Khulna University of Engineering and Technology,
Khulna-9203, Bangladesh.

Abstract

Subscriber satisfaction and maximum radio resource utilization are the pivotal criteria in communication system design. In multi-Carrier CDMA system, different paging algorithms are used for locating user within the shortest possible time and best possible utilization of radio resources. Different paging algorithms underscored different techniques based on the different purposes. However, low servicing time of sequential search and better utilization of radio resources of concurrent search can be utilized simultaneously by swapping of the algorithms. In this paper, intelligent mechanism has been developed for dynamic algorithm assignment basing on time-varying traffic demand, which is predicted by radial basis neural network; and its performance has been analyzed are based on prediction efficiency of different types of data. High prediction efficiency is observed with a good correlation coefficient (0.99) and subsequently better performance is achieved by dynamic paging algorithm assignment. This claim is substantiated by the result of proposed intelligent paging strategy.

Keywords: Concurrent Search, Multi-Carrier CDMA, Paging, Radial Basis Neural Network, Sequential Search.

1. Introduction

In mobile communication, performance of a system depends on the call processing. It is required to identify or page the user to initiate a call for a particular user. The forward-link communication channels which are used to

page and transmit system overhead messages to the MS is called paging channel [1].

As personal communication service demands are growing day by day, so it is mandatory to utilize the limited radio resources efficiently. Many researchers have been investigating different mechanisms and techniques to utilize the paging channel efficiently. Among them Rung-Hung Gue and his colleagues [2] proposed concurrent search algorithm. This algorithm was used for paging in cells. Another concept described in [3] and [4] which were related to divide a location area into paging zones. A different type of paging technique has been described in [5] where MS will be paged on the last registered cell first and then other cells in the location area if necessary. A proposal of mobility tracking scheme that merges a movement-based location update policy with a selective paging scheme was proposed in [6]. For efficiently using the resources, probability based priority paging used by other researchers [7]. Neural network has been used for load management in different purpose. A hybrid Bayesian neural network predicting locations on cellular networks was used in [8] and [9]. Neural network for creating personal mobility profile of individual user was used in [10]. Fuzzy logic for paging mobile user was observed in [11]. A Dynamic Location Management for reducing the location update and paging cost was investigated in [12].

In multi carrier CDMA system, an MS is capable of tuning to only one of the carrier frequencies at a time. That's why, it is required that the paging message must be sent on each

and every paging channel of all of the carrier frequencies. It is done by duplicating the paging message for each of the paging channels on multicarrier system [13]. As there are only seven paging channels for each carrier in a multi carrier CDMA system, therefore the existing system can't search more than seven users at a time. This result is an inefficient utilization of radio resources. Concurrent Search Algorithm solves this problem requiring slightly more servicing time compared to sequential search [14].

In this paper, intelligent paging strategy is proposed which can be used in busy hours to search more than seven users simultaneously in multiple carriers, like Concurrent Search Algorithm, and servicing time is as low as Sequential Search Algorithm in normal traffic. We adopt radial basis function (RBF) neural network for predicting paging and choosing the suitable algorithm between two. The system performance has been analyzed by using Erlang C formula.

2. Sequential Search Algorithm vs. Concurrent Search Algorithm

This paper is based on the limitation of concurrent search algorithm and Sequential Search Algorithm. In concurrent search algorithm system parallel paging rather than forward sequential paging scheme has been used to locate user i among k number of users.

Presuming there are only two carriers in the system with same number of users. It is also assumed that the system has no previous knowledge about the location of the users: so each user can exist in any of the two carriers with probability 0.5. in sequential paging scheme four pages will be required to locate two users where concurrent search need three messages. Overall 25% saving of paging message means 25% more channel for paging.

If the servicing time is calculated by absorbing Markov chains [15] considering that Search Algorithm has one unit paging (service) time than Concurrent Search Algorithm will have 1.5 units [14]. And if Concurrent system is imagined with 14 parallel servers having queue than Sequential system can be imagined with 7 parallel servers having queue [16]. In this case its average time in the system can be found from Erlang C formula [17], [14] is,

$$T = \frac{A^C}{\mu \times (C - A) \times \left\{ A^C + C! \left(1 - \frac{A}{C} \right) \sum_{k=0}^{C-1} \frac{A^k}{k!} \right\}} + \frac{1}{\mu} \quad (1)$$

And the blocking probability is

$$P_0 = \frac{A^C}{A^C + C! \left(1 - \frac{A}{C} \right) \sum_{k=0}^{C-1} \frac{A^k}{k!}} \quad (2)$$

where C is the number of channels. A is total offered traffic which is λ/μ , assuming λ is incoming traffic rate and μ is the service rate.

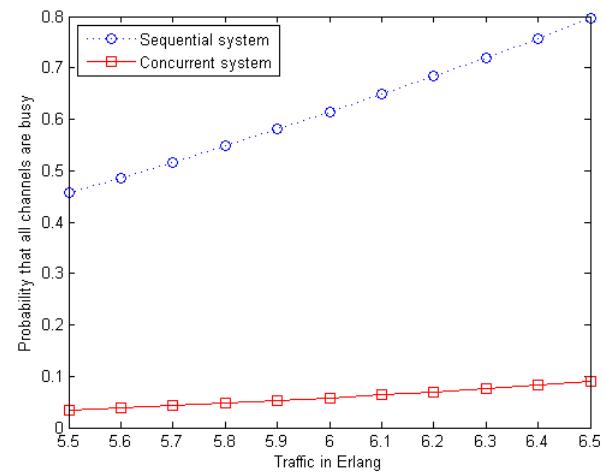


Fig. 1 Blocking probability of Concurrent Search and Sequential Search [14].

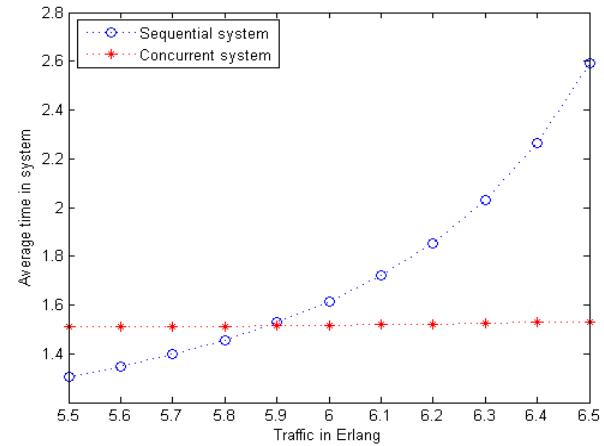


Fig. 2 Servicing time of Concurrent Search and Sequential Search [14].

From Fig. 1 and Fig. 2 it is clear that the blocking probability of concurrent system is always lower than that of sequential system. Considering the system performance in terms of average servicing time, Concurrent Search algorithm will provide low servicing time only after threshold traffic demand, which is six Erlang. Real time traffic analysis may be a possible way out of this problem, but it will introduce computational complexity at the

servicing end. Swapping of algorithms based on real time traffic cannot provide feasible solution because the services would be hampered during the switching process of the algorithms. So previously predicted traffic would solve the stated problem. By forecasting the load, the system will know which algorithm is to be used before the arrival of new traffic.

3. Method of Intelligent Paging Strategy

3.1 Radial Basis Neural Network

The incoming traffic of the mobile system usually follows a specific pattern depending on the characteristics of area and time. A reliable time series prediction method is needed for forecasting this type of system behaviors. In recent years, neural networks gain its recognition among different types of time series prediction methods. Radial basis neural network was used in [18], [19] and [20] for time series prediction problem. Moreover, radial basis function networks have rapid training time and have no local minima problem as back propagation does. Instead of using weighted-sum mechanism or sigmoid activation for hidden-unit outputs, radial basis neural network (RBNN) uses vector distance between its weight vector and the input vector [21]. Radial basis networks consist of two layers of neural (Fig. 3): a radial basis layer (hidden layer), and linear layer.

Radial Basis Layer: Each predictor variable defines a neuron in input radial basis layer. In the case of categorical variables, $N-1$ neurons are used where N is the number of categories. Each neuron's weighted input is the distance between the input vector and its weight vector. Net input is the product of its weighted input with its bias. Each neuron's output $h(x)$ is its net input x passed through the following function.

$$h(x) = e^{\frac{(x-u)^2}{2\sigma^2}} \quad (3)$$

Where x is the input vector, u is the mean and σ is the standard deviation. [21]

Linear Layer: The linear neuron uses a linear transfer function.

$$O = g \times I \quad (4)$$

Where I is the value coming out of a neuron in the hidden layer is multiplied by a weight associated with the neuron and passed to the summation which adds up the weighted

values and presents this sum as the output of the network. For classification problems, there is one output for each target category.

In this paper, three different types of traffic namely type-1 (T1), type-2 (T2) and type-3 (T3) have been considered to simulate the varying characteristics of traffic. The bias in the radial basis layer was set to 0.8326; and neuron will respond with 0.5 or more to any input vectors within a vector distance of one from their weight vector. Mean Square Error (MSE) goal was set to 0.02 to train the radial basis neural network. It was required 50 Neurons to achieve the defined Mean Square Error goal.

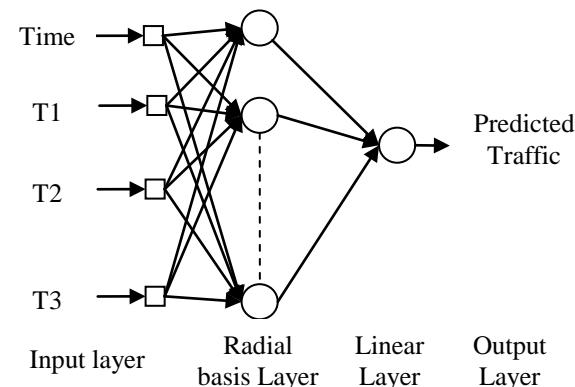


Fig. 3 Radial basis neural network for intelligent paging

3.2 Predictor Performance Evaluators

The performance of a predictor can be determined mathematically by calculating the correlation coefficient and error of actual and predicted traffic.

Supposing, $x(n)$ be the actual traffic, and $x_m(n)$ be the predicted traffic generated by the radial basis neural network model, then Mean Square Error (MSE) can be defined by the following equation incorporating them [22].

$$MSE = \frac{1}{N} \sum_{n=0}^{N-1} [x(n) - x_m(n)]^2 \quad (5)$$

The normalized form of MSE is

$$NMSE = \frac{\sum_{n=0}^{N-1} [x(n) - x_m(n)]^2}{\sum_{n=0}^{N-1} [x(n)]^2} \quad (6)$$

Another measurement is Root Mean Square Error, which is

$$RMSE = \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} [x(n) - x_m(n)]^2} \quad (7)$$

The Normalized version of RMSE is

$$NRMSE = \sqrt{\frac{\sum_{n=0}^{N-1} [x(n) - x_m(n)]^2}{\sum_{n=0}^{N-1} [x(n)]^2}} \quad (8)$$

Percent Root Mean Difference (PRD) can be determined by Eq. (9).

$$PRD = \sqrt{\frac{\sum_{n=0}^{N-1} [x(n) - x_m(n)]^2}{\sum_{n=0}^{N-1} [x(n)]^2}} \times 100\% \quad (9)$$

Cross Correlation can be defined as a sequence of r_{xx_m} ,

Where;

$$r_{xx_m}(l) = \sum_{n=-\infty}^{\infty} x(n)x_m(n-l), \quad l=0, \pm 1, \pm 2, \dots \quad (10)$$

or equivalently,

$$r_{xx_m}(l) = \sum_{n=-\infty}^{\infty} x(n+l)x_m(n), \quad l=0, \pm 1, \pm 2, \dots \quad (11)$$

The index l is the time shift parameter and the subscripts xx_m is the correlation sequence and $r_{xx_m}(l)$ the sequences being correlated [23].

4. Result and Performance Evaluation

System performance is highly depended on algorithms swapping efficiency, while it depends on the reliability of traffic prediction mechanism. Therefore, the performance of intelligent paging strategy has been analyzed considering the traffic prediction and total system performance which is measured by blocking probability and average servicing time of the system.

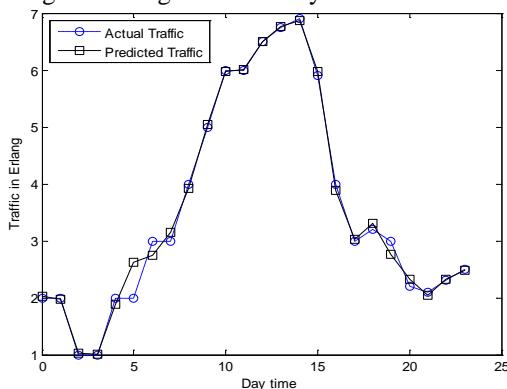


Fig. 4 Prediction of traffic by using RBNN for Type-1 traffic

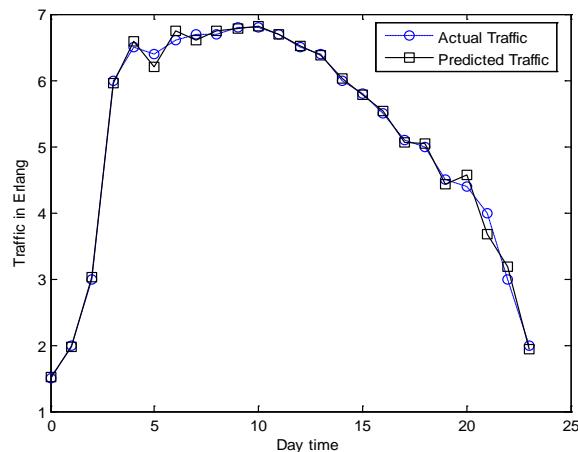


Fig. 5 Prediction of traffic by using RBNN for Type-2 traffic

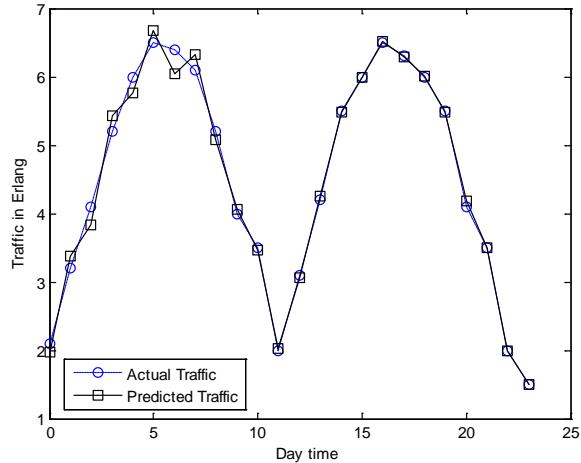


Fig. 6 Prediction of traffic by using RBNN for Type-3 traffic

Since prediction capability will determine the total system performance of the system. So, reliable predictor is required to get a realistic prediction. The Predictor, Neural Network, is trained with random data to get maximum performance. The prediction is almost identical to the simulated traffic with few exceptions and it is substantiated by the Fig 4, Fig 5 and Fig 6.

Mathematical error performance of predictor is presented in Table I. It is clearly depicted the performance of predictor by the high Cross correlation value and low MSE, RMSE and PRD.

Table I: Performance parameters of the predictor

	MSE	NMSE	RMSE	NRMSE	PRD	Correlation Coefficient
T1	2.6e-2	1.6e-3	1.6e-1	2.5e-6	2.5e-4	0.9992
T2	1.1e-2	3.7e-4	1e-1	1.4e-7	1.3e-5	0.9998
T3	2e-2	8.8e-4	1.4e-1	7.8e-7	7.8e-5	0.9996

System Performance depends on timely swapping of algorithm. This is achieved due to good prediction of the predictor and successful swapping is seen in the Fig 7, Fig 8 and Fig 9.

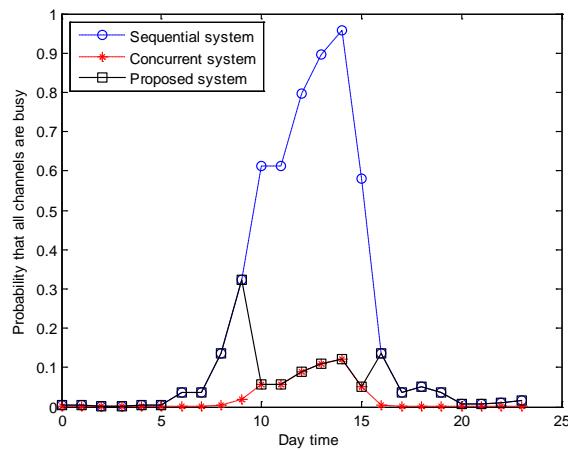


Fig. 7 Blocking probability comparison of proposed, concurrent and sequential search for Type-1 traffic.

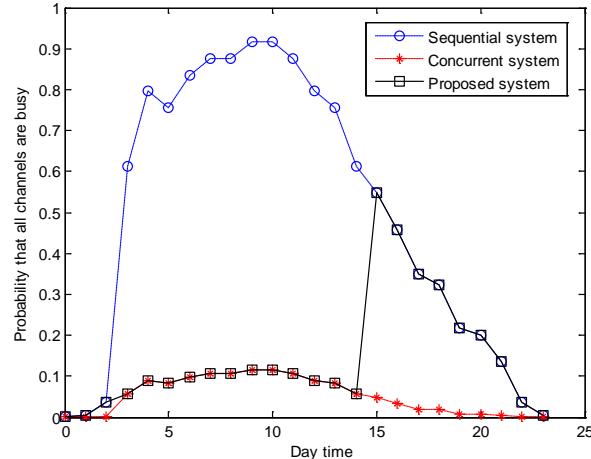


Fig. 8 Blocking probability comparison of proposed, concurrent and sequential search for Type-2 traffic

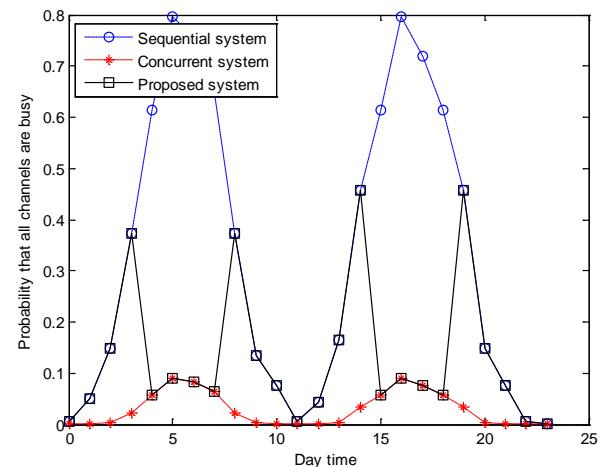


Fig. 9 Blocking probability comparison of proposed, concurrent and sequential search for Type-3 traffic

Performance of the system, which is least possible service time, can be achieved by efficient swapping of the two algorithms. When traffic is lower than the threshold traffic, sequential search will require minimal time. That is why our intelligent system will follow sequential search algorithm. When traffic is greater than the threshold traffic, then the minimal accessing time is provided by concurrent search algorithm. From that time our intelligent paging switches to concurrent search algorithm. The reverse pattern is also observed in the Fig 10, Fig11and Fig 12.

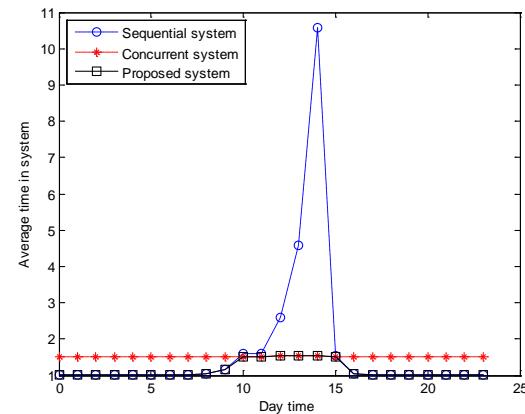


Fig. 10 Servicing time comparison of proposed, concurrent and sequential search for Type-1 traffic.

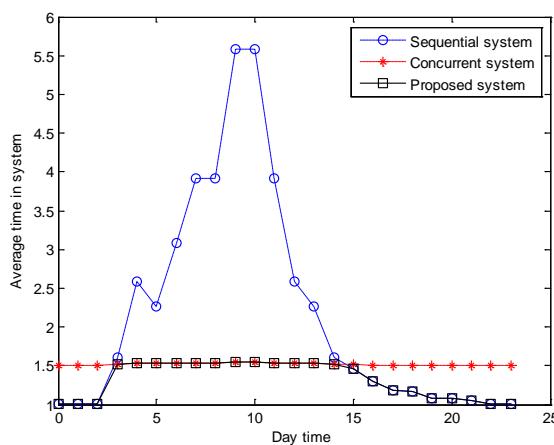


Fig. 12 Servicing time comparison of proposed, concurrent and sequential search for Type-3 traffic.

5. Conclusion

In this paper, an intelligent paging strategy has been developed assuming that traffic follows some deterministic pattern which is predicted by radial basis neural network. It not only forecast the traffic pattern with less than 0.15 RMSE error but also increases the system performance in varying traffic demand which is rather common in wireless system. Due to offline training method implemented in this strategy, the training time of the network has no significant effect on the system. It requires very limited resources of the implementing terminal. However, this strategy has been applied on simulated traffic with good result. The model can be further studied comprehensively by using real traffic data.

References

- [1] S. Saleh Faruque, Cellular Mobile System Engineering, Artech House Publishers, New ed., 2007-2008.
- [2] Rung-Hung Gue and Zygmunt J. Haas “Concurrent search of mobile users in cellular networks”, IEEE/ACM Transactions on Networking, vol. 12, no. 1, Feb 2004.
- [3] D. Munoz-Podrguez, “Cluster paging for traveling subscribers,” 40th IEEE Vehicular Technology Conf. 1990, May, 6-9, pp. 748-753.
- [4] D. Plassmann, “Location management strategies for mobile cellular networks of 3rd generation,” 44th IEEE Vehicular Technology Conf., June 1994, pp. 649-653.
- [5] G. L. Lyberopoulos, J.G. Markoulidakis, D.V. Polymeros, D.F. Tsirkas, and E.D. Sykas, “Intelligent paging strategies for 3rd generation mobile telecommunication systems,” IEEE Transaction on Vehicular Technology, vol 44 issue 3, Aug. 1995, pp. 573-553.
- [6] I.F. Akyildiz, J.S. M. Ho, and Y.-B. Lin, “Movement-based location update and selective paging for PCS networks,” IEEE /ACM Transaction on Networking, vol. 4, pp. 629-638, August 1996.
- [7] C. rose and R. Yates, “Minimizing the average cost of paging under delay constraints,” ACM /Kluwer Wireless Networks, vol. 1, no.2, pp. 211-219, 1995.
- [8] S. Akoush and A. Sameh, “The use of bayesian learning of neural networks for mobile user position prediction”, 7th International Conference on Intelligent Systems Design and Applications, 2007, pp. 441 – 446.
- [9] S. Akoush and A. Sameh, “Bayesian learning of neural networks for mobile user position prediction” 16th International Conference on Computer Communications and Networks, 2007, September 2007, pp. 1234 – 1239.
- [10] Kausik Majumdar and Nabanita Das, “Mobile user tracking using a hybrid neural network”, Wireless Networks, Volume 11 Issue 3, May 2005, pp. 275-284.
- [11] Sajal Saha, Raju Dutta, Soumen Debnath, and Asish K. Mukhopadhyay, “Intelligent paging based mobile user tracking using fuzzy logic”, International Conference on Methods and Models in Science and Technology, November 2010, pp. 19-23.
- [12] J. Amar Pratap Singh and M. Karnan,“A dynamic location management scheme for wireless networks using cascaded correlation neural network,” International Journal of Computer Theory and Engineering, Vol. 2, No. 4, August, 2010, pp. 581-585.
- [13] William E. Illidge, “CDMA multiple carrier paging channel optimization”, Patent number: US 6,542,752, Apr. 1, 2003.
- [14] Sheikh Shanawaz Mostafa, Khondker Jahid Reza , Gazi Maniur Rashid, Muhammad Moinuddin, Md. Ziaul Amin and Abdullah Al Nahid, “An efficient paging algorithm for multi-carrier CDMA system”, International Journal of Computer Science Issues, Vol. 8, Issue 3, No. 2, May 2011, pp. 421-426.
- [15] Charles M. Grinstead, J. Laurie Snell, Introduction to Probability, 2nd ed., American Mathematical Society, 1991.
- [16] CDMA2000 1X paging Optimization Guide, Huawei Technologies Co., Ltd., 2004.
- [17] Theodore S. Rappaport, Wireless Communications, 2nd ed., Prentice-Hall of India, 2002.
- [18] H. Leung, T. Lo, and S. Wang, “Prediction of noisy chaotic time series using an optimal radial basis function neural network,” IEEE Transaction on Neural Networks, vol. 12, no. 5, Sept. 2001, pp. 1163 – 1172.
- [19] Dongqing Zhang, Xuanxi Ning, Xueni Liu and Yubing Han; “Nonlinear time series forecasting with dynamic RBF neural networks”, 7th World Congress on Intelligent Control and Automation, June 2008, pp. 6988 – 6993.
- [20] Parras-Gutierrez, E.; Rivas, V.M.;” Time series forecasting: Automatic determination of lags and radial basis neural networks for a changing horizon environment,” The 2010 International Joint Conference on Neural Networks , July 2010, pp. 1 – 7.
- [21] Vallura Rao and Hayagriva Rao, C++ Neural Networks and Fuzzy Logic, 2nd ed , BPB publication,1996, pp. 114-115.
- [22] M.S.Manikandan, S.Dandapat,“Wavelet energy based diagnostic distortion measure for ECG”, Biomedical Signal Processing and Control, vol 2, issue 2, pp 80-96, Apr 2007.

- [23] John G. Proakis and Dimitris G. Manolakis, Digital Signal Processing, Prentice-Hall of India, 4th ed, 2008-2009.

Sheikh Shanawaz Mostafa completed B.Sc. Engineering in Electronics and Communication, from Khulna University-9208, Khulna, Bangladesh. His current research interests are: Wireless communication, Modulation techniques and Biomedical signal processing. He has more than seven papers, published in different local and international recognized journal and proceedings of conference.

Khondker Jahid Reza has completed his B.Sc. in Electronics and Communication Engineering Discipline in Khulna University, Khulna, Bangladesh. His current research interest is wireless communication, modulation and sensor networks. He has three papers, published in international recognized journal.

Md. Ziaul Amin is currently working as an Assistant Professor at the Khulna University, Khulna, Bangladesh. He obtained his B.Sc. in Electronics and Communication Engineering from same University. Previously, he worked as a System Engineer, planning, at RanksTel Bangladesh Ltd. since Nov 07 Aug 08. His current research interests are: Digital Signal Processing, Radio Network Planning and cognitive radio network. He has four papers, published in international recognized journal.

Dr. Mohiuddin Ahmad received his BS degree with Honors in Electrical and Electronic Engineering from Chittagong University of Engineering and Technology (CUET), Bangladesh and his MS degree in Electronics and Information Science from Kyoto Institute of Technology of Japan in 1994 and 2001, respectively. He received his PhD degree in Computer Science and Engineering from Korea University, Republic of Korea. From August 1995 to October 1998, he served as a lecturer in the Department of Electrical and Electronic Engineering at Khulna University of Engineering and Technology, Bangladesh. In June 2001, he joined the same Department as an Assistant Professor. In May 2009, he joined the same Department as an Associate Professor and now he is a full Professor. His research interests include biomedical signal and image processing, computer vision and pattern recognition, human motion analysis, circuits and energy conversion.

An Efficient and Secure Protocol for Ensuring Data Storage Security in Cloud Computing

Syam Kumar P, Subramanian R
Department of Computer Science, School of Engineering & Technology
Pondicherry University, Puducherry-605014, India

Abstract

Currently, there has been an increasing trend in outsourcing data to remote cloud, where the people outsource their data at Cloud Service Provider(CSP) who offers huge storage space with low cost. Thus users can reduce the maintenance and burden of local data storage. Meanwhile, once data goes into cloud they lose control of their data, which inevitably brings new security risks toward integrity and confidentiality. Hence, efficient and effective methods are needed to ensure the data integrity and confidentiality of outsource data on untrusted cloud servers. The previously proposed protocols fail to provide strong security assurance to the users. In this paper, we propose an efficient and secure protocol to address these issues. Our design is based on Elliptic Curve Cryptography and Sobol Sequence (random sampling). Our method allows third party auditor to periodically verify the data integrity stored at CSP without retrieving original data. It generates probabilistic proofs of integrity by challenging random sets of blocks from the server, which drastically reduces the communication and I/O costs. The challenge-response protocol transmits a small, constant amount of data, which minimizes network communication. Most importantly, our protocol is confidential: it never reveals the data contents to the malicious parties. The proposed scheme also considers the dynamic data operations at block level while maintaining the same security assurance. Our solution removes the burden of verification from the user, alleviates both the user's and storage service's fear about data leakage and data corruptions. Through security analysis, we prove that our method is secure and through performance and experimental results, we also prove that our method is efficient. To compare with existing schemes, our scheme is more secure and efficient.

Keywords: data storage, integrity, confidentiality, Elliptic Curve Cryptography(ECC), Sobol Sequence, Cloud Computing.

1. Introduction

Cloud storage becomes an increasing attraction in cloud computing paradigm, which enables users to store their data and access them wherever and whenever they need using any device in a pay-as-you-go manner[1]. Moving data into cloud offers great conveniences to the users since they do not have to care about the large capital investment in both the maintenance and management of the hardware infrastructures. Amazon's Elastic Compute Cloud (EC²) and Amazon Simple Storage Service (S3) [2] and apple icloud[3] are well known examples of cloud data storage. However, once data goes into cloud, the users lose the control over the data. This lack of control

raises new formidable and challenging issues related to confidentiality and integrity of data stored in cloud [4].

The confidentiality and integrity of the outsourced data in clouds are of paramount importance for their functionality. The reasons are listed as follows [5]:

- 1) the CSP, whose purpose is mainly to make a profit and maintains a reputation, has intentionally hide data loss an incident which is rarely accessed by the user's
- 2) The malicious CSP might delete some of data or is able to easily obtain all the information and sell it to the biggest rival of Company.
- 3) An attacker who intercepts and captures the communications is able to know the user's sensitive information as well as some important business secrets.
- 4) Cloud infrastructures are subject to wide range of internal and external threats.

The examples of security breaches of cloud service providers appear from time to time [6, 7]. The users require that their data remain secure over the CSP and they need to have a strong assurance from the cloud servers that CSP store their data correctly without tampering or partially deleting because the internal operation details of service providers may not be known to the cloud users. Thus, an efficient and secure scheme for cloud data storage has to be in a position to ensure the data integrity and confidentiality.

Encrypting the data before storing in cloud can handle the confidentiality issue. However, verifying integrity of data is a difficult task without having a local copy of data or retrieving it from the server. Due to this reason, the straightforward cryptographic primitives cannot be applied directly for protecting outsourced data. Besides, a naive way to check the data integrity of data storage is to download the stored data in order to validate its integrity, which is impractical for excessive I/O cost, high communication overhead across the network and limited computing capability. Therefore, efficient and effective mechanisms are needed to protect the confidentiality and integrity of user's data with minimum computation, communication and storage overhead.

Remote data integrity checking is a protocol that focuses on how frequently and efficiently we verify whether cloud server can faithfully store the user's data without retrieving it. In this protocol, the user generates some metadata. Later, he can challenge the server for integrity of certain file blocks through challenge-response

protocol. Then the server generates responses that the server still possesses the data in its original form to corresponding challenge sent by the verifier who may be original user or trusted third party entity. Recently, several researchers have proposed different variations of remote data integrity checking protocols under different cryptography schemes [8-21]. However, all these protocols focus on static data verification.

One of the design principles of cloud storage is to provide dynamic scalability of data for various applications. This means, the data stored in cloud are not only accessed by the users but also frequently updated through block operations such as modification, insert and delete operations. Hence, it is crucial to develop more secure and efficient mechanism to support dynamic audit services. The protocols to verify dynamic data in cloud are proposed in [22-27].

Although the existing schemes aim at providing integrity verification for different data storage systems, but problem of confidentiality of data has not been fully addressed.

The protocols [28-35] have been proposed to ensure the confidentiality and integrity of remote data. But, all these schemes are unable to provide strong security assurance to the users, because these schemes verifying integrity of outsourced data based on pseudorandom sequence, which does not cover the whole data while computing the integrity proof. Therefore, probabilistic verification schemes based on pseudorandom sequence does not give guarantee to the users about security of their data. Syam et al. [27] proposed a distributed verification protocol using Sobol sequence to ensure availability and integrity of data, but it is also not addressed the data confidentiality issue. How to achieve a secure and efficient design to seamlessly integrate these two important components for data storage service remains an open challenging task in Cloud Computing.

In this paper, we propose an efficient and secure protocol to ensure the confidentiality and integrity of data storage in cloud computing using Elliptic Curve Cryptography(ECC) [30, 37, 38] and Sobol Sequence [39]. The ECC can offer same levels of security with small keys comparable to RSA and other PKC methods. It is designed for devices with limited computing power and/or memory, such as smartcards, mobile devices and PDAs. In our design, first the user encrypts data to ensure the confidentiality, then, compute metadata over encrypted data. Later, the verifier can use remote data integrity checking protocol to verify the integrity. The verifier should able to detect any changes on data stored in cloud. The security of our scheme relies on the hardness of specific problems in Elliptic Curve Cryptography. Compared to existing schemes, our scheme has several advantages: 1) it should detect all data corruption if anybody deletes or modifies the data in cloud storage,

since we are using Sobol sequence instead of pseudorandom sequence for challenging the server for the integrity verification. 2) Our scheme achieves the confidentiality of data 3) It is efficient in terms of computation, storage, because its key size is low compared to RSA based solutions.

Main Contributions:

- 1) We propose an efficient and secure protocol. This protocol efficiently provides the integrity assurance to the users with strong evidence that the CSP is in faithfully storing all data and this data cannot be leaked to malicious parties. Our protocol also supports public verifiability and dynamic data operations such as modification, insertion and deletion.
- 2) We prove the security (integrity and confidentiality) of proposed scheme against internal and external attacks. Cloud server can provide valid response to the verifier challenges only if they actually have all data in an uncorrupted and update state.
- 3) We justify the performance of proposed protocol through concrete analysis, experimental results and comparison with existing schemes.

The rest of paper is organized as follows: **Section 2** describes the related works. **Sections 3** introduce the system model: including: cloud storage model, security threats, design goals and notations and permutations. In **Section 4**, we provide the detailed description of our scheme. **Section 5** gives the security analysis and **Section 6** gives the performance and experimental results and in **Section 7** we give conclusion to our work.

2. Related Work

The security of remote storage applications has been increasingly addressed in the recent years, which has resulted in various approaches to the design of storage verification primitives. The literature distinguishes two main categories of verification schemes [30]: Deterministic verification schemes check the conservation of a remote data in a single, although potentially more expensive operation and probabilistic verification schemes rely on the random checking of portions of outsourced data.

2.1. Deterministic Secure Storage

Deterministic solutions are verifying the storage of the entire data at each server. Deswart et al. [8] and Filho et al.[9] are firstly proposed a solution to remote data integrity. Both use RSA-based functions to hash the whole data file for every verification challenge. They require pre-computed results of challenges to be stored at verifier, where a challenge corresponds to the hashing of the data concatenated with a random number. However, both of them are inefficient for the large data files,

which need more time to compute and transfer their hash values. Carmoni et al. [10] described a simple deterministic approach with unlimited number of challenges is proposed, where the verifier like the server is storing the data. In this approach, the server has to send MAC of data as the response to the challenge message. The verifier sends a fresh unique random value as the key for the message authentication code to prevent the server from storing only the result of the previous hashing of the data. Golle et al. [11] proposed a SEC (Storage Enforcing Commitment) deterministic verification approach. This approach uses homomorphic verifiable tags, whose number is equal to two times of number of data chunks and the verifier choose a random value that will be used to shift indexes of tags to be associated with the data chunks when the integrity proof constructed by the server. Sebe et al. [12] presented a remote data checking protocol such that it allows an unlimited number of verifications and the maximum running time can be chosen at setup time and traded off against storage at verifier. However, none of the schemes were considered the problem of remote data confidentiality and dynamic data verifications.

To ensure the confidentiality of remote data, Shah et al. [27, 28] proposed a privacy-preserving audit protocol, which allows a third party auditor to keep online storage honest. In their schemes, the client first encrypts the data file and pre-computes a hash value over encrypted data using keyed hash function and sends it to the auditor. But, their schemes may potentially bring online burden to the users when the keyed hashes are used up. Oualha et al. [30] described a secure protocol for ensuring self organizing data storage (P2P) through periodic verifications, these verifications used for the integrity checks since each holder generates a response that they still having the data safely. In particular a data owner can prevent data damage at a specific holder by storing encrypted replicas crafted the use of elliptic curve cryptography. Wang et al. [31] proposed a privacy-preserving public auditing scheme for data storage security in cloud computing by using homomorphic authenticator and random masking. This scheme conceals the content of the original data from the TPA but not from the malicious servers. Similarly, Hao et al. [32] introduced the multiple replicas remote data possession checking protocol with public verifiability. However, this scheme does not support to dynamic data operations. In their subsequent work, Hao et al. [33] proposed a RSA-based privacy-preserving data integrity checking protocol with data dynamics in cloud computing. Their scheme extended the sebe's protocol [12] to support public verifiability. It does not leak any information to third party auditors. However, like[31] it is also not protecting data leakage from the malicious servers.

2.2. Probabilistically Secure Storage

The probabilistic verification schemes verify the specific portions of data instead of entire data at servers.

Ateniese et al. [13] proposed a RDC using PDP. In their system, the client pre-computes the tags for each block of a file using homomorphic verifiable tags and stores the file and its tags with the server. Then, the client can verify that server integrity of the file by generating a random challenge, which specifies the selected positions of file blocks. Using the queried blocks and their corresponding tags and the server generates a proof of integrity. Juels et al. [14] proposed a formal definition of POR and its security model. In this model, the encrypted data is being divided into small data blocks, which are encoded with Reed-Solomon codes. The "sentinels" are embedded among encrypted data blocks to detect whether it is intact. However, this can verify only limited number of times because this scheme has only finite number of "sentinels" in the file. When the finite "sentinels" are exhausted, the file must be sent back to the owner to re-compute new "sentinels". Ateniese et al. [15] proposed a new scheme with homomorphic linear authenticators (HLA) of which communication complexity is independent of the file length. This scheme supports unlimited number of verification, but it cannot verify publicly. Later, Shacham et al. [16] proposed the two POR protocols: The first one built from BLS signatures and has the shortest query and response with public verifiability. The second one is based on pseudorandom functions (PRFs) with private verifiability, but it requires a longer query. Both schemes rely on the homomorphic property-aggregating verification proofs into a small value. Dodis et al [17] first formally define the POR code, this construction improves the prior POR constructions. The main insight of their work comes from the simple connection between POR schemes and the notion of hardness amplification, extensively studied in complexity theory. Browers et al. [18] introduced a theoretical framework for previous POR protocols [14-16] using integrated forward error-correcting codes. In their subsequent work, Browers et al. [19] described a HAIL (High-Availability and Integrity Layer), in which the key insight is to embed MACs in the parity blocks of the dispersal code. As both MACs and parity blocks can be based on universal hash functions. Schwarz et al. [20] used a XOR-based, parity m/n erasure codes to create n shares of a file that stored at multiple sites. Curtomola et al. [21] extended the PDP [13] to the multiple servers, which are called Multiple Replica-Provable Data Possession (MR-PDP), it is aimed to ensure availability and reliability of data across distributed servers. In this scheme, the user stores multiple replicas of a single file across distributed servers, thus we can get an original file from any one of the servers even if any server fails.

Although all these schemes are aim at providing integrity verification for different data storage systems but the problem of data dynamics has not at fully addressed.

For the dynamic data integrity verification, Ateniese *et al.* [22] have designed a highly efficient and provably secure PDP with data dynamics is called “Scalable Data Possession”. It was based on symmetric key cryptography, while not requiring any bulk encryption. It improves the RDC[13] in terms of storage, bandwidth and computation overheads. However, it cannot perform block insertions anywhere because each update requires re-computing the all the remaining tokens, which is problematic for large files. In addition, it does not support public verifiability. Similarly, Wang *et al.* [23] discussed the problem of ensuring the availability and integrity of data storage in cloud computing. They utilized the homomorphic token and error correcting codes to achieve the integration of storage correctness insurance and data error localization, but like[22] their scheme do not support an efficient insert operation due to the index positions of data blocks. To overcome this probem, Erway *et al.* [24] firstly proposed a scheme to support dynamic data operations effieciently at block level instead of index positions[22, 23] by using rank-based verification skip list in the cloud servers. Later, Wang *et al.* [25] described a BLS based homomorphic authenticator with public verifiability and supports of data dynamics using Merkle Hash Tree (MHT) to verify the data integrity checking in cloud computing. They achieved the data integrity assurance with high efficiency. Similarly, Zhu *et al.* [26] proposed a dynamic auditing service for verification of integrity of outsourced data in cloud. Their design is based on fragment structure, random sampling and index-hash table. Their scheme achieved the integrity assurance with low computation, storage and computation overhead.

However, none of the schemes were address the problem of outsorced data confidentiality.

Ayad *et al.* [34] proposed a Provable Possession and Replication of Data over Cloud Servers with dynamic data support. This scheme achieves the availability, integrity and confidentiality of data storage in cloud. Chen *et al.* [35] described an efficient remote data possession in cloud computing. It has several advantages while achieving security of remote data as follows: First, it is efficient in terms of computation and communication. Second, it allows verification without the need for the challenger to compare against the original data. Third, it uses only small challenges and responses, and users need to store only two secret keys and several random numbers. Yang *et al.* [36] proposed a Provable Data Possession of Resource-constrained Mobile Devices in Cloud Computing. In this framework, the mobile terminal devices only need to generate some secret keys and random numbers with the help of trusted

platform model (TPM) chips, and the needed computing workload and storage space is fit for mobile devices. Like [25], by using bilinear signature and Merkle hash tree (MHT), this scheme aggregates the verification tokens of the data file into one small signature to reduce communication and storage burden.

All these schemes are unable to provide strong security assurance to the users because all these schemes are verifying integrity of data using pseudorandom sequence. It does not cover the whole data while computing integrity proof. Therefore, probabilistic verification schemes based on pseudorandom sequence does not give strong guarantee to the users about security of their data.

To overcome this problem, Syam *et al.* [27] proposed a homomropic distributed verification protocol to ensure data storage security in cloud computing using Sobol Sequence instead of pseudorandom sequence, which is more uniform than pseudorandom sequence. Their scheme achieves the availability and integrity of outsourced data in cloud but similar [23], it is also not addressing data confidentiality issue.

To achieve all these security and performance requirements of cloud storage, we propose an efficient and secure protocol in section 4.

3. System Model

3.1. Cloud Data Storage Model

The cloud storage model considering here is consists of three main components as illustrated in Fig. 1.

- 1) **Cloud User:** the user, who can be an individual or an organization originally storing their data in cloud and accessing the data.
- 2) **Cloud Service Provider (CSP):** the CSP, who manages cloud servers (CSs) and provides a paid storage space on its infrastructure to users as a service.
- 3) **Third Party Auditor (TPA) or Verifier:** the TPA or Verifier, who has expertise and capabilities that users may not have and verifies the integrity of outsourced data in cloud on behalf of users. Based on the audit result, the TPA could release an audit report to user.

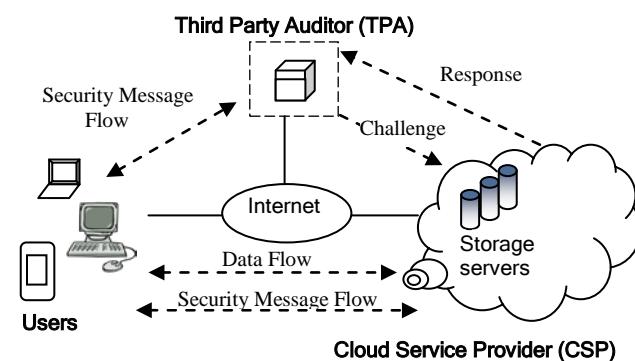


Fig.1. Cloud Data Storage Model

Throughout this paper, terms verifier or TPA and cloud server or CSP are used interchangeably

In cloud data storage model, the user stores his data in cloud through cloud service provider and if he wants to access the data back, sends a request to the CSP and receives the original data. If data is in encrypted form that can be decrypted using his secret key. However, the data is stored in cloud is vulnerable to malicious attacks; it would bring irretrievable losses to the users, since their data is stored at an untrusted storage servers. It doesn't matter that whether data is encrypted or not before storing in cloud and no matter what trust relations the client and the server may have a priori share. The existing security mechanisms need to reevaluate. Thus, it is always desirable to need an efficient and secure method for users to verify that whether data is intact? If user does not have the time, he assigns this task to third party auditor. The auditor verifies the integrity of data on behalf of users.

3.2. Security Threats

In this paper, we are considering two types of attacks for cloud data storage those are: Internal Attacks and External Attacks.

3.2.1. Internal Attacks: These are initiated by malicious Cloud Service Provider (CSP) or malicious users. Those are intentionally corrupting the user's data inside the cloud by modifying or deleting. They are also able obtain all the information and may leaked it to outsiders.

3.2.2. External Attacks: these are initiated by unauthorized parties from outside the cloud. The external attacker, who is capable of comprising cloud servers and can access the user's data as long as they are internally consistent i.e. he may delete or modify the customer's data and may leaked the user private information.

3.3. Design Goals

We have designed an efficient and secure storage protocol to ensure the following goals. These goals are classified into two categories: Efficiency and Security Goals.

3.3.1. Efficiency

The following efficiency requirements ought to be satisfied for a proposed scheme of practical use of cloud storage:

Low computation overhead: It includes the initialization and verification overheads of the verifier and the proof generating overheads of the server. It means that the proposed scheme should be efficient in terms of computation.

Less communication overhead: It refers to the total data transferred between the verifier and server. It means that the amount of communication should be low.

Low storage cost: It refers to the additional storage of client and server required by the scheme. It means that the additional storage should be low as possible.

3.3.2. Security

In this paper, we are considering two security requirements, which are needs to be satisfied for the security of proposed scheme:

Confidentiality: Confidentiality refers to only authorized parties or systems having the ability to access protected data.

Integrity: Data Integrity refers to the protection of data from unauthorized deletion, modification or fabrication. Further, detects any modifications to data stored in cloud.

3.4. Notations and Permutations

- F - the data file to be stored in cloud, the file F is divided into n blocks of equal length: m_1, m_2, \dots, m_n , where $n = [|m|/l]$.
- $f_{key}(\cdot)$ - Sobol Random Function (SRF) indexed on some key, which is defined as $f: \{0,1\}^* \times \text{key} - \{0,1\}^{\log_2 n}$.
- $\pi_{key}(\cdot)$ - Sobol Random Permutation (SRP) which is defined as $\pi: \{0,1\}^{\log_2(l)} \times \text{key} - \{0,1\}^{\log_2(l)}$.

Elliptic Curve Cryptography over ring Z_n :

Let n be an integer and let a, b be two integers in Z_n such that $\gcd(4a^3 + 27b^2, n) = 1$. An elliptic curve $E_n(a, b)$ over the ring Z_n is the set of points $(x, y) \in Z_n \times Z_n$ satisfying the equation: $y^2 \equiv ax + b$, together with the point at infinity denoted as O_n .

4. Efficient and Secure Storage Protocol

To ensure the confidentiality and integrity of data stored in cloud, we propose an Efficient and Secure protocol. Our scheme is designed under the Elliptic Curve Cryptography [30, 38] construction and use of Sobol sequence to verify the integrity of storage data randomly. This protocol consists of three phases, namely Setup, Verification and Dynamic Data Operations and Verification. The three process model is depicted in fig.2. The construction of these phases is presented briefly as follows:

4.1. Setup

In this phase, the user pre-processes the file before storing in cloud. The Setup phase consists of three algorithms, those are: 1) KeyGen
2) Encryption 3) MetadataGen.

4.1.1. KeyGen

In this algorithm, the user generates private key and public key pair using **algorithm 1**, it takes k as input and generates private key and public key pair as output as follows: the given security parameter k ($k > 512$), user chooses two large primes p and q of size k such that $p \equiv q \equiv 2 \pmod{3}$. Then compute

$$n = pq \quad (1)$$

and

$$N_n = \text{lcm}(p+1, q+1). \quad (2)$$

where N_n is a order of elliptic curve over the ring Z_n denoted by $E_n(0, b)$, and b is a randomly chosen integer such that $\text{gcd}(b, n) = 1$ and compute P is a generator of $E_n(0, b)$. It outputs public key $PK = \{b, n, p\}$ and private key $PR = \{N_n\}$.

Algorithm 1: KeyGen

1. **Procedure:** KeyGen($k \leftarrow \{PK, PR\}$)
 2. Take security parameter k ($k > 512$)
 3. Choose two random primes p and q of size k :
 $p \equiv q \equiv 2 \pmod{3}$
 4. Compute $n = pq$
 5. Compute $N_n = \text{lcm}(p+1, q+1)$
 6. Generate random integer $b < n$, $\text{gcd}(b, n) = 1$
 7. Compute P , is a generator of $E_n(0, b)$
 8. Private key $PR = \{N_n\}$
 9. Public key $PK = \{n, b, P\}$
 10. **end procedure**
-

4.1.2. Encryption

To ensure the confidentiality of data, the user encrypts the each data block m_i in the file F using **algorithm 2**, it takes m_i , keyed Sobol Random Function(SRF) and secrete random parameter s as inputs and produce m'_i as output as follows:

$$F = \{m_1, m_2, \dots, m_n\} = \{m_i\}_{1 \leq i \leq n} \quad (3)$$

$$F' = m'_i = m_i + f_k(s) \quad (4)$$

where s is random of size l .

Algorithm 2: Encryption

1. **Procedure :** Encryption($m_i, s \leftarrow m'_i$)
 2. **for** i to n
 3. Compute $m'_i = m_i + f_k(s)$
 4. **end for**
 5. **end procedure**
-

4.1.3. MetadataGen:

After encrypting the data, the user computes a metadata over encrypted data to verify the integrity of data using **algorithm 3**, which takes m'_i , public key and private key as inputs and produce metadata T_i as output:

$$T_i \leftarrow m'_i P(\text{mod } N_n) \quad (5)$$

where $P \in E_n(0, b)$

Algorithm 3: MetadataGen

1. **Procedure:** MetadataGen($m'_i, n, b, P \leftarrow T_i$)
 2. **for** i to n
 3. Compute $T_i \leftarrow m'_i P(\text{mod } N_n)$
 4. **end for**
 5. **end procedure**
-

After computation of metadata, the user sends metadata, public key to the TPA for later verification and sends file F' to cloud servers for storage.

4.2. Verification Phase

Once data has stored in cloud, in order to ensure the integrity of data, our scheme entirely relies on verification phase. To verify the integrity of data, the verifier first creates a challenge and sends to the server. Upon receiving a challenge from the verifier, the server computes a response as integrity proof and return to the verifier. It consists of three algorithms: 1) Challenge, 2) ProofGen 3) CheckProof .

4.2. 1. Challenge

The verifier creates a challenge by running **algorithm 4**, it takes k_{SRF}, j , and Q as input and return $chal$ as output as follows: the verifier chooses a random keys k_{SRF} and k_{SRP} using Sobol sequence and computes random indices $1 \leq j \leq n$ ($j = 1, \dots, c$) of the set $[1, n]$, where $c = \pi_{k_{SRP}}(c)$ (6) which prevents the server from anticipating which blocks will be queried in each challenge. The verifier also generates a fresh random value r to guarantee that the server does not reuse any values from the previous challenge and computes

$$Q = rP. \quad (7)$$

Then, verifier creates the challenge $chal = \{k_{SRF}, j, Q\}$, and sends to the server.

Algorithm 4: Challenge

1. **Procedure:** Challenge($k_{SRF}, j, Q \leftarrow chal$)
 2. Generates a random keys k_{SRF} , k_{SRP} and fresh random value using Sobol Sequence.
 3. Compute $c = \pi_{k_{SRP}}(c)$
 4. Compute $Q = rP \in E_n(0, b)$
 5. Create challenge $chal = \{k_{SRF}, j, Q\}$
 6. **end procedure**
-

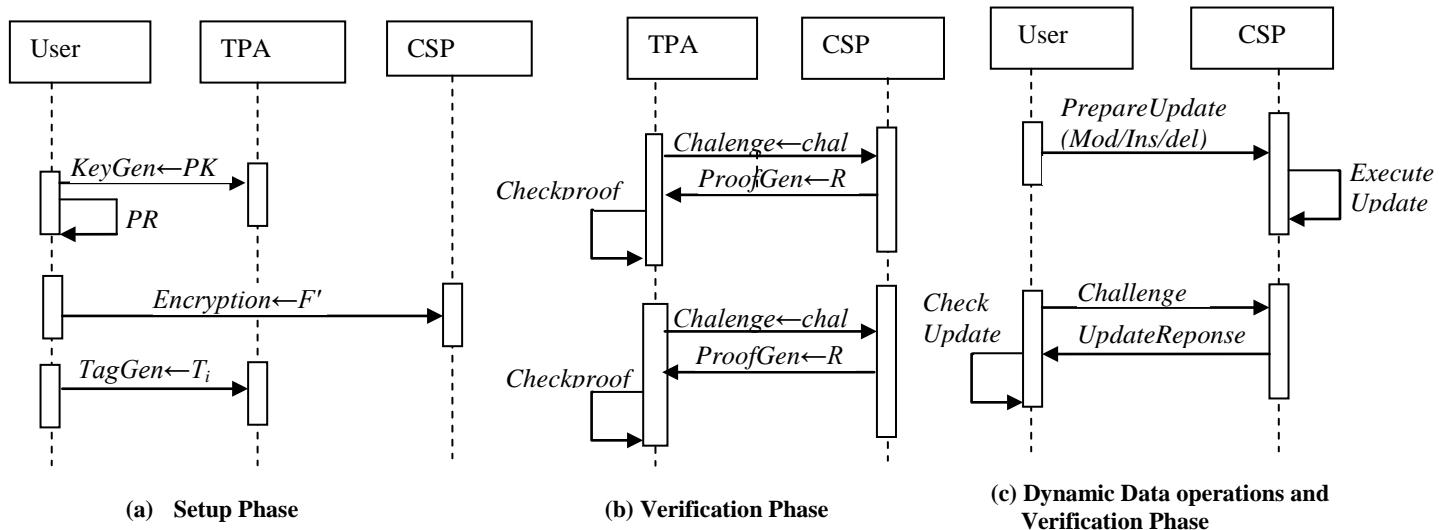


Fig. 2. Efficient and Secure Storage Processing Model

4.2.2. ProofGen

Upon receiving a challenge from the verifier, each server computes a response as integrity proof using **algorithm 5**, it takes encrypted data m'_i , challenge $chal$ as inputs and produce response R as output as follows: first, it generates random numbers using Sobol random Function (SRF) i.e.

$$a_j = f_{k_{SRF}}(j) \quad (8)$$

$$\text{Then compute } b = \sum_{j=1}^c a_j m'_{i_j} \quad (9) \quad \text{where } 1 \leq i_j \leq n$$

$$\text{Later, computes a response } R = bQ \bmod n \quad (10)$$

$$\begin{aligned} &= \sum_{j=1}^c a_j m'_{i_j} Q \bmod n \\ &= \sum_{j=1}^c a_j m'_{i_j} rP \bmod n \\ &= r \left(\sum_{j=1}^c a_j m'_{i_j} P \bmod n \right) \end{aligned}$$

Algorithm 5: ProofGen

1. **Procedure:** $\text{ProofGen}(m'_i, k_{SRF}, Q) \leftarrow R$
 2. Generates a n random numbers using k_{SRF}
 3. **for** 1 to n
 4. Generate $a_j = f_{k_{SRF}}(j)$
 5. **end for**
 6. compute $b = \sum_{j=1}^c a_j m'_{i_j}$
 7. compute $R = bQ \bmod n$
 8. **end procedure**
-

4.2.3. CheckProof

After receiving a response from the server, the verifier checks the integrity using **algorithm 6**, it takes public key pk , challenge query $chal$, and proof R as inputs and return output as 1 if the integrity of file is verified as successfully or 0 as follows: the verifier re-generates random numbers using Sobol Random function i.e.

$$a_j = f_{k_{SRF}}(j)$$

$$\text{Then compute } S = \prod_{j=1}^c a_j T'_{i_j} \bmod n \quad (11)$$

$$R' = rS \bmod n \quad (12)$$

Now, verifier checks whether

$$R' = R,$$

if response is valid, then it returns 1 otherwise 0.

Algorithm 6: CheckProof

1. **Procedure:** $\text{CheckProof}(T'_i, r, k_{SRF}, n) \leftarrow R'$
 2. Generates a n random numbers using key k_{SRF}
 3. **for** 1 to n
 4. Generate $a_j = f_{k_{SRF}}(j)$
 5. **end for**
 6. compute $S = \prod_{j=1}^c a_j T'_{i_j} \bmod n$
 7. compute $R' = rS \bmod n$
 8. verify if $(R' = R)$
 9. **return true**
 10. **else**
 11. **return false**
 12. **end if**
 13. **end procedure**
-

4.4. Dynamic Data Operations

The proposed scheme also supports dynamic data operations at block level [33] while maintaining same security assurance, such as Block Modification (BM), Block Insertion (BI) and Block Deletion (BD). These operations are performed by the server based on the user request in the general form (BlockOP, j, m_i), where BlockOp indicates the block operation such as BM, BI and BD. The parameter *j* indicates the particular block to be updated and m_i^{*} is the new block.

In order to update data in cloud, the user creates a request and sends to the server. Upon receiving an update request from the user, the server performs the particular update operation (modification/insert/delete).

Here, we show that how our scheme supports dynamic data operations efficiently:

Algorithm 7: PrepareUpdate

1. **Procedure:** PrepareUpdate \leftarrow (BM/BI/BD, j, m_i)
 2. Select a update block m_j
 3. **if**(update==modification/insert)
 4. Encrypt $m'_j \leftarrow m_j + f_k(s)$
 5. Compute $T_j \leftarrow m'_j P \bmod N_n$
 6. Update=(BM/BI, j, m_i)
 7. **else if**(update==deletion)
 8. Update =((BD, j))
 9. Send update request to the server
 10. **end if**
 11. **end procedure**
-

4.4.1. Block Modification (BM):

Data modification is one of the frequently used operations in cloud data storage. Suppose, the user wants to modify the block m_j with m_i^{*}, then the user runs the **algorithm 7** to do the following:

- 1) Create a new block m_j
- 2) Encrypt the new block using equation (2)

$$m'_j \leftarrow m_j + f_k(s) \quad (14)$$
- 3) Compute new metadata using equation

$$T_j \leftarrow m'_j P \bmod N_n \quad (15)$$
- 4) Create update request (BM, j, m_i) and sends to the server.
- 5) The Metadata sends to TPA for later verification

Upon receiving an update request, the server replace the block m_i with m_j^{*} and construct update version of the file F" by running **algorithm 8**.

Algorithm 8: ExecuteUpdate

1. **Procedure:** ExecuteUpdate \leftarrow {F"}
 2. **if**(update==modification)
 3. replace m_i with m_j^{*} in the file F"
 4. update file F"
 5. **else if**(update==insert)
 6. insert m_x^{*} before m_i or append
 7. **else if**(update==deletion)
 8. delete m_i from file F"
 9. update the file F"
 10. move all blocks backward after ith block
 11. **end if**
 12. **end procedure**
-

4.4.2. Block Insertion (BI)

In this operation, the user wants to insert a new block m^{*} after position j in the file F'= {m₁', ..., m_n'}. The block insertion operation changes the logical structure of the file; the proposed scheme can perform the block insertion operation without re-computing metadata of all blocks that have been shifted after inserting a block, because block index is not included in the metadata. To perform an insertion of a new block m^{*} after position j in a file, the user runs **algorithm 7** to do the following:

1. Create a new block m_j^{*}
2. Encrypt the new block

$$m'_j \leftarrow m^*_j + f_k(s) \quad (16)$$
3. Compute new metadata

$$T^*_j \leftarrow m'_j P \bmod N_n \quad (17)$$
4. Create update request (BI, j, m_i) and sends to the server.
5. The Metadata sends to TPA for later verification

Upon receiving the update request, the server replace the block m_j^{*} with m_j and construct update version of the file F" by run the **algorithm 8**.

4.4.3. Block Deletion (BD)

The Block deletion operation is the opposite of insertion operation. When one block is deleted, all subsequent blocks are moved one step forward. Suppose, the user wants to delete a specific data block at position j from the file F', creates a delete request (BD, j), sends to the server and also sends request to the TPA to delete corresponding block metadata. Upon receiving a delete request from the user, the server deletes the block m_j from the file and constructs update version of the file F". Similarly, the TPA deletes corresponding metadata. Here, deletion of metadata do not depends on other block metadata. The detail of delete operation is given in **algorithm 8**.

4.4.4. Verification

To ensure the security of dynamic data operations, the user verifies the integrity of updated block immediately after updating as follows:

- 1) The user challenges the server immediately for the proof of update operation i.e.

$$Q = rP \quad (18)$$

- 2) Upon receiving a request from the user, the server computes a response for updated block and returns to the user:

$$R_j \leftarrow m'_j P \bmod n \quad (19)$$

- 3) After receiving an update response from the server, the user verifies whether response is matched with metadata of particular block by running **algorithm 9**, if it returns true, server has been updated data successfully otherwise not.

Algorithm 9 : VerifyUpdate

```

1. Procedure: VerifyUpdate(pk, Q, R') → {1,0}
2.   if(update==modification/insert)
3.     if(Tj=Rj)
4.       return 1
5.     else
6.       return 0
7.     end if
8.   else if(update==deletion)
9.     verification directly starts from static case
10.  end if
11. end procedure

```

In next section, we analyze the security of our scheme.

5. Security Analysis

In this section, we present the formal security analysis of the proposed scheme. That means integrity and confidentiality of data stored in cloud.

5.1. Integrity

To ensure the integrity, we need three properties: Completeness, Soundness and Probability Detection. Here, we define these terms as follows: for completeness, soundness [30] and Probability Detection [24]

Completeness: After receiving a challenge from the verifier, if server honestly computes a correct integrity proof, the verifier always accepts the proof as valid.

Soundness: After receiving a challenge from the verifier, the server dishonestly computes the integrity proof by missing some data bits, the verifier accepts with negligible probability.

Probability Detection: After receiving a response from the server, the verifier check whether response is valid or not? If it is not valid, then the verifier detects the corruptions with high probability.

In our integrity analysis, we have depended on the Finding order of elliptic curve and Elliptic curve discrete logarithm problem denoted by ELDL problems.

1) Finding the order of elliptic curves:

The order of elliptic curve over the ring Z_n is: let $n=pq$ is defined in [38,] as $N_n = \text{lcm}(\#E_p(a, b), \#E_q(a, b))$. N_n is the order of the curve, i.e. for any $P \in E_n(a, b)$ and any integer k , such that

$$(k N_n + 1)P = P. \quad (20)$$

If $(a=0 \text{ and } p \equiv q \equiv 2 \pmod{3})$ or $(b=0 \text{ and } p \equiv q \equiv 3 \pmod{4})$, the order of $E_n(a, b)$ is equal to N_n . The given $N_n = \text{lcm}(\#E_p(a, b), \#E_q(a, b)) = \text{lcm}(p+1, q+1)$

$$(21)$$

Solving N_n is computationally equitant to factoring the corresponding number n .

2) Elliptic Curve Discrete Logarithm Problem(ECDLP)

Consider the equation $Q = rP$ where $Q, P \in E_n(a, b)$ and $r < n$. It is relatively hard to determine r given Q and P .

Theorem 1. The proposed protocol is complete

Proof: Here, we are proving this theorem according to the definition of sound and commutative property of point multiplication in an elliptic curve [30].

we have $R' = R$

$$R' = rS \bmod n$$

$$S = \prod_{j=1}^c a_j T_{i_j} \bmod n \text{ where } a_j = f_k(j)$$

$$= \prod_{j=1}^c (a_j m'_{i_j} P \bmod N_n) \bmod n$$

$$= \sum_{j=1}^c a_j m'_{i_j} P \bmod n$$

$$R' = rS \bmod n$$

$$= r \left(\prod_{j=1}^c (a_j m'_{i_j} P \bmod n) \right)$$

$$= r \left(\sum_{j=1}^c a_j m'_{i_j} P \bmod n \right)$$

$$= R$$

From the equation (13), the protocol is **complete or valid**. Then the verifier is “probabilistically” assured that server still holds data safely. In reality, verifier only verifies that server holds the j [1, c] selective blocks where j is chosen randomly.

Theorem 2: The proposed protocol is sound

Proof: In this proof, we show that our protocol is sound against dishonest server based on previous transactions and pre-computed metadata. There are four

possibilities that the server can compute the integrity proof without storing the user's data:

- 1) The server guessed or use pre-computed value. However, guessing occurs with negligible probability and pre-computing the correct response is not possible because each time the verifier challenge the server with a fresh challenge.
- 2) Other option is to cheat user, the server replayed the previous response. In this case, the server would have to find r from challenge $chal$ to compute the correct proof. since r is chosen randomly, finding r is hard based on ELDL problem.
- 3) Another option for the server to cheat user, he has an algorithm to compute $m'_i \bmod N_n$ with inputs instead of storing $m'_i [1 \leq i \leq n]$. But this option is not possible, because, the server cannot compute N_n based on the hardness of solving the order of elliptic curve $E_n(0, b)$ as we discussed above.
- 4) Last option for server is, if the server does not store the data $\{m'_i\}$ and it may try to collude with the other servers for storing the same data. However, this option is not feasible, since data stored at each server is securely encrypted using Sobol Random Function (SRF). The f is a keyed one-way function and s is a secrete parameter, so, no one except the user can retrieve the original data m_i from m'_i .

All these options lead to contradiction; so the server cannot compute response without storing the data. Hence, our protocol is complete.

Probability Detection

Here, we investigate how the probabilistic nature of the proposed protocol makes it possible to enforce integrity. We are making the following assumptions:

- The verifier's random selection of indexes is uniform, i.e., for n blocks, the probability to pick any block is $1/n$.
- If an attacker removes a portion d/n of data from the storage file, this portion is referred to as the corruption of the file.
- The verifier performs on average c challenges: $1 \leq c \leq n$ (n is the number of blocks) and detects the corruption with high probability.

The detection probability P_d of disrupted blocks is an important parameter to guarantee that these blocks can be modified or detected in a time. We have detection of probability is:

$$P_d = 1 - \left(1 - \frac{d}{n}\right)^c \quad (22)$$

For a given probability of detection of a data corruption, it is possible to probabilistically determine the average number of challenges that the verifier should perform to achieve the probability of detection. The number of challenges c can be derived as follows[30]:

$$c = \log_{1 - \frac{d}{n}} (1 - P_d) \quad (23)$$

Fig. 3 plots P_d for different values of n , c , d . To understand the importance of Fig. 3, suppose that if a fraction of the data file is corrupted, the Sobol' sequence achieves detection with high probability in a few number of blocks to challenge, while pseudorandom data requires more blocks and even sometimes it may not detect the corruptions, since it do not covers the whole data in the file while verifying the data integrity. For example if $d=1\%n$ (data corruption) is corrupted, the proposed scheme using sobol sequence detects corruption with 99% in $4\%n$ blocks whereas existing probabilistic checking methods using pseudorandom sequence requires $10\%n$ blocks and sometimes these blocks may not detect the corruption. Since, sobol sequence is more uniform than pseudo random sequence.

Therefore, the proposed method is more secure and efficient than existing probabilistic remote data checking methods.

Monte-Carlo Results.

Now, we turn to Monte-Carlo simulation to determine uniformity of random sequences. For the goodness of random numbers, we calculated the Monte Carlo integration using random numbers. The integration of a function $f(x)$ in the s -dimensional unit cube I^s . we are in fact calculating the average of the function at a set of randomly sample points. Where there are N sample points in the integral is:

$$V = \frac{1}{N} \sum_{i=1}^N f(x') \quad (24)$$

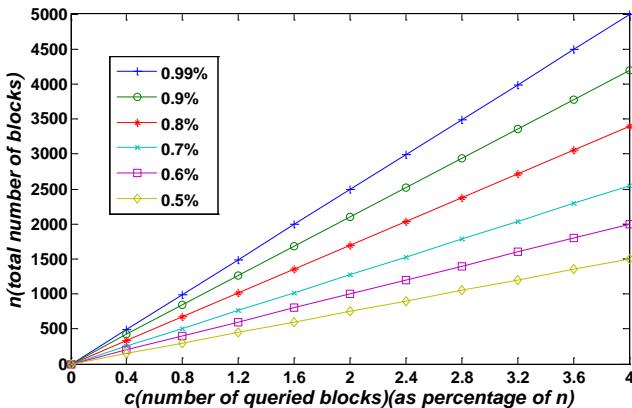
Where V is used to denote the approximation to the integral and x^1, x^1, \dots, x^N are the N , s -dimensional sample points. The Monte Carlo integration of V sampling in the region $-1 < x' < 1$ for the two cases: uncorrelated random numbers (pseudorandom sequence) and Sobol sequence. If pseudo-random sequence is used, the points x' will be independently and identically distributed, the estimate the expected error of integral is $N^{-1/2}$, while sobol sequence is used, whose fractional error decreases of N^{-1} . In Figure 4 we presented for calculation of six dimensional integral is:

$$I = \iiint \prod_{i=1}^6 (i \cos(ix_i)) dx_1 dx_2 dx_3 dx_4 dx_5 dx_6 \quad (25)$$

The exact value of integral is:

$$I = \prod_{i=1}^6 \sin(i) \quad (26)$$

Fig.4 shows that the pseudorandom sequence gives worst performance, whilst Sobol Sequence gives rapid convergence to the solution. To conclude that it has been shown that Sobol sequence can evaluate integrals more efficient than pseudorandom sequences.



(a) $z=1\%l$ using Sobol Sequence

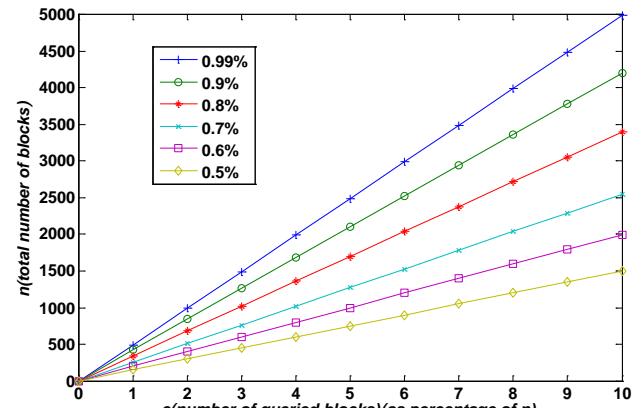


Fig. 3: The detection probability p_d against data corruption. We show p_d as a function of l (total number of rows) and r (the number of rows queried by the user, shown as percentage of l) for value of z (the number of rows modified by the adversary).

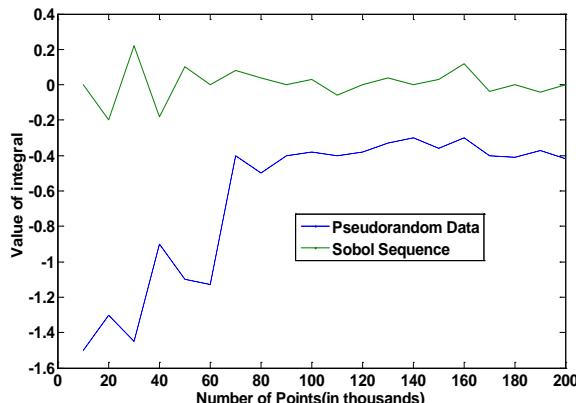


Fig. 4. Monte Carlo simulation using random numbers

5.2. Confidentiality

Now, we analyze the confidentiality of our scheme: The stored data in cloud cannot be leaked to an malicious attackers (servers and TPA). In this analysis, we depend on the hardness of the Elliptic Curve Diffie-Hellman (ECDH) and Elliptic Curve Discrete Logarithm (ECDL) problems.

Theorem 3: *The proposed protocol is confidential against data leakage to attacker.*

We prove this theorem under different attacks:

- 1) The secret parameter s cannot be derived by a malicious user eavesdropping on the communication link between the user and server because of Elliptic Curve Diffie-Hellman (ECDH) problem. The public parameter $\{b, n, P\}$ cannot help the adversary to infer or calculate any useful information that can reveal the shared key between the user and server.
- 2) Suppose, If the malicious server wants to access the data from the encrypted file $F'=m_i'$. But it is not possible, because in order to access the encrypted data, he should need a secrete parameter, this secrete key chosen by user randomly. If server try to get the

secret key by using different combinations of public parameters but fail to do so due to the ECDL problem. Hence, the server cannot learn anything from F' .

3) The TPA has $T_i \leftarrow m'_i P \pmod{N_n}$. If he tries to access data content from metadata, the user computes metadata over encrypted the data using secrete key. However, it is not possible because the secrete parameter chosen by the user from random. So there is no chance to TPA get secrete parameter using public key and metadata. Hence, The TPA cannot learn anything from metadata T_i .

Therefore on the basis of ECDH and ECDL problems, our protocol is confidential against data leakage.

6. Performance Analysis and Experimental Results

6.1 Performance Analysis

In this section, we analyze the performance of our scheme in terms of storage, communication and computation complexity.

Storage cost:

Here, we detail the storage cost required by the client, TPA and server.

User Side: The user needs to store the only secrete parameter. The storage cost for that is $O(1)$.

Server Side: the server needs to be store the complete file, the cost for storage file is $O(n)$ bits.

TPA or Verifier: the verifier needs to store metadata and public key. The metadata is a relatively smaller than original file, so storage cost for metadata is $O(1)$.

Communication Cost:

Here, we consider the communication cost between the server and verifier during verification phase. The challenge sent by the verifier to the server, which consists of $O(1)$ and the response(it is a small size compare to original file) sent by server to the verifier, which consists of $O(1)$. Thus, total communication cost is $O(1)$.

Computation Cost:

We analyze the computation cost of the user, verifier and server as follows:

User: during the setup phase, the user generates a private key and public key whose cost is $O(1)$. Then, to encrypt a file, the user needs to perform integer addition, its cost is $O(n)$. Finally, computes the metadata by performing n-bit point multiplications whose cost is $O(1)$. Hence, total computation cost of the user is: $O(1)$.

Verifier: During the verification phase, the TPA or verifier needs to generate three random numbers $\{k_{SRF}, r\}$, then compute $c = \pi_{k_{SRP}}(c)$ and $Q = rP$, whose cost is $O(1)$. Again, after receiving the response, the verifier re-generates $\{a_j\}_{j=1}^c$, the computation cost of each $a_j m'_{i_j}$ corresponds to the sum of point multiplication of two bits. Finally, the verifier computes R' , the cost of R' is a two point multiplications plus sum of 2 bit integer plus generating random numbers cost, which is $O(1)$ respectively. Hence, the total computation cost at verifier side is $O(1)$.

Server Side: During the verification phase, the server needs to generate n-Sobolrandom b-bit integers a_i , then

$$\text{it computes } b = \sum_{j=1}^c a_j m'_{i_j} \quad R$$

$= r \sum_{i=1}^n a_j m'_{i_j} P \bmod n$ The computation of each $a_j m'_{i_j}$ corresponds to the sum of point multiplication of two bits. The computation cost of $a_j m'_{i_j}$ is $O(1)$. Next,

the server computes a proof, which consists of point multiplications in **ProofGen** algorithm, its cost is $O(1)$. The total computation cost of server for generating integrity proof (response) is $O(1)$.

In table 1, we summarized the storage, communication and computation costs.

Table 1: Summary of Storage, Communication and Computation cost of Proposed Protocol

Storage Cost			Communication Cost		Computation Cost	
Verifier	Server	Verifier	Server	User	Verifier	Server
O(1)	O(n)	O(1)	O(1)	O(1)	O(1)	O(1)

6.2. Experimental Results

In this section, we present the experimental results of our protocol. All experiments conducted using C++ on system with dual core 2-GHZ processor and 4GB RAM running Windows 2007. In our implementation, we use MIRACL library version 5.4.2 to achieve better security work on elliptic curve with 160-bit group order instead of RSA on 1024 bits. Here, we are measuring total

time for computation cost of the verifier and server using ECC and RSA respectively.

$$\text{Speedup} = \frac{RSA - ECC}{RSA} * 100 \quad (27)$$

Then, we compare computation cost of our protocol with RSA-based remote data checking protocols, which includes the verifier, server and user computation costs and presented results in table 2, 3 & 4.

Table 2: Computation Cost at Verifier using RSA [13] and ECC based schemes.

File Size	Verifier side using RSA[33]	Verifier Side using ECC	Speedup
10MB	424.37 ms	316.26 ms	25%
20MB	482.81 ms	342.43 ms	29%
30MB	561.62 ms	376.03 ms	32%
40MB	641.46 ms	415.09 ms	35%
50MB	743.64 ms	465.13 ms	38%

Table 2 shows that the total computation cost of verifier for our proposed scheme is faster than existing RSA based scheme [33].

Table 3: Computation Cost at Server with RSA based scheme and ECC scheme

l(bits)	Server Side with RSA[33]	Server Side with ECC	Speedup(%)
10MB	388.01 ms	275.11 ms	29%
20MB	447.62 ms	312.43 ms	30%
30MB	508.39 ms	348.21 ms	31%
40MB	562.67 ms	381.21 ms	32%
50MB	625.16 ms	418.76 ms	33%

Table 3 shows that the total computation cost of the server for proposed scheme is faster than existing RSA based scheme [33].

Table 4: Metadata Computation Cost at user with RSA and ECC based schemes

l(bits)	Server Side with RSA[33]	Server Side with ECC	SpeedUp(%)
10MB	244.11 ms	183.06 ms	25%
20MB	296.41 ms	218.32 ms	26%
30MB	352.53 ms	253.38 ms	28%
40MB	403.17 ms	289.63 ms	29%
50MB	467.26 ms	323.92 ms	30%

Table 4 shows that the total computation cost of metadata at user side in our scheme is faster than existing RSA based scheme [33].

6.3. Comparison with Existing Schemes

We compared our scheme with existing RSA based verification schemes and probabilistic verifications schemes

Most of the schemes that use RSA based verification but the key length for secure RSA use as increased over recent years and this put a heavier processing burden on applications using RSA. To avoid this problem, we proposed an ECC based verification scheme. The principal of ECC compared to RSA is that it

appear to offer equal security for a far smaller key size,

thereby it reduced the computation overhead.

Table 5: Comparisons between Proposed Protocol and selective Existing Protocols

	RDC [13]	C.wang [21]	Q.wang [23]	Yan[24]	C.wang [30]	Hao[32]	Syam[27]	Proposed protocol
Type of Guaranty	Prob	Prob	Prob	Prob	Prob	Deter	Prob	Prob
Integrity	Partial	Partial	Partial	Partial	Partial	Yes	Yes	Yes
Confidentiality	no	no	no	no	Partial	Partial	no	Yes
Public Verifiability	no	Yes	Yes	Yes	Yes	Yes	no	Yes
Data Dynamics	no	Partial	Yes	Yes	Yes	Yes	Partial	Yes
Communication complexity	O(1)	O(clogn)	O(logn)	O(s)	O(logn)	O(1)	O(1)	O(1)
Server Computation	O(1)	O(clogn)	O(logn)	O(c+s)	O(nlogn)	O(n)	O(clogn)	O(1)
Verifier computation	O(1)	O(clogn)	O(logn)	O(c+s)	O(logn)	O(n)	O(clogn)	O(1)
Probability Detection	$O(N^{-1/2})$	$O(N^{-1/2})$	$O(N^{-1/2})$	$O(N^{-1/2})$	$O(N^{-1/2})$	$O(N^{-1/2})$	$O(N^{-1})$	$O(N^{-1})$

Prob: Probabilistic **Deter:** Deterministic

Next, we compare our scheme with probabilistic verification schemes. These schemes verify the integrity of outsourced data based on pseudorandom sequence but they do not provide satisfactory integrity assurance to the users i.e. sometimes they may not detect the data corruptions in cloud. Because, pseudorandom sequence is not uniform (uncorrelated random numbers), and it will take more time to detect data corruption, so its time consuming whereas proposed protocol verifies the integrity of the data using Sobol sequence. Our scheme should detect all data corruptions with less number of blocks since sobol sequence covers the entire data in the file more uniformly than pseudorandom sequence.

Finally, the proposed protocol is private against unauthorized data leakage because, we are encrypting the data before storing in cloud. In Table 5, we summarize the comparison between the selective existing protocols and proposed protocol.

7. Conclusion

In this paper, we have studied the problem of Integrity and Confidentiality of data storage in cloud computing and proposed an efficient and secure protocol using ECC and Sobol sequence. The proposed method is mainly suitable for thin users who have less resources and limited computing capability. It satisfies the all security and performance requirements of cloud data storage. Our method also supports public verifiability that enables TPA to verify the integrity of data without retrieving original data from the server and probability detects data corruptions. Moreover, our scheme also supports dynamic data operations, which performed by the user on data stored in cloud while maintaining same security assurance. We have proved that proposed scheme is secure in terms of integrity and confidentiality through security analysis. Through, performance analysis and experimental results proved that proposed scheme is efficient. Compared with

previously proposed protocols, we have also proved that proposed scheme is more secure and efficient.

Acknowledgments

The first author would like to acknowledge Prof. J.A.K. Tareen, vice-chancellor for availing good research environment and providing scholarship. The First author also would like to thank HRD ministry of India for providing the UGC-RGNF Fellowship.

References

- [1] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic. "Cloud Computing and Emerging IT Platforms: Vision, Hype, and Reality for Delivering Computing as the 5th Utility," Future Generation Computer Systems, vol. 25, no. 6, June 2009, pp 599–616.
- [2] Amazon.com, "Amazon Web Services (AWS)," Online at <http://aws.amazon.com>, 2008.
- [3] Apple "iCloud" Online at <http://www.apple.com/icloud/what-is.html> 2010.
- [4] T Mather, S Kumaraswamy, and S Latif "Cloud Security and Privacy", O'REILLY Publication, first edition, sep-2009.
- [5] H. Takabi, J.B.D. Joshi, and G. Ahn, "Security and Privacy Challenges in Cloud Computing Environments", Article in IEEE Security and Privacy, vol. 8, no.6, Nov-Dec. 2010, pp. 24-31.
- [6] N. Gohring, "Amazon's S3 down for several hours," Online at http://www.pcworld.com/businesscenter/article/142549/amasons_down_for_sever_hours.html, 2008 .
- [7] J. Kincaid, "MediaMax/TheLinkup Closes Its Doors",Online at <http://www.techcrunch.com/2008/07/10/mediamaxthelinkup-closes-its-doors/>, July 2008.
- [8] Y. Deswarte, J.-J. Quisquater, and A. Saidane. "Remote integrity checking". In Proc. of Conference on Integrity and Internal Control in Information Systems (IICIS'03), November 2003. lausanne, Switzerland.
- [9] D. L. G. Filho and P. S. L. M. Barreto, "Demonstrating Data Possession and Uncheatable Data Transfer," cryptology ePrint Archive, Report 2006/150, 2006, <http://eprint.iacr.org/>.
- [10] G. Caronni and M. Waldvogel, "Establishing Trust in Distributed Storage Providers", In Third IEEE P2P Conference, Linkoping 03, 2003.

- [11] P. Golle, S. Jarecki and I. Mironov, "Cryptographic Primitives Enforcing Communication and Storage Complexity", In proc. of Financial Crypto 2002. Southampton, Bermuda.
- [12] F. Sebe, J. Domingo-Ferrer, and A. Martinez-Balleste, Y. Deswarte, and J.-J. Quisquater, "Efficient Remote Data Possession Checking in Critical Information Infrastructures", IEEE Trans. Knowledge and Data Engineering, vol. 20, no. 8, aug-2008, pp. 1034-1038
- [13] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, "Remote Data Checking using Provable Data Possession," ACM Trans. ACM Transactions on Information and System Security, Vol. 14, No. 1, Article 12, may 2011, pp. 12.1-12.34.
- [14] A. Juels and J. Burton S. Kaliski, "PORs: Proofs of Retrievability for Large Files," Proc. of CCS '07, pp. 584-597, 2007. Alexandria, Va, USA.
- [15] G Ateniese, S. Kamara, J. Katz, "Proofs of Storage from homomorphic identification protocols". In Proc. of ASIACRYPT '09, 2009, pp. 319-333.Tokyo, Japan.
- [16] H. Shacham and B. Waters, "Compact Proofs of Retrievability", Proc. 14th Int'l Conference Theory and Application of Cryptology and Information Security: Advances in Cryptology (ASIACRYPT), LNCS 5350,2008, pp.90-107. Melborne, Australia.
- [17] Y. Dodis, S. Vadhan, D. Wichs. "Proofs of retrievability via hardness amplification". In: Proc. of TCC '09, 2009, pp.109--127.CA, USA.
- [18] K. D. Bowers, A. Juels, and A. Oprea, "Proofs of Retrievability: Theory and Implementation," Cryptology ePrint Archive, Report 2008/175, 2008 <http://eprint.iacr.org/>.
- [19] K. D. Bowers, A. Juels, and A. Oprea, "HAIL: A High-Availability and Integrity Layer for Cloud Storage," Cryptology ePrint Archive, Report 2008/489, 2008, <http://eprint.iacr.org/>.
- [20] T. S. J. Schwarz and E. L. Miller, "Store, Forget, and Check: Using Algebraic Signatures to Check Remotely Administered Storage," Proc. of ICDCS'06,2006,pp.12-21, Lisboa, Portugal.
- [21] R. Curtmola, O. Khan, and R. Burns. "Robust remote data checking". In: Proc. of StorageSS '08, 2008, pp.63-68, Virginal, USA.
- [22] G. Ateniese, R. D. Pietro, L. V. Mancini, and G. Tsudik, "Scalable and Efficient Provable Data Possession," Proc. of SecureComm '08, pp. 1-10, 2008.Istanbul, Turkey.
- [23] C. Wang, Q. Wang, K. Ren, N. cao and W. Lou , "Towards Secure and Dependable Storage Services in Cloud Computing". Accepted for publication in future issue of IEEE Trans. Service Computing. DOI:10.1109/TSC.2011.24.
- [24] C. Erway, A. Kupcu, C. Papamanthou, and R. Tamassia, "Dynamic provable data possession," in CCS'09, ACM, 2009, pp. 213-222, Chicago, USA.
- [25] Q. Wang, C. Wang, K. Ren W. Lou, and J. Li, "Enabling public verifiability and data dynamics for storage security in cloud computing," IEEE Trans. Parallel and Distributed Computing.VOL.22, NO.5, May 2011, pp.847-859
- [26] Yan Zhu, Huaixi Wang, Zexing Hu, Gail-J. Ahn, Hongxin Hu, Stephen S. Yau, "Dynamic Audit Services for Integrity Verification of Outsourced Storages in Clouds," Proc. of the 26th ACM Symposium on Applied Computing (SAC), Tunghai University, TaiChung, Taiwan, March 21-24, 2011.
- [27] P. Syam Kumar, R. Subramanian, "Homomorphic Distributed Verification Protocol for Ensuring Data Storage in Cloud Computing". International Journal of Information, VOL. 14, NO.10, OCT-2011, pp.3465-3476.
- [28] M. A. Shah, M. Baker, J. C. Mogul, and R. Swaminathan, "Auditing to Keep Online Storage Services Honest," Proc. 11th USENIX Workshop on Hot Topics in Operating Systems (HOTOS '07), 2007, pp. 1-6, CA, USA.
- [29] M. A. Shah, R. Swaminathan, and M. Baker, "Privacy-preserving audit and extraction of digital contents," Cryptology ePrint Archive, Report 2008/186, 2008, <http://eprint.iacr.org/>.
- [30] N. Oualha, M. Onen, Y. Roudier, "A Security Protocol for Self-Organizing Data Storage". Tech. Rep. EURECOM-2399, Institut Eurecom, 2008, France.
- [31] C. Wang, Q. Wang, K. Ren, and W. Lou, "Privacy-preserving public auditing for data storage security in cloud computing," In Proc. of IEEE INFOCOM'10, , March 2010. San Diego, CA, USA.
- [32] Z. Hao and N. Yu, "A multiple-replica remote data possession checking protocol with public verifiability," in Second International Symposium on Data, Privacy, and E-Commerce , 2010.Buffalo, Niagara Falls.
- [33] Z. Hao, S. Zhong, and N. Yu, "A Privacy-Preserving Remote Data Integrity Checking Protocol with Data Dynamics and Public Verifiability", Accepted for publication in future issue of IEEE Trans. Knowledge and Data Engineering, DOI: 10.1109/TKDE.2011.62
- [34] A. F. Barsoum and M. A. Hasan, "Provable possession and replication of data over cloud servers," Centre For Applied Cryptographic Research (CACR), University of Waterloo,Report2010/32,2010.<http://www.cacr.math.uwaterloo.ca/techreports/2010/cacr2010-32.pdf>.
- [35] L. Chen, G. Guo, "An Efficient Remote Data Possession Checking in Cloud Storage", International Journal of Digital Content Technology and its Applications. Volume 5, Number 4, April 2011, pp.43-50.
- [36] J. Yang, H. Wang, J. Wang1, C. Tan and D. Yu, "Provable Data Possession of Resource-constrained Mobile Devices in Cloud Computing" JOURNAL OF NETWORKS, VOL. 6, NO.7, JULY 2011, pp.1033-40.
- [37] V. Miller, "Uses of elliptic curves in cryptography", advances in Cryptology, Proceedings of Crypto'85, Lecture Notes in Computer Science, 218 Springer-Verlag, pp.417-426. 1986.
- [38] K. Koyama, U. Maurer, T. Okamoto, and S. Vanstone, "New Public-Key Schemes Based on Elliptic Curves over the Ring Z_n ", Advances in Cryptology - CRYPTO '91, Lecture Notes in Computer Science, Springer-Verlag, vol. 576, Aug 1991, pp. 252-266..
- [39] Brately P and Fox B L "Algorithm 659: Implementing Sobol's Quasi-random Sequence Generator" ACM Trans. Math. Software 14 (1), (1988) , pp. 88-100.

¹P Syam Kumar: is currently Ph.D student (Computer Science), in department of Computer Science, School of Engineering and technology, Pondicherry University, Puducherry, India. He received M.Tech Degree in 2006 from the department of Computer Science and Technology in Andhra University, India, and received B.Tech (graduation) in 2003 from CSE department, Vagdevi college of Engineering, JNTU Warangal, India. He is especially interested on Network Security, Cloud Computing, Distributed Systems etc.

²R. Subramanian: is Professor of Computer Science department, School of Engineering and technology, Pondicherry University, Puducherry, India. He received his Ph.D degree of The Department of Mathematics, IIT Delhi, India in 1989. He received a M.Sc . degree in 1984 from IIT Delhi, India and he received B.Sc. form Madurai Kamaraj University in 1982, Tamilnadu, India. He is especially interested in Parallel & Distributed Algorithms, Cloud Computing, Evolutionary Algorithms, Robotics, etc.

Performance Evaluation of Fingerprint Identification Based on DCT and DWT using Multiple Matching Techniques.

Lavanya B N¹ and K B Raja¹

¹ Department of Computer Science and Engineering, Bangalore University, University Visvesvariah College of Engineering Bangalore, Karnataka 560 001, India

Abstract

The fingerprint is a physiological trait used to identify a person. In this paper, Performance Evaluation of Fingerprint Identification based on DCT and DWT using Multiple Matching Techniques (FDDMM) is proposed. The fingerprint is segmented into four cells of each size 150*240. The DCT is applied on each cell. The Harr Wavelet is applied on DCT coefficient of each cell. The directional information features and centre area features are computed on LL sub band. The final Feature Vector is obtained by concatenating Directional Information and Centre Area Features. The matching techniques viz., ED, SVM, and RF are used to compare test image feature with database image features. It is observed that the values of TSR and FRR are better in the case of proposed algorithm compared to existing algorithm.

Keywords: Fingerprint, Directional Information Features, Centre Area Features, FRR, TSR.

1. Introduction

The Biometrics is used to recognizing a person and the term is derived from the greek word *bio* (life) and *metric* (measurement). The biometric parameters are broadly classified as Physiological characteristics of a person such as face, fingerprint, palmprint, Iris, DNA etc. and behavioral characteristics of human being like signature, voice, keystroke, gait etc. Biometric system operates as verification mode or identification mode depending on the requirement. The verification mode validates a person with readymade template. The identification mode recognizes a person's identity by performing matching against multiple biometric templates. The biometrics is more secure compared to the traditional methods such as PIN, smartcard for verifying a person. Some of the biometric applications are financial transaction, access to computer, access to confidential documents.

Fingerprints have been used as most popular biometric authentication and verification measure because of high acceptability, immutability and uniqueness. Immutability refers to the persistence of the fingerprints over time

whereas uniqueness is related to the individuality of ridge details across the whole fingerprint image. Fingerprint classification is an important step in any fingerprint identification system because it reduces the time taken in identification of fingerprints. Classification allows test fingerprint to be matched against a database. The disadvantages of fingerprint are (i) User acceptance is not guaranteed due to temporary or permanent injury to fingerprint. (ii) Due to distortion on fingerprint or dirt the success rate of a person is reduced and (iii) Malicious fingerprint is produced with the help of special material to match with some person. The different techniques used to identify a person with fingerprint are (i) Minutiae based technique: Ridge ending and ridge bifurcation are minutiae of fingerprint and represented by location (x, y) and orientation Θ. The preprocessing on fingerprint is required to extract a true minutia which is time consuming, but accuracy rate of recognition is high. (ii) Image based technique : The fingerprint preprocessing is optimal and the spatial domain fingerprint image is processed to generate features using mean, standard deviation, variance, energy, gradient, directional features etc. (iii) Transformed domain technique – The fingerprint is converted into frequency domain using Fast Fourier Transform (FFT), Discrete Cosine Transform (DCT), Dual Tree Complex Wavelet Transform (DTCWT), Discrete Wavelet Transform (DWT), Principal Component Analysis, Singular Value Decomposition (SVD). The features are generated using transformed domain coefficients. (iv) Multiple Feature Fusion Technique – The feature are created using fusion of minutiae features, ridge features, image based features and transformed domain feature. (v) Multiple Classifier Technique – The test fingerprint features are compared with features in the database using multiple classifiers such as ED, SVM, RF etc.

Contribution: In this paper, the Fingerprint is segmented and DCT is applied on each segmented region. The DWT is applied on DCT coefficients to get low and high frequency components. The directional information features and centre area features are computed and concatenated to get final features. The ED, SVM and RF

are used to verify test fingerprint with database fingerprint.

Organization: The paper is organized into the following sections. Section 2 is an overview of related work. The FDDMM model is described in Section 3. Section 4 is the algorithm for FDDMM system. Performance analysis of the system is presented in Section 5 and Conclusions are contained in Section 6.

2. Related Work

Deepak Kumar Karna et al., [1] has proposed a fingerprint recognition method based on normalized cross correlation. The method correlates the common region of two fingerprint images by the image rotation and scaling. Honglie Wei et al., [2] has proposed quadrangle based matching algorithm. The quadrangle is composed of two ridge lines that connect two minutiae points with the associated two ridge points. The line matching has been emphasized. The local quadrangle set is formed using the type, the length of the basic ridge lines and angle between the basic edge and the ridge. The fake pairs are removed by considering correlation among pairs. The fingerprint matching is done using the shrunken set. Xiao long Zheng et al., [3] has presented a minutiae scoring method which considered various aspects influencing the quality of minutiae. The image quality is described from three aspects viz., the global factor, the block image quality factor and the neighborhood detailed structure. These three factors are investigated and combined by the product rule and linear weight. The minutiae score is incorporated into fingerprint matching by checking the scores of corresponding minutiae pairs.

Chengming Wen and Tiande Guo [4] presented fingerprint matching algorithm based on convex hull matching principles. The query fingerprint is matched with the template fingerprint using local features. Using global features the matching minutiae pairs are achieved with the help of maximum weight bipartite matching model. Spurious matching is eliminated using the convex hull similarity checking process. Chengming wen et al., [5] has proposed an algorithm for one-to-one matching of minutiae points using motion coherence methods. They have used the K-plet to describe local structure. Chongwen wang and Gangyi Ding [6] has presented a fingerprint representation method, which combines adjacent features of minutiae with curvature of ridges. These features are used to identify corresponding minutiae between two fingerprint impressions by computing the Euclidean distance between vectors. Only the position and direction are used to define minutiae. The algorithm uses the curvature of ridges to alignment and delivers a matching

score that expresses the degree of similarity between the two fingerprints.

Zhong Wei-bo et al., [7] used the relative topological relationship among the minutiae for the fingerprint matching. The minutiae type and the topological relationship among the minutiae and its neighbor with the absolute coordinate information gives the changeless characteristics in the fingerprint distortions. Jiong Zang et al., [8] proposed an algorithm which extracts a fingerprint reference point and reference direction based on the orientation field of the image is proposed. Then they take the reference point as the origin and the reference direction as the polar axis to establish a polar coordinate system. The fingerprint minutiae information which is constituted by minutiae point is independent of the shift and rotation of the fingerprint. Overall matching is employed here to match two fingerprint images with shift and rotation. Khuram Yasin Qureshi and Shoab A. Khan [9] proposed fingerprint matching using five neighbor of one single minutiae i.e., center minutiae. The authentication of a minutia is assured by the characteristics of its five neighbors. The characteristics are minutiae type, distance to the central minutiae, relative angle calculated by the coordinates of central point, coordinates of neighbor, direction of central point and ridge count. The special matching criteria incorporate fuzzy logic to select final minutiae for matching score calculation.

JuCheng Yang et al., [10] proposed fingerprint matching using invariant moments which are preprocessed for minutiae and the reference point determination. The reference point is then aligned. These features are invariant to translations and rotation. Anil K. Jain et al., [11] proposed algorithm to compare the latent fingerprint image with that of the stored in the template. From the latent fingerprint minutiae orientation field and quality map are extracted. Both level 1 and 2 features are employed in computing matching scores and at each feature level both quantitative and qualitative scores are computed. Xuzhou Li and Fei Yu [12] proposed fingerprint matching algorithm that uses minutiae centered circular regions. The circular regions constructed around minutiae are regarded as a secondary feature. This feature is tolerant to the linear distortion. Other features like ridge count, distance, relative angle and the minutiae type are also considered to construct the local features. The degree of similarity is obtained based on the local features matching. The minutiae pair that has the higher degree of similarity than the threshold is selected as reference pair minutiae.

Danese et. al., [13] proposed fingerprint matching in the field of embedded systems for real-time authentication. FPGA based architecture is used that employs the phase only correlation algorithm. Sevng -Hoon Chae et al., [14] used the technique of obtaining the distance between the two ridges to compare the two ridges. Initially the compare

points are selected on the neighboring ridges, then the distance between the pairs of comparison points are calculated. The mean distance and the standard deviation are also determined. Jian-De Zheng et al., [15] introduced fingerprint matching based on minutiae. The algorithm uses a method of similar vector triangle. The ridge end points are considered as the reference points and the vector triangles are constructed. The fingerprint matching is performed by comparing the vector triangles.

Minakshi Gogoi and Bhattacharya [16] proposed an effective method of minutiae clustering for fingerprint verification and graph theoretic approach for fingerprint comparison. Hausdorff distance is the invariance measure of each fingerprint. Zhi Hao Lo [17] proposed an enhanced minutiae based fingerprint feature extraction approach for use in fingerprint identification system. Ramaswamy et al., [18] discussed two types of system, they are Automatic Fingerprint Authentication System (AFAS) and Automatic Fingerprint Identification System (AFIS) which are used for manual matching and classification of fingerprints

M. Rajinikannan et al., [19] described three models of minutiae detection system which differently enhances the input image for minutiae detection. FFT and Gabor filter based on frequency and orientation image enhancement technique are used. Conti et al., [20] proposed a fingerprint recognition approach which is based on core and delta points. A Pseudo singularity point is extracted . In automatic fingerprint recognition fingerprint classification and matching are the two key issues. Pseudo singularity points were detected and extracted for fingerprint recognition. Neil Yager and Ted Dunstone [21] discussed the existence of goats, lambs, wolves for biometric applications and presented methods for dealing with their existence. A differing degree of accuracy is seen within the biometric system. Goats, wolves and lambs are labels commonly applied to problem users. Mohammed S Khalil et al., [22] proposed fingerprint verification based on statistical analysis. This method incorporates two techniques detecting the reference point and analyzing the fingerprint image. A biometric fingerprint images are statistically analyzed for personal identification. A sub image of 129 * 129 was extracted from original image and transformed into co-occurrence matrix. The results have been analyzed by program for rate estimation and statistical summaries.

3. Model

The definitions of performance parameters and the proposed FDDMM model for fingerprint recognition based on Directional Information, Centre and Edge Features of DWT (FDDMM) is discussed in detail.

3.1 Definitions

3.1.1 False Rejection Rate

It is the measure of the biometric security system that incorrectly rejects an access attempt by an authorized user. A FRR is the ratio of the number of false rejections to the total number of identification attempts is given in the Equation 1.

$$FRR = \frac{\text{Number of rejected persons}}{\text{Total number of persons}} \quad (1)$$

3.1.2 Total Success Rate (TSR)

It is the rate at which match occurs successfully.

$$TSR = \frac{\text{Number of matched persons}}{\text{Total number of persons}} \quad (2)$$

3.1.3 Gradient

Gradient is a directional change in the intensity or colour in an fingerprint image. It can also be used to extract directional information from fingerprint images.

3.1.4 Variance

The variance for a block indicates how much each individual element in the block deviates from the sample mean. It gives better result as compared to standard deviation.

3.2 Proposed FDDMM model

Fingerprint recognition system based on directional Information, Centre Area Features, DCT and DWT is as shown in the Figure 1.

3.2.1 Fingerprint Database

Fingerprint images are considered from the data base of FVC 2004[23]. The fair and distinct fingerprint image databases DB1, DB2, DB3 and DB4 are created with different scanners and time as shown in the Figure 2. Each data base has 110 fingers with 8 samples per finger leading to 880 fingerprint images. DB3 data base is considered to test our algorithm and it is decomposed into two parts viz., DB3_A and DB3_B having first 100 fingers and last 10 fingers respectively. The source data base consists of DB3_A with first seven samples of every fingerprint constituting 700 samples. The test fingerprint data base consists of DB3_A with the last eighth sample of each fingerprint leads to 100 fingerprint samples. An eight fingerprint samples of a Person in DB3_A database is

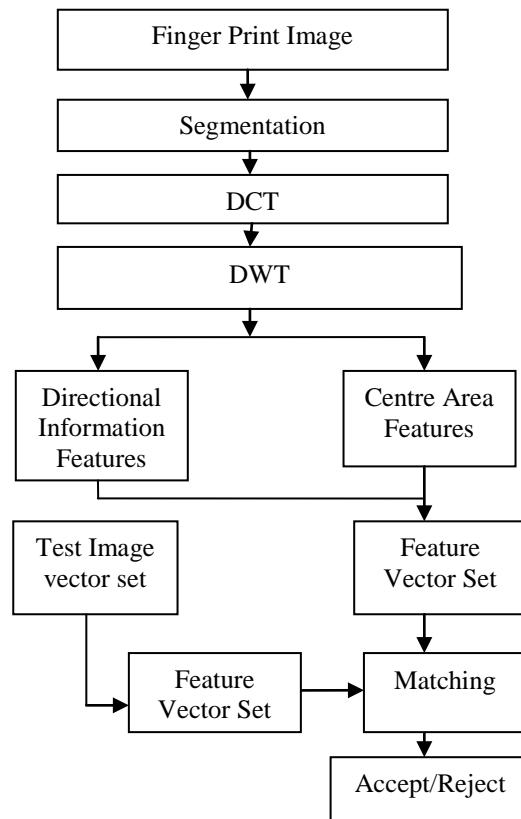


Fig 1: Block Diagram FDDMM



Fig 2. Fingerprint image samples

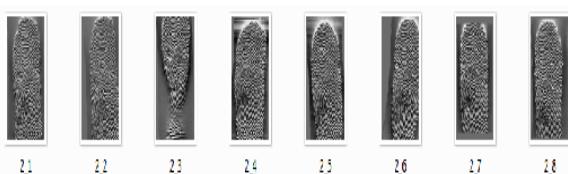


Fig 3. A sample of finger print of DB3_A

shown in the Fig 3. Using the source data base and test data base, FRR can be calculated. The DB3_B data base is considered for second test data base having 80 samples which are used to compute FAR.

3.2.2 Segmentation

This function divides the two-dimensional matrix into adjacent sub matrices. The original fingerprint size of 300×480 is converted into four sub cells of each size 150×240 is shown in the Figure 4. The final features of each cell are compared between Test-image and database image to get better performance results.

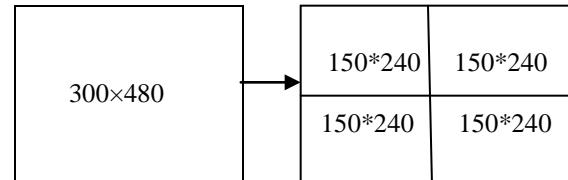


Fig 4. Segmentation

3.2.3 Discrete Cosine Transform (DCT)

Fingerprint representation in DCT [24] is a sum of sinusoids of varying magnitudes and frequencies. Each cell is converted into DCT co-efficient and few DCT co-efficient has significant information of an image. DCT is used to convert spatial domain image into frequency domain image which differentiate whole image into high and low frequency components. The DCT is used in data compression as reconstruction of original image from frequency domain is possible with few DCT coefficients.

The DCT Equation is given in an Equation 3

$$Y(k) = w(k) \sum_{n=1}^N x(n) \cos \frac{\pi(2n-1)(k-1)}{2N}$$

Where

$$K=1, 2, 3, \dots, N$$

$$W(k) = \begin{cases} \frac{1}{\sqrt{N}} & k=1 \\ \sqrt{\frac{2}{N}} & 2 \leq k \leq N \end{cases} \quad (3)$$

The DCT has better energy compaction properties with just a few of the transform coefficients representing the majority of the energy in the sequence. The properties of the DCT make it useful in applications such as data communications, signal coding, etc.

3.2.4 Discrete Wavelet Transform (DWT)

The one level Daubechies wavelet [25] is applied on DCT of segmented portion of Fingerprint and features are extracted from LL, LH, HL and HH sub bands for the verification of fingerprint. LL sub band gives significant information, LH sub band represents vertical information, HL gives horizontal details and HH gives diagonal details of DCT coefficients in the fingerprint image.

3.2.5 Directional Information Features.

The LL sub band of the DWT is considered for directional information features. The gradient of LL is computed using LH and HL sub bands ie. the gradient G_{mn} and corresponding angle θ_{mn} at the position (m,n) is computed using Equations 4 and 5 respectively.

$$G_{mn} = (|G_{mn}^x| + |G_{mn}^y|) \quad (4)$$

$$\theta_{mn} = \tan^{-1} (G_{mn}^x / G_{mn}^y) \quad (5)$$

The quantities G_{mn}^x and G_{mn}^y represent the components of G_{mn} in horizontal and vertical directions, respectively. The coherence is determined using gradient and angle as given in the Equation. 6 using the window size of (5 x 5).

$$\delta_{mn} = \frac{\sum G_{ij} \cos(\theta_{mn} - \theta_{ij})}{\sum G_{ij}} \quad (6)$$

Where i = 1 to 5

and j = 1 to 5

The dominant local orientation is calculated from the gradient and coherence. The dominant local orientation θ is defined in Equation 7.

$$\theta = \frac{1}{2} \tan^{-1} \frac{\sum_{m=1}^N \sum_{n=1}^N \delta_{mn}^2 \sin 2\theta_{mn}}{\sum_{m=1}^N \sum_{n=1}^N \delta_{mn}^2 \cos 2\theta_{mn}} + \frac{\pi}{2} \quad (7)$$

Where N = 8.

Thus, each 8x8 size window represents one directional information.

3.2.6 Center Area Features

The LL sub band of wavelet of size 75*120 is considered for centre area features. The centre point for LL, sub band is fixed by considering the pixel with maximum variance among rows and columns. The 16 X 16 window is considered around the centre point. The Correlation, Contrast, Homogeneity and Energy are determined for 16 X 16 windows around the centre point for LL sub band to derive the second set of fingerprint features.

Correlation: It is a measure of how correlated a pixel is to its neighbor over the whole image. The range for Gray-

Level Co-occurrence Matrix (GLCM) is given by [-1 1]. Correlation is 1 or -1 for a perfectly positively or negatively correlated image. Correlation is NaN (Not-a-Number) for a constant image. It is given by the Equation 8.

$$\text{Correlation} = \sum_{i=1}^N \sum_{j=1}^M \frac{(i-\mu_i)(j-\mu_j) P(i,j)}{\sigma_j} \quad (8)$$

Where μ = mean,

σ_j = standard deviation.

Contrast: It is a measure of the intensity contrasts between a pixel and its neighbor the whole fingerprint image; it is given by Equation 9. The range for Gray-Level Co-occurrence Matrix is given by [0, (size (GLCM,1)-1)^2], Contrast is zero for a constant image.

$$\text{Contrast} = \sum_{i=1}^N \sum_{j=1}^N (|i - j|)^2 P(i,j) \quad (9)$$

Energy: It is the sum of squared elements in the GLCM. The range for GLCM is given by [0 1], Energy is 1 for a constant image. By using Equation 10 Energy of center area in fingerprint image is given by equation 10

$$\text{Energy} = \sum_{i=1}^N \sum_{j=1}^N P(i,j)^2 \quad (10)$$

Homogeneity: It is a value that measures the closeness of the distribution of elements in the GLCM to the diagonal. The range for gray-level co-occurrence matrix is [0 1], homogeneity is 1 for a diagonal GLCM. By using Equation 11 Homogeneity of center area in fingerprint image is given by Equation 11

$$\text{Homogeneity} = \sum_{i=1}^N \sum_{j=1}^N p(i,j)/(1 + (i - j)) \quad (11)$$

3.2.7 Feature Vectors

Directional Information features and Centre Area Features are concatenated to constitute Feature Vector set to increase correct recognition rate and decrease FRR.

3.2.8 Matching

The final features of test fingerprints are compared with the final features of database fingerprints to verify a person using matching techniques such as (i) Euclidean Distance, (ii) Support Vector Machines and (iii) Random Forests.

(i) Euclidean Distance [26]

It includes three consecutive stages viz., Euclidean Distance determination of feature vectors, the second is sum of matched dominant features and the third is match stage.

Euclidean Distance determination: The two fingers features are converted to single matrix, and a feature point of first row matched with associated feature point of second row using standard Euclidian function.

Sum of matched dominant features: Dominant features are extracted features, obtained by concatenation of correlation, energy, homogeneity and contrast of one level DWT features. Given two fingers dominant features are to be matched, choose any one feature from each finger dominant features, and calculate the similarity of the two features associated with the two referenced feature points. If the similarity is larger than a threshold, assign value one to that feature and finally get the sum of that features as value one.

Match stage: It uses the elastic match algorithm to decide whether the two fingers are matched which is based on Euclidean distance and Sum of matched dominant features.

(ii) Support Vector Machines (SVMs) [27]

It takes a set of input data and predicts which of the possible classes the input belongs to. The input data is treated as an x-dimensional vector. It builds a model by constructing a set of hyper planes in a high dimensional space. For an x-dimensional vector, (x-1) hyper planes are created. A good separation is achieved by the hyper plane that has the largest distance to the nearest training data points of any class and is the functional margin. The larger the margin, the better is the classification

(iii) Random Forests (RF) [28]

The training set consisting of N cases is sampled at random to form a new training set, which is used for growing the decision trees. If there are M variables, then $m < M$ variables are selected which are used to split the node. The tree is grown without pruning. For a given input vector each tree gives its classification, this is called as voting. The forest chooses the class which has maximum number of votes. The error is calculated during the training. The feature vectors are sampled and a few vectors are left out and are called OOB (out-of-bag) data. The size of OOB data is about $N/3$. The classification error is estimated by using this OOB data. The classification error is calculated as follows: A prediction is obtained for each vector which is OOB relative to the i^{th} tree. The class winner (one with majority votes) is found from the vectors which are OOB and compared to ground-truth response. The ratio of misclassified OOB vectors to all vectors in the original data is equal to the classification error. The forest error depends on two things such as: (i) The correlation between any two trees in the forest. Increasing the correlation increases the error. (ii) The strength of each individual tree in the forest. Increasing the strength decreases the error.

4. Algorithm

PROBLEM DEFINITION: The Fingerprint is verified using FDDMM algorithm. The objectives are

1. To increase Total Success Rate of verification
2. To reduce False Rejection Ratio
3. Decrease number of features to reduce execution time

The dual transform i.e., DCT and DWT are applied on fingerprint to derive transform domain coefficients. The directional features and centre area features are computed on transform domain coefficients to derive features. The final feature vector is formed by concatenating two features resulting in unique features is given in the Table 1

Assumptions: The finger print database DB3_A of FVC 2004 having size of 300x480 with 512 dpi is considered for performance analysis

Table 1. FDDMM Algorithm

• Input: Finger print Database, Test Fingerprint.
• Output: Verified Finger print image
1) The database is created for 100 persons with 7 images per persons. i.e. Total number of images $100*7=700$
2) The fingerprint image is segmented into Four cells of size 150×240 each
3) DCT is applied on each cell
4) DWT is applied on cell of DCT coefficient
5) Directional Information Features and Centre Area Features are derived from LL band
6) The final feature vector set obtained by concatenating of Directional Information Features and Centre Area Features
7) The final Feature Vector set for test fingerprint image is created using step2 to step6.
8) The given test image is compare with database using Euclidean Distance, Random Forest and Support Vector Machine

5. Performance Analysis

The database FVC 2004 is considered to verify the performance analysis. The database is created using different sensors and timing which leads to four types of fingerprint images DB1, DB2, DB3 and DB4. To test the proposed algorithm, DB3 set consisting of 100 person's database with 8 fingerprint samples per person results in 800 fingerprint images are considered.

The values of TSR and FRR with number of features and threshold is given in the Table 2. It is seen that the values of TSR and FRR increases and decreases respectively as threshold increases. The value of threshold increases with the number of features

The values of TSR and FRR for different matching methods ED, RF and SVM is given in the Table 3. It is noticed that the performance is better in the case of ED matching compared to RF and SVM classifiers

The comparison of performance parameters such as number of features, FRR, TSR and Elapsed time for existing technique. Fingerprint recognition based on Directional, Centre and Edge features of DWT (FRDCE) [29] and the proposed technique FDDMM is given in the Table 4. It is observed that the number of features in the case of proposed algorithm reduces compared to existing technique since dual transformation and no edge features are used which results in less elapsed time. The values of FRR and TSR are improved in the case of proposed

technique since directional features and centre area features are computed on dual transform coefficients to obtain unique features.

Table 2: The values of TSR and FRR for different threshold values

No. of features	Threshold	TSR	FRR
200	4.5.5	0	100
	5.75	1	99
	7.75	71	29
	9	94	6
	9.25	97	3
	9.5	98	2
	9.75	99	1
	10	100	0
300	5-6.5	0	100
	6.75	1	99
	9	77	23
	10.5	98	2
	13	100	0
400	7-8	0	100
	8.5	3	97
	10.5	79	21
	12	99	1
	13	100	0
572	10-15	0	100
	16	3	97
	24	54	46
	48	98	2
	54	100	0

Table 3: The values of TSR and FRR for different classifiers.

Methods	ED	RF	SVM
TSR	100	99	99
FRR	0	1	1

Table 4: Performance comparison between existing FRDCE and proposed FDDMM method.

	Existing FRDCE [29]	Proposed FDDMM
Number of features	5846	556
FRR	3	0
TSR	97	100
Elapsed time	20 min	2 min 35 sec

6. Conclusions

The Biometrics is used in almost all areas of applications. In this paper FDDMM algorithm is proposed in which DCT is applied on segmented portion of fingerprint. The DWT is applied on DCT to generate four sub bands. The features such as directional information features and centre area features are computed and concatenated for verification using ED, SVM and RF. The success rate of recognition and FRR values are better in the case of proposed algorithm compared to existing algorithm.

References

- [1] Deepak Kumar Karna, Suneeta Agarwal and Shankar Nikam, "Normalized Cross Correlation Based Fingerprint Matching," *Fifth International Conference on Computer Graphics, Imaging and Visualization*, 2008, pp 229-232.
- [2] Honglei Wei, Danni Liu and Changyou Guo "FingerprintMatching Based on Quadrangle," Second International Congress on Image and Signal Processing, 2009, pp 1-4.
- [3] Xiaolong Zheng, Yangsheng Wang, Xuying Zhao and Zheng Wei, "A Scheme of Minutiae Scoring and its Application to Fingerprint," *Proceedings of the Seventh World Congress on Intelligent Control and Automations*, 2008, pp 5917-5921.
- [4] Chengming Wen and Tiande Guo, "An Efficient Algorithm for Fingerprint Matching based on Convex Hulls," *IEEE International Conference on Computational Intelligence and Natural Computing*, 2009, pp 66-69.
- [5] Chengming Wen, Tiande Guo and Shuguang Wang "Fingerprint Feature-Point Matching based on Motion Coherence," *Second International Conference on Future Information Technology and Management Engineering*, 2009, pp 226-229.
- [6] Chonwen Wang and Gangyi Ding, "Fingerprint MatchingCombining the Adjacent Feature with Curvature of Ridges," *Proceedings of the Seventh World Congress on Intelligent Control and Automation*, 2008, pp 6811-6816.
- [7] Zhong Wei-Bo, Ning Xin-Bao and Wei Chen-Jian, "A Fingerprint Matching Algorithm Based on Relative Topological Relationship Among Minutiae," *IEEE International Conference on Neural Networks and Signal Processing*, 2008, pp. 225-228.
- [8] Jiong Zang, Jie Yuan, Fei Shi and Si-dan Du, "A Fingerprint Matching Algorithm of Minutiae Based on Local Characteristic," *Fourth International Conference on Natural Computation*, 2008, pp. 13-17.
- [9] Khurram Yasin Qureshi and Shoab A Khan, "Effectiveness of Assigning Confidence Levels to Classifiers and a Novel Feature in Fingerprint Matching," *IEEE International Conference on Systems, Man, and Cybernetic* , 2009, pp. 2181-2185.
- [10] JuCheng Yang, JinWook Shin, ByoungJunMin, JoonWhoan Lee, and DongSun Park, SookYoon, "Fingerprint Matching using Global Minutiae and Invariant Moments," *IEEE Congress on Image and Signal Processing*, 2008, pp. 599-602.
- [11] Anil K Jain, Jianjiang Feng, Abhishek Nagar, and Karthik Nandakumar, "On Matching Latent Fingerprints," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2008, pp. 1-8.
- [12] Xuzhou Li and Fei Yu, "A New Fingerprint Matching Algorithm Based on Minutiae," *Proceedings of International Council of Chemical Trade Associations*, 2009, pp. 869-873.
- [13] G Danese, M. Giachero, F Leporati, G Matrone, and N Nazzicari, "An FPGA-based Embedded System For Fingerprint Matching Using Phase-Only Correlation Algorithm," *Twelfth Euromicro Conference on Digital System Design / Architectures, Methods and Tools*, 2009, pp. 672-679.
- [14] Seung-Hoon Chae, Jong Ku Kim, Sung Jin Lim, Sung Bum Pan, Daesung Moon, and Yongwha Chung, "Ridge-based Fingerprint Verification for Enhanced Security," *International Conference on Consumer Electronics*, 2009, pp. 1-2.
- [15] Jian-De Zheng, Yuan Gao, and Ming-Zhi Zhang, "Fingerprint Matching Algorithm based on Similar Vector Triangle," *Second International Congress on Image and Signal Processing*, 2009, pp 1-6.
- [16] Minakshi Gogoi and D K Bhattacharyya, "An Effective Fingerprint Verification Technique," *Journal of Computer Science and Engineering*, vol 1, May 2010, pp 27-35.
- [17] Zhi Hao Lo "An Approach for Minutiae Based Fingerprint Feature Extraction," *Journal Of Computer Science and Engineering*, vol 1, June 2010, pp 28-33.
- [18] G Ramaswamy, Vuda Sreenivasarao, P Ramesh and D Ravi Kiran, "A Novel A pproach for Human Identification Through Fingerprints," *International Journal of Computer Applications*, vol 4, July 2010, pp 35-42.

- [19] M Rajinikannan, D Ashok Kumar, R Muthura, "Estimating the Impact of Fingerprint image," *International Journal of Computer Applications*, vol. 2, no. 1, May 2010, pp 36-42.
- [20] V Conti, C Militello, S Vitabile and F Sorbello, "Introducing Pseudo Singularity Points for Efficient Fingerprints Classification and Recognition," *International Conference on Complex, Intelligent and Software Intensive Systems*, vol. 10, July 2010, pp 368-375.
- [21] Neil Yager and Ted Dunstone, "The Biometric Menagerie", *Proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 2, February 2010, pp 220-230.
- [22] Mohammed S Khalil, Dzulkifli Muhammad, Muhammad Khurram Khan and Khaled Alghathbar, "Fingerprint Verification based on Statistical Analysis," *Journal of Computer and Information Science and Engineering*, vol. 10, 2010, pp. 1657-1672.
- [23] <http://biometrics.cse.msu.edu/fvc04db/index.htm>.
- [24] Andrew B Watson "Image Compression using the Discrete Cosine Transform," *Journal of Mathematics*, vol 4, 1994, pp. 81-88.
- [25] Kingsbury N G "Complex wavelets for shift invariant analysis and filtering of signals," *Journal of Applied and Computational Harmonic Analysis* vol. 10, 2001, pp. 234-253.
- [26] J C Gower, "Properties of Euclidean distance and Non Euclidean distance linear algebra," vol. 67, 1985, pp. 81-97.
- [27] Robert Freund, "Training Support Vector Machines," IEEE conference on Computer Vision and Pattern Recognition , 1997, pp. 130-136.
- [28] Liaw, Andy & Wiener, Matthew "Classification and Regression by random Forest" vol. 2/3, 2002, p. 18.
- [29] Shashi Kumar, K B Raja, Laxmana K K, R K Chhotaray and Sabyasachi Pattanaik "Fingerprint Recognition based on Directional, Center and Edge Features of DWT," *Third IEEE International Conference on Computer Modelling and Simulation*, 2011, pp. 38-42.

Lavanya B.N. received her Bachelor of Engineering in Computer Science and Engineering from Sri Jayachamarajendra College of Engineering Mysore and Master of Engineering in computer science from Dr.Ambedkar Institute of Technology Bangalore. She is pursuing her Ph.D in Computer Science and Engineering in JNTU Hyderabad and her area of interest is biometrics, image processing.

K B Raja is an Assistant Professor, Dept. of Electronics and Communication Engg, University Visvesvaraya college of Engg, Bangalore University, Bangalore. He obtained his Bachelor of Engineering and Master of Engineering in Electronics and Communication Engineering from University Visvesvaraya College of Engineering, Bangalore. He was awarded Ph.D. in Computer Science and Engineering from Bangalore University. He has over 75 research publications in refereed International Journals and Conference Proceedings. His research interests include Image Processing, Biometrics, VLSI Signal Processing, computer networks.

FLoMSqueezer: An Effective Approach For Clustering Categorical Data Stream

M Sora¹, S Roy² and S I Singh³

¹Department of Computer Science & Engineering,
Rajiv Gandhi University, Doimukh 791112, Arunachal Pradesh, INDIA

²Department of Information Technology,
North Eastern Hill University, Shillong 793022, Meghalaya, INDIA

³Department of Computer Science & Engineering,
Tezpur University, Tezpur 7984028, Assam, INDIA

Abstract

Squeezer is an effective histogram based approach for categorical data stream clustering. Drawback of Squeezes is that it is not scalable in terms of memory. The size of histogram increases with the increase in records in the dataset. Accommodation of unpredictably large histogram in the main memory is not always feasible. To handle the bottleneck, a modified version of Squeezes, FLoMSqueezer, is proposed in this paper. It uses concise sampling technique for handling increasing memory requirement by the Squeezes. Experimental results shows that proposed approach scales better in terms of quantitative cluster, memory as well as execution time.

Keywords: Cluster analysis, data stream, histogram, sampling, quantitative cluster.

1. Introduction

With the advent of storage technology, internet and network technology, enormous amount of incremental data generated every day in a timely fashion in the form of data stream. Handling such a huge stream of data gives rise to a new data processing model [1]. Unlike traditional data processing, which is normally static in nature, stream data arrives in the form of continuous, high-volume, fast, and time-varying streams and the processing of such streams entail a near real-time constraint. Data streams knowledge discovery systems are usually constrained by three limited resources: time, memory and sample size. Nowadays, time and memory seem to be the bottleneck for machine learning applications, mainly the last one. Many important applications, ranging from network security, sensor data

processing, to stock analysis, climate monitoring, are a part of the data stream model. From the last decade, data mining [3] meaning extracting useful information or knowledge from large amounts of data, has become the key technique to analyze and understand data. Typical data mining tasks include association mining, classification, and clustering. These techniques help find interesting patterns, regularities, and anomalies in the data. However, traditional data mining techniques cannot directly apply to data streams. This is because mining algorithms developed in the past target disk resident or in-core datasets, and usually make several passes of the data. Mining data streams are allowed only one look at the data, and techniques have to keep pace with the arrival of new data. Furthermore, dynamic data streams pose new challenges, because their underlying distribution might be changing. The problem of clustering becomes more challenging when the data is categorical, that is, when there is no inherent distance measure between data values. This is often the case in many domains where data is described by a set of descriptive attributes, some of which are neither numerical nor inherently ordered in any way. A data stream is a massive unbounded sequence of data elements continuously generated at a rapid rate. Consequently, the knowledge embedded in a data stream is more likely to be changed as time goes by. Identifying the recent change of a data stream, especially for an online data stream, can provide valuable information by analysis of the data stream [4].

2. Data stream mining

Data stream are continuous flow of data. A data stream mining is ordered sequence of points that can read only one or small number of times. Formally, a data stream is a sequence of point $x_1, \dots, x_i, \dots, x_n$ read in increasing ordered of indices i. according data stream model[2]. The performance of an algorithm that operates on data stream is measured by number of passes that the algorithm must make over stream, when constrained in terms of available memory, in addition to the more conventional measures. The data stream model is motivated by emerging application involving massive data sets, e.g customer click streams, telephone records; large set of web pages, multimedia data, and sets of retail chain transaction can be modeled as data streams. These data are far too large to fit in main memory and are typically stored in secondary storage devices, making access, partially random access very expensive. Data stream algorithms access the input only a linear scan without random access and require to scan only once over the data. Furthermore, since amount of data far exceeds the amount of space (main memory) available to the algorithm, it is not possible to remember to much of data scanned in the past. These scarcity of space necessities the design of novel kind of algorithm that stores only the past data, leaving enough memory for future data. Clustering has recently been studied across several disciplines, but only few techniques have developed scalable very large datasets. In more recently years, a few categorical stream clustering algorithms has formulated such as Squeez [5] algorithms for categorical data streams.

2.1 Categorical Domains and Attributes

Categorical Domains and Attributes: Let A_1, \dots, A_m be a set of categorical attributes with domains D_1, \dots, D_m respectively. Let the dataset $D = \{X_1, X_2, \dots, X_n\}$ be a set of objects described by m categorical attributes, A_1, \dots, A_m . The value set V_i of A_i is set of values of A_i that are present in D. For each $v \in V_i$, the frequency $f(v)$, denoted as f_v , is number of objects $O \in X$ with $O.A_i = v$. Suppose the number of distinct attribute values of A_i is p_i , we define the histogram of A_i as the set of pairs: $h_i = \{(v_1, f_1), (v_2, f_2), \dots, (v_{p_i}, f_{p_i})\}$. The histogram of the data set D is defined as: $H = \{h_1, h_2, \dots, h_m\}$.

2.2 Dissimilarity Measures: Let X, Y be two categorical objects described by m categorical attributes. The dissimilarity measure between X and Y can be defined by the total mismatches of the corresponding attribute values of the two objects. The smaller the number of mismatches is, the more similar the two objects. Formally,

$$d_1(X, Y) = \sum_{j=1}^m \delta(x_j, y_j) \quad (1)$$

$$\text{where } \delta(x_j, y_j) = \begin{cases} 1 & (x_j = y_j) \\ 0 & (x_j \neq y_j) \end{cases} \quad (2)$$

Given the dataset $D = \{X_1, X_2, \dots, X_n\}$ and an object Y, The dissimilarity measure between X and Y can be defined by the average of the sum of the distances between X_i and Y.

$$d_2(D, Y) = \frac{\sum_{j=1}^n d_1(x_j, y)}{n} \quad (3)$$

If we take the histogram $H = \{h_1, h_2, \dots, h_m\}$ as compact representation of the data set D, eq (3) can be redefined as

$$d_2(H, Y) = \frac{\sum_{j=1}^m \phi_1(x_j, y)}{n} \quad (4)$$

Where

$$\phi(h_j, y_j) = \sum_{l=1}^{p_j} f * \delta(v_l, y_j) \quad (5)$$

In some cases, it is convenient to use similarity rather than distance. Similarity between Y and H is define as

$$Sim(H, Y) = \frac{\sum_{j=1}^m \psi(h_j, y_j)}{n} \quad (6)$$

Where

$$\psi(h_j, y_j) = \sum_{l=1}^{p_j} f * (1 - \delta(v_l, y_j)) \quad (7)$$

From the implementation efficiency viewpoint, eq. (6) can be computed more efficiently because it only requires computing the frequencies of matched attribute value pairs [1].

The clustering accuracy for measuring the clustering results was computed as follows. Given the final number of clusters, k, clustering accuracy r was defined as:

$$r = \frac{\sum_{i=1}^k a_i}{n}$$

where n is the number of record in the dataset, a_i is the number of instances occurring in both cluster i and its corresponding class, which had the maximal value. In other words, a_i is the number of records with the class label that dominates clusters i. Consequently, the clustering is defined as $e=1-r$. Furthermore, we define the absolute clustering error ace as: $ace=e*n$.

3. Related Works

Below we present some existing works for handling categorical data as well as data streams.

3.1 STIRR [11]: (Sieving through Iterated Relational Reinforcement) is an algorithm based on non-linear dynamical systems. The database is represented as a graph where each distinct value in the domain of each attribute is represented by a weighted node. Thus if there are N attributes and the domain size of i -th attribute is d_i . Then the number of nodes is the graph is $\sum_i d_i$. For each tuple in the database, an edge represents a set of nodes which participate in that tuple. Thus a tuple is collection of nodes one from each attribute type. The set of weights of all the nodes define a configuration of this structure. The initial weights of all the nodes can be either assigned uniformly or randomly or by focusing technique. STIRR iteratively changes the configuration by updating weight of any node. The new weight of the node is calculated based on a combiner function, which combines the weights of other nodes participating in any tuple with given node for which the weight to be updated. Thus it moves from one configuration to the other till it reaches a stable point, called as basin. The convergence is dependent on the combiner function. Analyzing the stability is hard for any arbitrary combiner function. However, for simple combination function like sum of multiplication, the system definitely converges to a fixed point. It is easy to see that for categorical attributes, the values which are related through common tuples influence each other during weight modification. Thus one does not require any similarity metric to be defined for categorical attributes. Interestingly, in order to cluster the set of tuples, STIRR maintains multiple copies of weights. When the fixed point is reached, the weight is one or more of the basins isolate two groups of attribute values on each attribute the first with large positive weights and second with small negative weights. The nodes with large positive weights and second with small negative weights are grouped to determine cluster. These groups correspond intuitively to projections of clusters on the attribute. However, the automatic identification of such sets of closely related attribute values from their weights requires a non-trivial post-processing step; such a post-processing step was not addressed in their work. Moreover, the post-processing step will also determine what 'cluster' are output. The underlying idea of STIRR is unique but it may be hard to analyze the stability of the system for any useful combiner function. One requires rigorous experimentation and fine tuning of parameters to arrive at a meaningful clustering[4].

3.2 CACTUS[8]: Catagorical data ClusTering Using Summersies is a sort of subspace clustering. CACTUS attempts to split the database vertically and tries to cluster the set of projection of these tuples to only a pair of attributes. Its basic principle can be described as follows: let us consider two attributes values of two different attributes in the database. Say, a_i of attribute type A and a_j of attribute type B. there may be tuples where a_i and a_j co-occur. The support of these two values in the database is the proportion of tuples in which they appear together. If this support exceeds a pre-specified value, we say these values are strongly connected. This concept can be to compute the inter-attribute and intra-attribute summaries of the given data set. Most interesting aspect of these steps are that these can be computed using-attribute and intra-attribute summary. It is not necessary to refer to the original data base. CACTUS first identifies the cluster projections on all pairs of attributes by fixing one attribute. Then it generates an interesting set to represent the cluster projection on this attribute for n-cluster(involving all the attributes). Once all the cluster projections on individual attributes are generated, these are synthesized to get the clusters of the database. The major steps of CACTUS are: Finding cluster projection on given attribute A_i with respect to another attribute A_j Intersecting all the cluster projection for any given A_i to get the cluster projection A_i with respect to all attribute. Synthesizing the resulting cluster projections to the main clusters.

3.3 COOLCAT [6]: It is capable of efficiently cluster large data sets of records with categorical attributes. COOLCAT's clustering results are very stable for different sample sizes and parameter settings. Also, the criteria for clustering are a very intuitive one, since it is deeply rooted on the well-known notion of entropy. Entropy and Clustering Entropy is the measure of information and uncertainty of a random variable. Formally, if X is a random variable, $S(X)$ the set of values that X can take, and $p(x)$ the probability function of X , the entropy $E(X)$ is defined as shown in Equation:

$$E(X) = - \sum_{x \in S(X)} p(x) \log(p(x))$$

The entropy of a multivariate vector $\vec{x} = \{X_1, \dots, X_n\}$ can be computed as shown in Equation

$$E(\vec{x}) = - \sum_{X_i \in S(X_i)} p(x_1 \dots x_n) \log(x_1 \dots x_n)$$

Entropy is sometimes referred to as a measure of the amount of "disorder" in a system. It is novel method which uses the notion of group records.

COOLCAT initially finds a suitable set of clusters out of a sample, $|S|$ taken from the dataset ($|S| < N$, where N is the size of entire dataset). Form the sample dataset it first find k most "dissimilar" records by maximizing the minimum

pairwise entropy of chosen points. That is COOLCAT starts by finding the two points P_{s1} and P_{s2} that maximize $E(P_{s1}, P_{s2})$ and placing them into two separate clusters(C1, C2) making records. From there it proceeds incrementally .

3.4 ROCK [10]: Robust hierarchical clustering with links uses a novel concept of links to measure the similarity/proximity between a pair of data points. The number of links between two tuples is the number of common neighbors they have in the data set. It is an agglomerative hierarchical clustering algorithm that employs links and not distances when merging clusters. Methods in ROCK naturally extend to non-metric similarity measures that are relevant in situations where a domain expert/similarity table is the only source of knowledge.

Starting with each tuple in its own cluster, the two closest cluster are merged until the required number of clusters are obtained. ROCK, instead of working on the whole data set, clusters a sample randomly drawn from the data set, and then partitions the entire data set on the clusters from the sample. The basic idea of ROCK is based on the following definition.

Neighbor: An object's neighbors are those objects that are considerably similar to it. Let $sim(O_i, O_j)$ be similarity function that is normalized and captures the closeness between the pair of objects O_i and O_j . The similarity function sim assumes values between 0 and 1. Given a threshold θ (between 0 and 1), a pair of objects O_i and O_j are defined as neighbours if $sim(O_i, O_j) \geq \theta$.

Link: The $link(O_i, O_j)$ between the object is defined as the number of common neighbors between O_i and O_j . ROCK attempts to maximize the sum of $link(O_q, O_r)$ for pairs of objects O_q and O_r belonging to single cluster and at the same time to minimize the sum of the $link(O_q, O_r)$ for object pairs belonging to other cluster.

Link between clusters: It is a summation of *links* of all pairs, where each pair is formed by taking one object from each cluster.

Goodness measure: The goodness measure between two cluster is the result obtained after dividing the number of cross-links between the clusters by the expected number of cross-links between the cluster.

The link based adopts the global perspective of the clustering problem. It captures the global knowledge of neighboring data points into the relationship between the individual pair of points. After drawing the random sample from the database, a hierarchical clustering algorithm that employs links is applied to the sample objects. It follows the standard principle of hierarchical clustering. It starts with singleton objects as an individual class and progressively merges the two clusters based on the *goodness criteria*, determined by link structure. The

merging is continued till one of the following two criteria is met: (1) a specified number of clusters is obtained or, (2) no links remain between the clusters [4].

3.5 CLOPE [7] : Clustering with CLOPE algorithm is developed starting from heuristic method of increasing the height-to-width ratio of the cluster histogram. While being quite effective, CLOPE is very fast and scalable when clustering large transactional databases with high dimensions, such as market basket data and web server logs.

To construct cluster histogram, occurrence of every distinct item is counted for each cluster, and then height (H) and width (W) of the cluster is obtained. For example, cluster {ab, abc, acd} has occurrences of a :3,b :2,c :2 and d:1. Therefore, its corresponding cluster histogram's (H), width (W) and H/W ratio are 2.0, 4.0 and 0.5 respectively.

CLOPE uses a global criterion function that tries to increase the intra-cluster overlapping of transaction items by increasing the height-to-width ratio of the cluster histogram. Moreover, it generalizes the idea by introducing a parameter to tightness of the cluster. Different number of clusters can be obtained by varying parameter. A larger height-to-width ratio of the cluster histogram means better intra-cluster similarity. And, the global criterion function is defined using the geometric properties of the cluster histogram [5].

3.6 SQUEEZER [5]: It is an efficient algorithm for clustering categorical data, Squeezing, which can produce high quality clustering results and at the same time deserve good scalability. The Squeezing algorithm reads tuples from dataset one by one. When the first tuple arrives, it forms a cluster alone. The consequent tuples are either put into existing cluster or rejected by all existing clusters to form a new cluster by given similarity function defined between tuple and cluster.. It is obvious that the Squeezing algorithm only makes one scan over the dataset, thus, highly efficient for disk resident datasets where the I/O cost becomes bottleneck of efficiency. It is suitable for clustering data streams, where given a sequence of points, the objective is to maintain consistently good clustering of the sequence so far, using a small amount of memory and time. It can also handle outliers efficiently and directly in Squeezing. Squeezing achieves both high quality of clustering results and scalability. The summary of Squeezing algorithm as follows:

A novel algorithm for clustering categorical data, Squeezing combines both efficiency and quality of clustering results. The algorithm is extremely suitable for clustering data streams, where given a sequence of points, the objective is to maintain consistently good clustering of the sequence so far, using a small amount of memory and time. Outlier can be handled efficiently and directly. The algorithm does not require the number of desired clusters as an input parameter.

This is very important since the user usually does not know this number in advance. The only parameter to be specified is the value of similarity between the tuple and cluster, which incorporates the user's exception that how closely the tuples in a cluster should be.

SQUEEZER Algorithm:

Inputs :

DS = Categorical Data Stream

St = user defined similarity threshold

Algorithm

for each record t in DS **do Begin**

for each existing cluster Ci

begin

 Measure the similarity between t and Ci
 end

 Smax = Max(all similarity measures)

 Tc = target cluster with Smax

if smax > = St **then**

 Include t in Tc

Update the histogram of Tc

else

 Create a new cluster with t

 Create its histogram

end if

end

The time and space complexities of the Squeezing algorithm depend on the size of dataset and the number attributes.

4. Motivation

4.1 Unbounded Memory Requirements

Since data streams are potentially unbounded in size, the amount of storage required to compute an exact answer to a data stream query may also grow without bound. While external memory algorithms for handling data sets larger than main memory have been studied, such algorithms are not well suited to data stream applications since they do not support continuous queries and are typically too slow for real-time response. The continuous data stream model is most applicable to problems where timely query responses are important and there are large volumes of data that are being continually produced at a high rate over time. New data is constantly arriving even as the old data is being processed; the amount of computation time per data element must be low, or else the latency of the computation will be too high and the algorithm will not be able to keep pace with the data stream. For this reason, we are interested in algorithms that are able to confine themselves to main memory without accessing disk. In the data stream model of computation, once a data element has been streamed by, it cannot be revisited. A few algorithms have been

proposed in recent years for clustering categorical data stream. Squeezing is one of the promising technique, however its performance degrades as the size of the histogram increase. Below we present a new sampling based approach to overcome the bottleneck of Squeezing.

5. FLoMSqueezing: A sampling based approach

The aim was to improve Squeezing to make suitable for clustering categorical data stream and to make less error, reduce the size of histogram and computational time. It is found that if concise sampling technique is applied the new algorithm gives less memory footprint.

5.1 Definition and Notations

1) *Similarity Measure:* Similarity between an object Y and D is now defined as $\text{Sim}(H,Y) = \text{SUM}(f(h_i,y_i)/n ; I = 1, 2, \dots, M)$ Where $f(h_i,y_i) = \text{SUM}(f_j * (1 - \text{distance}(v_j, y_i)))$

2) *Histogram:* Histograms are commonly-used summary structures to succinctly capture the distribution of values in a data set. A compact representation of a cluster, H = Histogram of D = {h₁, h₂, h₃, ..., h_n} where h_i = {(v₁, f₁), (v₂, f₂), ..., (v_p, f_p)} where p is the no of distinct value of attribute A_i and f_i is the no of objects in D having this value v_i. For every record processed, we have to calculate the similarity with existing clusters, update or create a new histogram, and do the pruning if it's at footprint bound.

3) *Cluster error:* $e = 1 - r$ where $r = \text{SUM}(A_i/n) I = 1, 2, \dots, k$ where n = total no of data points in the data set and A_i is the no of instances occurring in both the cluster I and its corresponding class. Absolute cluster error ace = e*n.

4) *Sampling:* Sampling is the most versatile approximation technique available. Most data processing algorithms can be used on a random sample of a data set rather than the original data with little or no modification. Sampling-based algorithms can produce approximate answers that are provably close to the exact answer.

5.2 Processing step of FLoMSqueezing

Sampling techniques have been introduced into the update histogram module of Squeezing algorithm. For every incoming tuple t in target cluster Tc, for each attribute value t.A of tuple t in target cluster Tc. It checks the corresponding dimensions of the histogram tc with a probability of 1/p. Whether t.A is present in the histogram

Tc. If t.A present in the histogram of corresponding dimensions, that simply increment the count of attribute value. Otherwise t.A remains singleton entry. After every operation on the histogram the histogram for particular dimension is greater some pre-specified Footprint bound "g" then it first raise the value of (P) to (P') sampling rate and sample the histogram for that dimensions in terms of P'.

FLoMSqueezer Algorithm

Inputs :

DS = Categorical Data Stream
St = user defined similarity threshold
p = initial sampling rate
g = pre specified footprint bound

Algorithm

```

for each record t in DS do
    begin
        for each existing cluster Ci
            begin
                Measure the similarity between t and Ci
            end
                Smax = Max(all similarity measures)
                Tc = target cluster with Smax
                if smax >= St then
                    Include t in Tc
                    for each t.A in t begin
                        with a probability (1/p) check
                        if t.A is in Histogram for Tc.
                        if t.A is present then
                            Increase its count in its <value,count> pair
                        else
                            t.A remain as a singleton
                        end
                        if S.A >= g
                        then Raise p to p'
                    Subject each sample points in S.A to p'
                    end if
                    p = p'
                else
                    Create a new cluster with t
                    Create its histogram
                end if
    end

```

5.3 Analysis

Firstly, It is shown that the FLoMSqueezer algorithm can be proved to have space guarantee on main memory consumed.

1) Theorem 1: A concise sample method is a uniform random sample of the data set such that values appearing more than once in the sample are represented as <value, frequency> pairs. S =

{<v1,f1>,<v2,f2>,...<vj,fj>, ...vj+1,...,vl} Sample size represented by S = 1 - j + SUM(fk) k = 1...j. Memory footprint of S=I+j.

2) Theorem 2: FLoMSqueezer algorithm uses 'g' pre specified footprint bound. If S.A >= g then sampling start. Initial sampling raise p to p' each sample point in S.A to p' it terminates if p=p'.

6. Results and Discussion

6.1 Quality of Clustering with Real-life Datasets

Comparison of FLoMSqueezer with Squeezer is done with real-life Mushroom dataset, which are obtained from UCI Machine Learning Repository and have been tested both algorithms. The Mushroom dataset has 22 attributes with 8124 tuples. Each tuple records physical characteristics of a single mushroom. A classification label of poisonous or edible is provided with each tuple. The numbers of edible and poisonous mushrooms in the dataset are 4208 and 3916, respectively.

6.2 Tuning Parameter

The footprint bound 'g' and initial sampling 'p' rate makes the FLoMSqueezer algorithm differ from the squeezer algorithm. These parameters can affect the results of clustering and speed of the algorithm. In the sequel, an empirical result shows how they can affect the FLoMSqueezer algorithm Mushroom dataset.

Firstly, each histogram Hi is pruned by deleting some entries at bucket boundaries. Apparently, large 'g' results more pruning. When the initial sampling is increased and the similarity was set to 7,8,9,10,11,12,13,14,15,16 respectively, the processing time decreases. This algorithm tends to produce more stable cluster when similarity threshold become larger and footprint bound become low and initial sampling rate increases.

6.3 Scalability Evaluation

1. Scalability with Synthetic Dataset: The stream size (i.e number of rows), the number of attributes and number of classes are major parameters in the synthetic categorical data streams. The experiments were carried out with synthetic datasets. The scalability of the algorithm is determined by sample size, by taking dataset with 6 attributes and 8 tuples then the performance is same in small dataset. But Squeezing algorithm deteriorates with increase of database size.

2. Scalability with Real Dataset: The experiment is carried out with real-life dataset the Mushroom dataset [9], which are obtained from UCI Machine Learning Repository. Both the algorithms clusters based on the threshold value, initial sampling rate, footprint bound are taken. Table shows that comparison between time, histogram clustering errors and number of clusters produce by two algorithms. The proposed algorithm performs better than Squeezer algorithm in handling histogram size, time and clustering error.

7. Experimental Results

A comprehensive performance study has been conducted to evaluate above method. Both the quality of the clustering results and the efficiency are examined. Synthetic dataset and Real-life dataset are used to evaluate the quality of clustering results. The experiments were divided into three parts:

- 1) Firstly we tested with execution time on the data stream of different sizes.
- 2) Secondly, we studied the effect the histogram size.
- 3) Thirdly, we demonstrated the changes of error value, comparing error value.

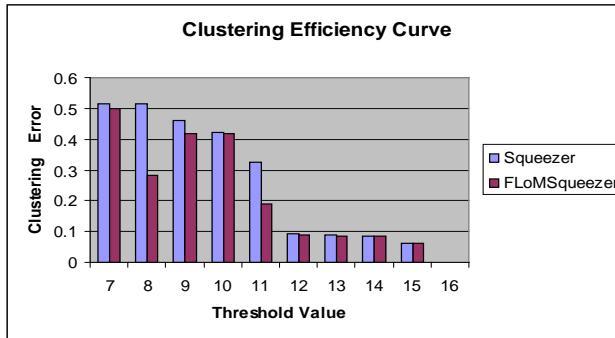


Fig. 1 Clustering result with error comparison

The relative performance of the two algorithms with clustering errors and threshold value shows FLoMSqueezer algorithm out performed Squeezer algorithm.

TABLE I: Relative performance of Squeezer and FLoMSqueezer

Ranking	Average clustering Error
Squeezer	0.256696206
FLoMSqueezer	0.213429346

TABLE I : Histogram and time comparison

Threshold	Squeezer			FLoMSqueezer		
	Histogram Size	Time in seconds	No. of clusters	Histogram Size	Time in seconds	No. of Clusters
7	183	1	2	140	1	2
8	237	1	3	299	1	5
9	289	1	4	279	1	5
10	357	1	6	333	1	7
11	401	2	8	450	2	10
12	583	4	14	556	3	14
13	550	4	15	542	3	15
14	598	5	17	583	3	17
15	637	5	20	633	4	20
16	697	7	23	697	4	23

Fig. 2 shows the results of histogram comparisons between Squeezer and FLoMSqueezer algorithm. Out of 10 different threshold values FLoMSqueezer outperform Squeezer algorithms. It performs best in 6 cases, shows less performs in two cases and perform equal in 2 cases. It never performed worst in other cases.

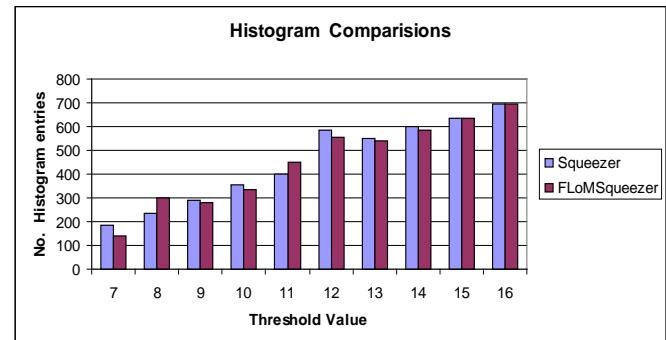


Fig. 2 Comparisons of histogram with different threshold values.

It implies that the FLoMSqueezer can produce good clustering output with limited memory in the data stream environment.

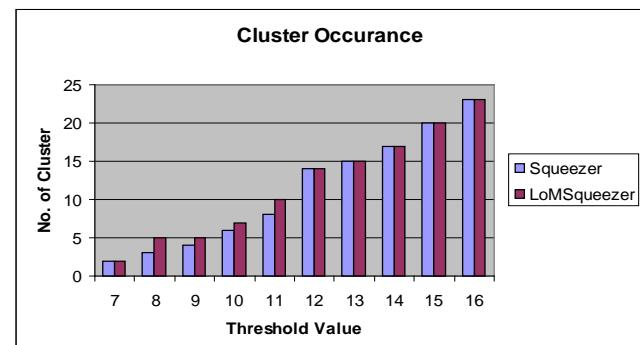


Fig. 3 Comparison of cluster occurrences

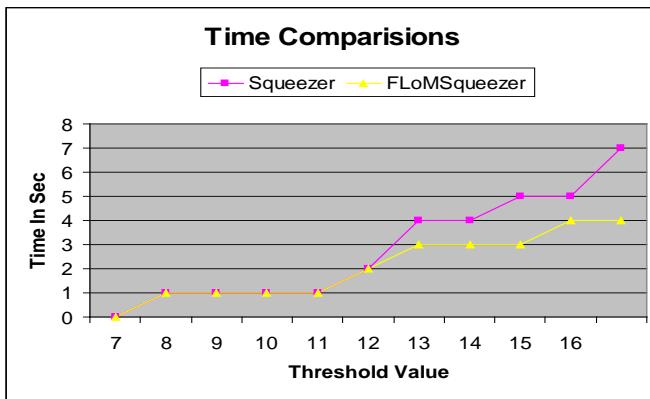


Fig.4 The execution time with different threshold values

The result of Fig.4 shows that the choice of St(threshold value) can affect both the quality of clustering and execution time for FLoMSqueezer. Thus, choosing a proper parameter value is one of the important tasks that must be considered in the FLoMSqueezer algorithm. The threshold St value footprint bound ‘g’ and initial sampling rate controls the decision to merge a new tuple into an existing cluster, or to place it in a cluster by itself. A good choice for ‘g’, p , and St is necessary to produce a concise and useful summarization of the data set.

8. Conclusions

An efficient algorithm, FLoMSqueezer is proposed, which produces high quality of clusters from categorical data streams using sampling technique. The performance of algorithm is tested with synthetic dataset and real-life dataset and found effective in terms of handling large stream of categorical data.

References

- [1] Babcock, S. Babu, M. Datar, R. Motwani, and J. Widom, “Models and issues in data stream systems”. In Proc of PODS, 2002.
- [2] Ding Q and Perrizo W, “Decision Tree Classification of Spatial Data Streams Using Peano Count Trees”, Proc of the ACM Symposium on Applied Computing, Madrid, Spain, March 2002.
- [3] Hand D J, “Statistics and Data Mining: Intersecting Discipline”, ACM SIGKDD Explorations, 1, 1, pp. 16-19, June 1999.
- [4] Golab L and Ozu M. “Issues in data stream management”. In SIGMOD Record, Vol. 32, No.2, pages 5–14, June 2003.

[5] He Z, Xu X and Deng S, “Squeezer: an efficient algorithm for clustering categorical data”, Journal of Computer Science & Technology, 17(5), pp.611-624(2002).

[6] Yi Li, Barbara D, Couto J, “Coolcat: An entropy-based algorithm for categorical Clustering”, In Proc of CIKM’02, pp. 582-589, 2002.

[7] You J, Yang Y, Guan X, ”Cloppe: A fast and effective clustering algorithm for transactional data”. In Proc of SIGKDD’02, Edmonton, Canada, 2002.

[8] Ganti V, Gehrke J, Ramakrishnan R. “CACTUS –clustering categorical Using Summaries” In Proc. of ACM SIGKDD, San Diego, California, USA, pp. 73—83, 1999.

[9] UCI Repository of Machine Learning Databases. (<http://www.ics.uci.edu/~mlearn/MLRRepository.html>)

[10] Guha S, Rastogi R and Shim K, “ROCK: A robust clustering algorithm for categorical attributes”, Proc of the IEEE Intl Conf on Data Engineering, Sydney, March 1999.

[11] Gibson D, Kleiberg J, Raghavan P. ”Clustering categorical data: An approach based on dynamic systems”. In Proc. Int. Conf. Very Large Databases, New York, August, 1998, pp.311-322.

[12] G.S Maku and R.Motwani, “Approximate frequency count over data stream”, Proceeding of 28th VLDB conference, Hong kong, China, 2002.



Marpe Sora obtained B.Tech form NERIST and M.Tech from Tezpur University. Presently he is Assistant professor in Rajiv Gandhi University, Department of Computer Science and Engineering Arunachal Pradesh. His main research interests include signal and speech processing and Data mining.



Swarup Roy did his M.Tech. in Information Technology and pursuing his Ph.D in Comp Sc & Engg. from Tezpur University. Presently he is an Assistant Professor in the department of Information Technology at North Eastern Hill University, Shillong. He is a recipient of university gold medal for securing first position in M.Tech. His research interest includes Data mining and Computational Biology. S Roy has published a number of papers in different International Journals and refereed Int'l. Conf. Proceedings and authored a book. He is a reviewer of few International Journals.



Sarangthem Ibotombi Singh obtained MCA from Manipur University. He is presently working as Assistant Professor in the Department of Computer Science & Engineering at Tezpur University. His current area of interest is Data mining, Spatial Database and Web Services.

iAgile: A Tool for Database Generation Guided by Graphical User Interface

Shaimaa Galal¹ and Ehab Hassanein²

¹ Information systems department, Cairo university
Cairo, Egypt

² Information systems department, Cairo university
Cairo, Egypt

Abstract

The agile development of the database and software systems is highly productive activity; it reduces time consumed, cost and effort invested in project development, but many agile projects do not apply agile practices to database development and still consider it in a serial manner as heavy-weight methodologies exactly work, while agile methodologies were introduced to overcome the problems experienced with the heavy-weight methodologies. The Enhanced Early Development of Graphical User Interface Practice Framework was introduced to enable performing the database development process in an evolutionary manner. In this article a proposed tool will be presented to help generating the final software product through applying and automating this framework to support agility in both directions of coding and data modeling as well, using such a tool will provide a high level of customer collaboration and help data professionals to work in an agile manner to avoid the problem of having overbuilt systems along with automatically generating portions of the software code based on available modern software architecture models.

Keywords: Agile database development; Agile data modeling; Graphical user interface; Data access Layer generation.

1. Introduction to Agile Database Development

Most data-oriented techniques are serial in nature, requiring the creation of fairly detailed models before implementation is “allowed” to begin. These models are often base line and put under change management control to minimize changes (at the end, this will be actually called a change prevention process) [3]. Agile data modeling allows data professionals to adopt evolutionary approaches to all aspects of their work; examples of those techniques are [3]:

1. **Database refactoring:** evolve an existing database schema by implementing few changes at every time to improve the quality of its design without changing its semantics.

2. **Evolutionary data modeling:** Model the data aspects of a system iteratively and incrementally, just like all other aspects of a system, to ensure that the database schema evolves in step with the application code.
3. **Database regression testing:** Ensure that the database schema actually works.
4. **Configuration management of database artifacts:** Your data models, database tests, test data, and so on are important project artifacts that should be managed just like any other artifact.
5. **Developer sandboxes:** Developers need their own working environments in which they can modify the portion of the system that they are building and get it working before they integrate their work with that of their teammates.

The Enhanced Early Development of Graphical User Interface practice Framework (EEUID framework for short) applied the **evolutionary data modeling** technique on the Early Development of Graphical User Interface practice [8, 9] to provide data professionals with an agile data modeling technique that will have the following advantages:

1. Remove frustration that happens during the development process, as developers ignore data professionals’ advice, standards, guidelines, and enterprise models, While developers often do not even know about these people and things in the first place, this problem illustrated in[3].
2. Accelerate the development process by automatically generating code portions that formulate essential parts of the system.
3. Increases the documentation level for the agile methodologies, which is considered a competitive advantage.

In this article, a tool is proposed to put the EEUID Framework in action, which in turn will help to reach an

evolutionary final product that is based on different available modern software architecture models.

2. Introduction to the Enhanced Early Development of Graphical User Interface

Agile methodologies have arisen since the last decade to fulfill the need of developing information systems more quickly in a competitive business environment [7].

There is a framework recently introduced – “the Enhanced Early Development of Graphical User Interface practice framework” [6] – the framework main objective is to generate a final software product through applying agility in both directions of coding and data modeling as well.

The EEUID process works in an evolutionary manner as shown in fig.1.

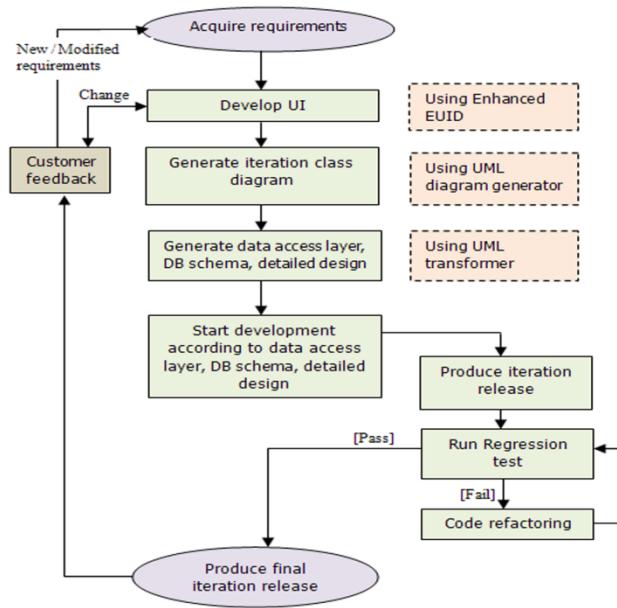


Figure 1: Enhanced EUID process

As shown in fig. 1, through each iteration start, requirements acquired from the customer, then GUI will be developed using the EEUID generator producing three outputs (HTML GUI structure, XML behavior file, XML class diagram description), hence class diagram is generated using UML diagram generator. From this point, the framework automatically generates the data access layer, DB schema and detailed design documents needed for this iteration, then developers can start the development process and produce release for this iteration. For each successive iteration a complete regression test should be applied for the iteration release if it fails, refactoring should be done, if it passes the test, iteration release is

presented to the customer for feedback and starting the new iteration.

The framework supports four types of available modern software architecture models; the proposed tool in this article, will apply only one model which is the **GUI layer and object relational access layer above relational DB model** [6].

3. Tool Overview

The iAgile studio is a proposed tool to implement the EEUID framework components; it will help system analysts to automatically generate the database schema, the data access layer and the quality documentation needed for future maintenance and testing purposes.

According to the iAgile Layout and content editor displayed in fig. 2, the visual specification of the project pages is constructed through three sets of controls:

1. The Early Development of Graphical User Interface practice components (EUID components for short), which are components of several types that are used to build main parts of the web page. In the next sub section, further details about these components will be illustrated.
2. Typical web development controls: Examples including Textboxes, Comboboxes and images. These are used for further addition\modification of the web page controls. Meanwhile, when a user drags and drop any of these, it must be attached to a particular page component, then events can be added to different types of controls. The tool offers a set of specified predefined actions such as saving, searching, adding, deleting data or moving to another page as an initial set of most frequently used actions. For further actions, the system analyst should write the Test First examples that will be implemented manually in that iteration development phase.
3. iAgile studio tools: help to build more interactions on the page, such as *simple math calculator* that is used to produce an output in a field from performing simple math calculations on other inputs of the same page, e.g. calculating a person's age from his/her date of birth.

According to fig. 1, at the beginning of each project iteration, the system analyst should start to construct the web pages GUI through the previously mentioned sets of tools; upon completion of this step, the tool will be able to generate the system accumulated class diagram up to this project iteration making use of the generated XML behavior file and XML class diagram description file.

From this point, the system analyst can generate the database schema, the data access layer and the detailed design documents for this iteration according to the chosen software architecture model; consequently a phase of further development should start to complete the missing functionalities using the Test First Examples generated with the web pages GUI and finally proceed to next iteration after considering the customer feedback.

be used for viewing data only or for editing or deleting. The most famous form for editing or deleting records in that component is through adding links for each record in the datagrid.

- 4. Navigator component:** There are several actions occurs when a link is clicked. For example, on clicking a link the user can move to another page or make changes on the same page.

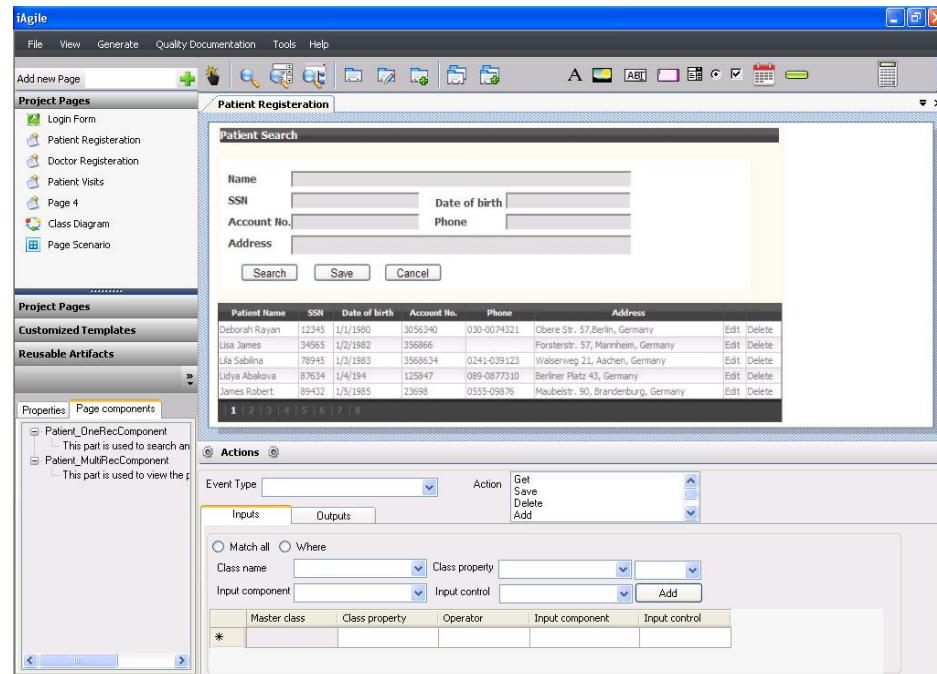


Figure 2: iAgile layout and content editor

3.1 The tool support for the early development of graphical user interface practice components

According to similarities between information systems web applications, the EUID components were proposed as essential constructs for each web page applying the EUID practice, there are four types of these components, which are [8, 9]:

- 1. Search component:** It allows users to search for certain data. This may come in different forms, for example, a combobox or a set of editable controls with a search button.
- 2. One record component:** It is responsible for adding, editing, or viewing a one record data. One famous form for that component is a set of controls in columnar representation.
- 3. Multi record component:** The Multi record component can be referenced to be a datagrid where data is represented in tabular form. It can

iAgile supports these types of components via the idea of that for every screen there is at least a conceptual master class and potential set of dependant classes. The properties of each class will be added from every screen to those classes in order to form the final class definition.

4. iAgile Software Architecture Design

One extreme importance is illustrating the relationship between engineering principles and architectural view; this section will illustrate a number of insights into what iAgile software architecture might be. Fig.3 shows the component diagram of iAgile architecture, which implements the EUID framework.

The software architecture is composed of:

- 1. The Enhanced Early Development of Graphical user Interface package:** designed for designing the pages' UI via constructing each page of the system in the form of one or several components

and it stores the classes' details associated with each screen. This package will generate the pages' user interface in the form of HTML, the XML class description files and the XML behavior files. The XML class description file structure is shown in fig. 4, and the XML behavior file structure is shown in fig. 5.

2. *The UML diagram generator component:* processes the “XML class description files” in order to generate the current system class diagram that is crucial part of the system which will be used mainly to generate the database schema.
3. *The UML transformer package:* will use the class diagram and XML behavior file to generate the final expected outputs of the system.

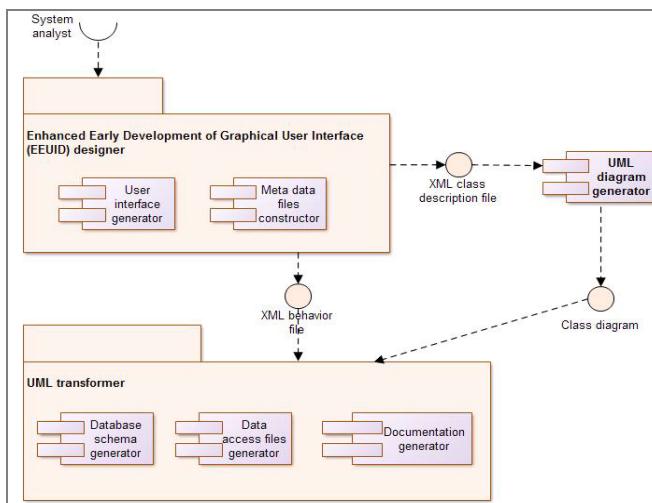


Figure 3: iAgile software architecture's component diagram

```

<?xml version="1.0" encoding="utf-8" ?>
<masterclass name="patients" classidentifier="patientsIID">
    <classproperties>
        <property name="Name" datatype="String"/>
        <property name="SSN" datatype="String"/>
        <property name="Dateofbirth" datatype="DateTime"/>
        <property name="accountno" datatype="Double"/>
        <property name="Phone" datatype="String"/>
        <property name="Address" datatype="String"/>
    </classproperties>
    <classoperations>
        <operation name="Search">
            <input datatype="String"/>
            <input datatype="String"/>
            <input datatype="DateTime"/>
            <input datatype="Double"/>
            <input datatype="String"/>
            <output datatype="Patients[]"/>
        </operation>
    </classoperations>
    <dependantclasses>
        <class name="PatientVisits"/>
    </dependantclasses>

```

Figure 4: EEUID designer's XML class description file structure

```

<?xml version="1.0" encoding="utf-8"?>
<components>
    <component type="searchbutton" id="component1">
        <active_event id="event1" control="button1" active="onclick">
            <action id="action1">
                <desc>
                    when button 1 is clicked the patient data will be fetched from the DB and viewed in the datagrid
                </desc>
                <input> component1.name_1</input>
                <input> component1.SSN_2</input>
                <input> component1.AccountNo_3</input>
                <input> component1.Dateofbirth_4</input>
                <input> component1.Phone_5</input>
                <input> component1.Address_6</input>
                <output>component2.datagrid1</output>
            </action>
        </active_event>
    </component>
    <component type="multirecord" id="component2">
        <active_event>
            <action id="action2" control="linkbutton2" active="onclick">
                <desc>
                    when edit link button is clicked the patient data for the clicked row in thisdatagrid should be saved.
                </desc>
                <input> component2.datagrid1.row1.PatientName_7</input>
                <input> component2.datagrid1.row1.SSN_8</input>
                <input> component2.datagrid1.row1.Dateofbirth_9</input>
                <input> component2.datagrid1.row1.AccountNo_10</input>
            </action>
        </active_event>
    </component>

```

Figure 5: EEUID designer's XML behavior file structure

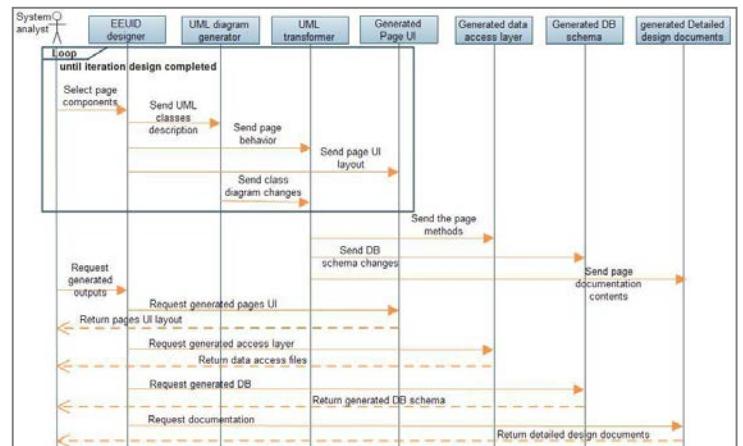


Figure 6: iAgile UML sequence diagram

A UML sequence diagram is shown in fig. 6 to illustrate iAgile process sequence. First, the system analyst starts to design the system pages, each time a component is added in any page, a class description for this page must be provided to indicate whether this class is master or dependant class and whether it have been used previously in other page or not, this is the most critical task to be done as all later steps will depend on this major step. Once the pages for this project iteration are completed, the system will be able to generate the final three outputs. All successive iterations are done the same way and modified outputs for the previous iteration will be generated smoothly.

4. iAgile in Action

This section will demonstrate a simple example of how iAgile will reach the expected final results. Consider a web page for “patients’ registration” as shown in fig. 7.

To construct the web page in fig. 7, the page will contain two components: Insert_OneRecord component and Navigation_Edit_Multirecord component.

Patient Name	SSN	Date of birth	Account No.	Phone	Address	Edit	Delete
Deborah Rayan	12345	1/1/1980	3056340	030-0074321	Obere Str. 57, Berlin, Germany	Edit	Delete
Lisa James	34565	1/2/1982	356866		Forsterstr. 57, Mannheim, Germany	Edit	Delete
Lila Sabrina	78945	1/3/1983	3568634	0241-039123	Walserweg 21, Aachen, Germany	Edit	Delete
Lidya Abakova	87634	1/4/1994	125847	089-0877310	Berliner Platz 43, Germany	Edit	Delete
James Robert	89432	1/5/1985	23698	0555-09876	Maubelstr. 90, Brandenburg, Germany	Edit	Delete

Figure 7: Patients’ registration web page

First, the user drags Insert_OneRecord component to the content editor and a screen for master class details will appear. If the data on the web page appears for the first time in the system, then the web page contains a *new master class*, details for the class properties that will appear on the screen should be completed e.g. to add patient name class property, then details should be as follow: Control Type → Text Box, name → Patient Name, Data Type → String, Length → thousand. The rest properties should be filled as shown in fig. 8.

and delete link buttons. Finally to add the extra button for the search criteria and view the results in the data grid, the user should drags a button and sets its action details from the panel in the most bottom of the screen shown on fig.2 to define the action inputs and outputs. After the system analyst finish constructing the web pages for this project iteration, a class diagram will be generated automatically making use of the iAgile metadata files represented in the XML behavior file and the XML class diagram description. iAgile will thereafter generate the three expected outputs, which are database schema, data access layer and detailed design documents, further details illustrated in next sections.

5.1 The tool support for the early development of graphical user interface practice components

Starting from the class diagram classes we need to transform objects into tables in the database. To store an object in a relational database, it should be flatten, i.e. create a data representation for this object. To retrieve the object, data retrieved from the database and then object is created — often referred to as restoring the object — based on that data [2]. There are existing approaches to generate the relational database (RDB) schema from a given class diagram [1, 2], but still it is complicated task as mappings can be done in several ways and choices for each mapping way to use should be defined.

From the previous generated class diagram, we can have RDB schema according to the following mapping choices [6]:

- a) *Mapping Meta-data:* Class properties are mapped to table columns:

Control type	Name	Data type	Length	Unique object identifier
Text Box	Patient Name	String	1000	<input checked="" type="checkbox"/>
Text Box	SSN	String	100	<input type="checkbox"/>
Text Box	Date of birth	DateTime		<input type="checkbox"/>
Text Box	Account No.	String	50	<input type="checkbox"/>
Text Box	Phone	String	50	<input type="checkbox"/>
Text Box	Address	String	1000	<input type="checkbox"/>

Figure 8: master class details

Second, the user drags Navigation_Edit_Multirecord component to have a data grid in the web page with edit

- b) *Inheritance Relationships*: are best mapped by mapping each class to table, due to the simplicity in implementation and understanding.
- c) *Relationships between classes*: will be maintained in relational databases through the use of foreign keys.
- d) *Object Identifiers*: will be maintained in relational databases through the use of primary keys.

5.2 Data Access Layer

To generate the appropriate classes and files for object relational (OR) data access layer an adequate ORM tool should be chosen first, large number of ready-made tools (O/R mapping tools) are present in the market due to the popularity of using such technologies according to its advantages discussed in [6, 5, 12]. These tools allow mapping between objects and RDB. Examples of these tools are Hibernate/NHibernate, Entity Framework, Open Access, Subsonic, Light Speed, Data Objects.Net and Hera Framework. iAgile here implements the NHibernate tool as it is the most usable ORM tool for .NET and java as well [11].

iAgile automatically generates the object relational access layer files that fit for the relational database schema using the NHibernate engine, the following files will be generated for each web page in the project:

- *Domain object (classes)* – contains class properties getters and setters and main class operation.
- *Hibernate/NHibernate hbm files* – these files represent Mapping Metadata between objects properties and relational table columns along with relationships between classes.

And only one file for the *physical database driver configuration* that contains necessary pieces of information in order to connect to data source. For the web page in fig. 6, two files will be generated which are Patients.cs and Patients.hbm.

5.3 Detailed Design Documents

In Agile, documentation is sparse – often limited to the source code and a set of user stories or UML diagrams. While reducing the amount of documentation can increase productivity, it does come at some risk and cost. Documentation serves as a way to bring new members up to speed. It is useful when transitioning the project to a maintenance team. From a business perspective, documents form the basis for audits assuring proper quality procedures are followed. Documentation serves as a domain knowledge repository and is necessary to retain critical information over time [4, 10].

iAgile builds quality document templates according to the information stored during the web pages GUI building process, one of the templates that can be produced is the detailed design document for the designed web pages. This will help more documentation and give agile methodologies a competitive advantage. Sample of the detailed design document is shown in fig. 9.

Detailed Design Form											
Web Page Name: Patients.html		Date: 01/05/2011									
System Version No.: 1		Screen Shot: E:\Project1\Patients.html									
Database Type: SQL server 2005		Database Name: EMR									
General Page Description: This page for searching patients information to start editing the electronic medical record											
<u>Page Events Description</u>											
Control name	Event type	Inputs	Outputs	Mapping to DB	Description	Test first examples					
Search button	OnClick	component1.name_1<<TextBox>> component1.SSN_2<<TextBox>> component1.AccountNo_3<<TextBox>> component1.Dateofbirth_4<<TextBox>> > component1.Phone_5<<TextBox>> component1.Address_6<<TextBox>>	component2.datagrid1	Inputs: Table: Patients Columns: Name, SSN, AccountNo., Dateofbirth, Phone, Address Outputs: Table: Patients Columns: Name, SSN, AccountNo., Dateofbirth, Phone, Address.	When button 1 is clicked the patient data will be fetched from the DB and viewed in the data grid.						

Figure 9: Detailed Design Document

The proposed detailed design document contains two main sections. The first section contains general information as web page name, creation date, iteration number, database engine type, etc. The second section explains page events by stating each control causing event for this page and full details for inputs, outputs, mapping to database, description and associated test first cases [6].

6. Conclusion and Future Work

There are proven practices employed in Agile Methodologies that, when applied under the right circumstances result in lower-risk projects and ultimately better productivity and quality. In this article, a proposed tool – iAgile – was presented to automatically generate a full software product using the EEUID framework that is based on robust software architecture in which developers can refactor later in order to finalize the needed product through several iterations; this will lead to even more productivity and speed up the project delivery time.

iAgile did not stop at the software product generation stage, it also included automatic documentation generation as documentation is a part of the agile software development, but as all agile approaches include a minimum of documentation, it makes documentation the responsibility of the development team to ensure enough

documentation is available for future maintenance and testing.

Future work should include the following items for better development process:

- a. Generation of the rest software architecture models as for any project there is not something called “best architecture model”, but there is “adequate architecture model” according to the project circumstances.
- b. Generation of different documentation types used for quality and testing purposes that can be formulated from the iAgile studio meta-data files.
- c. Supporting the bidirectionality of the Enhanced EUID framework illustrated in [6].
- d. It is highly recommended integrating the tool with other automated testing tools as NUnit and JUnit, this will lead the agile family to gain more competitive advantages.
- e. Enhancing the GUI generation part by integrating it with additional application of a picture editing tool, e.g. those used for sketching UI widgets. When more sophisticated UI behavior is necessary, embeddable objects such as Adobe Flash need to be generated separately. iAgile will help to model ordinary UIs, but otherwise needs to become part of an interrelated tool-chain.
- f. Finally integrating the tool with task management systems will allow easier project planning and task assignments for the project managers, as after finishing each iteration GUI design, the project manager can assess how much work still need to be done and start assigning missing functionalities to developers.

Acknowledgments

I am truly and deeply grateful to my professor Dr.Ehab Ezzat who encouraged and helped me to produce this work and enhanced my personality as well.

References

- [1] S. A., “Mapping objects to relational databases - What you need to know and why”, 2000.
- [2] S. A., “Agile Database Techniques—Effective Strategies for the Agile Software Developer”, New York: 2003.
- [3] S. A., and P. J. S., “Refactoring Databases: Evolutionary Database Design”, Addison-Wesley, 2006.
- [4] C. M., and B. S., “The Impact of Agile Methods on Software Project Management”, in Engineering of Computer-Based Systems conference, 2005.
- [5] A. D., and V. R., “Object-Relational Mapping Techniques for .Net Framework”, 2004.
- [6] Sh. G., E. H., “Applying Agile Methodology on Database Generation Guided by Graphical User Interface”, in

international conference on computational intelligence and software engineering, 2011.

- [7] J. H., “Agile software development ecosystems”, Boston: 2002, Addison Wesley.
- [8] C. L., E. H., and O. H., “Early development of graphical user interface (GUI) in agile methodologies”, in Informatics and Systems Conference, 2010.
- [9] C. L., E. H., and O. H., “Early Development of Graphical User Interface (GUI) in Agile Methodologies”, Journal of Computational Methods in Sciences and Engineering, V. 9, No. 1, 2009.
- [10] M. L., et al., “Empirical Findings in Agile Methods”, in Universe and First Agile Universe Conference on Extreme Programming and Agile Methods, 2002.
- [11] NHibernate 3.1.0, <http://www.hibernate.org/>, Last date accessed 5.2011.
- [12] W. Z.; N. R., “The Real Benefits of Object-Relational DB-Technology”, in British National Conference on Databases: Advances in Databases.

ENHANCING E-LEARNING WITH VRML TECHNIQUES

Sangeetha Senthilkumar¹, Dr.E.Kirubakaran²

¹ Research Scholar

Department of Computer Science

Vinayaka Missions University,

Salem, TamilNadu, India-636308

² Additional General Manager

Outsourcing Department

Bharat Heavy Electricals Limited,

Tiruchirappalli, TamilNadu, India-620014

Abstract - Virtual Reality (VR) is a computer-generated three-dimensional space that is multi-sensorial, interactive and engaging. Virtual reality is an artificial environment that is created with software and presented to the user in such a way that the user suspends belief and accepts it as a real environment. On a computer, virtual reality is primarily experienced through two of the five senses: sight and sound. This research paper is focused on enhancing E-Learning using the three dimensional Web Techniques namely the Virtual Reality Techniques. The present state of Techniques in Virtual Reality is static in nature.

Index terms -Virtual reality, multi-sensorial, modelling, dynamism, Blood pressure

I. INTRODUCTION

In this paper 'Dynamism' is imparted in the Virtual reality Learning Objects. The parameters are picked up dynamically from the Data base (which gets updated dynamically according to the real time situation).

The following three domains are taken up for implementing these techniques

- 1) Educational Field
- 2) Industrial applications
- 3) Medical Field

1) In the field of education, the characteristics of basic Learning Objects (for example basic geometrical shapes and colours) can be altered dynamically by the learner and the same is displayed in a 3 dimensional web.

2) In the field of Industrial applications, the basic manufacturing methodologies are created using a Virtual reality modelling. However the characteristics (for example the dimensions of the job, the speed of Welding the intensity of the welding arc etc.) are altered dynamically, depending on the specific job

3) Medicine is beginning to see the potential of VR technology to revolutionize the practice of medicine. One such application is Blood Pressure (BP) training tool. "Blood pressure" is the force of blood pushing against the walls of the arteries as the heart pumps blood. If this pressure rises and stays high over time, it can damage the body in many ways. Simulation of the BP training tool is made (using Virtual Reality techniques) and provision is made for altering the various parameters for studying.

II. SCOPE OF THE WORK

Virtual reality can work for educators as a tool in assisting students to become immersed in a learning environment where they can participate in their own learning in a technology based environment. Virtual Reality allows us to experience a body of knowledge interactively. Students learn while they are situated in the context where what they learn is to be applied. They get immediate feedback as they explore their understanding of the material.

Virtual Reality (VR) is becoming the enabling technology for Industrial developments. Designers can make alterations without having to scrap the entire model, as they often would with physical ones. The development process becomes more efficient and less expensive as a result.

In medicine, Virtual Reality based Blood Pressure tool is used for analysing the blood pressure. Blood is carried from the heart to all parts of your body in vessels called arteries. Blood pressure is the force of the blood pushing against the walls of the arteries. Each time the heart beats (about 60-70 times a minute at rest), it pumps out blood into the arteries. Various factors, such as age and gender influence average values, influence a person's average BP and variations. Also, an individual's BP varies with exercise, emotional reactions, sleep, digestion and time of day.

III. OBJECTIVES AND HYPOTHESES

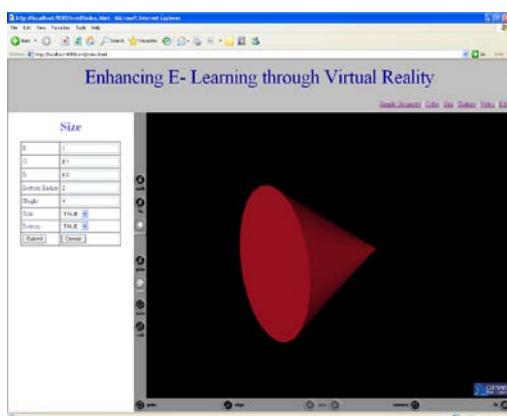
Educational Field

As human mind can comprehend 3 D more effectively than 2 Dimension, Web based Learning using 3 D is more effective than 2 D. The concept of both 2 D and 3 D Web is used by dividing the Web page into two frames one consisting of conventional 2D and the other one 3D. User gives his required input through the 2D Screen and the output is rendered in 3D, on the other window.

Study of relative sizes: The parameters for the size is given on the 2D and according to the size, 3D objects are formed. Different operations are chosen from the data base. For each of the records of the choice, different values are picked up from the data base and the VR is produced on the fly.

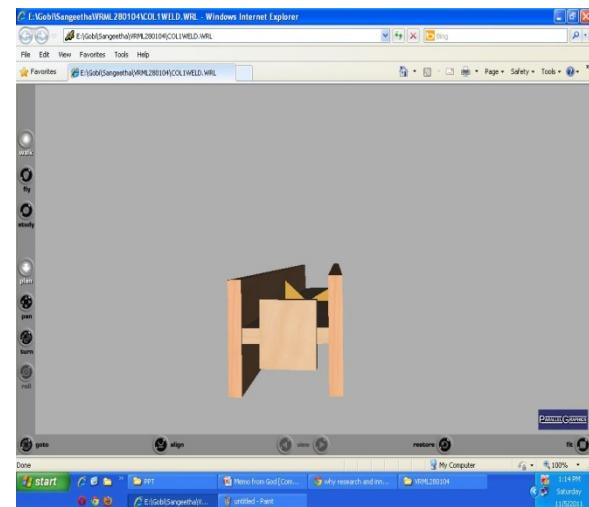
The conventional type of E-Learning limits to a single machine , but the Web based E-Learning extends to cover multiple users and the Web Based VR enhances Web Based Learning and Training.

Example1- Studying Geometric Shapes - Colours, Shapes



Industrial applications

In the Industrial front, if the beam welding is to be explained, the VR application is the best choice as this method gives the users an opportunity to get involved as if they are involved in the real life. They can view the welding process from different perspectives as they would like to and learn the technique better.



Medical Field

VR based Blood Pressure tool illustrates the Blood Pressure findings. Blood pressure is measured as systolic (sis-TOL-ik) and diastolic (di-a-STOL-ik) pressures. "Systolic" refers to blood pressure when the heart beats while pumping blood. "Diastolic" refers to blood pressure when the heart is at rest between beats.

Often the blood pressure numbers written with the systolic number above or before the diastolic number, such as 120/80 mmHg. (The mmHg is millimeters of mercury—the units used to measure blood pressure.)

The table below shows normal blood pressure numbers for adults. It also shows which numbers put you at greater risk for health problems.

Categories for Blood Pressure Levels in Adults (measured in millimeters of mercury, or mmHg)

Classification of blood pressure for adults		
Category	systolic, mmHg	diastolic, mmHg
Hypotension	< 90	< 60
Desirable	90–119	60–79
Prehypertension	120–139	or 80–89
Stage1 Hypertension	140–159	or 90–99
Stage2 Hypertension	160–179	or 100–109
Hypertensive Crisis	≥ 180	or ≥ 120

The ranges in the table apply to most adults (aged 18 and older) who don't have short-term serious illnesses.

IV. METHODOLOGY ADOPTED

These scenes are usually created with VRML. VRML 2.0 has many features like additional light effect, sound effect, fog effect, sensors to have interactions with outside world, VRML script to create additional events or objects or properties etc to make the interactions between the VR world and the user more effective.

VR applications play an important role by increasing the understanding of concepts and the techniques. This is possible because of the involvement of the user with the 3-D world and because of body centred interactions. Users can feel the situation as they are immersed within the virtual world. Percentage of immersion depends on many parameters like the created scene, network speed, computer's ability to render the scene smoothly etc.

But the scenes created by VRML lacks dynamism i.e, the scene presented will always be static. Users will get same scene again and again and it is up to the users explore the contents differently as they are free to navigate in different directions. This will give them a sense of dynamism but actually dynamism is not part of the VR applications, created with VRML.

V. RESULTS AND DISCUSSION

As VR applications enable the user to have interaction with the virtual world, their involvement is higher and as a result their understanding also is higher. Because of this reason, VR applications are used very much in different domains.

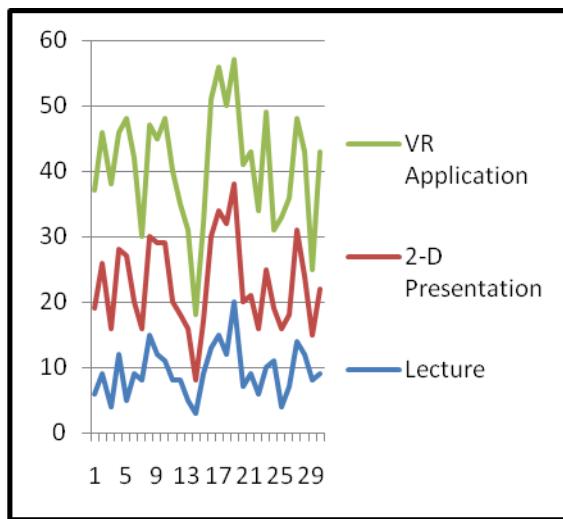
In this report, only three domains viz education, industry and medical, are considered and a study of improving the effectiveness of the scenes.

If any concept is to be explained, it is possible to do it with explanation, demo with 2-D and with 3-D scenes. Among the three the third one is always better because of its impact on the user and the understanding of the users. This is proved with the study conducted on three groups with same intelligence level. These groups were administered with questionnaire and the third group comes up to the top among the three levels.

The following tabular column illustrates the three groups of same intelligence level undergoing training by various methods.

Roll No	Lecture	2-D Presentation	VR Application
101	6	13	18
102	9	17	20
103	4	12	22
104	12	16	18
105	5	22	21
106	9	11	22
107	8	8	14
108	15	15	17
109	12	17	16
110	11	18	19
111	8	12	20
112	8	10	17
113	5	11	15
114	3	5	10
115	9	8	15
116	13	17	21
117	15	19	22
118	12	20	18
119	20	18	19
120	7	13	21
121	9	12	22
122	6	10	18
123	10	15	24
124	11	8	12
125	4	12	17
126	7	11	18
127	14	17	17
128	12	12	19
129	8	7	10
130	9	13	21

Graphical Representation



In the Industrial front, if beam welding is to be explained, the VR application is better as this method gives the users an opportunity to get involved as if they are involved in the real life. They can view the welding process from different perspectives as they would like to and learn the technique better.

Same thing applies to medical field also. For example, doctors can perform the operation on a virtual body and they can monitor the blood pressure level, heart beats, general improvement or the deterioration of the body etc.

They can do the operation any number of times till they are equipped to perform the operation on the real body. Thus a VR model is developed to train the health care providers to find the abnormalities in the Blood Pressure.

Issues connected with VR applications are:

- Multi-user environment
- Multi-server environment
- Shared behaviours
- Persistence

To overcome these drawbacks, a new method of dynamic generation of virtual scenes for use in Internet virtual reality applications is presented. The virtual scenes are dynamically generated from virtual scene models coded in a high-level XML-based language called X-VRML. X-VRML combines the features of VRML and XML. VRML takes care of the 3-D presentation whereas the XML takes care of the data meant for it.

The X-VRML language consists of a set of XML tags introducing new elements to programming of virtual reality. These new elements enables the VR application creation with object-orientation, access to databases, and programming techniques known from procedural languages like variables, conditions, and loops. The X-VRML language overcomes the main limitations of current virtual reality systems.

Use of X-VRML simplifies the code of virtual scene models, allows retrieval of data from databases, selection of virtual scene contents, customization of virtual scenes, and efficient coding of elements that have repetitive structure. Applications of X-VRML include on-line data visualization, geographical information systems, scientific visualization, virtual games, and e-commerce applications such as virtual shops.

In the current scenario, XML plays many roles viz creation of new languages like CML, MML etc, data exchanges among the application components, dynamic data exchanges between server and the client (AJAX). These features are merged with VRML to make 3-D scene creation easy and effective as many of the application servers are capable of interpreting and processing of data in XML format, by using XML parsers.

VI. CONCLUSION

On the whole, VR systems are much safer and, in the long run, less expensive than alternative training methods. This VR based Blood Pressure training tool helps the trainer to study about various alterations in the blood pressure. The biggest challenge in using VR technology to perform procedures is latency, since any delay in such a delicate procedure can feel unnatural to the user. Such systems also need to provide finely-tuned sensory feedback to the user. Thus our tool is developed considering all constraints such that it provides a convenient training methods.

ACKNOWLEDGEMENT

The authors would like to thank the Lord God Almighty for pouring his grace and strength upon us. Next, we would like to thank technical staff members of Vinayaka Misson University for their valuable contribution and continuous motivation.

BIBLIOGRAPHY / REFERENCES

- [1] Roudavski, S. (2010). Virtual Environments as Techno-Social Performances: Virtual West Cambridge Case-Study, in CAADRIA2010: New Frontiers, the 15th International Conference on Computer Aided Architectural Design Research in Asia, ed. by Bharat Dave, Andrew I-kang Li, Ning Gu and Hyoung-June Park, pp. 477–486
- [2] Robles-De-La-Torre G. The Importance of the Sense of Touch in Virtual and Real Environments. IEEE Multimedia 13(3), Special issue on Haptic User * Interfaces for Multimedia Systems, pp. 24–30 (2006).
- [3] Stanney, K. M. ed. (2002). Handbook of Virtual Environments: Design, Implementation, and Applications. Lawrence Erlbaum Associates, Inc., Mahwah, New Jersey
- [4] Hayward V, Astley OR, Cruz-Hernandez M, Grant D, Robles-De-La-Torre G. Haptic interfaces and devices. Sensor Review 24(1), pp. 16–29 (2004).
- [5] “A meta analysis of the training effectiveness of Virtual Reality Surgical Simulators “, Syed Haque, Member IEEE & Shankar Srinivasan, 2006.
- [6] wordnetweb.princeton.edu/perl/webwn
- [7] www.dcri.duke.edu/patient/glossary.jsp
- [8] www.mywhatever.com/cifwriter/content/66/4620.html

Knowledge Acquisition and Knowledge Management through E_Learning

V.Thennarasu¹, Dr.E.Kirubakaran²

¹ Research Scholar

Department of Computer Science
Vinayaka Missions University, Salem
TamilNadu, India, 636308

² Additional General Manager

Outsourcing Department

Bharat Heavy Electricals Limited,
Tiruchirappalli, TamilNadu, India, 620014

Abstract:

E-Learning has become increasingly hot in the past decade but is still on its way to a new pinnacle and now E-learning comprises all forms of electronically supported learning and teaching. The Information and communication systems, whether networked or not, serve as specific media to implement the learning process. The term will still most likely be utilized to reference out-of-classroom and in-classroom educational experiences via technology, even as advances continue in regard to devices and curriculum.

Key words: E_Learning, Knowledge Management, Knowledge Acquisition,

1. INTRODUCTION

Human beings are designed to learn from one another with both verbal and visual hints in order to retain new knowledge. In this process e-learning has been playing an important part of late. With more and more youngsters using the internet these days, e-learning is becoming popular in countries like India. This seems to be a blessing of the new-generation technology which makes teaching possible anytime and anywhere.

E_learning is essentially the computer and network-enabled transfer of skills and knowledge. E_learning applications and processes include Web-based learning, computer-based learning, virtual classroom opportunities and digital collaboration

In E_learning, small group collaborative learning has been shown to result in higher achievement, less stress and greater student satisfaction, and greater appreciation for diversity. Some educators suggest that it may be particularly important and well suited to the online environment as a way of incorporating the learning process into a virtual environment

2. E-LEARNING

The delivery of a learning, training or education program by electronic means. E_learning involves the use of a computer or electronic device (e.g. a mobile phone) in some way to provide training, educational or learning material.

3. E-LEARNING STANDARDS

Provide fixed data structures and communication protocols for E-Learning objects and cross-system workflows. E-Learning standards can be categorized as:

- a) Metadata: how to describe e-Learning resources in a consistent manner
- b) Content packaging: how to gather the resource useful bundles
- c) Communication Interface or API: how e-Learning resources can exchange information dynamically
- d) Learners Profile: how to share information about learners

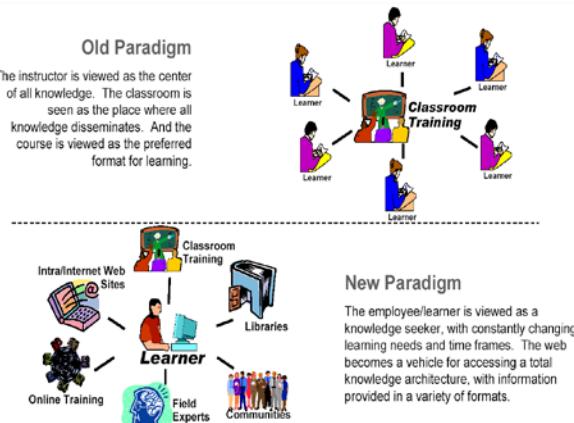


Figure-1: Learning Paradigm Old Vs New

4. REQUIREMENTS IN E-LEARNING

To assess the knowledge level of the students who are all completed their specific courses through E_Learning methodology. Knowledge acquisition can be achieved rapidly through E_Learning rather than traditional method of teaching.

Employees of any manufacturing companies can easily learn their task within the stipulated time frame and also completed their work with more accuracy. Comparison between traditional class room teaching and online teaching and also analyzing how much the receiver can achieve their goals.

In Expert-driven Knowledge Acquisition experts and/or the knowledge engineer's roles are minimized (or eliminated). They may provide useful preliminary knowledge discovery and acquisition, Increase the productivity of knowledge engineering and Computer support can eliminate some limitations. Students of the educational institutions and an employee of an organization can test their knowledge level and their learning performance through this system.

Dependency is very less whenever they have chosen E_learning methodology for their career. E_Learners are more competent than the traditional method learners. Going forward to overcome the difficulties of knowledge gaining each and every one should adopt E_learning methodology.

5. OBJECTIVES

The main aspiration of this research is to develop an E_learning tool to assess the knowledge level of the students and make the comparative analysis between traditional teaching and online teaching. This research investigates the likelihood of using the E_Learning tool to classify the different categories of students in terms of their knowledge level. The intention of this research is

- a) To design an E_learning tool and make it user friendly to learn their task in a short span of time.
- b) To establish an E-Learning structure which best go with C# and .Net using class room environment and find the efficient features for improving the performance of E_Learning system.
- c) To successfully implement the E Learning tool in the class room environment and the industry and also identify their performance.
- d) To prove that E- Learning gives supporting environment to all individuals for learning methods.
- e) To demonstrate that students can learn independently in any time and place. E

Learning is self-paced and the learning sessions are available all time.

- f) To build a database and use this database to store the knowledge level of the user and prepare the comparative chart with traditional teaching
- g) To prove that always online teaching is having more value than our traditional method in so many ways.

6. METHODOLOGY ADOPTED

E_Learning tool was designed in C# and .NET Microsoft SQL Server2008 is the database which is used to store the data whenever the user accesses this tool for their learning purpose.

E_learning technologies are generally categorized as asynchronous or synchronous. Asynchronous activities use technologies such as blogs, wikis, and discussion boards. The idea here is that participants may engage in the exchange of ideas or information without the dependency of other participant's involvement at the same time. Electronic mail (Email) is also asynchronous in that mail can be sent or received without having both the participants' involvement at the same time. Asynchronous learning also gives students the ability to work at their own pace. This is particularly beneficial for students who have health problems. They have the opportunity to complete their work in a low stress environment.

7. BASIC REQUIREMENTS FOR E-LEARNING EVALUATION

The purpose of E-Learning performance evaluation is to promote learning, whose requirements are mainly as follow:

- E-Learning performance evaluation is formative evaluation, which means that the focus of evaluation shifts from the learning results to the learners' learning process. This is the basic idea of E-Learning performance evaluation, which has been widely accepted by the education circle.
- The combination of autonomous evaluation and multi-object, that is, the evaluation of E-Learning performance should focus on the returning of autonomy, combining the self-evaluation with teacher evaluation and group evaluation.
- The content of the evaluation should be comprehensive, that is, increasing the effectiveness of online learning performance evaluation through evaluation of different perspectives and at different levels.
- Emphasize the principle of diverse evaluation technology, combining digitization and humanization.

8. RESULTS AND DISCUSSION

In order to examine the Knowledge level of the student through E_Learning a study was conducted with Computer Science students were involved in this study. Data was collected through the administration of a detailed 40 question survey distributed in-person in hard copy form to 60 students. The survey was designed to assess students' technology access, skills, and usage; prior experiences with e-learning, course delivery preferences, perceived satisfaction with e-learning, and perceptions of, and preferences towards various e-learning components.

Response Category	SA	A	N/U	D
Overall I was satisfied with the course Website.	42.9 %	47.9%	8.6%	0.7%
I found the course Website to be a helpful resource.	40.0 %	50.0%	9.3%	0.7%
I used the course Website to help me understand course information.	37.0 %	39.1%	16.7%	7.2%
I regularly used the course Website to answer my questions.	26.1 %	40.6%	23.2%	10.1%
I believe that course Websites enhance learning	25.7 %	35%	26.4%	12.9%
I would like to see course Websites added to all of my courses.	44.1 %	25%	14%	16.9%
I believe that course Websites will play an important role in college education in the future.	44.2 %	40.6%	8.7%	6.5%

SA= Strongly Agree, A=Agree,

N/U=Neutral/Undecided, D=Disagree,

Table-1: Knowledge level of the student

This study examines student's E_Learning experiences, Knowledge level and their preferences. Overall, students were satisfied with the course Website (47.9% agree and 42.9% strongly agree), found the course Website to be a helpful resource (50% agree and 40% strongly agree), used the course Website to understand course information (39.1% agree and 27.0% strongly agree), and regularly used the course Website to answer their questions (40.6% agree and 26.1% strongly agree). Most students felt that course Websites enhance learning (35.0% agree and 25.7% strongly agree), should be added to all of their courses (44.2% agree and 31.2% strongly agree), and will play an increasingly more important role in college education in the future (40.6% agree and 44.2% strongly agree). These findings are depicted in Table1

Knowledge level of the students from various semester in the department of MCA were assessed through the traditional method and E_Learning method. Chart1 shows the result.

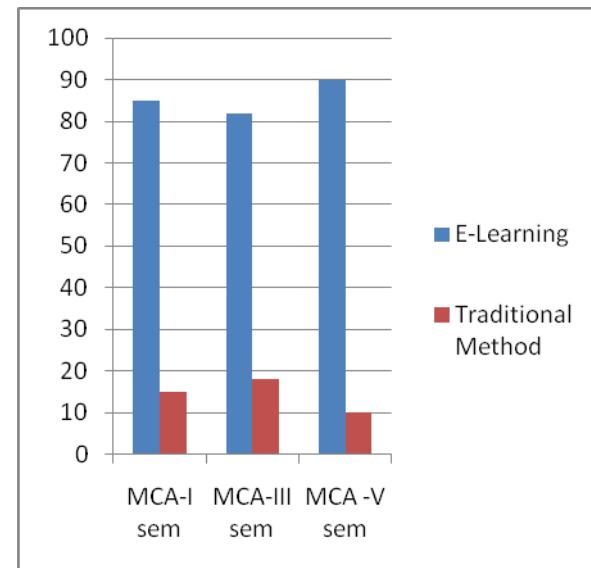


Figure-2: Comparison from E_learning and Traditional Method

We could see the tremendous changes in the knowledge level of the students those who were completed the specific task in both instructor Led Training method and Web based Tool method. Chart 2 shows the result of the students from various semesters in Computer Science and Engg department.

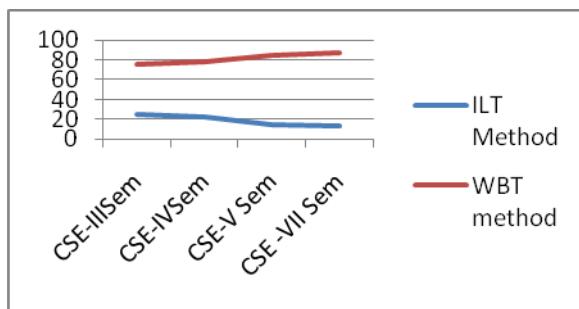


Figure-3: Comparison between Instructor Led Training and Web Based Training

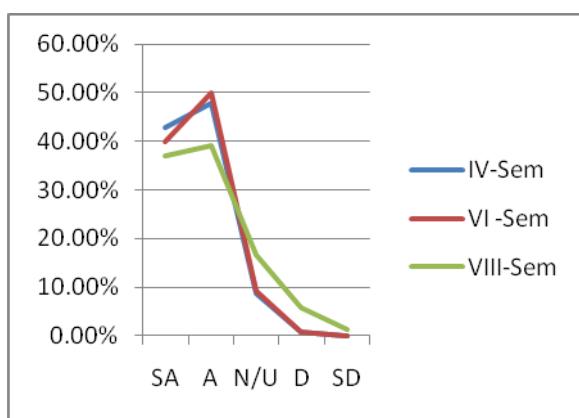


Figure-4: Acceptance level of E_Learning

The above chart shows the acceptance level of the E_Learning system by the non computer science students.

Benefits of using E-Learning component are

- Increased interaction by the students with the material outside of class
- Increased access to materials provided to the students
- Increased discussion among students and between students and instructor
- Increased efficiency for the instructor in time management
- Increased creative outlet for the instructor, and increased freedom for instructor to travel or reside away from the home institution.
- One can log-in and get all blended learning courses on the World Wide Web environment.

Students can learn independently in any time and place. E-Learning is self-paced and the learning sessions are available all time. Students can customize the course material as per their own needs. They have added control over their learning process and are able to better understand the subject.

E-Learning gives supporting environment to all individuals for learning methods.

Students get single central location for all course materials. Students get a chance for enhanced exchange with other students and qualified teachers which are based on communication and information technologies.

E-learning provides improved organization for regular studies like meeting assignment deadlines, homework etc. Students can deal with teachers who are highly qualified, but cannot reach because of distance barriers, now with e-learning coming in scene

Interaction with students becomes more appropriate. When students are sharing their problems with teachers, since it's not face to face they fear less and can ask their problems freely

9. CONCLUSION

Teaching aspects completely achieved by virtual teaching. The real teacher profits from the direct physical contact with the student to identify the needs of the latter, to answer these questions, the real teacher, by sharing time and space with the student, may make his course dynamic according to the face-to-face interaction with the student.

The virtual course is, by nature, lifeless if the powers of e-learning are not used. We can call for an invisible teacher who should continuously supervise the scene between the Web-based e-learning platform and its user (student). This teacher, who may be a neural system, provides real-time assistance by offering demonstrations, modifying the course, clarify confusing points, asking question progressing in complexity, reminding fundamental aspects and previous explanations.

We could see a tremendous improvement in the knowledge level of the students when they are using this E_Learning tool for their learning purpose. Thus the system was designed and implemented successfully and proved that always online teaching is more effective than traditional face-to-face teaching method.

ACKNOWLEDGEMENT

The authors would like to thank the Lord God Almighty for pouring his grace and strength upon us. Next, we would like to thank Vinayaka Misson University. We would also like to thank Jayaram College of Engineering and Technology staff and students for its valuable contribution and continuous motivation.

REFERENCES

- [1] Basir, H.M., Ahmad, A.,Noor, N.L.M, " Institutional strategy for effective blended e-learning: HCI perspective of sustainable embedding" User Science and Engineering (i-USer), 2010
- [2] Introduction to Knowledge Management by Todd R.Groff and Thomas P.Jones.
- [3] Knowledge Management in the Intelligence Enterprise by Edward Waltz
- [4] Basics of E-Learning by Elisabeth Rossen and Darin Hartley
- [5] The e-Learning Handbook: Past Promises, Present Challenges - by Saul Carliner and Patti Shank
- [6] The New Virtual Classroom: Evidence-based Guidelines for Synchronous e-Learning by Ruth Colvin Clark and Ann Kwinn
- [7] Study was made on e-Learning—Today's Challenge, Tomorrow's Reality.
- [8] Barron, T. "A Smarter Frankenstein: The Merging of E-Learning and Knowledge Management", ASTD Learning Circuits,
- [9] Gerry Stahl, Timothy Coachman, Dan Suthers, "computer-supported collaborative learning: an historical perspective", Cambridge handbook of the learning sciences, 2006.
- [10] Schmidt, A., 2005. Knowledge Maturin and the Continuity of Context as a Unifying Concept for Knowledge Management and ELearning, Proceedings of I-KNOW '05, Special Track on Integrating Working and Learning (IWL), Graz, Austria

A Survey of Web Crawler Algorithms

Pavalam S M¹, S V Kashmir Raja², Felix K Akorli³ and Jawahar M⁴

¹ National University of Rwanda
Huye, RWANDA

² SRM University
Chennai, INDIA

³ National University of Rwanda
Huye, RWANDA *Email address*

⁴ National University of Rwanda
Huye, RWANDA

Abstract

Due to availability of abundant data on web, searching has a significant impact. On-going researches place emphasis on the relevancy and robustness of the data found, as the discovered patterns proximity is far from the explored. Inspite of their relevance pages for any search topic, the results are huge to be explored. Also the users' perspective differs from time to time from topic to topic. Usually ones' want is others unnecessary. Crawling algorithms are thus crucial in selecting the pages that satisfies the users' needs. This paper reviews the researches on web crawling algorithms used on searching.

Keywords: *web crawling algorithms, crawling algorithm survey, search algorithms*

1. Introduction

These are days of competitive world, where each and every second is considered valuable backed up by information. Timely Information retrieval is a solution for survival. Due to the abundance of data on the web and different user perspective, information retrieval becomes a challenge .

When a data is searched, hundreds and thousands of results appear. The user's don't have persistence and stretch to go through each and every page listed. So the search engines have a bigger job of sorting out the results, in the order of interestingness of the user within the first page of appearance and a quick summary of the information provided on a page.

Web crawlers are programs which traverse through the web searching for the relevant information [1] using algorithms that narrow down the search by finding out the most closer and relevant information. This process is iterative, as long the

results are in closed proximity of user's interest. The algorithm determines the relevancy based on the factors such as frequency and location of keywords.

Web pages needs not only relevance but also authoritativeness – from a trusted source of strong, precise information [2]. Search engines uses algorithms which sorts, ranks the result in the order of authority, that is closer to the user's query. Many algorithms are in use - Breadth first search, Best first search, Page Rank algorithm, Genetic algorithm, Naïve Bayes classification algorithm to mention a few.

There are chances that the website may not contain the keyword, but they are completely relevant website. For example if the user is searching for cars, then the result returned are information about used cars for sale rather than information about cars manufacturers website.

Not all information represented are useful. The search engine techniques may become useless or junky if the information it draws are not attracting users, especially if the malicious user who are trying to attract more traffic in to their site by embedding the most used keywords invisibly in to their site. The challenges are relevancy, robustness and the ability to download large number of pages.

2. Fundamentals of web crawling

Crawlers have bots that fetches new and recently changed websites, and indexes them. By this process billions of websites are crawled and indexed using algorithms (which are usually well

guarded secrets) depending on a number of factors. Several commercial search engines change the factors often to improve the search engines process.

It generally starts with a set of URLs from the previous crawl, visits each of these websites, detects links and adds it to the list of links to crawl. It also notes whether there is any new website or website that has been recently changed (updated), websites that are no more in use and accordingly index is updated.

The indexer compiles the list of words it sees and its location on each page for future consultation. The information compiled are mostly because crawlers are majoritively text based.

When a user initiates a search, the key words are extracted and searches the index for the websites which are most relevant. Relevancy is determined by a number of factors and also it differs for the different search engines.

2.1 How the targets are selected?

The size of the web is huge, search engines practically can't be able to cover all the websites. Only 60 percentage are the indexed web [3]. There is a high chance of the relevant pages in the first few downloads, as the web crawler always download web pages (in fractions). This calls for measures for prioritizing Web pages. The importance of a page is a function of its essential quality, its reputation in terms of links or visits, and even of its URL. Different researchers used different strategies such as breadth first, depth first, page rank for selecting the websites to be downloaded.

2.2 Where to start?

We have to start from any URL (Seed), but imagine that the starting URL couldn't reach all the web pages or even the pages referenced by seed URL doesn't reference it back, which eventually makes us to restart the crawl. It is always better to have a good seed URL – pages that has been submitted to them by majority users around the world. For example yahoo or Google.

2.3 Any restrictions on the number of pages to follow (Link)

There is a cost associated with crawling, indexing and storing the results. When the web gets bigger and bigger, the “better” pages are downloaded. So there needs to be a scheduling strategy to minimize crawling time and to reduce cost [4] and it differs from one search engine to another. As the web is huge and to download as many pages as possible, parallel crawlers are distributed so that multiple downloads can be carried out in parallel [5]

2.4 Freshness of a page and revisiting policy

When the same copy exists in the local as well as the remote sources, then it is considered to be the “fresh” page. Cho and Garcia [6] calculated the freshness of a page as

$$F(e_i; t) = \begin{cases} 1 & \text{if } e_i \text{ is up-to-date at time } t \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

Where e_i is the element of database

And the age of a page as

$$A(e_i; t) = \begin{cases} 0 & \text{if } e_i \text{ is up-to-date at time } t \\ t - t_m(e_i) & \text{otherwise.} \end{cases} \quad (2)$$

Where $t_m(e_i)$ is the time of first modification of e_i after the most recent synchronization. The freshness drops to zero when the real-world element changes and the age increase linearly from that point on. When the local element is synchronized to the real-world element, its freshness recovers to one, and its age drops to zero.

Two types of visiting policy has been proposed – Uniform change frequency - the revisiting is done at the uniform regardless of its change and non uniform change frequency – the revisiting is not uniform and the revisiting is done more frequently and the visiting frequency is directly proportional to the change frequency.

3. Web crawler strategies:

3.1 Breadth First Search Algorithm:

This algorithm aims in the uniform search across the neighbour nodes. It starts at the root node and searches the all the neighbour nodes at the same level. If the objective is reached, then it is reported as success and the search is terminated. If it is not, it proceeds down to the next level sweeping the

search across the neighbour nodes at that level and so on until the objective is reached. When all the nodes are searched, but the objective is not met then it is reported as failure.

Breadth first is well suited for situations where the objective is found on the shallower parts in a deeper tree. It will not perform so well when the branches are so many in a game tree, especially like chess game and also when all the path leads to the same objective with the same length of the path [7] [8].

Andy yoo et al [9] proposed a distributed BFS for numerous branches using Poisson random graphs and achieved high scalability through a set of clever memory and communication optimizations.

3.2 Depth First Search Algorithm

This powerful technique of systematically traverse through the search by starting at the root node and traverse deeper through the child node. If there are more than one child, then priority is given to the left most child and traverse deep until no more child is available. It is backtracked to the next unvisited node and then continues in a similar manner [10].

This algorithm makes sure that all the edges are visited once breadth [11]. It is well suited for search problems, but when the branches are large then this algorithm takes might end up in an infinite loop [8].

3.3 Page Rank Algorithm

Page rank algorithm determines the importance of the web pages by counting citations or backlinks to a given page [12]. The page rank of a given page is calculated as

$$PR(A) = (1-d) + d \left(\frac{PR(T1)}{C(T1)} + \dots + \frac{PR(Tn)}{C(Tn)} \right)$$

PR(A) → Page Rank of a Website,

d → damping factor

T₁,...,T_n → links

Yongbin Qin and Daoyun Xu [13] proposed an algorithm, taking the human factor into consideration, to introduce page belief recommendation mechanism and brought forward a balanced rank algorithm based on PageRank and

page belief recommendation which ultimately attaches importance into the subjective needs of the users; so that it can effectively avoid topic drift problems. Tian Chong [14] proposed a new type of algorithm of page ranking by combining classified tree with static algorithm of PageRank, which enables the classified tree to be constructed according to a large number of users' similar searching results, and can obviously reduce the problem of Theme-Drift, caused by using PageRank only, and problem of outdated web pages and increase the efficiency and effectiveness of search. J.Kleinberg [15] proposed a dynamic page ranking algorithm. Shaojie Qiao [16] proposed a new page rank algorithm based on similarity measure from the vector space model, called SimRank, to score web pages. They proposed a new similarity measure to compute the similarity of pages and apply it to partition a web database into several web social networks (WSNs)

3.4 Genetic Algorithm

Genetic algorithm is based on biological evolution whereby the fittest offspring is obtained by crossing over of the selection of some best individuals in the population by means of fitness function. In a search algorithm solutions to the problem exists but the technique is to find the best solution within specified time [17].

[18] shows the genetic algorithm is best suited when the user has literally no or less time to spend in searching a huge database and also very efficient in multimedia results. While almost all conventional methods search from a single point, Genetic Algorithms always operates on a whole population. This contributes much to the robustness of genetic algorithms. It reduces the risk of becoming trapped in a local stationary point [19]. The applicability of Genetic Algorithms by various researchers [20], [21], [22], [23], [24], [25], [26] has been depicted in [27].

3.5 Naïve Bayes classification Algorithm

Naïve Bayes algorithm is based on Probabilistic learning and classification. It assumes that one feature is independent of another [28]. This algorithm proved to be efficient over many other approaches [29] although its simple assumption is not much applicable in realistic cases [28].

Wenxian Wang et al [30] proposed an efficient crawler based on Naïve Bayes to gather many relevant pages for hierarchical website layouts. Peter Flach and Nicolas Lachiche [31] presented

Naïve Bayes classification of structured data on artificially generated data.

3.6 HITS Algorithm

This algorithm put forward by Kleinberg is previous to Page rank algorithms which uses scores to calculate the relevance [32]. This method retrieves a set of results for a search and calculate the authority and hub score within that set of results. Because of these reasons this method is not often used [2].

Joel C. Miller et al [33] proposed a modification on adjacency matrix input to HITS algorithm which gave intuitive results.

4. Conclusion:

The main objective of the review paper was to throw some light on the web crawling algorithms. We also discussed the various search algorithms and the researches related to respective algorithms and their strengths and weaknesses associated. We believe that all of the algorithms surveyed in this paper are effective for web search, but the advantages favors more for Genetic Algorithm due to its iterative selection from the population to produce relevant results .

References:

- [1] Pavalam S M, Jawahar M, Felix K Akorli, S V Kashmir Raja “ Web Crawler in Mobile Systems” International Conference on Machine Learning (ICMLC 2011), Vol. , pp
- [2] Alessio Signorini, “A Survey of Ranking Algorithms” retrieved from <http://www.divms.uiowa.edu/~asignori/phd/report/a-survey-of-ranking-algorithms.pdf> 29/9/2011
- [3] Maurice de kunder, “Size of the world wide web”, retrieved from <http://www.worldwidewebsize.com/> 8/8/11
- [4] Ricardo Baeza-Yates, Ricardo Baeza-Yates “Crawling a Country: Better Strategies than Breadth-First for Web Page Ordering” , Proc. WWW 2005.
- [5] Marc Najork, “Web Crawler Architecture” retrieved from <http://research.microsoft.com/pubs/102936/EDS-WebCrawlerArchitecture.pdf> accessed on 10/8/11
- [6] Junghoo Cho and Hector Garcia-Molina “Effective Page Refresh Policies for Web Crawlers” ACM Transactions on Database Systems, 2003.
- [7] Steven S. Skiena “The Algorithm design Manual” Second Edition, Springer Verlag London Limited, 2008, Pg 162.
- [8] Ben Coppin “Artificial Intelligence illuminated” Jones and Barlett Publishers, 2004, Pg 77.
- [9] Andy Yoo,Edmond Chow, Keith Henderson, William McLendon, Bruce Hendrickson, AUnit CatalyAurek “A Scalable Distributed Parallel Breadth-First Search Algorithm on BlueGene/L” ACM 2005.
- [10] Alexander Shen “Algorithms and Programming: Problems and solutions” Second edition Springer 2010, Pg 135
- [11] Narasingh Deo “Graph theory with applications to engineering and computer science” PHI, 2004 Pg 301
- [12] Sergey Brin and Lawrence Page “Anatomy of a Large scale Hypertextual Web Search Engine” Proc. WWW conference 2004
- [13] Yongbin Qin and Daoyun Xu “A Balanced Rank Algorithm Based on PageRank and Page Belief recommendation”
- [14] TIAN Chong “A Kind of Algorithm For Page Ranking Based on Classified Tree In Search Engine” Proc International Conference on Computer Application and System Modeling (ICCASM 2010)
- [15] J.Kleinberg “Authoritative sources in a hyperlinked environment”, Proc 9th ACM-SIAM Symposium on Discrete Algorithms, 1998.
- [16] Shaojie Qiao, Tianrui Li, Hong Li and Yan Zhu, Jing Peng, Jiangtao Qiu “SimRank: A Page Rank Approach based on similarity measure” 2010 IEEE
- [17] S. N. Sivanandam, S. N. Deepa “Introduction to Genetic Algorithms” Springer, 2008, pg 20
- [18] S.N. Palod, Dr S.K.Shrivastav,Dr P.K.Purohit “Review of Genetic Algorithm based face recognition” International Journal of Engineering Science and Technology (IJEST) Vol. 3 No. 2 Feb 2011
- [19] Deep Malya Mukhopadhyay, Maricel O. Balitanas, Alisherov Farkhad A., Seung-Hwan Jeon, and Debnath Bhattacharyya “Genetic Algorithm: A Tutorial Review” International Journal of Grid and Distributed Computing Vol.2, No.3, September, 2009.
- [20] Shian-Hua Lin, Jan-Ming Ho, Yueh-Ming Huang ,ACRID ,intelligent internet document organization and retrieval ,IEEE Transactions on Knowledge and data engineering, 14(3),559-613, 2002

- [21] D.H. Kraft, F.E. Petry, B.P. Buckes, T. Sadasivan, Genetic algorithm for query optimization in information retrieval: relevance feedback, in: E. Sanchez, T. Shibata, L.A. Zadeh (Eds.), *Genetic Algorithms and Fuzzy Logic Systems*, 1997, pp. 155–173.
- [22] E. Sanchez, H. Miyano, J. Brachet, Optimization of fuzzy queries with genetic algorithms. Applications to a database of patents in biomedical engineering, in: Proc. VI IFSA Congress, Sao- Paulo, Brazil, 1995, pp. 293–296.
- [23] Zacharis Z. Nick and Panayiotopoulos Themis, Web Search Using a Genetic Algorithm, IEEE Internet computing, 1089-7801/01c2001, 18-25, IEEE
- [24] Ramakrishna Varadarajan, Vagelis Hristidis, and Tao Li , Beyond Single-PageWeb Search Results,IEEE Transactions on knowledge and data engineering, 20(3) ,411 - 424, 2008
- [25] Judit Barllan, Comparing rankings of search results on the Web, Information Processing and Management 41 (2005) 1511–1519
- [26] Adriano Veloso, Humberto M. Almeida, Marcos Goncalves, Wagner Meira Jr., Learning to Rank at Query-Time using Association Rules, SIGIR'08, 267-273 , 2008, Singapore.
- [27] S.Siva Sathya and Philomina Simon," Review on Applicability of Genetic Algorithm to Web Search" International Journal of Computer Theory and Engineering, Vol. 1, No. 4, October2009
- [28] Harry Zhang "The Optimality of Naive Bayes" American Association for Artificial Intelligence 2004.
- [29] Rich Caruana, Alexandru Niculescu-Mizil "An Empirical Comparison of Supervised Learning Algorithms" Proc 23 rd International Conference on Machine Learning, Pittsburgh, PA, 2006.
- [30] Wenxian Wang, Xingshu Chen, Yongbin Zou, Haizhou Wang, Zongkun Dai "A Focused Crawler Based on Naive Bayes Classifier" Third International Symposium on Intelligent Information Technology and Security Informatics, 2010
- [31] Peter A. Flach and Nicolas Lachiche "Naive Bayesian Classification of Structured Data" Machine Learning, Kluwer Academic Publishers
- [32] Kleinberg, John "Hubs, Authorities, and Communities" ACM computing survey, 1998.
- [33] Joel C. Miller, Gregory Rae, Fred Schaefer "Modifications of Kleinberg's HITS Algorithm Using Matrix Exponentiation and Web Log Records" Proc. SIGIR'01, ACM 2001.

Construction of Learning Path Using Ant Colony Optimization from a Frequent Pattern Graph

Souvik Sengupta¹, Sandipan Sahu², Ranjan Dasgupta³

^{1,2}Bengal Institute of Technology, Kolkata, India

³National Institute of Technical Teachers Training and Research, Kolkata, India

1. Abstract :

In an e-Learning system a learner may come across multiple unknown terms, which are generally hyperlinked, while reading a text definition or theory on any topic. It becomes even harder when one tries to understand those unknown terms through further such links and they again find some new terms that have new links. As a consequence they get confused where to initiate from and what are the prerequisites. So it is very obvious for the learner to make a choice of what should be learnt before what. In this paper we have taken the data mining based frequent pattern graph model to define the association and sequencing between the words and then adopted the Ant Colony Optimization, an artificial intelligence approach, to derive a searching technique to obtain an efficient and optimized learning path to reach a unknown term.

2. Key words : e-Learning, Learning Path, Data Mining, Frequent Pattern , AI, ACO

3. Introduction :

The self regulated learning for definition of a term is where the learners hold incremental beliefs about intelligence within their control. In the hypertext document, links have been established in such a way that the user can explore, browse and search for not only a particular term but can also get information regarding relevant and associated issues [3]. The problem with self-regulated definition learning is that learner often gets confused where to start from because the term definition itself contains multiple keywords which are known or unknown. So prerequisites required to understand a particular term are most important consideration here. That is, ultimately the learner always tries to find out a learning path which is a sequence of words like $a_1 a_2 \dots a_n$. This sequence from source to destination may

include several unknown terms which indicate the need of prerequisite knowledge required to understand the target term. In this paper first we have created a graph to show the interconnection between the terms. Then we have adopted a hybrid model in the sense that we take data from direct mode classroom question answer session between students and teachers and then train the graph data based on our experience. We have used the association rule of data mining to create the frequent pattern among the links of the graph and then derived the learning path using ACO technique.

4. Related works:

Many research works have been made to establish an individual learning path for the learners in which the application of Artificial Intelligence, Fuzzy-neural and Data Mining techniques are most commonly used. An AI based tutoring system [5] has been proposed to identify, monitor and adapt the student's learning path, according to the students actual knowledge, learning habits and preferred learning style, whereas personalized curriculum sequencing is another important research issue for Web-based learning systems because no fixed learning paths will be appropriate for all learners. Therefore, many researchers focused on developing e-learning systems with personalized learning mechanisms to assist online Web-based learning and adaptively provide learning paths in order to promote the learning performance of individual learners [6].

Experimental results indicated that applying a genetic-based personalized e-learning system for Web-based learning is superior to the freely browsing learning mode because of high quality and concise learning path for individual learners. In an another approach working on design of Learners' Quanta (LQ) based efficient and adaptive learning system, it has been observed that a generic platform is required to design and develop an LMS with the adaptive algorithm already proposed in [8], [9] and [10].

A graph based frequent pattern model:

Let in an e-Learning context a term ‘A’ is defined with help of terms ‘B’, ‘C’, ‘D’. These terms may not be not in any order that means the sequence of these terms may come differently in a different context. So the order of the terms is not an important consideration here we have only concentrated on their presence as a prerequisite to describe a keyword. So we create a sequence like this: B-C-D-A and join this branch with root.

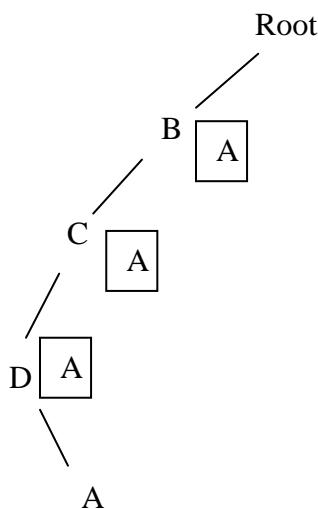


fig 5.1 : branch BCDA

We introduce an additional data structure called a datalist associated with each node that contains information about the term for which it has been used. The datalist keeps on updating as new branches are introduced because same term can be reused for describing multiple keywords. For instance if another term definition ‘G’ involves ‘E’, ‘F’, ‘C’ then we create another branch E-F-C-G and append it to the graph like fig 5.2 . now the node ‘C’ has ‘A’ and ‘G’ in its data list which indicates that ‘C’ is used to describe both these terms.

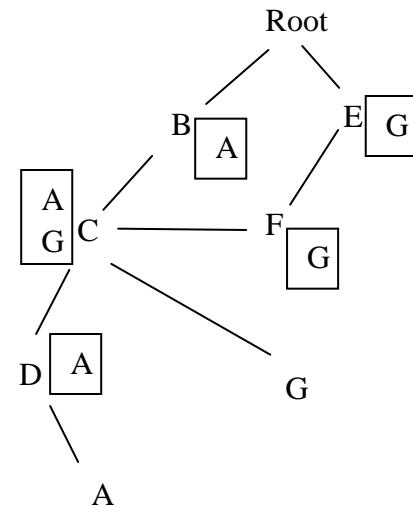


fig 5.2: Appending EFCCG

This way we go on expanding the graph with new terms and link them with the existing terms.

5. Data Mining :

A popular and effective way of discovering new knowledge from large and complex data sets is data mining [1]. The association rule and the frequent pattern tree of data mining are the major efficient techniques in e-Learning data [4]. We will use association rule to reveal the dependencies of a term on the other term, the FP graph shows the potential usage of the other terms to learn any particular ‘term’ by the learner.

6.1 Creating frequent pattern and association :

We have taken the following examples of definition from a well known science dictionary website.

DNA: Deoxyribonucleic acid or DNA is a double-stranded **nucleic acid** that contains the genetic information for **cell** growth, division, and function.

Nucleic Acid: Nucleic acids direct the course of **protein** synthesis, thereby regulating all **cell** activities. The two main types, **DNA** and **RNA**, are composed of similar materials but differ in structure and function. Both are long chains of repeating **nucleotides**.

Nucleotides: Nucleotides are molecules that, when joined together, make up the structural units of **RNA** and **DNA**. A nucleotide consists of a sugar molecule attached to a phosphate group and a nitrogen-containing base. The bases are adenine, cytosine, guanine, and thymine or uracil (for RNA).

RNA: Ribonucleic acid or RNA a **nucleic acid** that is generally single stranded and plays a role in transferring information from **DNA** to **protein**-forming system of the **cell**.

Protein: Proteins are natural polymer molecules consisting of **amino acid** units. The number of amino acids in proteins may range from two to several thousand.

Amino acid: A **molecule** consisting of the basic amino group (NH_2), the acidic carboxylic group (COOH), a hydrogen atom (H), and a side-chain(R) that varies between different amino acids attached to the carbon **atom**.

Cell: The structural, functional and biological unit of all organisms. It is the smallest unit of life that is classified as a living thing.

Eukaryotic: A **cell** that contains membrane-bound compartments in which specific **metabolic** activities take place. Most important among these compartments is the **nucleus**.

Nucleus: The nucleus is a membrane-enclosed **organelle** found in **eukaryotic cells**. It contains most of the cell's **genetic material**, organized as multiple long linear **DNA** molecules in complex with a large variety of **proteins**.

Organelle: An organelle is a specialized subunit within a **cell** that has a specific function.

Genetic material: The genetic material of a cell or an organism refers to those materials found in the nucleus, mitochondria and cytoplasm, which play a fundamental role in determining the structure and nature of cell substances.

Mitochondria: Mitochondria are the powerhouses of the human **cell**; mitochondrion is a membrane-enclosed **organelle** found in most **eukaryotic** cells. They generate most of the cell's supply of adenosine triphosphate (**ATP**), used as a source of chemical energy.

Cytoplasm: The cytoplasm is a thick liquid residing between the cell membrane holding all **organelles**, except for the **nucleus**. All the contents of the **cells** of **eukaryotic** organisms (which lack a **cell nucleus**) are contained within the cytoplasm.

Ribosome: A ribosome is the component of a biological **cell** that creates proteins from all **amino acids** and **RNA** representing the **protein**.

Next we create the transaction out of these terms by fetching out the keywords from the definitions.

DNA:	<i>Cell, Nucleic acid.</i>
Protein:	<i>Amino acid.</i>
Eukaryotic:	<i>Cell, Metabolism, Nucleus.</i>
Mitochondria:	<i>Cell, Eukaryotic, organelle.</i>
Nucleic acid:	<i>Cell, DNA, Nucleotide, Protein, RNA.</i>
RNA:	<i>Cell, DNA, Nucleic acid, Protein.</i>
Nucleotide:	<i>DNA, RNA</i>
Amino acid:	<i>atom, Molecule</i>
Nucleus:	<i>DNA, eukaryotic, genetic material, organelle, proteins</i>
Organelle:	<i>Cell</i>
Genetic material:	<i>Cell, Cytoplasm, Mitochondria, Nucleus, Organism.</i>
Cytoplasm:	<i>Cell, Eukaryotic, Nucleus, Organelle,</i>
Ribosome:	<i>Amino acid, Cell, Protein, RNA.</i>

6.2 Construction of the graph

We will now simulate the above sequences to form the branches of the graph. For instance the sequence of DNA will be Root--Cell--Nucleic Acid--DNA. Now as the branches are joined by the common terms, ultimately we form a frequent pattern graph by showing the frequency on individual edges. Fig 6.1 describes this graph, the frequencies are not shown to make the graph look simple.

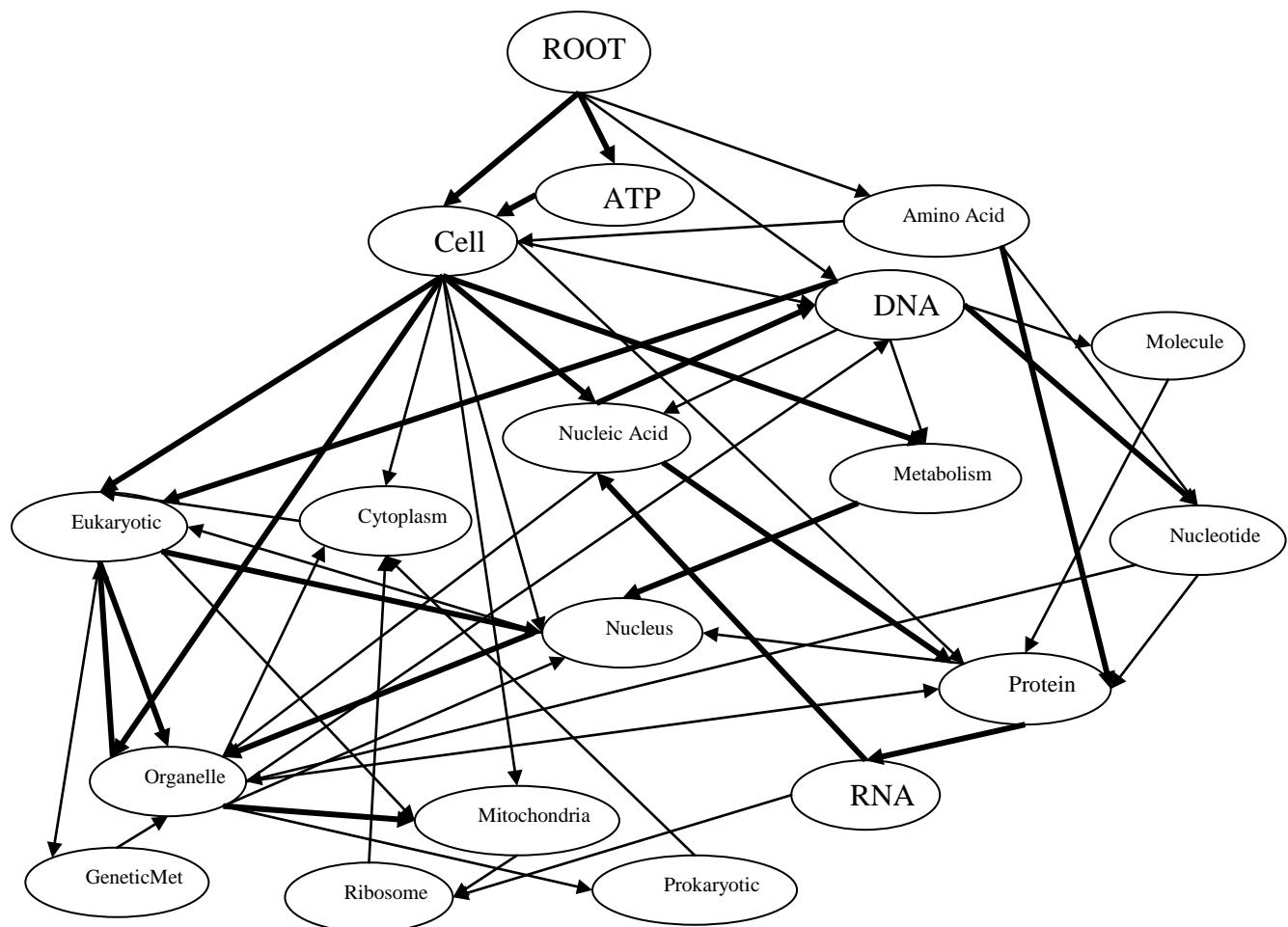


Fig 6.1: Frequent Pattern Graph

Next we will try to find association between the nodes or the terms by applying the knowledge obtained from the question answer session of the contact mode classes. We will create transaction from each question answer and map the frequency of these transactions into the branches to form the associations. Let $S = \{s_1, s_2, \dots, s_k\}$ is the set of all terms used by the teacher while delivering the entire lecture on a particular topic. The ‘what is’ type questions are commonly answered by the term definitions, so in our approach we have taken only such answers to create transactions. $T = \{t_1, t_2, \dots, t_m\}$ is the set of all successful answers made by the teacher on $Q = \{q_1, q_2, \dots, q_m\}$ which is set of questions made by the students and t_i and q_i has 1 to 1 correspondence. We consider all t_i as set of keywords taken from S . So $t_i \subseteq S$. Any term is said to be independent if it does not have

other keyword term, belonging to S , is used to define it. Independent keywords are generally easily understandable by its own definition so they have the higher probability of staying nearer to the root. For example say question t_1 is satisfactorily answered with the transaction $\{t_2, t_3, t_4\}$ then we search for a branch in the FP Graph with the sequence $\{t_2 \rightarrow t_3 \rightarrow t_4 \rightarrow t_1\}$. The frequency of $t_i \rightarrow t_{i+1}$ will be increased by 1 if it is a existing edge in graph. If the entire transaction does not match to any branch the we look for match with all possible sub transaction with t_1 like here we can create $\{t_3 \rightarrow t_4 \rightarrow t_1, t_4 \rightarrow t_1\}$. Now we will keep on updating the FP Graph based on the records stored from different classes taken to different set of students on that particular topic. For every round we will calculate the frequency value of each edges so that if frequency of

any link increases to such a level that the frequency reaches or crosses σ , a predefined threshold value, then convert sequence (\rightarrow) into association (\Rightarrow). Any keyword which is amalgamated in nature [set of words] always preceded by its subset hence there is always a trivial association from subset to superset. In our example we have converted only some of the links into association based on our assumptions, shown in fig 6.1, but the actual implementation of this model should have a very large number of classroom data of question answer session recorded and then analyzed to form the transactions. Then if the repeated mapping of these transactions increased the frequency of any link above the threshold value then only that link is converted into an association.

6. Artificial Intelligence:

The use of swarm intelligence techniques in e-learning scenarios provides a way to combine simple interactions of individual students to solve a more complex problem. Genetic algorithm and Ant Colony Optimization are two commonly used approaches. In our proposal we have suggested Ant Colony Optimization-based inductive planning for recommending learning paths as it is a heuristic method that has been successfully used to deal with this kind of problems. The ants construct a solution by starting from a given node which is a unknown term. Then, at each construction step it moves along the edges of the graph. Each ant checks the data list in subsequent steps to understand which node is used to define the target term and it chooses among the edges that do not lead to vertices that it has already visited. An ant has constructed a solution once it has reached a known word or the root. At each construction step, an ant probabilistically chooses the edge to follow among those that lead to yet unvisited vertices. The probabilistic rule is biased by pheromone values and heuristic information: the higher the pheromone and the heuristic value associated to an edge, the higher the probability of an ant to choose that particular edge. Once all the ants have completed their tour, the pheromone on the edges is updated. Each edge then receives an amount of additional pheromone proportional to the quality of the solutions to which it belongs.

7.1 ACO adaptation:

Each ant starts from a query node (unknown term) q_k and iteratively adds nodes till unvisited node is available in its partial tour. The solution construction terminates once a path from q_k to Root node or any node with known keyword is found. Optimal path should contain maximum number of association in it. If the feasible neighborhood N_i^k does not contain any association, then it chooses the edge which has maximum frequency value and data list of which contains the target term. Each term should come at most once in a whole searching process. This constraint is enforced by the rule that an ant at each construction step chooses the next node only among those it has not visited yet. Each ant a_k contain a memory M^k which contain the nodes already visited, in the order they are visited. This memory is used to define the feasible neighbourhood N_i^k in node i , comprises all nodes that are unvisited. This feasible neighbourhood concept we have implemented via introduction of the data lists with each term, by which one ant can understand whether a particular edge takes part in the solution path or not. The pheromone trail value τ_{ij} indicate how proficient it has been in the earlier iterations to make a particular move from node s_i to node s_j ($s_i \rightarrow s_j$). In this paper the pheromone represents the occurrence of association of a term in describing other terms. The trail value is therefore a posterior indication of the desirability of that move. Trail value of a particular move will increase if the number of edges which form the association a_{ij} , is increased in tour T^k where $k = 1, 2, \dots, m$, k is identification number of ant, and m is total number of ant which contain the move $s_i \rightarrow s_j$. The attractiveness η_{ij} is the heuristic information of the move $s_i \rightarrow s_j$. Trail value η_{ij} is proportional to frequency value c_{ij} of the edge e_{ij} .

In an Ant System (AS), m (artificial) ants concurrently build path from query node q_k to the root node or any node with known keyword is found. Initially, ants are put on q_k . At each construction step, ant a_k applies a probabilistic action choice rule, called random probability rule, to decide which node to visit next.^[7] In particular, the probability with which ant a_k , currently at node s_i , chooses to go to node s_j is

$$P_{ij}^k = \begin{cases} \frac{[\tau_{ij}]^\alpha [\eta_{ij}]^\beta}{\sum_{l \in N_i^k} [\tau_{il}]^\alpha [\eta_{il}]^\beta} & \text{if } s_j \in N_i^k \\ 0 & \text{Otherwise} \end{cases}$$

where α and β are two parameters which determine the relative influence of the pheromone trail and attractiveness. In our example we have set both of them to 1 for the sake of simplicity.

After each iteration t , i.e., when all ants have completed a solution, trail values are updated by:

$$\tau_{ij}(t) = \rho\tau_{ij}(t-1) + \Delta\tau_{ij}$$

Where $0 < \rho \leq 1$ is the pheromone evaporation rate. The parameter ρ is not used in our approach as the edge frequency once achieved will never be reduced. $\Delta\tau_{ij}$ represents the sum of contributions of all ants that used move $s_i \rightarrow s_j$ to construct their solution. All ants deposit pheromone on the edge they have used in their solution path.

$$\Delta\tau_{ij} = \sum_{k=1}^m \Delta\tau_{ij}^k$$

Where m is number of ants and $\Delta\tau_{ij}$ is amount of trail deposited on e_{ij} by ant a_k , which can be computed as

$$\Delta\tau_{ij}^k = \begin{cases} Q * C^k & \text{if } s_{ij} \in T^k \\ 0 & \text{otherwise} \end{cases}$$

Where C^k is represented by the number of association presents at T^k . Q being a constant parameter which we have set to 1.

A Learning path is constructed by applying the constructive procedure to each ant: (1) at the start each ant positioned at the query node q_k ; (2) use pheromone value and heuristic value to probabilistically construct a path by iteratively adding the node that the ant has not visited yet and has the target term in its datalist ,until a path from q_k to Root node; and (3) if a term q_u in the solution is not well understood by the learner and requires extra clarification then the same algorithm will be applied with target term q_u in place of q_k . It is then repeatedly applied until a termination criterion is satisfied.

7.2 Searching Algorithm:

The performance of the algorithm depends on the correct tuning of the parameters, and the relative importance of

trail and attractiveness. The algorithm mainly works on two procedures 1) LEARNING_PATH 2) UPDATE_TRAIL

Initialize: $\eta_{ij} = c_{ij}$ and $\tau_{ij} = \tau_{ij}(0) \quad \forall e_{ij} \in E$,
 Target_Term = Query_Term;

```
Procedure LEARNING_PATH (Node Target_Term)
{
    FOR each ant k      DO
        Repeat
            WHILE Not (ROOT is Reached) DO
            {
                Choose Next_Term using Probability
                Distribution  $P_{ij}^k$  ;
                Target_Term = Next_Term;
                 $T^k := T^k +$  Next_Term;
                Update  $M^k$  &  $N_i^k$ ;
                IF (Next_Term == Association)
                     $C^k = C^k + 1$ ;
                }
            END FOR
             $T := \min(T^k)$  Where  $k = 1, 2, \dots, m$ 
            UPDATE_TRAIL ();
            IF NOT(END TEST)
                LEARNING_PATH (Node Target_Term);
            Else
                Return  $T$ ;
            }
}
```

```
Procedure UPDATE_TRAIL ()
{
    FOR each edge  $e_{ij}$  DO
        Repeat
            FOR each ant k      DO
                Repeat
                    IF( $e_{ij} \in T^k$ )
                         $\Delta\tau_{ij}^k = Q * C$  ;
                    Else
                         $\Delta\tau_{ij}^k = 0$  ;
                     $\Delta\tau_{ij} = \Delta\tau_{ij} + \Delta\tau_{ij}^k$  ;
                // Update the TRAIL MATRIX
                END FOR
            END FOR
        }
}
```

8. Case study :

Based on our above example we tried to find out a recommended learning path for two terms: ‘Mitochondria’ and ‘Eukaryotic’. The last updated data list that we have used is given below .

Term	Data List
ATP	Mitochondria
Amino Acid	Ribosome, Protein
cell	Mitochondria, Eukaryotic, Organelle, DNA, Nucleic acid, Cytoplasm, Ribosome
Cytoplasm	Mitochondria
DNA	Nucleic acid, Eukaryotic, Nucleus
Organelle	Mitochondria, Eukaryotic, DNA, Cytoplasm, Nucleus, Ribosome
Eukaryotic	Mitochondria, Cytoplasm, Nucleus
Metabolism	Eukaryotic
Mitochondria	Cytoplasm
Molecule	Nucleic acid
Nucleus	Eukaryotic, DNA, Cytoplasm
Nucleic acid	DNA
Prokaryotic	Cytoplasm
Ribosome	Cytoplasm
Nucleotide	Nucleic acid, Ribosome
Protein	Nucleic acid, Nucleus, Ribosome
RNA	Nucleic acid, Ribosome
GeneticMet	Nucleus

Table 8.1 : Data list of the terms

Results:

- I. For *Mitochondria*: the recommendation is {ATP, Cell, Eukaryotic, Organelle }
- II. For *Eukaryotic*: the recommendation is {Cell, Metabolism, Nucleus, Organelle}

7. Conclusion :

In this paper we have tried to propose an artificial intelligent based approach to find a learning path which may be helpful for a learner to understand the prerequisite terms, after going through which the unknown target term is more easily understandable. The

flexibility of the algorithm is that the solution can be changed at learner’s choice of keyword or term which is not known, so the solution is personalized. In this paper we have identified the problem of a self learner to learn a term in a e-Learning environment. Then our approach towards solution is three folded; first we have proposed a frequent graph model which can be created from repository of e-learning contents and then we create transaction from real world classroom session of questions answers and updated the frequencies to identify the associations. Finally we have applied the ant colony optimization technique to find a learning path between two nodes one unknown and the other is either known or comprehensive. The possibility of wrong path selection which is generally biased by the presence of significant number of association in the branches, is nullified by the introduction of the data list in each node.

10. Reference :

- [1] “Mining Association Rule between Sets of Items in Large Database”, Rakesh Agrawal, Tomasz Imielinski, Arun Swami, Proceedings of ACM sigmod Conference Washington DC,USA,May 1993
- [2] “A FP Tree based approach for mining all strongly correlated pairs without candidate generation”, Zengyou He, Xiaofei Xu, Shengchun Deng, Kluwer Academic Publishers’ Data Mining and Knowledge Discovery, 8, 53–87, 2004
- [3] “Design of a Learning Management System on LTSA “, S. Sengupta, N. Chaki, ,R. Dasgupta, International Journal of Education and International Technololgies,Issue 1, Volume 3, 2009
- [4] “A Data Mining Approach to Determine an Efficient Learning Path.” S. Sengupta, R. Dasgupta CSREA EEE 2010: page 59-64
- [5] “An Artificial Intelligence - Multi Agent based Adaptive Learning Environment”, Proceedings of the World Congress on Engineering and Computer Science 2009 Vol I,WCECS 2009, October 20-22, 2009, San Francisco, USA
- [6] “Intelligent Web-based Tutoring System with Personalized Learning Path Guidance”, ISBN: 0-7695-2916-X, Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007)

[7] "Ant Colony Optimization Theory: A Survey", M. Dorigo, C. Blum, Theoretical Computer Science vol. 344, 2005, page 243-278.

[8] "Design of an adaptive web-based courseware", S Ray, N Chaki, R Dasgupta, IASTED International Conference on Intelligent Systems & Control (ISC 2004), Honolulu, Hawaii, USA, pp266-271, August 23-25,2004

[9] " A learners' Quanta Based Framework for Identification of Requirements and Automated Design of Dynamic Web-based Courseware", N Chaki & R Dasgupta,14th International Monterey Workshop held at the Monterey Conference Center in Monterey,California, Sept. 10-13,2007

[10] " A Learner's Quanta Model based Framework Towards Building Dynamic web-based Courseware ",S Ray , N Chaki , R Dasgupta 4th International Conference on multimedia and ICTs in Education (mICTE 2006) ,Seville, Spain pp 238-242, Nov 22-25, 2006

Souvik Sengupta is an Assistant Professor in the Computer Science Department of Bengal Institute of Technology, India . He has obtained his Master of Technology form West Bengal University of Technology, India in 2008. He is currently performing his research work under the PhD program of

the Department of Information Technology of the university of Calcutta ,India. He has 9 international journal and conference papers published in last 3 years. He has also served as a visiting faculty in the MTech program of Department of Computer Science of National Institute of Technical Teachers' Training and Research, Kolkata, India. His areas of research interests include Web-based Learning, Multimedia Database, Data mining, Software engineering, Artificial Intelligence etc.

Sandipan Sahu is an Assistant Professor in the Computer Science Department of Bengal Institute of Technology, India . He has obtained his Master of Technology form West Bengal University of Technology, India in 2007. He has 2 international and national conference paper published. His areas of research interests include Web-based Learning, Data mining, Artificial Intelligence, Image Processing etc.

Ranjan Dasgupta is Professor & Head, Dept. of CSE. National Institute of Technical Teachers' Training & Research, Kolkata, India, Previously he was with Jadavpur University, and also served several Indian IT companies, for five years. He received his B. Tech from,Dept. of Radio Physics & Electronics, M. Tech, & Ph.D from Dept. of Computer Science, University of Calcutta,India. His areas of research interests include software,engineering, databases, GIS, distributed computing. Dr.Dasgupta has published around 30 research papers in different International Conferences & Journals. He has also served several other Universities & national & international bodies, World Bank Assisted Projects in various capacities.



Developing an Intelligent User Interaction Development Engine

Ashit Kumar DUTTA

Department of Computer Science and Information System., Shaqra University, KSA

Abstract

This paper presents an intelligent user interaction development (IUID) engine that helps to enhance the structure of the relational database by using artificial intelligence. The IUID consists of two phases, the first phase enhances the user database by comparing it with a well structured pre-defined database, and the second phase gives the user the ability to use the artificial intelligence by writing java code and organize it in a tree structure. The main objectives of the IUID engine are to allow user to use the expert knowledge to upgrade his/her database, and increasing the speed of the development process by appending a new artificial intelligence layer at the user application that is represented by a package to run in the application

1. Introduction

Databases are gaining prime importance in a huge variety of application areas employing private and public information systems. Databases are built with the objective of facilitating the activities of data storage, processing, and retrieval associated with data management in information systems. An intelligent database is a full-text database that employs artificial intelligence (AI), interacting with users to ensure that returned items (hits) contain the most relevant information possible. This is in contrast to a traditional database, which is searchable only by keywords and verbatim phrases connected by Boolean operations such as AND, OR, and NOT. Intelligent database technology is in its infancy, and is evolving as AI becomes more advanced. An Intelligent Database System is endowed with a data management system able to manage large quantities of persistent data to which various forms of reasoning can be applied to infer additional data and information. This includes knowledge representation techniques, inference techniques, and intelligent user interfaces - interfaces which extend beyond the traditional query language approach by making use of Artificial Intelligence. This technique plays an important role in enhancing databases systems. In this paper we have

designed an engine with artificial intelligence algorithm which can upgrade the user database.

2. Related Work

User interface design has been a topic of considerable research, including on its aesthetics. In the past standards have been developed, as far back as the eighties for defining the usability of software products. One of the structural basis has become the IFIP user interface reference model. The model proposes four dimensions to structure the user interface:

- The input/output dimension (the look)
- The dialogue dimension (the feel)
- The technical or functional dimension (the access to tools and services)
- The organizational dimension (the communication and co-operation support)

This model has greatly influenced the development of the international standard ISO 9241 describing the interface design requirements for usability. The desire to understand application-specific UI issues early in software development, even as an application was being developed, led to research on GUI rapid prototyping tools that might offer convincing simulations of how an actual application might behave in production use.^[3] Some of this research has shown that a wide variety of programming tasks for GUI-based software can, in fact, be specified through means other than writing program code.

Research in recent years is strongly motivated by the increasing variety of devices that can, by virtue of Moore's Law, host very complex interfaces.

There is also research on generating user interfaces automatically, to match a user's level of ability for different kinds of interaction.

3. The Intelligent User Interaction Development Engine

The IUID consists of two stages the database stage and artificial intelligence algorithms (AI) stage, each stage consists of 3 phases as shown in figure 1.

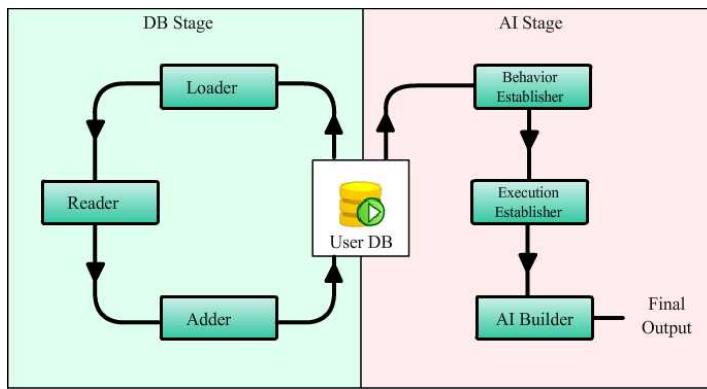


Figure 1: The architecture of the IUID engine

The database stage consists of Loader, Reader and Adder. It starts by loading the user database into java scope as objects. The Reader concerns about reading the user database and understand every table and its columns. The table and all column names must be known by the reader. At this phase the user asked to modify any unknown table or column names. In the adder phase, the user database will be compared with a well structured relational database recommended some suggestions about the user database related of adding or deleting some features from it. If the user doesn't have a database then our engine will help him/her to build a new a well structured relational database.

The output of the first stage is a well-structured identified database uses as input to the second stage. The three stages appear in figure 2.

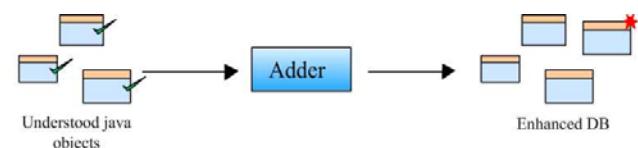
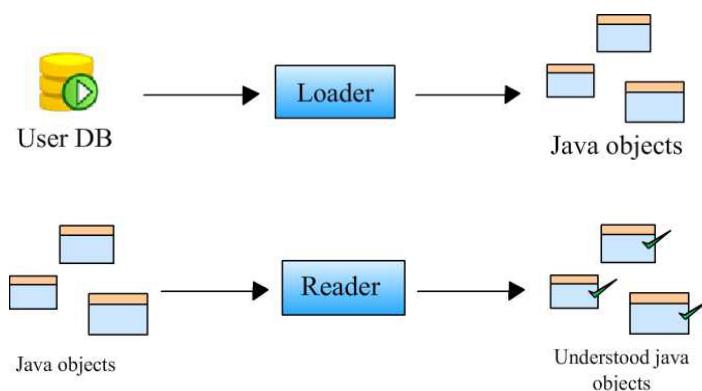


Figure 2: The structure of Loader, Reader, and Adder

The AI stage consisted of three phases: Behavior Establisher, Execution Establisher, and the AI builder. The behavior establisher starts working on the user database after enchantment by the previous phase. The user can create nodes called behaviors; these behaviors are java codes that are connected to entities from the user database, all the behaviors calculation will reflect on the user DB. The behavior establisher presented in Figure 3.

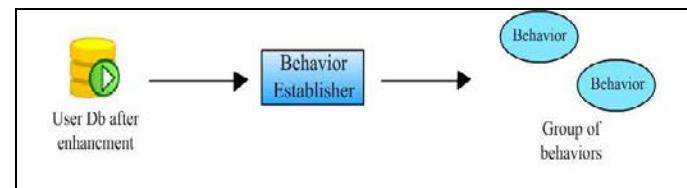


Figure3 Creating the behaviors in Behavior Establishment phase.

The execution Establisher phase main purpose is to group the behaviors made by user in a tree structure that the user will have the ability to define some rules to navigate throw it. Many types of trees are available like: list, Recursion tree and conditional tree. The execution establisher presented in figure 4.

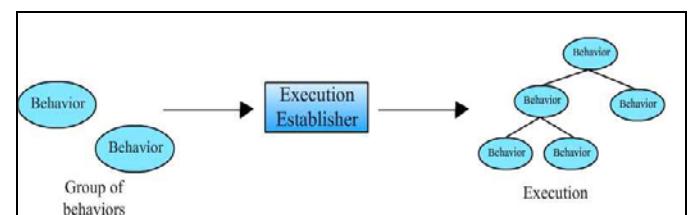


Figure 4 Building the execution tree by using group of behaviors in Execution Establishment.

At the AI stage the user can design tree algorithms to use in his/her application. In order to establish the tree algorithm, the user has to build it step by step starting with:

- i. Creating an execution which represent the tree of the algorithm
- ii. The tree of the execution contains nodes that hold the user written-by-wizard code.
- iii. These node
- iv. s defined in the execution called AI-behaviors.
- v. Each AI-behavior (node) should have a parameters list and a return variables list.

- vi. Each of these lists have a group of variables that represent a cell in the user database, these variables are called AI-cells.

There are three types of executions (tree); conditional node tree, recursion tree and execution list.

The conditional node tree should have conditional nodes that are used to control the run path of the tree.

Each conditional node should have a group of enter points⁽¹⁾, one for each child behavior, that are represented by a Boolean method that the user should code, this enable the user to use his logic(variables and methods in his own application) to make the conditional node decision. The AI builder stage presented in figure 5.

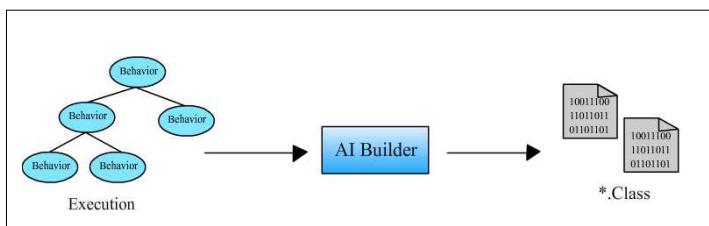


Figure 5 Generating .class files in AI Building Phase

The AI builder phase is the last phase that works on the executions and transforms them into java classes; each class represents a tree of execution.

Finally the output classes will be grouped into a package and will be given to the user, so he can use these classes in his project.

4. Implementation

4. Implementation
Our system was developed in Oracle 10g enterprise edition release , MySQL Server , and NetBeans were implemented on an Intel® Pentium® M dual core with speed 1.8 GH processor and DDR@ I GBRAM, running on the Microsoft Windows XP Professional Service Pack 2.

Our system work space arranged to design and developed a high quality structural database by using the AI structure algorithm. Figure 1 presents the workspace that includes menus, variety of tools and palettes for viewing, editing and adding new elements to your project.

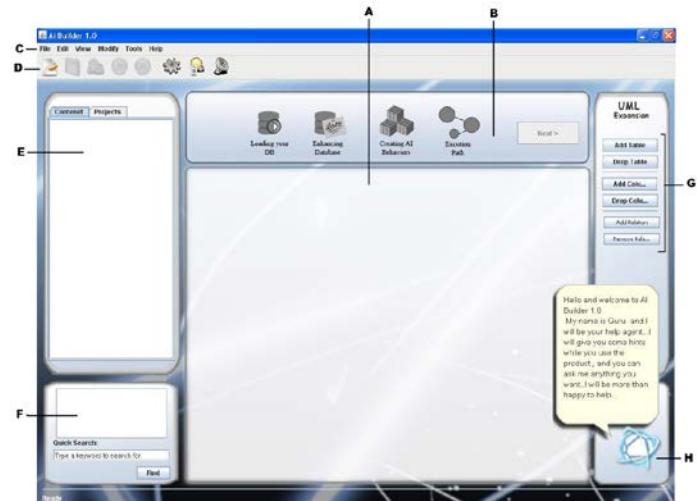


Figure 6: Main screen **A**. Project work Area **B**. Progress palette **C**. Menu bar **D**. Tools bar **E**. Project contents **F**. Search palette **G**. Tools palette **H**. Help agent

When the user starts a new project or open an old one, the user selects the database category type and import his\her database that should be installed in a database service provider (like oracle...). After the user has started his\her project the project shall have a directory folder saved in the pre-defined work space of AI-Builder.

The project starts with loading the user database very efficiently from Oracle, MySql or SqlServer into java scope (project scope) using the Loader Class, some tests has been made in a modern pc to insure the speed reliability of Loader and one of the results was :

167 table each with (on average) 10 columns and at least 50 rows have been loaded successfully in 3.8 seconds.

The Loader class has the ability to load a database from different database service providers including: Oracle, MySQL and SqlServer

Which are the most commonly used around the world and that insures the flexibility of the resulted database. The Loader is a dynamically implemented means that it can load the tables whatever its structure design was or its number of tables or columns, and this is a needed feature in AI-Builder to be able to handle any user data base. However the class was built in a specific design so it can store the tables in a way that separate the structure ⁽¹⁾ and data from each other insuring Loader ease of use and performance stability. (See figure 7)

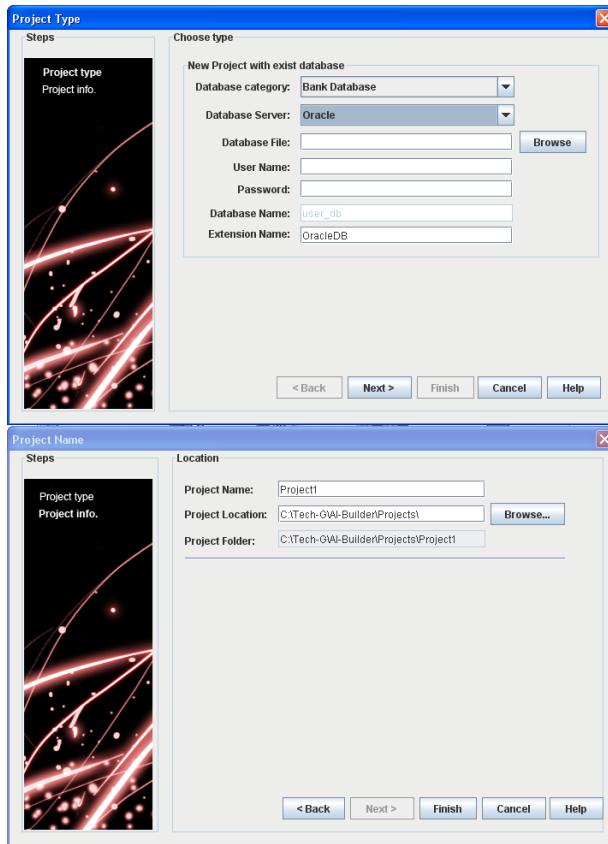


Figure 7: these figures shows some information the user has to fill to complete the connection to the database like database type, database provider, user name, and password.

The Reader starts with reading all the loaded tables, this done so AI-Builder can recognize what does the user mean (its type) by any selected cell in the user database.

When reading a table the reader starts with identifying its name by a text recognition engine ⁽²⁾ that investigate if the table is known to Reader or not known, so AI-Builder can ask the user to identify it. I must point here to the flexibility of the Reader design, the Reader can be bound easily with any text recognition algorithm with minimum code changes that need only one hour effort.

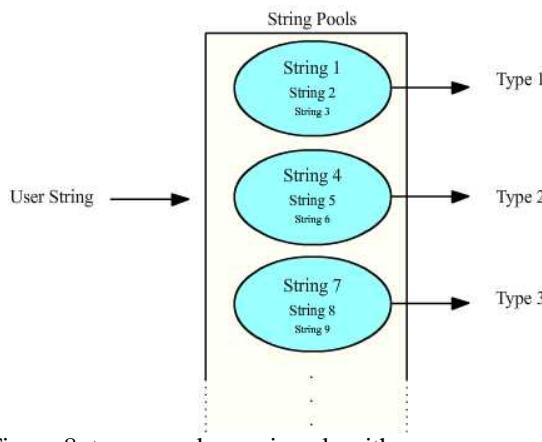


Figure 8: types-pool mapping algorithm

After all the table's names are identified, the Reader starts to read the columns inside the table using text recognition engine in the same way as the tables' names identification. Every table name or column name are identified by the types-pool mapping algorithm⁽³⁾, and its works in this way: if the name (table's or column's) given from the user is not recognized by text recognition the name is added to a pool of names that are all bound to a specific general type (an answer to the following question: what does the user mean by this table and this column so data in cell are meaningful) that is used later by AI-Builder. (See figure 5.3)

An important feature of Reader is the ability to learn from the user, in other words it never ask the user to identify something twice as Reader store it in the first user identification. (See Figure 5.4)

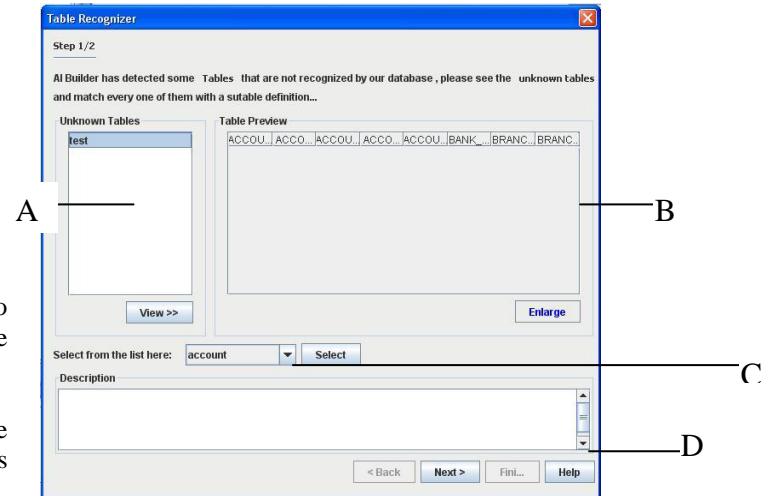


Figure 9: Table recognizer
A. list of unknown tables **B.** preview of the unknown table
C. Table type **D.** Description palette

Before the reading operation the Reader checks the tables and asks the AIHandler⁽¹⁾ to load the needed information from the recognition database, and then start the reading operation, that uses the text recognition engine.

While the reading operation, names that are not understood by Reader are enquired in an updates buffer and treated as new knowledge that needed to be add to the AI database, these updates are stored in the updater class.

The adder engine main job is to enhance the user tables after they are read by the reader engine, all the tables at this point are well understood and ready to enter the adder phase.

It starts by comparing the structure of the user tables with a structure of a general tables in the general database⁽¹⁾ stored in our program, the design of the general database passed throw several steps to insure its effectiveness; this is has been accomplished by an intensive study to database principles and collecting information from deferent resources.

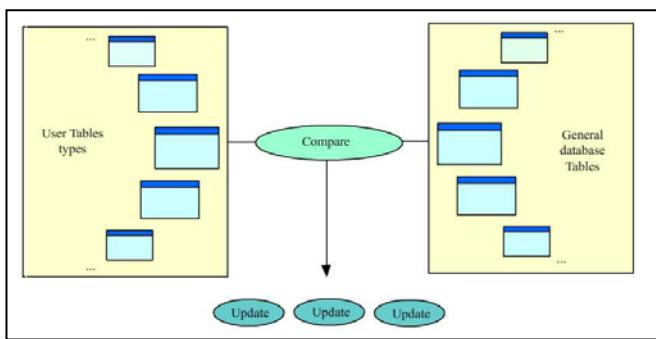


Figure 10: Comparing the user tables with the general database tables.

The comparison is done in two phases; the first one includes comparison the user tables with the general database tables, that means the columns is not included in this phase. (See Figure 5.5)

The next phase of comparison is comparison the columns of each table with another set of columns from the general database, the result from the both phases will be a group of updates that will be given to the user. (See figure 5.6)

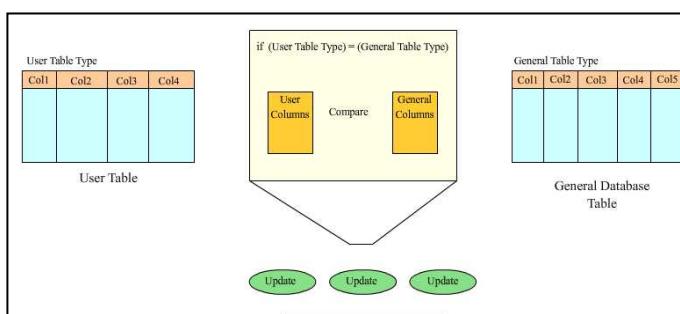


Figure 11: comparing the columns in each table

After the adder has done the comparison, Adder will suggest group of updates that are listed in a buffer and these updates represents the result of the Adder comparing the general database with the user database. Any new addition on the user database will be grouped in a list and will be given to the user to determine if he/she wants to use any of them or not,

The addition to the user database includes adding new tables, new columns and creating new relations between tables; deletion is not an option to the adder so the user will be sure that no lost of tables will happen to his\her database.

The user can add a new general database as a new category that is added to the adder as easy as one click; the user can import his\her database structure into the program and make it as a general database so AI-Builder can use it later to enhance any future work, also he/she can import the database that a viable in any of the most commonly used database service provider in the world; MySql, oracle and SQL Server.

The idea of adding new general database will give the adder a huge advantage by making it able to learn. Every new general database means a new field of experience added to its huge

previous experiences, this feature gives a really good advantage at development time, the developers can use, work-on and benefit, so the experience of a single expert is effectively reused.

The ability of learning that is represented in the reader and the adder engines makes AI-Builder a huge resource of valuable and precious information collected throw every use of the program, the more you use the program the cleverer it will be. After the database has been well defined and enhanced, the user can start to build his\her AI algorithms that he will use later in his\her application, the final output of this stage is a package of classes that represents AI executions (the AI algorithms) so the user can simply execute any of them anywhere in his\her application by a method call.

The major advantage of this binding is that the user can build complex AI algorithms and implement its code and use it in a short time, and the reader may ask what is the benefit ?, what is the main advantage? And the answer is simply that your developers will do things faster easier and the AI expert doesn't have to worry about doing the structure design that will hold his\her AI code.

Behavior Establishment:

The user can build an AI algorithm by first constructing some behaviors so they can be used in the algorithm, a behavior is a list of statements (code lines) that are treated as a method in a programming language, the user write code in a behavior to a specific thing that he want, remember that a behavior is only a part of the code because it represent a node in a tree and one path in that tree will be the code that will execute.

each behavior have two lists of variables that are accessible inside the behavior , first the parameter variables list which represents the parameters used in the behavior and second the return variables list that represent the variables that will be affected after the behavior has completes.

The variables that are defined in the two lists each represents a cell in the user scanned database (this is done in the Reader phase) so the user can write the behavior code to use the values of the parameters and to return new values to the return variables (as update these database cells at user application run time).

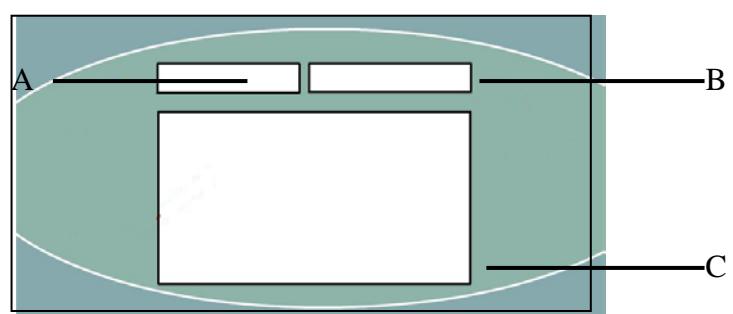


Figure 12: Behavior content

A. Return list B. Parameter List C. User code

The variables scope is the whole user database; this is means that the user can define a variable on any cell in the database. This is insures that AI-Builder can be used for any application

that depends on the database as major part of design. A variable definition in most development IDEs is represented by declaring the variable name and type and maybe an initial value, in AI-Builder you can define the variable in the same way by giving a name and selecting the table and column of the cell that represent the variable and the cell value will be the variable initial value.

If you look closely u can see that the table type and the column type are bound to gather ⁽¹⁾ so they represent the variable type, by other words the cell type as an answer of the following question : what does the user mean by this cell?. This feature enables AI-Builder to have ready-for-use behaviors that are generally used on deferent projects that have the same category ⁽²⁾,and this insures the reusability of behaviors (built-in or user-defined)in deferent project that are enhanced using AI-Builder.

The binding of behaviors in an AI algorithm (called Execution in AI-Builder) depends on the algorithm type , behaviors are sometimes bounded directly as nodes in a tree⁽³⁾ or bounded to a conditional node ⁽³⁾ that are used to enable the user to control the run path of the tree algorithm.

Execution Establishment

When the user have an AI idea that he think will make his\her application better ,as for example he won't his\her application to make some shortcuts to have some the arrangement in the his\her client interface, create a new execution and start to design the AI algorithm so it can do what he want.

When the user have an AI idea that he think will make his\her application better, as for example he won't his\her application to make some shortcuts to have some the arrangement in the his\her client interface, create a new execution and start to design the AI algorithm so it can do what he want, after the user has finished building his\her algorithm he can use it by using its execution file.

for every algorithm there is a .java file that holds the behaviors code and have the execute method that is used to execute the algorithm , that file is call an execution file and is used when the user has finished his\her algorithm design and has established and compiled his\her code.

The types of algorithms that can be built in AI-Builder are structured as a tree of behaviors, by other words AI-Builder gives the ability to write code sections and distribute them on the tree nodes from the root node down to all leaves and all of these nodes will be treated as behaviors (as mentioned before a behavior is a node the contain code to be executed) so he can control what code to be executed, what to pass over. The control decisions can be made at run time as I will explain shortly.

The types of algorithms that can be built in AI-Builder are structured as a tree of behaviors, by other words AI-Builder gives the ability to write code sections and distribute them on the tree nodes from the root node down to all leaves and all of these nodes will be treated as behaviors (as mentioned before a behavior is a node the contain code to be executed) so he can control what code to be executed, what to pass over. The control decisions can be made at run time as I will explain

shortly.

There are three types of these tree algorithms:

1. Conditional tree algorithm.
2. Recursion tree algorithm.
3. Single path list algorithm.

Conditional tree algorithm:

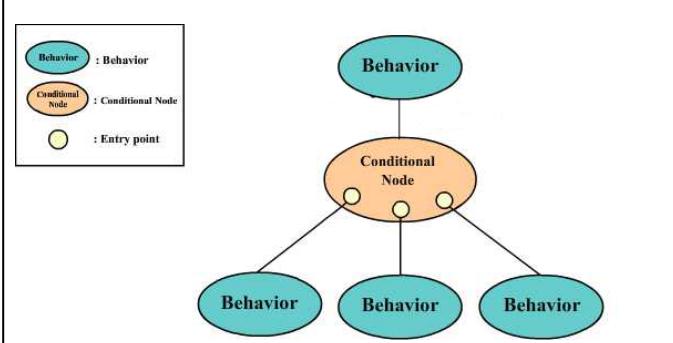


Figure13: Conditional tree algorithm.

This type of tree consists of behaviors and conditional nodes, the behaviors are used to write code in it and the conditional nodes are for control so at run time a single path in the tree is executed.

A conditional node main purpose is to determine the path for the execution, to determine the next node that will be executed. To make that possible a conditional node consists of enter points one of every behavior that is child to the conditional node, each enter point is bounded to a method that return a Boolean and it's the user job to code that method ,these methods are given to the user in a class (that comes with the execution class) that have empty methods that return true by default and every one of these methods must be coded in order to make the control possible⁽¹⁾, notice that AI-Builder sets the decisions coding in the user's application scope so the user control can be easily linked with his\her application logic. The enter points are checked from left to right at runtime and the first one that comes with a true result will be consider the valid enter point and its behavior will be the next one to be executed so after that its goon to another conditional node in a lower level of the tree, and its repeated until the last behavior have no conditional node linked to it as a child.

The reader can think of this as that each enters point has a priority, the first one has the highest priority and the last one has the lowest priority. There is a little trick that the user might consider, the user can add an enter point and make its method always return false so it is never entered so that the method can have some code to be executed before the decision process of the conditional node begins, this is can be extremely useful especially if the decision of the conditional node depends on some calculations the is needed to be done first.

Recursion Tree Algorithm:

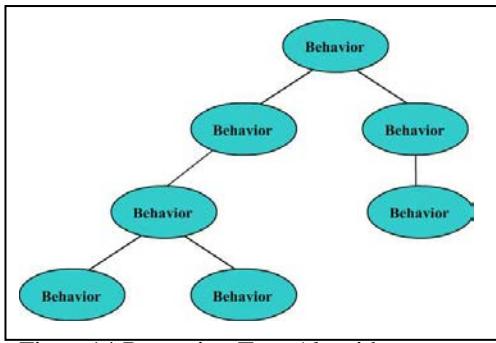


Figure14:Recursion Tree Algorithm

This type of tree consists only of behaviors, the behaviors are used to write code in it and all of the tree nodes (behaviors) will be called as the tree is a recursion tree.

Generally in recursive tree the root node will call the left child and the right child, in AL-Builder the recursion tree nodes are behaviors and the behaviors call each other recursively. Notice that in this type of tree all behaviors are executed, so the user job is to code his\her defined behaviors and the more important to control the order of execution of these behaviors. Recursion tree algorithms are generally building a single problem solution; the algorithm handles one main idea and they can be used when ever needed.

There is a little trick that might be very useful, the user can use his\her recursion algorithm (call it's execute method in its execution java file) inside another algorithm, maybe conditional node algorithm, the user can create an empty behavior and call the execute method of that algorithm inside the empty behavior. By this the user can create general algorithms and use it whenever he wants to, this insures that high reusability of the user work and is done by a single line of code.

Behaviors' List Algorithm:

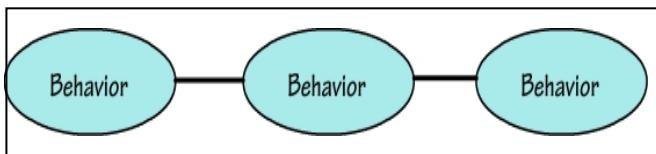


Figure15: Behaviors' List Algorithm

This is the simplest type of algorithm that the user can build; it's simply a list of behaviors that will execute in a linear order.

Generally the behaviors list algorithm is used when the user want a simple group of actions to be done when he call the execute method in his\her application, there is no conditions no recursive calls (handled by execution java file not by user; the user just call it).

Main Output:

The main output is a package of classes that represent the executions and there support classes, the user can use these executions by simply import the output package and call the execute method of a specific execution anywhere needed in his\her application.

The user can use any conditions in his\her application logic or any build in listeners in java libraries, this insure the usability and high compatibility of AI-Builder.

5. Conclusion

The intelligent user interaction development (IUID) engine makes the development of the artificial intelligence code and the addition to the product is possible and risk free. It provides the ability to use an artificial intelligence to improve the structure of the relational database. It consists of two phases, database phase that includes: loader, reader and adder and artificial intelligence algorithms phase that includes behavior establishment, execution establishment and the artificial intelligence builder.

The output of the first phase is a well-structured identified database used as input to the second phase. The output of the artificial intelligence algorithms phase is a tree structure that represented as java classes grouped together in a java package. It gives the user the power to develop some clever actions to use in his/her applications without any need to worry about the code structure or object oriented constraints in an easy way.

References

- [1]. Bertino, B. Catania, G.P. Zarri, "Intelligent database systems", Reading, Addison Wesley Professional, 2001.
- [2]. Kamran Parsaye, Mark Chignell, Setrag Khoshafian and Harry Wong, "Intelligent databases-object-oriented,deductive hypermedia technologies", New York, John Wiley& Sons, 1989.
- [3]. Androutsopoulos, G.D. Ritchie, and P. Thanisch, Natural Language Interfaces to Databases An Introduction, Journal of Natural Language Engineering 1 Part 1 (1995), 29–81
- [4]. Charniak E. 1993, "Statistical Language Learning", MITPress.
- [5]. Church K., Mercer R. 1993, "Introduction to the special issue on computational linguistics using large corpora", Computational Linguistics, 19 (1), pp. 1-24.
- [6]. Miikkulainen R. 1993, "Subsymbolic Natural Language Processing: An Integrated Model of Scripts, Lexicon, and Memory", MIT Press, Cambridge, MA.
- [7]. Marcus M., Santorini B., Marcinkiewicz M. 1993, "Building a large annotated corpus of English: The Penn Treebank", Computational Linguistics, 19 (2), pp. 313-330.
- [8]. McCarthy J., Lehnert W., 1995, "Using decision trees for coreference resolution", Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, pp. 1050-1055.
- [9]. Riloff E. 1993, "Automatically constructing a dictionary for information extraction tasks", Proceedings of the Eleventh National Conference on Artificial Intelligence, pp. 811-816.
- [10]. Riloff E. 1996, "Automatically generating extraction patterns from untagged text", Proceedings of the Thirteenth National Conference on Artificial Intelligence, pp. 1044-1049.
- [11]. Majumder P., Mitra M., Chaudhari B., 2002, "N-gram: A

- Language Independent Approach to IR and Natural Language Processing”, Lecture Notes.
- [12].Miikkulainen R., 1997,“Natural language processing with subsymbolic neural networks”, Neural Network Perspectives on Cognition and Adaptive Robotics.
- [13].Reilly R., Sharkey N. (Eds.),1992 “Connectionist Approaches to Natural Language Processing”, Lawrence Erlbaum and Associates, Hilldale, NJ.
- [14].Shashtri L.,1997, “A model of rapid memory formation in the hippocampal system”, Proceeding of Meeting of cognitive Science Society, Stanford.
- [15].Abrahams P. W. et al. “The LISP 2 Programming Language and System”, in proceedings of FJCC, No. 29, USA, 1966, pp. 661– 676.
- [16].J. McCarthy, “LISP Programmers Manual, Handwritten Draft” MIT AI Lab., Vambridge, USA, 1959.
- [17].T. Warren, “A Step toward Man-Computer Symbiosis”, Ph.D. Thesis, Massachusetts Institut of Technologie, Project on Mathematics and Computation (MAC), Technical Report MAC-TR-32, Cambridge, MA, USA, 1966.
- [18].Rohit J. Kate and Raymond J. Mooney, Using String-Kernels for Learning Semantic Parsers, COLINGACL (2006).
- [19].Woods, W. (1973). An experimental parsing system for transition network grammars. In Natural language Processing, R. Rustin, Ed.,Algorithmic Press, New York.
- [20].Woods, W., Kaplan, R. and Webber, B. (1972). The Lunar Sciences Natural Language Information System. Bolt Beranek and Newman Inc., Cambridge, Massachusetts Final Report. B. B. N. Report No 2378.
- [21].Hendrix, G. (1977). The LIFER manual A guide to building practical natural language interfaces. SRI Artificial Intelligence Center, Menlo Park, Calif. Tech. Note 138.
- [22].Hendrix, G., Sacrdoti, E., Sagalowicz, D. and Slocum, J. (1978). Developing a natural language interface to complex data. ACM Transactions on Database Systems, Volume 3, No. 2, USA, Pages 105 – 147
- [23].D.L. Waltz., “An English Language Question Answering System for a Large Relational Database
- [24].Yunyao Li, Huahai Yang, and H.V. Jagadish, Constructing a Generic Natural Language Interface for an XML Database, EDBT (2006).
- [25].Ana-Maria Popescu, Alex Armanasu, Oren Etzioni, David Ko, and Alexander Yates, Modern Natural Language Interfaces to Databases:Composing Statistical Parsing with Semantic Tractability, COLING (2004).
- [26].Yuk Wah Wong, Learning for Semantic Parsing Using Statistical Machine TranslationTechniques, Technical Report UT-AI-05-323, University of Texas at Austin, Artificial Intelligence Lab, October 2005.

An Integrative Study on Bioinformatics Computing Concepts, Issues and Problems

^{1*}Muhammad Zakarya, ²Izaz Ur Rahman, ¹Nadia Dilawar & ¹Reshma Sadaf

ABSTRACT: *Bioinformatics is the permutation and mishmash of biological science and ⁴IT. The discipline covers every computational tools and techniques used to administer, examine and manipulate huge sets of biological statistics. The discipline also helps in creation of databases to store up and supervise biological statistics, improvement of computer algorithms to find out relations in these databases and use of computer tools for the study and understanding of biological information, including ¹DNA, ²RNA, protein sequences, gene expression profiles, protein structures, and biochemical pathways [1]. The study of this paper implements an integrative solution. As we know that solution to a problem in a specific discipline may be a solution to another problem in a different discipline. For example entropy that has been rented from physical sciences is solution to most of the problems and issues in computer science. Another example is bioinformatics, where computing method and applications are implemented over biological information. This paper shows an initiative step towards that and will discuss upon the needs for integration of multiple discipline and sciences. Similarly green chemistry gives birth to a new kind of computing i.e. green computing. In next versions of this paper we will study biological fuel cell and will discuss to develop a mobile battery that will be life time charged using the concepts of biological fuel cell. Another issue that we are going to discuss in our series is brain tumor detection. This paper is a review on ³BI i.e. bioinformatics to start with.*

Keywords

¹DNA Deoxyribose Nucleic Acid

²RNA Rybo Nucleic Acid

³BI Bio Informatics

⁴IT Information Technology

I.

INTRODUCTION & CONCEPTS:

Bioinformatics is new research vicinity in Pakistan and Abdul Wali Khan University, Mardan (AWKUM) is one of those institutes who are interested and initiating this activity. SAIMS (Society for Advancement & Integration of Multiple Sciences), a very babyish research squad i.e. a sub domain of iFuture (a leading researcher's forum) in the Department of Computer Science (AWKUM), aims to set up Bioinformatics activity, in partnership with Department of Botany, Biology, Physics, Chemistry, Mathematics, Statistics, and even social sciences.

Bioinformatics is the arithmetical, statistical, algebraic and computing routines that endeavor to explain biological issues and problems using DNA, amino acid sequences and interrelated information. Usually speaking bioinformatics is the establishment and development of sophisticated information and computational technologies for troubles in biological science most frequently molecular biology and gradually more in other areas of biology as well. As such it deals with schemes for storing, accessing and analyzing biological data and information such as DNA, RNA, protein sequences, protein structures & functions and genetic interactions. Computer databases play a chief role in bioinformatics. Similarly the information retrieval mechanisms, techniques and algorithms that are applied in mathematical, statistical and computer sciences are also applicable to biological data.

Molecular biology is the understanding of biological processes at the molecular level i.e. it deals with the biological activity at the molecular level. Molecular biology is an amalgamation of genetics and biochemistry. Without Molecular biology, there would not have been bioinformatics. DNA holds the genetic information of all plants, animals, bacteria and some viruses. A property of living organism is that DNA is reproduced and migrated on to the subsequent generations. It contains instruction for making proteins. DNA contains a base plus a deoxyribose sugar and phosphate. The four bases of DNA are cytosine (C), adenine (A), thymine (T) and guanine (G). RNA contains ribose instead of deoxyribose and uracil (U) instead of thymine (T). The three major classes of RNA are transfer RNA (tRNA), messenger RNA (mRNA) and ribosomal RNA (rRNA). Gene is a segment of DNA

¹ Department of Computer Science, Abdul Wali Khan University (AWKU), Mardan, Khyber Pakhtunkhwa (KPK), Pakistan

² School of Information Systems, Computing and Mathematics, Brunel International University, London, UK

* Corresponding Author

representing nucleotides that requisite for the fabrication of a functional protein or a functional DNS, RNA molecule.

The rest of the paper is structured as follows. In next section i.e. Section II we are going to define the terms like DNA, RNA, and Gene etc. In section III we have a brief discussion on bioinformatics and bioinformatics computing. In section IV we give an overview of the use of computing methods and application to biological data like databases. In section V a sketch is given on issues, hurdles and problems that computer scientist or biologists are or will face in bioinformatics. In section VI, the role of computer networks and distributed computing is discussed in term of biological data. In subsequent section VII an overview of bioinformatics literacy in Pakistan is sketched. We conclude in section VIII, with some future directions and work in subsequent section IX.

II. RELATED WORK & GENERAL TERMS

i. Biology

Biology is the study of different living organisms like plants, humans etc. Biology is the learning of life. In conjunction with chemistry & physics, biology is one of the biggest and most imperative branches of science. At the uppermost level, biology is divided into different sections based on the kind of mortal being considered for study. For example zoology is the study of animals, microbiology is the study of microorganisms and botany is the study of plants. More short and snappy classifications based on the grouping of organism being studied, biology restrains numerous other expert sub-disciplines, which may spotlight on just one kind of organism or address organisms from diverse categories [5]. This take account of biochemistry that is the crossing point and edge between biology and chemistry; ecology which studies the interactions between organisms themselves, molecular biology which looks at life on the molecular level, cellular biology which studies diverse types of cells and how they work, physiology which looks at organisms at the level of tissue and organs, ethology which studies the deeds of animals particularly complex animals, and genetics overlapping with molecular biology, which studies the rules & regulations of life i.e. DNA, RNA etc [7].

ii. Molecular Biology

Molecular biology is the understanding of biological processes at the molecular level i.e. it deals with the biological activity at the molecular level. Molecular biology is an amalgamation of genetics and biochemistry. Without Molecular

biology, there would not have been bioinformatics. It deals with the study of molecular building blocks such as nucleic acid and proteins [5].

iii. Cell

Cell is the structural or functional unit of the body or living organism. There are two components of cell i.e. nucleus and cytoplasm. Nucleus has nuclear materials which have chromosomes. There are 46 chromosomes which has 23 pairs in a human body. Chromosomes are the combination of DNA. Some organisms are unicellular like bacteria and protozoa, which contain only a single cell. Others are multi cellular like plants, animals and fungi, which carry out thousands of biochemical reactions in each minute and produce a new cell each time.

iv. Enzymes

Special organic substance composed of polymers and amino acid which acts as a catalyst to police and regulates the speed of chemical reactions that are implicated in the metabolism process of living organisms.

v. DNA

DNA holds the genetic information of all plants, animals, bacteria and some viruses. Deoxyribo means sugar, nuclei means nitrogenous base and acid stands for phosphoric acid. DNA is the structural or functional unit of the cell. A property of living organism is that DNA is reproduced and migrated to the next generation. It contains instruction for making proteins. DNA contains a base plus a deoxyribose sugar and phosphate. The four bases of DNA are cytosine (C), adenine (A), thymine (T) and quinine (G). The chromosome composed of single molecule of DNA shaped like a twisted ladder that is composed of linked chemical compounds known as nucleotide, which consists of sugar, phosphates and bases. Bases consist on purine and pyronidine. Pyronidine consists of cytosine (C), thymine (T) while purine is a group of adenine (A) and quinine (G), where adenine (A) usually pairs with thymine (T) and guanine (G) pair with cytosine(C). Purine and pyronidine are called nitrogenous bases. A nucleotide may also consist of pyronidine, sugar and phosphate. The way in which all these bases or two groups i.e. purine and pyronidine are combined and integrated to make a DNA are called pairing rules. It states that always purine make pair with pyronidine i.e. adenine (A) with

thymine (T) and guanine (G) with cytosine(C). Always two nitrogenous bases make double or triple bond i.e. hydrogen bonds. In summary DNA is double standard and ladder type and it is composed of large number of nucleotides or nitrogen bases. The structure of the integration of DNA is called DNA sequence and it is unique in nature. DNA is also called chromosomal DNA.

DNA is the crucial foundation of genetic information in cells. Humans, plants, animals, and bacteria all living things contain DNA. DNA is physically passed from generation to generation, bestowing definite persona and character of parents to their kids. The cause why kids have bodily characteristics from each of their parents e.g. an infant or teen may have their mother's eye color and father's hair color is for the reason that they received half their DNA from each parent.

vi. RNA

RNA contains ribose instead of deoxyribose and uracil (U) instead of thymine (T). The three major classes of RNA are messenger RNA (mRNA), ribosomal RNA (rRNA) and transfer RNA (tRNA). Ribo means sugar, again nucleic is nitrogenous base and acid stands for phosphoric acid. RNA is single standard and it is synthesized from the DNA structure. RNA is important for transmitting of message or information within the cell and outside the cell. Bases consist on purine and pyronidine. Pyronidine consists of cytosine (C), uracil (U) while purine is a group of adenine (A) and quinine (G), where adenine (A) usually pairs with uracil (U) and guanine (G) pair with cytosine(C). Purine and pyronidine are called nitrogenous bases. The three different types of RNA have different functionality. For example mRNA transmit message to ribosome and rRNA reads the mRNA. tRNA makes proteins and transfers or sends it to ribosome. RNA is used in protein synthesis or production.

vii. Proteins

Proteins are fundamental parts of organism like nucleic acids that participate in nearly every process within the cell. Proteins define the structure and function of living materials. It contains four elements, carbon, hydrogen, oxygen and nitrogen. Most proteins restrain some sulphur. These elements are bonded together as a compound called amino acid. So proteins are made up of amino acids. So far 20 amino acids are commonly founded in proteins. Each amino acid has two groups, carboxyl group i.e. COOH and

amino group i.e. NH₂. These two groups are bonded jointly by peptide bond which produces polypeptide chain where the number of amino acid is from 40 to 500. Proteins are biochemical compounds consist of one or supplementary polypeptides usually folded into fibrous form or globular form facilitating a biological function.

a) Proteins Structure

The majority of proteins fold into distinctive 3 dimensional structures. The form into which a protein logically folds is known as its native conformation. The structure of the protein depends on the number of polypeptides, amino acid type and sequences. Proteins molecules may be fibrous as keratins that are founded in hair, fur, nails, claws, outer skin, cornea of eye, bones and ligaments. Others are globular e.g. hemoglobin globular proteins often consist of two or more polypeptides.

b) Protein Sequences

Protein sequencing is a procedure to conclude the amino acid sequence of a protein in addition to which conformation the protein agreed to and the scale or degree to which it complexes with any non peptide molecule recovering the structure and function of proteins in living being. It is a significant tool & gizmo for understanding cellular process and allows drugs that target definite metabolic to be invented more easily. The two major direct methods of proteins sequencing are edman degradation reaction and the mass spectrometry. It is also achievable to produce an amino acid sequence from DNA & mRNA sequence encoding the proteins.

c) Functions of proteins

Proteins are extremely essential, significant and vital molecules in our cells. Every protein within the body has a definite unique function. Some proteins are involved in bodily movement while others are involved in structural support or in defense against germs and diseases. Antibodies proteins make the defense system, contractile proteins are responsible for movement, hormonal proteins are messenger proteins help in coordination of certain bodily activities, enzymes proteins facilitate biochemical reactions, structural proteins provide support, storage proteins stores amino acids and transport proteins migrate molecules from one position to another around in the body.

- Proteins make the major component of cell.
- Proteins make muscles that help in movement.
- Proteins perform the function to carrying oxygen from lungs to different parts of body.
- Proteins help in food digestion
- Proteins are important source of energy
- It is the main component of some seeds e.g. pulses, peas etc
- Proteins help in blood clotting
- Many proteins act like enzymes that catalyze biochemical reactions and are fundamental to the process of metabolism.

viii. Gene

Gene is a segment of DNA representing nucleotides required for the fabrication and invention of a functional protein or a functional RNA molecule. Gene is the basic unit of heredity found in the cell of living organisms e.g. from bacteria to human. It determines the physical characteristics that an organism inherits such as color of hairs or shape of face. Genes composed of small segments of DNA (a molecule that form thread like structures called chromosome). The study of functions and behavior of genes is called genetics. A sequence of DNA is called gene and different cell DNA; they are different on the basis of arrangement not on the basis of their structure. Any change in of disarrangement of gene is called mutation. Mutations are of two types i.e. useful and abnormality.

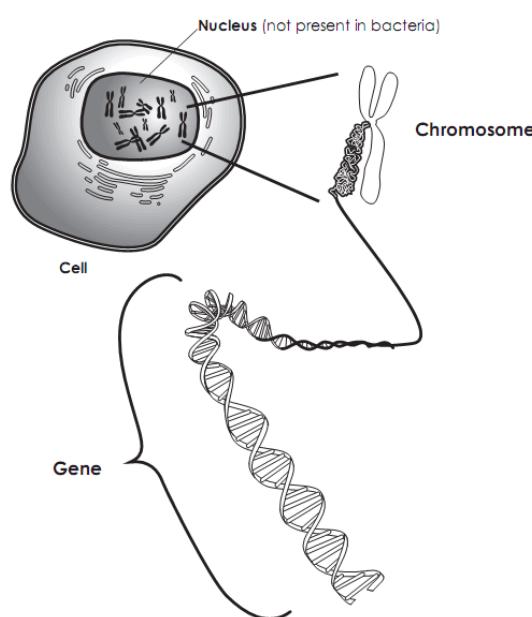


Fig 1: DNA, Cell, Chromosome, Gene [6]

ix. Chemoinformatics

Chemoinformatics is the application of computational, statistical and mathematical methods to chemical problems. Computer technology, particularly computational power, networking & storage space and capacity, has sophisticated to an arena that it is proficient of managing some of the recent disputes and hurdles posed by chemistry and biology [2]. This makes it achievable to grip the enormous quantity of data or information that is being generated as a result of the international genome project [3, 17]. JMOL [18] is a simulation environment that researchers from biochemistry, chemistry and bioinformatics departments use for 3D representation of molecular structure of different compounds. The tool provides an easy use for the beginners. Although all functionalities are not GUI based, but command prompt can be used to execute all commands, which are available. JMOL is a java based application that is portable and can be used on any platform or architecture.

III. BIOINFORMATICS:

Up-to-the-minute molecular biology and medical research encompasses an increasing quantity of information, as well as an increasing diversity of information. The usage of informatics to organize, manage, and scrutinize these information has subsequently become a significant element of biology and biomedical research. Bioinformatics is the mixture of computing, mathematics, and biology to address this requisite [4]. Data management and knowledge discovery are the two most important issues that are addressed in [4]. The integration of biological and computational sciences have resulted in this new era and story of bioinformatics, which is an interested area for students, researchers and scientist at most teaching departments and knowledge places and institutors of repute. Following are some common applications of bioinformatics.

- Drug design process
- Heart modeling
- DNA sequencing
- Biological fuel cells
- Plants pathology
- Criminal investigations
- Health care
- Agriculture
- Brain tumor detection
- Environmental protection
- Crops / development of crops varieties
- Industrial applications etc

Thus bioinformatics is an interdisciplinary vicinity at the meeting point of biological, mathematical, statistical, computer, and information sciences essential and crucial to handle, route, and recognize bulky amounts of information, for illustration from the sequencing of the human genome, or from bulky databases holding information about animals and plants for utilize in discovering and developing new drugs [13].

IV. BIOINFORMATICS COMPUTING:

Bioinformatics Computing (BC) means to apply informatics and computational methodologies and techniques on biological information and data to process, manage and understand large amount of data i.e. biological information and to store these information in data stores i.e. databases in computational terms. Machines have played a major role in human life al the time. When biologist felt that there research data in enormous and hence needs to be properly managed and stored, they find computer and information technology that can easily handle their issues. Hence a new field was introduced i.e. bioinformatics and bioinformatics computing.

V. ISSUES & PROBLEMS IN BIOINFORMATICS COMPUTING:

Bioinformatics is an integrated science and hence face a lot of problems and challenges. Computer scientists often have no knowledge of biology and biologist's lacks computer programming and information theory as well. Therefore there is a need to minimize the gap between these scientists that cover a lot of other disciplines as well. So we have to develop some special teaching mechanism and course contents in academia. In [8] the author has sketched a blueprint of teaching mechanism and course contents. To congregate the requirements of computer scientists, critical facets of the chemistry of existence & life have to be presented, together with an introduction to procedures, issues, hurdles & challenges and scientific idea behind genomics, molecular biology, and biotechnology. To assemble the desires of biologists, essential skills and ethics in information technology and computational science must to be offered, together with the development and application of computer programming proficiency. We need critical bioinformatics experts in our academia but it's too difficult as most experts are hired by private organization where there is a restriction on every information to be disclosed to audience.

Another possible solution is to initiate some integrated programs that will address multiple subjects and hence providing a broad background in all facets of bioinformatics. In such a program students have more choices and options to pursue in the field. Our university has also started a BS (Integrated) 4 years program that we think, will results in scientists that will have knowledge of

every subjects and discipline, but still they need some better curriculum to address our future needs and requirements. Following are some major issues currently researchers and academia need to focus.

- a) **Gap between biological and computer sciences**
Because biology and computer sciences are two different fields, biologists don't know the complex computer programming languages and computer experts needs biological data and knowledge before writing programs [8].
- b) **explosion of data**
The steep amount of information in Genbank is now rising on exponential velocity and as data type beyond RNA, DNA and protein sequences being to undergo the same kind of sudden increase, simple managing, monitoring, accessing and presenting this data to the user in an intellectual form is a decisive and critical task [15].
- c) **large computation**
It requires a large scale computations (for scheduling, exact monitoring and fault tolerance purpose), as several fields and technologies are merged in bioinformatics.
- d) **quality of data**
The quality of large scale data is also notoriously variable, with relative high error rates and low productivity.
- e) **algorithms issues**
Bioinformatics algorithms need to encompass complex scientific assumptions that can do complicated programming and data modeling.
- f) **complex biological data (an opportunity for computer scientists)**
Biological data is extremely difficult and interlinked. A speck on DNA array for example, is linked not only to instant information about its intensity but to layers of information about DNA sequence, genomic location, structure, function and many more. Creating information systems that permits biological scientists to effortlessly track these links exclusive of getting in a sea of information, is also a gigantic opportunity for computer scientist [13, 15].
- g) **overlap graphs**
A memory proficient data structure representing accurate match overlap graphs with the application for subsequent generation DNA assembly.
- h) **availability of large tools**
The wide variety of the analysis tools available today means that it is often difficult for the non specialist to choose an appropriate tool for his

specific problem, in order to select the most suitable method for the particular problem.

i) **education**

Development of suitable bioinformatics course curriculum for secondary, under graduate and graduate education level is a major problem in different academia and laboratories.

j) **database issues**

The number of publicly available data repositories has grown in an exponential rate in 2002. There are thousands of publicly bimolecular data sources. These data sources are developed independently, the data they contains are represented in a wide variety of formats, annotated using a variety of methods and may or may not be supported by a database management system.

k) **data integration issues**

Many of the problems facing genomic data integration are related to data semantics. For example the difference between the meaning of the data representing in a data source and the difference between the semantics within a set of sources is a major issue.

VI. ROLE OF COMPUTER NETWORKS IN BIOINFORMATICS:

New generations of networked computing systems, aided by the global existence of the Internet infrastructure, are playing an increasingly important role in the society and civilization. Distributed computing has altered the face of computing and presented rapid and clear-cut solutions for a diversity of complex issues & problems for diverse disciplines and fields. With the rise of new computing era distributed systems like Grid, Cloud systems, there is a need to implement cloud computing over the sheer amount of data in bioinformatics. Grid computing is term used for On Demand Computing, where services are provided to customers on demand base. The underlying communication media in Grid computing is local area networks, high speed Ethernet. It means Grid may span specific and limited geographic area. Cloud computing indicates the infrastructure as a "Cloud" from which businesses and customers are competent and capable to access & use applications from everywhere in the world using on demand techniques. Storage, processing, managing and monitoring of huge amount of data will be easy if Grid or Cloud i.e. distributed concepts are applied and implemented over bioinformatics [9, 10, 11, 12].

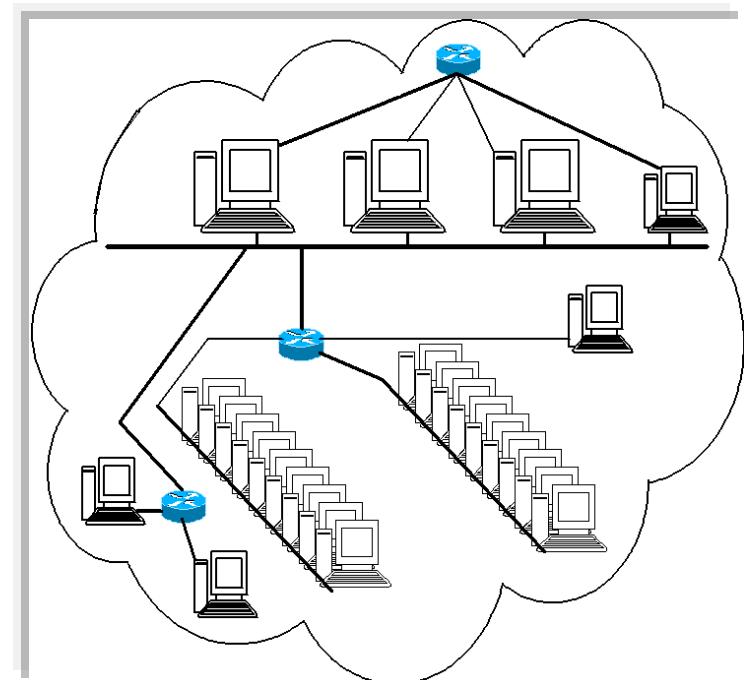


Fig 2. High Performance Computing

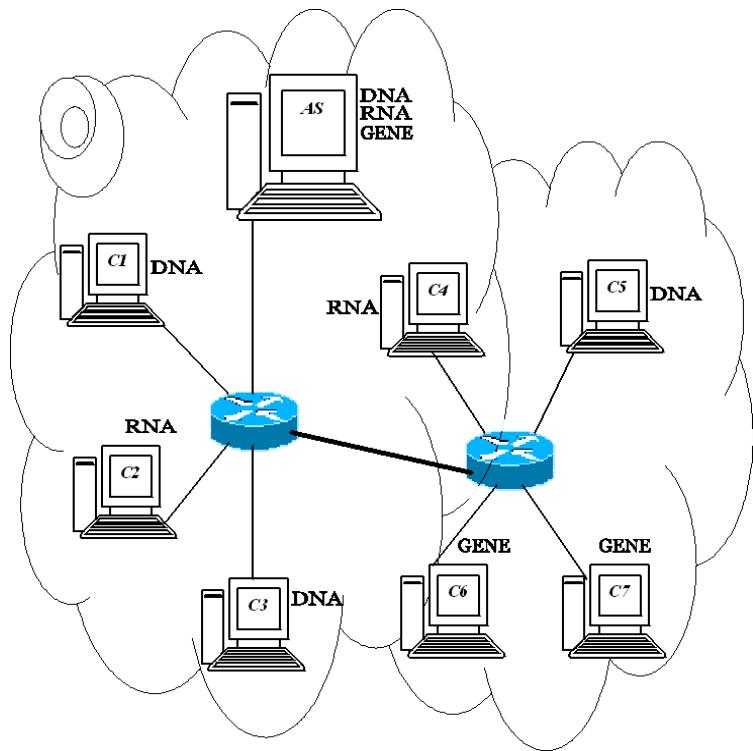


Fig 3. HPC in Bioinformatics

VII. HPC OVER BIOINFORMATICS:

High Performance Computing can be implemented over the huge amount of biological data, to achieve High availability and better performance. Fig 3 shows a HPC environment where DNA, RNA and GENE related database is replicated in the network. Fig 2 shows HPC environment where more than one systems are trying to find a solution for a specific problem. Fig 3 shows how biological databases are replicated over the network and how they interact each other to access and service user in a distributed fashion. The AS is responsible to authenticate and authorize users and to provide an easy way for accessing and storing biological data. AS maintains all the replication information and can disconnect users depending on server load to balance the workloads i.e. load balancing. The architecture provides the following benefits.

- *Limited & Local Security Policy*
- *Little computation as compared to Global security policy*
- *User accesses AS, strong authenticated & authorization check*
- *Performance Scalability*
- *Load balancing*
- *Quality of Service*
- *No need for resources to check the user identity*
- *Local & Quick allocation of resources by AS*
- *No Single point of failure, affect some part of the HPC*
- *Fault Tolerance*

One of the main disadvantage of this architecture is that AS are required to inform all corresponding AS in case of new node to any geographical community means if some database is updated, all the replicated databases are also needed to be updated as soon as possible. If [20] the authors have proposed a distributed architecture CISM, which is still under development. This agent based systems showed in its initial study and results that it is adequate for building and designing of complex systems taking the advantage of composite services. CISM provide a robust, flexible, modular and adaptable solution that can cover most requirements of a wide diversity of distributed systems. [19] describe in full detail the opportunities for the Pakistani IT industry within bioinformatics, and have sketched a very knowledgeable summary of current bioinformatics activities in Pakistani universities, academia and laboratories. Different institutes of repute are considered for their bioinformatics literacy and goals.

VIII. BIOINFORMATICS AND PAKISTAN:

Two fields biology and computer sciences merged with each other and introduced a new area called bioinformatics, which is mostly interested area for students,

researchers and teachers in academia, organizations and laboratories. In Pakistan bioinformatics was 1st introduced in 2003 with the help of two institutes i.e. COMSATS Institute of Information Technology and Muhammad Ali Jinnah University, but nowadays approximately 12 or more institutes are working in this field. At these institutes and universities researchers and bioinformatics expertise were focusing on guess & forecast of the functions of the newly recognized genes, finding the protein interaction, restriction analysis, identification of mutations, determination of the active sites of enzymes and etc. In Pakistan biotechnology is used from corner to corner to an extensive range of applications including agriculture, health care, crime investigation, environmental protection, industries, development of new crops varieties, drug designing, finding targets of drugs and many more. In Pakistan LIMS (laboratory Information Management Systems) and medical informatics and clinical trials management are used for laboratory tests to improve the results of databases.

As bioinformatics is mostly interested area for researchers, students and teachers in Pakistan. Students and researchers have made a group that has produced useful platform for young bioinformaticians. The group is called Regional Students Group (RSG-Pakistan), which came into being in January 2010. The aim of this platform is to develop the new generation, help students in their projects, encourages students and researchers to come & join, to train students and provides opportunities to commit with experts in computational biology.

A brief survey at different institutes, research forums and organization were taken into consideration. Fig 4 shows computer and internet availability and use to researchers, students and employers. In some cases we were unable to get the opinions. In Fig 5 we have sketched the bioinformatics tools that researchers, students and employers at different institutes, research forums and organizations uses for analysis of there problems and research issues. Fig 6 shows Jmol i.e. a java base simulator for representation of 3D molecular structure of different compounds. Mostly students related to biochemistry and chemical sciences were observed that were some what familiar with Jmol. Fig 7 is short survey of PROSPECT. PROSPECT stands for PROtein Structure Prediction and Evaluation Computer ToolKit that is a protein structure prediction system that make use of a computational procedure called protein threading to assemble a protein's 3-D model. We observed that students at most universities and institutes are not properly aware of bioinformatics literacy and its role in our society. Most of the students were in favor of complexity of bioinformatics. Students of computer literacy were founded that they have no interest in bioinformatics computing, although a few were well educated on bioinformatics.

iFuture a leading Research Group in Computer Science Department is also working on bioinformatics computing and arranging different seminars and trying to explore this new era of computational and biological

sciences. This work is a result of the efforts of this leading research group.

Government of Pakistan needs proper attention to the lose development in bioinformatics literacy. They needs to offer scholarship for higher study in bioinformatics which will encourage our students and will increase their number and interest in this new and immature field of science & technology in Pakistan. Educational institutes need to arrange workshops and seminars on this growing field. Similarly the experts are required to discuss some easy bioinformatics applications and their uses in rural areas, which will force the people on thinking and hence this field of biotechnology will make a place in our developing country.

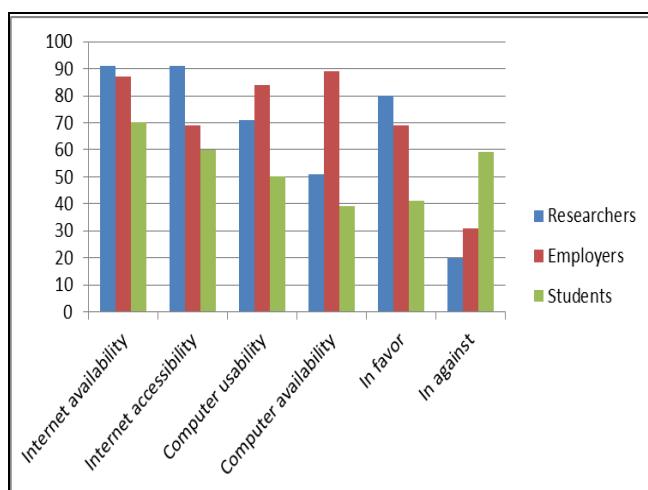


Fig 4. Bioinformatics Literacy ratio in Pakistan

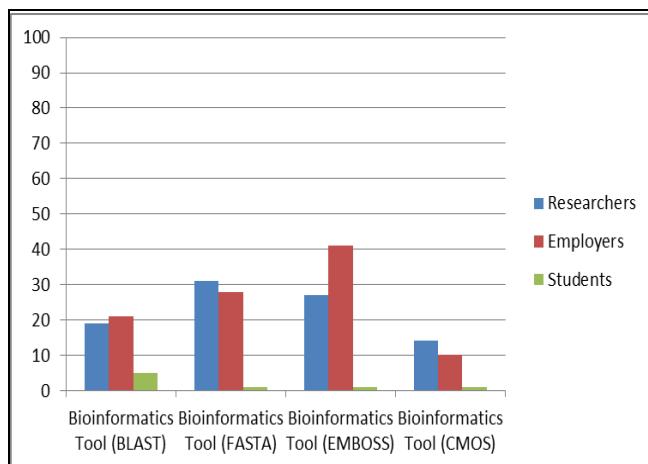


Fig 5. Bioinformatics Literacy ratio in Pakistan

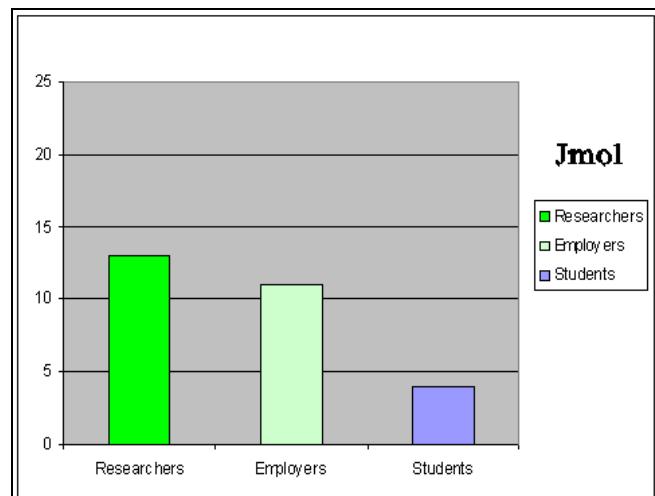


Fig 6. Jmol usage

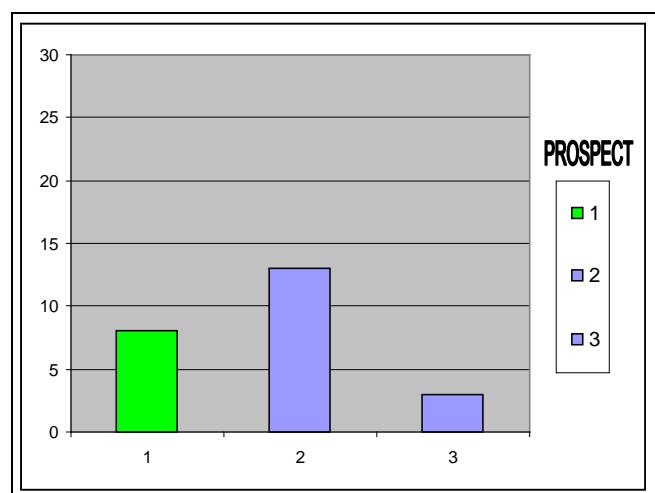


Fig 7. PROSPECT usage in Pakistan

IX. CONCLUSION:

Biology is problematical and complex field of science. In other fields such as physics, chemistry or engineering sciences and Information Technology, applications are developed from well-characterized ethics and ideology. With biotechnology on the leading edge of molecular biology research, it can be intricate or even impossible to predict the conclusions of manipulations and they can have surprising consequences and cost. For the reason that it is impossible to wholly guess the result of these events, scientists must carry out experiments, take observations, purify theories, and to finish develop functional applications. This is why biotechnology research is so difficult, time uncontrollable, and loaded with astonishing setbacks and frustrations.

In this article we have discussed some bioinformatics issues and proposed the need for HPC to efficiently maintain, access and store all biological data and information. Our next step is to implement HPC over bioinformatics and will study the proposed architecture statically and will have some simulations results.

X. FUTURE WORK:

This article discusses bioinformatics on a very basic level that is easily understandable by computer scientist. The purpose is to draw a sketch and some overview of the challenges that need to be addressed. This work is basically an initiative step to implement the cloud computing environment over bioinformatics. Biological data is so huge that a single processor cannot be addressed for its execution neither can be stored on a single system. Therefore High Performance Computing like Clusters, Grids and Clouds are needed to implement for easy data storage, access and management.

ACKNOWLEDGMENT

This work is fully supported by Abdul Wali Khan University, Mardan, Khyber Pakhtun Khwa (KPK), Pakistan. The author(s) of this article are greatly thankful to SAIMS i.e. Society for Advancement & Integration of Multiple Sciences for full guidance and major support. iFuture is also given a number of credits for arranging seminars on the subject matter. iFuture is a Research Group at the Department of Computer Science. The authors are very thankful to ¹Dr. Zidong Wang, who helps us and guided us for the entire period of this survey.

REFERENCES:

- [1] Bourne, Philip E., and Helge Weissig, eds. Structural Bioinformatics. Methods of Biochemical Analysis 44. Hoboken, NJ: Wiley-Liss, 2003. BA Call Number: 572.8733
- [2] Hwa A. Lim, Bioinformatics and Chemoinformatics in the Drug Discovery Cycle, Santa Clara, California 95051- USA 1997
- [3] Mapping and Sequencing the Human Genome, (National Research Council, National Academy Press, Washington, D.C., 1988).
- [4] Vladimir B. Bajic, Vladimir Brusic, Jinyan Li, See-Kiong Ng, Limsoon Wong, From Informatics to Bioinformatics, Australian Computer Society 2003
- [5] Nikola Kasabov, Igor A. Sidorov, Dimiter S. Dimitrov, Computational Intelligence, Bioinformatics and Computational Biology: A Brief Overview of Methods, Problems and Perspectives, Computational and Theoretical Nanoscience Vol.2, 473–491, 2005
- [6] Yali Friedman, Ph.D., Building Biotechnology, Third Edition, ISBN-13, Hardcover: 978-0-9734676-5-9, Sofcover: 978-0-9734676-6-6
- [7] <http://www.wisegeek.com/>
- [8] Gustavo Caetano-Anollés, Teaching Bioinformatics: Needs and Challenges, gca@uiuc.edu
- [9] Yanpei, Vern Paxson, Randy H.Katz, What's New About Cloud Computing Security, University of California at Berkeley, January 20, 2010
- [10] Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy H. Katz, Above the Clouds: A Berkeley View of Cloud Computing, University of California at Berkeley, February 10, 2009
- [11] Muhammad Zakarya, Izaz Ur Rahman, Mukhtaj Khan, Cloud QoS, High Availability & Service Security Issues with Solutions, BUJICT July 2011
- [12] Muhammad Zakarya, Ayaz Ali Khan, Hameed Hussain, Grid High Availability & Service Security Issues with Solutions, ICIIT 2010, IEEE
- [13] Bryan Bergeron, Bioinformatics Computing, Prentice Hall of India, 2006
- [14] Stephen A.Krawetz, David D.Womble, Introduction to Bioinformatics – A theoretical and practical approach, Humana Press 2009
- [15] See-Kiong Ng and Limsoon Wong, Accomplishments and Challenges in Bioinformatics, 1520-9202/04/\$20.00 © 2004 IEEE
- [16] Lawrence Hunter, Molecular Biology for Computer Scientists - ARTIFICIAL INTELLIGENCE & MOLECULAR BIOLOGY, Chapter 1
- [17] C. Stan Tsai, An Introduction to Computational Biochemistry, ISBN 81-265-1105-2
- [18] JMOL online & offline documentation
- [19] Dr. Ahmad Ijaz Gilani, Sadia Zaidy, Opportunities for the Pakistani IT industry within bioinformatics, Ministry of IT, Government of Pakistan, Islamabad, December 2005
- [20] Javier Bajo, Carolina Zato, Fernando de la Prieta, Ana de Luis, Dante Tapia, Cloud Computing in Bioinformatics, Springer-Verlag Berlin Heidelberg 2010

Keywords

¹School of Information Systems, Computing and Mathematics, Brunel International University, UK

- [21] Anthony Finkelstein, James Hetherington, Linzhong Li, Ofer Margoninski, Peter Saffrey, Rob Seymour, Anne Warner, Computational Challenges of Systems Biology, 0018-9162/04/\$20.00 © 2004 IEEE
- [22] Atul J. Butte, Challenges in bioinformatics: infrastructure, models and analytics, TRENDS in Biotechnology Vol.19 No.5 May 2001
- [23] Peter Tarczy-Hornoch, Mark Minie, BIOINFORMATICS CHALLENGES AND OPPORTUNITIES, MEDICAL INFORMATICS
- [24] See-Kiong Ng and Limsoon Wong, Accomplishments and Challenges in Bioinformatics, 1520-9202/04/\$20.00 © 2004 IEEE
- [25] N.M. Luscombe, D. Greenbaum, M. Gerstein, What is bioinformatics? An introduction and overview, Yearbook of Medical Informatics 2001
- [26] Mario Cannataro, Carmela Comito, Filippo Lo Schiavo, and Pierangelo Veltri, Proteus, a Grid based Problem Solving Environment for Bioinformatics: Architecture and Experiments, IEEE Computational Intelligence Bulletin 2004

ABOUT THE AUTHOR(S):



Currently the author of this paper is working as lecturer in computer science department of Abdul Wali Khan University Mardan, Pakistan. The author of this paper is a new researcher to the field of new emerging computing technologies like Grid, Cloud and Green Computing. He has done MS in Computer Science and is interested for a doctorate degree in computer engineering. Currently the author of this paper is working on security issues in Grid and Cloud computing i.e. distributed systems. The author is interested in implementing Cloud computing over bioinformatics. He is also working and interested to Green the Cloud.



Izaz Ur Rahman is working as Lecturer in the same university and department. He is enrolled for a PhD study in Brunel International University, UK. His research area is distributed systems and bioinformatics.

Nadia Dilawar and Reshma Sadaf are students of master in Computer Science, at Computer Science Department of Abdul Wali Khan University. Both of them are interested in bioinformatics computing. This work is an initiative step toward the implementation of a new research area i.e. implementation of Grid, Cloud on demand computing environment over bioinformatics

Automatic Fracture Detection Using Classifiers- A Review

S.K.Mahendran¹ and S.Santhosh Baboo²

¹ Doctoral student in the Department of Computer science ,Bharathiar University, Coimbatore, Tamil Nadu, India

² Reader in the Postgraduate and Research department of Computer Science at D G Vaishnav College, Chennai, Tamil Nadu. India

Abstract

X-Ray is one the oldest and frequently used devices, that makes images of any bone in the body, including the hand, wrist, arm, elbow, shoulder, foot, ankle, leg (shin), knee, thigh, hip, pelvis or spine. A typical bone ailment is the fracture, which occurs when bone cannot withstand outside force like direct blows, twisting injuries and falls. Fractures are cracks in bones and are defined as a medical condition in which there is a break in the continuity of the bone. Detection and correct treatment of fractures are considered important, as a wrong diagnosis often lead to ineffective patient management, increased dissatisfaction and expensive litigation. The main focus of this paper is a review study that discusses about various classification algorithms that can be used to classify x-ray images as normal or fractured.

Keywords: *X-ray, Classification, Machine Learning, Fusion classifier*

1. Introduction

Since Wilhelm Roentgen discovered the existence of X-rays in 1895, medical imaging has advanced at a tremendous rate and has become the fundamental diagnostic tool in modern healthcare. As a combination of radiation and computer image processing technologies, digital X-ray imaging device are being widely used in many medical applications. Image classification is an area in image processing where the primary goal is to separate a set of images according to their features into one of a number of predefined categories. It is the problem of finding a mapping from images to a set of classes, not necessarily object categories. Each class is represented by a set of features (feature vector) and the algorithm that maps these feature vectors to a class uses machine

learning techniques. The ability to perform binary-class image classification as an automatic task using computers is increasingly becoming important in fracture detection. This is due to the huge volume of image data available, which are proving to be difficult for manual analysis. The difficulty arises because of lack of human experts, poor quality images and time complexity. The current market need is to have techniques which can classify images as having normal or fracture, with minimum intervention from the users in an efficient and effective manner.

This paper presents a review of the various classification approaches that can be used to classify bone x-ray images as either normal or fractured. The rest of the paper is organized as follows. The working of a general classification system is presented in Section 2. Section 3 reviews the various machine learning classification methods, while Section 4 presents the concepts of fusion classification. Section 5 concludes the work.

2. General Approach to Classification

As mentioned earlier, classification, also known as pattern recognition, discrimination, supervised learning or prediction, is a task that involves construction of a procedure that maps data into one of several predefined classes [26]. It applies a rule, a boundary or a function to the sample's attributes, in order to identify the classes. Classification can be applied to databases, text documents, web documents, web based text documents, etc. Classification is considered as a challenging field and contains more scope for research. It is considered challenging because of the following reasons :

- Information overload –The information explosion era is overloaded with information and finding the required information is prohibitively expensive.
- Size and Dimension – The information stored is very high, which in turn, increases the size of the database to be analyzed. Moreover, the databases have very high number of “dimensions” or “features”, which again pose challenges during classification.

The input data for a classification task is a collection of features arranged as in row-wise fashion (records). Each record, also known as an instance or example, is characterized by a tuple (X, y) where X is the attribute set and y is a special attribute, designated as the class label (also known as category or target attribute).

A classification technique, or a classifier, is a systematic approach to building classification models from an input data set. Examples include, Decision Tree Classifiers, Rule-Based Classifiers, Neural Networks, Support Vector Machines and Naïve Bayes Classifiers. Each technique employs a learning algorithm to identify a model that best fits the relationship between the attribute set and class label of the input data. The model generated by a learning algorithm should both fit the input data well and correctly predict the class labels of records it has never seen before. Therefore, a key objective of the learning algorithm is to build models with good generalization capability, i.e., models that accurately predict the class labels of previously unknown records. First, a training set consisting of records whose class labels are known must be provided. The training set is used to build a classification model, which is subsequently applied to the test set, which consists of records with unknown class labels.

3. Machine learning

Machine learning is the process of automating the development of some part of a system which performs some task. The algorithm, parameters to an algorithm or process can be learnt adaptively over a period of time. The overall structure of a machine learning approach to a problem involves three steps [32]:

- (i) The generation of some representation of a solution to the problem.
- (ii) The evaluation of the generated solution.
- (iii) If the evaluated solution is not good enough, the solution is iterated (i.e. almost always improved) and the machine learning process goes to step 2.

Given the selection of some representation of a solution to the problem, the initial generation is usually random but constrained by some parameters. For example, in a neural network the structure is fixed and the weight associated with each link is generated according to some algorithm which ensures that the initially generated solution will almost certainly not be the same from time-to-time. However, there are a wide variety of possible representations, including Feed-forward neural networks, Genetic algorithms, Support vector machines, Simulated annealing, Decision trees, Naïve Bayes Algorithm, Bayesian networks and Genetic programming.

There are, broadly, three ways in which the iteration from one solution to the next can be performed. This iteration is how the search for a good solution is carried out. Each of the three types of iteration will be summarized briefly in this section.

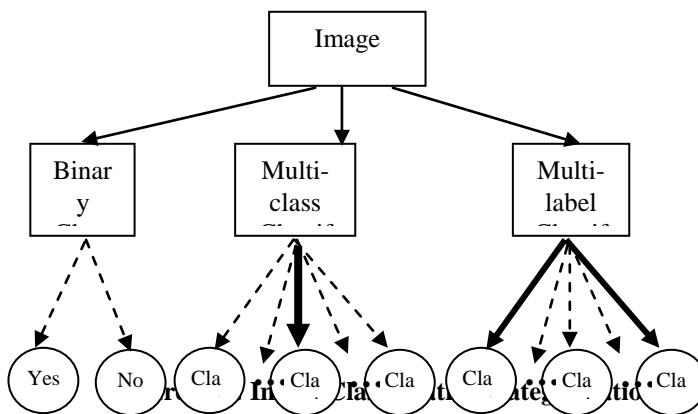
- (i) Unsupervised - Unsupervised learning is normally used to locate patterns in the input data. No information is given to the system, which finds the patterns as to the correctness or incorrectness of the patterns.
- (ii) Reinforcement - Reinforcement in terms of the quantity of information given to the system regarding the correctness of its output. Reinforcement learning is intermediary between supervised and unsupervised learning.
- (iii) Supervised - When supervised learning is used the precise, correct output which should have been given for any particular training input is known to the system and used by the system to adjust the answer it will give to other training examples.

The performance of the classifiers can be determined using various performance parameters like accuracy, speed and error rate.

4. Image Classification

Assigning images to pre-defined categories by analyzing the contents is defined as ‘Image classification’ or ‘Image categorization’ [4]. Image classification normally involves the processing of two main tasks, namely, feature extraction task (extracts image features and forms a feature vectors) and classification task (uses the extracted features to discriminate the classes). Three

paradigms can be identified during the classification (Figure 3.3).



The binary case classification classifies images into exactly two predefined classes. Here, a sample image belongs exactly to one of the two given classes. The classifier has to determine to which of the two sets the new image goes [25]. In multi-class case, an image belongs exactly to just one class of a set of 'm' classes ([7], [12]). Finally, in the multi-label case, an image may belong to several classes at the same time, that is, classes may overlap [16].

In binary classification a classifier is trained, by means of supervised algorithms, to assign a sample document to one of the two possible sets. These two sets are usually referred to as belonging samples (positive) and not belonging samples (negative) respectively. This method is otherwise termed as the one-against all approach or one-against one approach. Several algorithms exist for this type of classification. They are Naïve Bayes, Linear Regression, Support Vector Machines (SVM) [11] and LVQ [22]. The binary case has been set as a base case from which the other two cases, multi-class and multi-label, are built.

In multi-class and multi-label cases, the traditional approach consists on training a binary classifier for every class and then whenever the binary base case returns a measure of confidence on the classification, assigning either the top ranked one (multi-class assignment) or a given number of the top ranked ones (multi-label assignment). More details about these three paradigms can be found in [1]. The proposed fusion-based image classifier defines binary-class classifiers, which is used to decide whether a given input image has fracture or not.

5. Fusion Classifier

Broad classes of statistical classification algorithms have been developed and applied successfully to a wide range of real-world domains. In general, ensuring that the particular classification algorithm matches the properties of the data is crucial in providing results that meet the needs of the particular application domain. One way in which the impact of this algorithm/application match can be alleviated is by using group of classifiers, where a variety of classifiers (either different types of classifiers or different instantiations of the same classifier) are pooled before a final classification decision is made.

Intuitively, fusion classification allows the different needs of a difficult problem to be handled by classifiers suited to those particular needs. Mathematically, fusion classifier provide an extra degree of freedom in the classical bias/variance tradeoff, allowing solutions that would be difficult (if not impossible) to reach with only a single classifier. Because of these advantages, fusion classification has been applied to many difficult real-world problems.

Recently, many scholars make use of fusion of classifier to enhance the performance of classification. In the past several years, a lot of effort has been devoted to different fusion methods to achieve better performance. In reality, how to select appropriate classification methods towards image classification is an unsolved problem [29]. According to [30] when a perfect set of features that can describe the image data is given, the accuracy of the resultant classification depends on the classifier adopted. Several solutions have been proposed for this purpose. Among which, the usage Neural Network (NN), Support Vector Machines (SVM) and Naïve Bayes based classifiers are more prominent. The reasons behind this popularity are (i) easy of implementation procedures and (ii) accurate classification. As pointed out by [27], the success rate or accuracy of a classification problem can be improved by using multiple classifiers

6. Automatic Abnormality Diagnosis in X-Ray Images.

Research proposals with respect to automatic fracture detection are limited. Relevant work in osteoporosis has been proposed. Most research involving the analysis of orthopaedic X-ray images has been focused on detecting osteoporosis and determining fracture risk, using methods such as texture and fractal analysis. Some authors ([8], [20], [28]) have used first order statistics such as the

standard deviation and mean to measure texture, while others ([24], [37]) computed second order texture statistics like the co-occurrence matrix. Other methods such as surface area measurement [3], semi-variance ([14]) and power spectral analysis to determine the fractal dimension [2] have also been used to detect osteoporosis. Caligiuri *et al.* [3] found that in some cases their method was capable of distinguishing fractured specimens from normal specimens. Fractal analysis was applied to the micro x-ray images of human knees by [21], while a multi-resolution wavelet technique was used by [23] to analyse the micro X-ray CT images of rat lumbar vertebrae. While this work is related, both used micro X-ray images rather than normal diagnostic x-rays.

Other groups have attempted to detect fractures using non-visual techniques. Ryder *et al.* [33] analyzed acoustic pulses as they travelled along a bone to determine if a fracture was present, [13] analyzed mechanical vibrations in a bone using a neural network model, and [34] measured electrical conductivity. Unfortunately none of these techniques are as accurate as x-rays for the diagnosis, localization and classification of long-bone fractures, and as a result they are not used in a clinical setting.

The fracture detection techniques proposed can be loosely categorized into classification-based and transform-based. The first published work on the detection of fractures in x-ray images is that of [36]. The method detected femur fractures by computing the angle between the neck axis and shaft axis. Subsequently, Gabor, MRSAR, and gradient intensity were used for fracture detection, and a simple voting scheme was used to combine the individual classifiers that work on single features ([17], [5]). Since the individual classifiers tend to complement each other, the combined method improves both the accuracy and sensitivity significantly. A similar approach of combining classifiers was also proposed by [18] who combined probabilistic combination methods for segmentation.

Use of fuzzy index and reasoning is another area that is frequently used for defect detection in bones. A fuzzy index method was proposed by [19], while a fuzzy reasoning approach was proposed by [15].

Hough transforms have long been used as computationally efficient methods for detecting particular shapes in images. The use of hough transformation in identifying fractures have also been proved advantages. The Hough transform [6] is a feature extraction technique in image analysis, computer vision, and digital image processing. It is concerned with the identification of straight lines, position of arbitrary shapes, most circles or ellipses. The important case of Hough transform is the linear transform for

detecting straight lines. Compared with other algorithms that detect straight lines, Hough transform can be used to find and link segments in an image. A line in the image space is mapped to a point in the parameter space. Similarly, each pixel of the image space is transformed to a parameterized curve of the parameter space. Each transformed point in the parameter space is considered as a candidate for being a line and accumulated in the corresponding cell of an accumulator. Finally, a cell with a local maximum of scores is selected, and its parameter coordinates are used to represent a line segment in the image space. The main advantage of the Hough transform technique is that it is tolerant of gaps in feature boundary description and is relatively unaffected by image noise. However, using Hough transform introduces computation complexity, which in turn slows the feature extraction and fracture detection process.

Randomized Hough Transform (RHT) [38] is an improvised version of Standard Hough Transform (SHT) for line detection. The basic idea behind the RHT is that, instead of transforming one pixel from image space to parameter space, two or more pixels are randomly selected and mapped to a point in the parameter space. Ji and Xie [10] proposed a method for line detection and circle detection using Randomized Hough Transform based on error propagation which improved detection robustness and accuracy by analytically propagating the errors with image pixels to the estimated curve parameters. Ho and Chen [9] introduced a high speed method for line detection using the geometric property of a pair of parallel lines. Stephen [35] proposed a probabilistic Hough transform based method where it was proved a strong relationship between the Hough Transform and the Maximum Likelihood method. Rodrigo *et al.* [31] considered straight line detection as an energy minimization problem and proposed an energy based line detection.

7. Conclusion

This paper discussed about classification and the various methods that can be used to classify x-ray images. Fracture detection from X-ray images is a complex operation for which a limited algorithms have been proposed. Moreover, although many classification approaches have been developed, which approach is suitable for a given application area is not fully understood. Selection of a suitable classifier requires consideration of many factors, such as classification accuracy, algorithm performance, and computational resources. Multiple classification techniques are more popular with satellite or natural scene classification, where it has proved to be more efficient than

the usage of single classifier. The limited publications mostly use SVM and Bayes classifier. Moreover, the presented works use a set of feature vectors on multiple classifiers to detect fractures. Fusion of feature vectors has not been considered. Thus, in future, the usage of multiple classifiers to detect fractures in X-ray images is to be probed.

References

- [1] Allwein, R.L., Schapire, R.E. and Singer, Y. (2000) Reducing multiclass to binary: A unifying approach for margin classifiers, Proceedings of 17th International Conference on Machine Learning, San Francisco, CA, Pp.9-16.
- [2] Buckland-Wright, J.C., Lynch, J.A., Rymer, J. and Fogelman, I. (1994) Fractal signature analysis of macroradiographs measures trabecular organization in lumbar vertebrae of postmenopausal women, Calcified Tissue International, Vol. 54, No. 2, Pp. 106–112.
- [3] Caligiuri, P., Giger, M.L. and Favus, M. (2004) Multifractal radiographic analysis of osteoporosis, Medical Physics, Vol. 21, No. 4, Pp. 503–508.
- [4] Chang, C.H. (2006) Simulated annealing clustering of Chinese words for contextual text recognition, Pattern Recognition Letters, Vol. 17, No. 1, Pp.57-66
- [5] Chen, Y., Yap, W.H., Leow, W.K., Howe, T.S. and Png, M.A. (2004) Detecting femur fractures by texture analysis of trabeculae, Proc. Int. Conf. on Pattern Recognition.
- [6] Duda, R. and Hart, P. (1972) Use of the hough transformation to detect lines and curves in pictures, Comm. ACM, Vol. 15, Pp. 11–15.
- [7] Foody, G.M. and Mathur, A. (2004) Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification, Remote Sensing of Environment, Vol. 93, Pp. 107–117.
- [8] Geraets, W., Van der Stelt, P. and Elders, P. (1993) The radiographic trabecular bone pattern during menopause, Bone, Vol. 14, No. 6, Pp. 859–864.
- [9] Ho, C.T. and Chen, L.H. (1996) A High Speed Algorithm for Line Detection, Pattern Recognition Letters, vol. 17, pp. 467–473.
- [10] Ji, Q. and Xie, Y. (2003) Randomized Hough transform with error propagation for line and circle detection, Pattern Analysis Applications, Vol. 6, Pp. 55–64.
- [11] Joachims, T. (1998) Text categorization with support vector machines: learning with many relevant features, European Conference on Machine Learning (ECML), Springer, Berlin.
- [12] Joshi, A.J., Porikli, F. and Papanikopoulos, N. (2009) Multi-Class Active Learning for Image Classification, IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, Pp. 2372-2379.
- [13] Kaufman, J., Chiabrera, A., Hatem, M., Hakim, N., Figueiredo, M., Nasser, P., Lattuga, S., Pilla, A. and Siffert, R. (1990) A neural network approach for bone fracture healing assessment, IEEE Engineering in Medicine and Biology, Vol. 9, Pp. 23–30.
- [14] Khosrovi, P., Kahn, A.J., Genant, H.K. and Majumdar, S. (1994) American Society for Bone and Mineral Research, L. Raisz, Ed. Kansas, USA: Mary Ann Liebert, Inc., P. S156.
- [15] Lashkia, V. (2001) Defect detection in X-ray images using fuzzy reasoning, Original Research Article, Image and Vision Computing, Vol. 19, Issue 5, Pp. 261-269
- [16] Li, X., wang, L. and Sung, E. (2004) Multilabel SVM active learning for image classification, International Conference on Image Processing ICIP '04, Vol. 4, Pp.2207-2210.
- [17] Lim, S.E., Xing, Y., Chen, Y., Leow, W.K., Howe, T.S. and Png, M.A. (2004) Detection of femur and radius fractures in x-ray images, Proc. 2nd Int. Conf. on Advances in Medical Signal and Info. Proc..
- [18] Lim, V.L.F., Leow, L.W., Chen, T., Howe, T.S. and Png, M.A. (2005) Combining classifiers for bone fracture detection in X-ray images, IEEE International Conference on Image Processing, ICIP 2005, Vol. 1, Pp.I - 1149-1152
- [19] Linda, C.H. and Jiji, G.W. (2010) Crack detection in X-ray images using fuzzy index measure, Applied Soft Computing, Article in Press.
- [20] Link, T., Majumdar, S., Konermann, W., Meier, N., Lin, J., Newitt, D., Ouyang, X., Peters, P. and Genant, H. (1997) Texture analysis of direct magnification radiographs of vertebral specimens: Correlation with bone mineral density and biomechanical properties, Academic Radiology, Vol. 4, No. 3, Pp. 167–176.
- [21] Lynch, J.A., Hawkes, D.J. and Buckland-Wright, J.C. (1991) Analysis of texture in macroradiographs of osteoarthritic knees using the fractal signature, Physics in Medicine and Biology, Vol. 36, No. 6, Pp. 709–722.
- [22] Martin-Valdivia, M.R., Garcia-Vega, M. and Urena-Lopez, L.A. (2003) LVQ for text categorization using multilingual linguistic resource, Neurocomputing, Vol. 55, Pp. 665-679.
- [23] Matani, A., Okamoto, T. and Chihara, K. (1998) Evaluation of a trabecular structure using a

- [24] multiresolution analysis, Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Pp. 632–633.
- [25] Materka, A., Cichy, P. and Tuliszewicz, J. (2000) Texture analysis of x-ray images for detection of changes in bone mass and structure, *Texture Analysis in Machine Vision*, M. K. Pietikainen, Ed. Singapore: World Scientific.
- [26] Mehta, B., Nangia, S. and Gupta, M. (2008) Detecting image spam using visual features and near duplicate detection, Proceeding of the 17th international conference on World Wide Web (WWW '08) ACM, New York, NY, USA, 497-506.
- [27] Montejo-Raez, A. (2005) Automatic Text Categorization of documents in the High Energy Physics domain, CERN-THESIS-2006-008, ISBN:84-338-3718-4.
- [28] Neeba, N.V and Jawahar, C.V. (2009) Empirical evaluation of character classification schemes, *Seventh International Conference on Advances in Pattern Recognition (ICAPR)*, Pp. 310–313
- [29] Ouyang, X., Majumdar, S., Link, T.M., Lu, P.A.Y., Lin, J.C., Newitt, D.C. and Genant, H.K. (1998) Morphometric texture analysis of spinal trabecular bone structure assessed using orthogonal radiographic projections, *Medical Physics*, Vol. 25, No. 10, Pp. 2037–2045.
- [30] Oza, N.C. and Turner, K. (2008) Classifier ensembles: Select real-world applications, *Journal of Information Fusion*, Vol. 9, Issue 1, Pp. 4-20.
- [31] Park, D.C. (2010) Image Classification Using Partitioned-Feature based Classifier Model, *International Conference on Computer Systems and Applications (AICCSA)*, IEEE/ACS, Pp.1-6.
- [32] Rodrigo, R., Shi, W. and Samarabandu, J. (2006) Energy based line detection, *Canadian Conference on Electrical and Computer Engineering, CCECE '06*, Pp. 2061 - 2064
- [33] Russell, S. and Norvig, P. (2003) *Artificial Intelligence: A Modern Approach*, 2nd edition ed. Prentice-Hall, Englewood Cliffs, NJ.
- [34] Ryder, D., King, S., Olliff, C. and Davies, E. (1993) A possible method of monitoring bone fracture and bone characteristics using a non-invasive acoustic technique, *International Conference on Acoustic Sensing and Imaging*, Pp. 159–163.
- [35] Singh, V. and Chauhan, S. (1998) Early detection of fracture healing of a long bone for better mass health care. Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Pp. 2911–2912.
- [36] Stephen, R.S. (1991) A Probabilistic Approach to the Hough Transform, *Journal Image and Vision*, Vol. 9, Issue 1, Pp. 55-60.
- [37] Tian, T.P., Chen, Y., Leow, W.K., Hsu, W., Howe, T.S. and Png, M.A. (2003) Computing neck-shaft angle of femur for x-ray fracture detection, *Proc. Int. Conf on Computer Analysis of Images and Patterns (LNCS 2756)*, Pp. 82–89.
- [38] Veenland, J., Link, T., Konermann, W., Meier, N., Grashuis, J. and Gelsema, E. (1997) Unraveling the role of structure and density in determining vertebral bone strength, *Calcified Tissue International*, Vol. 61, No. 6, Pp. 474–479.
- [39] Xu, E.O.L. and Kultanen, P. (1990) A new curve detection method -randomized hough transform (rht), *Pattern Recognition Letters*, Vol. 11, Pp. 331–338.

S.K.Mahendran is a doctoral students in the department of Computer Science , Bharathiar University, Coimbatore, Tamil Nadu, India. He is working as HOD in Master of Computer Application Department, Sankara College of Science and Commerce, Saravanampatti, Coimbatore 641035.. His Research interests lies in the area of Medical Image Processing. He has authored three different technical books.

Lt.Dr.S.Santhosh Baboo, has around Nineteen years of postgraduate teaching experience in Computer Science, which includes Six years of administrative experience. He is a member, board of studies, in several autonomous colleges, and designs the curriculum of undergraduate and postgraduate programmes. He is a consultant for starting new courses, setting up computer labs, and recruiting lecturers for many colleges. Equipped with a Masters degree in Computer Science and a Doctorate in Computer Science, he is a visiting faculty to IT companies. It is customary to see him at several national/international conferences and training programmes, both as a participant and as a resource person. He has been keenly involved in organizing training programmes for students and faculty members. His good rapport with the IT companies has been instrumental in on/off campus interviews, and has helped the post graduate students to get real time projects. He has also guided many such live projects. Lt.Dr. Santhosh Baboo has authored a commendable number of research papers in international/national Conference/journals and also guides research scholars in Computer Science. Currently he is Reader in the Postgraduate and Research department of Computer Science at Dwaraka Doss Goverdhan Doss Vaishnav College (accredited at 'A' grade by NAAC), one of the premier institutions in Chennai.

Automatic Histogram Threshold with Fuzzy Measures using C-means

1. K.SRINIVAS

Research Scholor K.L. University
 Vaddeswaram,AP,India.

2. Dr. V. SRIKANTH

Dean-SED & HOD of IST
 K.L. University,Vaddeswaram,AP,India

Abstract:

In this paper, an automatic histogram threshold approach based on a fuzziness measure is presented. This work is an improvement of an existing method. Using fuzzy logic concepts, the problems involved in finding the minimum of a criterion function are avoided. Similarity between gray levels is the key to find an optimal threshold. Two initial regions of gray levels, located at the boundaries of the histogram, are defined. Then, using an index of fuzziness, a similarity process is started to find the threshold point. A significant contrast between objects and background is assumed. Previous histogram equalization is used in small contrast images. Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. This method is frequently used in pattern recognition. It is based on minimization of the objective function ! No prior knowledge of the image is required.

Keywords: fcm,threshold,histogram

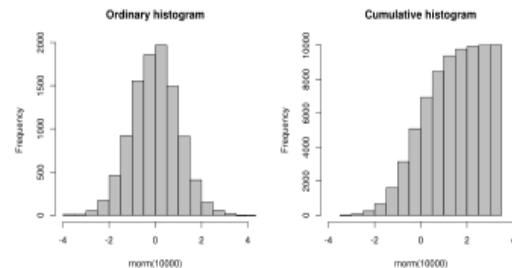
INTRODUCTION

IMAGE segmentation plays an important role in computer vision and image processing applications. Segmentation of nontrivial images is one of the most difficult tasks in image processing. Segmentation accuracy determines the eventual success or failure of computerized analysis procedures. Segmentation of an image entails the division or separation of the image into regions of similar attribute. For a monochrome image, the most basic attribute for segmentation is image luminance amplitude. Segmentation based on gray level histogram thresholding is a method to divide an image containing two regions of interest:object and background. In fact, applying this threshold to the whole image, pixels whose gray level is under this value are assigned to a region and the remainder to the other. Histograms of images with two distinct regions are formed by two peaks separated by a deep vale called bimodal histograms. In such cases, the threshold value must be located on the valley region. When the image histogram does not exhibit a clear separation, ordinary thresholding techniques might perform poorly. Fuzzy set theory provides a new tool to deal with multimodal histograms. It can incorporate human perception and linguistic concepts such as similarity, and has been successfully applied to image thresholding

An **image histogram** is a type of histogram that acts as a graphical representation of the tonal distribution in a digital image. It plots the number of pixels for each tonal value. By looking at the histogram for a specific image a viewer will be able to judge the entire tonal distribution at a glance. Image histograms are present on many modern digital cameras. Photographers can use them as an aid to show the distribution of tones captured, and whether image detail has been lost to blown-out highlights or blacked-out shadows. The horizontal axis of the graph represents

the tonal variations, while the vertical axis represents the number of pixels in that particular tone. The left side of the horizontal axis represents the black and dark areas, the middle represents medium grey and the right hand side represents light and pure white areas. The vertical axis represents the size of the area that is captured in each one of these zones. Thus, the histogram for a very bright image with few dark areas and/or shadows will have most of its data points on the right side and center of the graph. Conversely, the histogram for a very dark image will have the majority of its data points on the left side and center of the graph.

Mathematical definition



An ordinary and a cumulative histogram of the same data. The data shown is a random sample of 10,000 points from a normal distribution with a mean of 0 and a standard deviation of 1. In a more general mathematical sense, a histogram is a function m_i that counts the number of observations that fall into each of the disjoint categories (known as *bins*), whereas the graph of a histogram is merely one way to represent a histogram. Thus, if we let n be the total number of observations and k be the total number of bins, the histogram m_i meets the following conditions:

$$n = \sum_{i=1}^k m_i.$$

Cumulative histogram

A cumulative histogram is a mapping that counts the cumulative number of observations in all of the bins up to the specified bin. That is, the cumulative histogram M_i of a histogram m_j is defined

$$M_i = \sum_{j=1}^i m_j.$$

as:

Some theoreticians have attempted to determine an optimal number of bins, but these methods generally make strong assumptions about the shape of the distribution. Depending on the actual data distribution and the goals of the analysis, different bin widths may be appropriate, so experimentation is usually needed to determine an appropriate width. There are, however, various useful guidelines and rules of thumb. The number of bins k can be assigned directly or can be calculated from a suggested bin width h as:

$$k = \lceil \frac{\max x - \min x}{h} \rceil.$$

Implementation

Consider a discrete grayscale image $\{x\}$ and let n_i be the number of occurrences of gray level i . The probability of an occurrence of a pixel of level i in the image is

$$p_x(i) = p(x = i) = \frac{n_i}{n}, \quad 0 \leq i < L$$

L being the total number of gray levels in the image, n being the total number of pixels in the image, and $p_x(i)$ being in fact the image's histogram for pixel value i , normalized to $[0,1]$.

Let us also define the cumulative distribution function corresponding to p_x as

$$cdf_x(i) = \sum_{j=0}^i p_x(j),$$

which is also the image's accumulated normalized histogram. We would like to create a transformation of the form $y = T(x)$ to produce a new image $\{y\}$, such that its CDF will be linearized across the value range, i.e.

$$cdf_y(i) = iK$$

for some constant K . The properties of the CDF allow us to perform such a transform (see Cumulative distribution function#Inverse it is defined as

$$y = T(x) = cdf_x(x)$$

Notice that the T maps the levels into the range $[0,1]$. In order to map the values back into their original range, the following simple transformation needs to be applied on the result:

$$y' = y \cdot (\max\{x\} - \min\{x\}) + \min\{x\}$$

Histogram equalization of color images : The above describes histogram equalization on a grayscale image. However it can also be used on color images by applying the same method separately to the Red, Green and Blue components of the RGB color values of the image. However, applying the same method on the Red, Green, and Blue components of an RGB image may yield dramatic

changes in the image's color balance since the relative distributions of the color channels change as a result of applying the algorithm. However, if the image is first converted to another color space, Lab color space, or HSL/HSV color space in particular, then the algorithm can be applied to the luminance or value channel without resulting in changes to the hue and saturation of the image. There are several histogram equalization methods in 3D space. Trahanias and Venetsanopoulos applied histogram equalization in 3D color space. However, it results in "whitening" where the probability of bright pixels are higher than that of dark ones. Han et al. proposed to use a cdf defined by the iso-luminance plane, which results in uniform gray distribution

EXISTING METHOD:

This work is an improvement of an existing method based on a fuzziness measure to find the threshold value in a gray image histogram. The method incorporates fuzzy concepts that are more able to deal with object edges and ambiguity and avoids the problems involved in finding the minimum of a function. However, it has some limitations concerning the initialization of the seed subsets. To achieve an automatic process these limitations must be overcome. In order to implement the thresholding algorithm on a basis of the concept of similarity between gray levels, Tobias and Seara made the assumptions that there exists a significant contrast between the objects and background and that the gray level is the universe of discourse, a 1-D set, denoted by . The purpose is to split the image histogram into two crisp subsets, object subset and background subset , using the measure of fuzziness previously defined. The initial fuzzy subsets, denoted by and , are associated with initial histogram intervals located at the beginning and the end regions of the histogram. The gray levels in each of these initial intervals have the intuitive property of belonging with certainty to the final subsets object or background. For dark objects and , for light objects and . These initial fuzzy subsets, and , are modeled by the and membership functions, respectively. The parameters of the and functions are variable to adjust its shape as a function of the set of elements .These subsets are a seed for starting the similarity measure process. A fuzzy region placed between these initial intervals is defined as depicted in Fig. 2. Then, to obtain the segmented version of the gray level image, we have to classify each gray level of the fuzzy region as being object or background. The classification procedure is done by adding to each of the seed subsets a gray level picked from the fuzzy region. Then, by measuring the index of fuzziness of the subsets and , the gray level is assigned to the subset with lower index of fuzziness (maximum similarity). Applying this procedure for all gray levels of the fuzzy region, we can classify them into object or background subsets. Since the method is based on measures of index of fuzziness, these measures need to be normalized by first computing the index of fuzziness of the seed subsets and calculating a normalization factor according to where and are the IF's of the subsets and , respectively. This normalization operation ensures that both initial subsets have identical index of fuzziness at the beginning of the process. It is a necessary condition since the method is based in the calculation of similarity between gray levels. illustrate how the normalization works. For dark objects, the method can be described as follows.

1. Compute the normalization factor .
2. For all gray levels in the fuzzy region compute and .
3. If is lower than , then is included in set , otherwise is included in set .

For light objects the method performs similarly except for the

set inclusion in step 3. In this case, if is lower than , then is included in set , otherwise is included in set .

Threshold: It is the simplest method of image segmentation. From a grayscale image, thresholding can be used to create binary images Method. During the thresholding process, individual pixels in an image are marked as "object" pixels if their value is greater than some threshold value (assuming an object to be brighter than the background) and as "background" pixels otherwise. This convention is known as threshold above. Variants include

ISSN (Online) 1694-0814 which is opposite of threshold above; threshold [www.IJCSI.org](#), where a pixel is labeled "object" if its value is between two thresholds; and threshold outside, which is the opposite of threshold inside .Typically, an object pixel is given a value of "1" while a background pixel is given a value of "0." Finally, a binary image is created by coloring each pixel white or black, depending on a pixel's labels.

Threshold selection: The key parameter in the thresholding process is the choice of the threshold value (or values, as mentioned earlier). Several different methods for choosing a threshold exist; users can manually choose a threshold value, or a thresholding algorithm can compute a value automatically, which is known as automatic thresholding A simple method would be to choose the_mean_or median value, the rationale being that if the object pixels are brighter than the background, they should also be brighter than the average. In a noiseless image with uniform background and object values, the mean or median will work well as the threshold, however, this will generally not be the case. A more sophisticated approach might be to create a histogram of the image pixel intensities and use the valley point as the threshold. The histogram approach assumes that there is some average value for the background and object pixels, but that the actual pixel values have some variation around these average values. However, this may be computationally expensive, and image histograms may not have clearly defined valley points, often making the selection of an accurate threshold difficult. One method that is relatively simple, does not require much specific knowledge of the image, and is robust against image noise, is the following iterative method:

1. An initial threshold (T) is chosen, this can be done randomly or according to any other method desired.
2. The image is segmented into object and background pixels as described above, creating two sets:
 1. $G_1 = \{f(m,n) : f(m,n) > T\}$ (object pixels)
 2. $G_2 = \{f(m,n) : f(m,n) \leq T\}$ (background pixels) (note, $f(m,n)$ is the value of the pixel located in the m^{th} column, n^{th} row)
3. The average of each set is computed.
 1. $m_1 = \text{average value of } G_1$
 2. $m_2 = \text{average value of } G_2$
4. A new threshold is created that is the average of m_1 and m_2
 1. $T' = (m_1 + m_2)/2$
5. Go back to step two, now using the new threshold computed in step four, keep repeating until the new threshold matches the one before it (i.e. until convergence has been reached).

This iterative algorithm is a special one-dimensional case of the k-means clustering algorithm, which has been proven to converge at a local minimum—meaning that a different initial threshold may give a different final result.

Method selects the algorithm to be applied :

The Ignore black and Ignore white options set the image histogram bins for [0] and [255] greylevels to 0 respectively. This may be useful if the digitised image has under- or over-exposed pixels. White object on black background sets to white the pixels with values above the threshold value (otherwise, it sets to white the values less or equal to the threshold). Set Threshold instead of Threshold (single images) sets the thresholding LUT, without changing the pixel data. This works only for single images. If you are processing a stack, two additional options are available: Stack can be used to process all the slices (the threshold of each slice will be computed separately). If this option is left unchecked, only the current slice will be processed. Use stack histogram first computes the histogram of the whole stack, then computes the threshold based on that histogram and finally binarises all the slices with that single value. Selecting this option also selects the Stack option above automatically.

Important notes:

349

1. This plugin is accessed through the Image>Auto Threshold menu entry, however the thresholding methods were also partially implemented in ImageJ's thresholder applet accessible through the Image>Adjust>Threshold... menu entry. While the Auto Threshold plugin can use or ignore the extremes of the image histogram (Ignore black, Ignore white) the applet cannot: the 'default' method ignores the histogram extremes but the others methods do not. This means that applying the two commands to the same image can produce apparently different results. In essence, the Auto Threshold plugin, with the correct settings, can reproduce the results of the applet, but not the way round.

2. From version 1.12 the plugin supports thresholding of 16-bit images. Since the Auto Threshold plugin processes the full greyscale space, it can be slow when dealing with 16-bit images. Note that the ImageJ thresholder applet also processes 16-bit images, but in reality ImageJ first computes a histogram with 256 bins. Therefore, there might be differences in the results obtained on 16-bit images when using the applet and the true 16-bit results obtained with this plugin. Note that for speeding up, the histogram is bracketed to include only the range of bins that contain data (and avoid processing empty histogram bins at both extremes).

3. The result of 16 bit images and stacks (when processing all slices) is an 8 bit container showing the result in white [255] to comply with the concept of "binary image" (i.e. 8 bits with 0 and 255 values). However, for stacks where only 1 slice is thresholded, the result is still a 16 bit container with the thresholded phase shown as white [65535]. This is to keep the data untouched in the remaining slices. The "Try all" option retains the 16 bit format to still show the images with methods that might fail to obtain a threshold. Images and stacks that are impossible to threshold remain unchanged.

4. The same image in 8 and 16 bits (without scaling) returns the same threshold value, however Li's method originally would return different values when the image data was offset (e.g. when adding a fixed value to all pixels). The current implementation avoids this offset-dependent problem.

5. The same image scaled by a fixed value (e.g. when multiplying all pixels by a fixed value) returns a similar threshold result (within 2 greyscale levels of the original unscaled image) for all methods except Huang, Li and Triangle due to the way these algorithms work.

Proposed System: : The concept presented above sounds attractive but has some limitations concerning the initialization of the seed subsets. In these subsets should contain enough information about the regions and its boundaries are defined manually. The proposedmethod in this paper aims to overcome some of the limitations of the existing method. In fact, the initial subsets are defined automatically and they are large enough to accommodate a minimum number of pixels defined at the beginning of the process. This minimum depends on the image histogram shape and it is a function of the number of pixels in the gray level intervals and . It is calculated as follows:

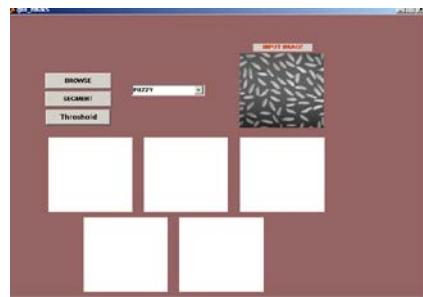
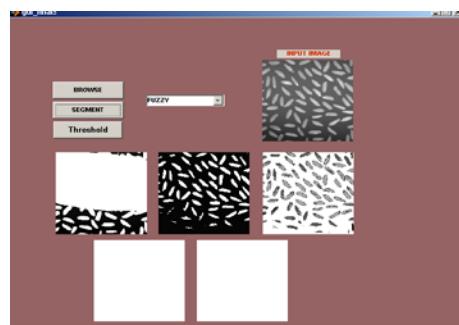
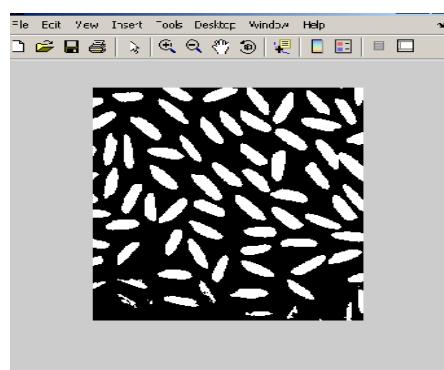
Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. This method is frequently used in pattern recognition. It is based on minimization of the objective function !

1. Initialize $U = [u_{ij}]$ matrix, $U(0)$
2. At k-step: calculate the centers vectors $C(k) = [c_j]$ with $U(k)$
The Algorithm (Contd...)
3. Update $U(k), U(k+1)$

Where,

U = Membership Matrix £ = Termination Criteria

C = Centroids

 X = Pixel Intensity**Results:****Input Image****Segmentation with fuzzy****Thresholding****Segmentation with fuzzy c means****Conclusion & Future work:****In this paper**

In this paper, an automatic histogram threshold approach based on index of fuzziness measure is presented. This work overcome some limitations of an existing method concerning the definition of the initial seed intervals. Method convergence depends on the correct initialization of these initial intervals. After calculating the initial seeds a similarity process is started to find the threshold point. This property of similarity is obtained calculating an index of fuzziness. To measure the performance of the proposed method the misclassification error parameter is calculated. For performance evaluation purposes, results are compared with two well established methods: the Otsu's technique and the Fuzzy C-means clustering algorithm. After results analysis we can conclude that the proposed approach presents a higher performance for a large number of tested images. Standard Fuzzy C-Mean is not suitable for the lip and skin region. The resulting regions are not spatially continuous, due to the fact that only gray level uniformity is checked. FW, In future we will implement the nuero fuzzy based histogram threshold approach. That will be over come the distortion level occur based on fuzzy.

References

- [1] W. K. Pratt, *Digital Image Processing*, third ed. New York: Wiley, 2001.
- [2] A. S. Pednekar and I. A. Kakadiaris, "Image segmentation based on fuzzy connectedness using dynamic weights," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1555–1562, Jun. 2006.
- [3] S. Sahaphong and N. Hiransakolwong, "Unsupervised image segmentation using automated fuzzy c-means," in *Proc. IEEE Int. Conf. Computer and Information Technology*, Oct. 2007, pp. 690–694.
- [4] O. J. Tobias, R. Seara, and F. A. P. Soares, "Automatic image segmentation using fuzzy sets," in *Proc. 38th Midwest Symp. Circuits and Systems*, 1996, vol. 2, pp. 921–924.
- [5] O. J. Tobias and R. Seara, "Image segmentation by histogram thresholding using fuzzy sets," *IEEE Trans. Image Process.*, vol. 11, 2002.
- [6] C. V. Jawahar, P. K. Biswas, and A. K. Ray, "Investigations on fuzzy thresholding based on fuzzy clustering," *Pattern Recognit.*, vol. 30, no. 10, pp. 1605–1613, 1997.
- [7] K. S. Chuang, H. L. Tzeng, S. Chen, J. Wu, and T. J. Chen, "Fuzzy c-means clustering with spatial information for image segmentation," *Comput. Med. Imag. Graph.*, vol. 30, no. 1, pp. 9–15, 2006.
- [8] L. K. Huang and M. J. J. Wang, "Image thresholding by minimizing the measures of fuzziness," *Pattern Recognit.*, vol. 28, no. 1, pp. 41–51, 1995.
- [9] H. R. Tizhoosh, "Image thresholding using type II fuzzy sets," *Pattern*

- [10] A. Rosenfeld and P. de la Torre, "Histogram concavity analysis as an aid in threshold selection," *SMC*, vol. 13, no. 3, pp. 231–235, Mar. 1983.
- [11] J. S. Wezka and A. Rosenfeld, "Histogram modification for threshold selection," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-9, pp. 38–52, 1979.
- [12] T. Ridler and S. Calvard, "Picture thresholding using an iterative selection method," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-8, pp. 630–632, Aug. 1978.
- [13] N. Otsu, "A threshold selection method from gray level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-9, pp. 62–66, 1979.
- [14] J. Kittler and J. Illingworth, "Minimum error thresholding," *Pattern Recognit.*, vol. 19, no. 1, 1986.
- [15] J. N. Kapur, P. K. Sahoo, and A. K. C. Wong, "A new method for graylevel picture thresholding using the entropy of the histogram," *Graph. Models Image Process.*, vol. 29, pp. 273–285, 1985.
- [16] T. Pun, "A new method for gray-level picture thresholding using the entropy of the histogram," *Signal Process.*, vol. 2, no. 3, pp. 223–237, 1980.
- [17] M. Sezgin and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation," *J. Electron. Imag.*, vol. 13, no. 1, pp. 146–165, Jan. 2004.
- [18] C. Murthy and S. Pal, "Fuzzy thresholding: Mathematical framework, bound functions and weighted moving average technique," *Pattern Recognit. Lett.*, vol. 11, pp. 197–206, 1990.
- [19] N. R. Pal and J. C. Bezdek, "Measuring fuzzy uncertainty," *IEEE Trans. Fuzzy Syst.*, vol. 2, 1994.
- [20] A. Kaufmann, *Introduction to the Theory of Fuzzy Subsets*. New York: Academic, 1975, vol. I.
- [21] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Reading, MA: Addison-Wesley, 1993.

Bibliography:



Author 1: K. SRINIVAS obtained his M.Tech. from Nagarjuna University and pursuing his PhD from KL University under the guidance of Dr. V. Srikanth. His interested research area are image processing.



Author 2: Dr. V.Srikanth obtained his B.Tech and M.E degrees from Madras University and PhD from Nagarjuna university. Presently working as Dean and HOD of IST, K.L. University,Vaddeswaram.AP.

Modified LRU Algorithm To Implement Proxy Server With Caching Policies

Jitendra Singh Kushwah¹

Jitendra Kumar Gupta²

Mahendra S. Yadav³

Hitesh Madan⁴

¹ Asst. Prof., School of Computer Application, ITM University
Gwalior (M.P), India

² Asst.Prof., Department of Computer Science & Engineering, SRGI,
CSE Campus, Jhansi (U.P), India

³ Asst. Prof., School of Computer Application, ITM University
Gwalior (M.P), India

⁴ Asst.Prof., Department of Computer Science & Engineering, SRGI,
CSE Campus, Jhansi (U.P), India

Abstract

In order to produce and develop a software system, it is necessary to have a method of choosing a suitable algorithm which satisfies the required quality attributes and maintains a trade-off between sometimes conflicting ones. Proxy server is placed between the real server and clients. Proxy server uses caching policies to store web documents using algorithms. For this, different algorithms are used but drawbacks of these algorithms are that it is applicable only for the video files not for other resource types. Second drawback is that it does not tell any thing about organizing the data on the disk storage of the proxy server. Third drawback is that it is difficult to implement. Fourth drawback is that they require the knowledge about the workloads on the proxy server. Major problems in previous described algorithms is that "Cold Cache Pollution".

As described in the previous description all the existing algorithms used for caching suffers from various disadvantages. This paper is proposing a technique to remove the problem of cold cache pollution which is proved mathematically that it is better than the existing LRU-Distance algorithm.

Keywords: LRU, Proxy Server, Cold Cache Pollution, Web Documents,

1. Introduction

Proxy server is placed between a client application such as a web browser, and a real server the request to the real server. Different key features are improving performance of proxy server like caching the documents and thread polling etc. It provides security like firewall and protection of local Area Network (LAN) from accessing of unauthorized users. Most important feature is caching the web documents. Caching refers to store copies of the popular documents in proxy memory and thus used it for

future references and reduces the bandwidth requirement. There are different available techniques but this paper is proposing the new and better technique. It intercepts all requests to the real server to see if it can fulfill the requests itself. If not, it forwards unique to improve the performance of Least Recently Used- Distance (LRU-D) of caching the web document in the proxy server. proxy server is a server (a computer system or an application program) that acts as an intermediary for requests from clients seeking resources from other servers. A client connects to the proxy server, requesting some service, such as a file, connection, web page, or other resource, available from a different server. The proxy server evaluates the request according to its filtering rules. For example, it may filter traffic by IP address or protocol If the request is validated by the filter, the proxy provides the resource by connecting to the relevant server and requesting the service on behalf of the client. A proxy server may optionally alter the client's request or the server's response, and sometimes it may serve the request without contacting the specified server. In this case, it caches responses from the remote server, and returns subsequent requests for the same content directly.

A proxy server has many potential purposes, including:

- To keep machines behind it anonymous (mainly for security)
- To speed up access to resources (using caching). Web proxies are commonly used to cache web pages from a web server
- To apply access policy to network services or content, e.g. to block undesired sites.
- To log / audit usage, i.e. to provide company employee Internet usage reporting.
- To bypass security/ parental controls.
- To scan transmitted content before delivery.
- To scan outbound content, e.g., for data leak protection.
- To circumvent regional restrictions.

A proxy server that passes requests and replies unmodified is usually called gateway sometimes tunneling proxy.

A proxy server can be placed in the user's local computer or at various points between the user and the destination servers on the Internet.

A revisers proxy is (usually) an Internet-facing proxy used as a front-end to control and protect access to a server on a private network, commonly also performing tasks such as load-balancing, authentication, decryption or caching.

2. Types and functions

Proxy servers implement one or more of the following functions:

2.1 Caching proxy server

A caching proxy server accelerates service requests by retrieving content saved from a previous request made by the same client or even other clients. Caching proxies keep local copies of frequently requested resources, allowing large organizations to significantly reduce their upstream bandwidth usage and cost, while significantly increasing performance. Most ISPs and large businesses have a caching proxy. These machines are built to deliver superb file system performance (often with RAID and journaling) and also contain hot-rod versions of TCP. Caching proxies were the first kind of proxy server.

2.2 Web proxy

A proxy that focuses on world wide wave traffic is called a "web proxy". The most common use of a web proxy is to serve as a web cache . Most proxy programs provide a means to deny access to URLs specified in a blacklist, thus providing content filtering. This is often used in a corporate, educational or library environment, and anywhere else where content filtering is desired. Some web proxies reformat web pages for a specific purpose or audience, such as for cell phones and PDAs.

2.3 Content-filtering web proxy

A content-filtering web proxy server provides administrative control over the content that may be relayed through the proxy. It is commonly used in both commercial and non-commercial organizations (especially schools) to ensure that Internet usage conforms to acceptable use policy. In some cases users can circumvent the proxy, since there are services designed to proxy information from a filtered website through a non filtered site to allow it through the user's proxy.

2.4 Anonymous proxy server

An anonymous proxy server (sometimes called a web proxy) generally attempts to anonymize web surfing. There are different varieties of anonymizers. One of the more common variations is the open proxy. Because they are typically difficult to track, open proxies are especially useful to those seeking online anonymity, from political dissidents to computer criminals.

2.5 Hostile proxy

Proxies can also be installed in order to eavesdrop upon the dataflow between client machines and the web. All accessed pages, as well as all forms submitted, can be captured and analyzed by the proxy operator.

2.6 Intercepting proxy server

An intercepting proxy combines a proxy server with a gateway or router (commonly with NAT capabilities). Connections made by client browsers through the gateway are diverted to the proxy without client-side configuration (or often knowledge).

2.7 Transparent and non-transparent proxy server

The term "transparent proxy" is most often used incorrectly to mean "intercepting proxy" (because the client does not need to configure a proxy and cannot directly detect that its requests are being proxied). Transparent proxies can be implemented using Cisco's WCCP (Web Cache Control Protocol). A 'non-transparent proxy' is a proxy that modifies the request or response in order to provide some added service to the user agent, such as group annotation services, media type transformation, protocol reduction, or anonymity filtering".

2.8 Forced proxy

The term "forced proxy" is ambiguous. It means both "intercepting proxy" (because it filters all traffic on the only available gateway to the Internet) and its exact opposite, "non-intercepting proxy" (because the user is forced to configure a proxy in order to access the Internet).

2.9. Open proxy server

Because proxies might be used to abuse, system administrators have developed a number of ways to refuse service to open proxies. Many IRC networks automatically test client systems for known types of open proxy. Likewise, an email server may be configured to automatically test e-mail senders for open proxies.

2.10 Reverse proxy server

A reverse proxy is a proxy server that is installed in the neighborhood of one or more web servers. All traffic coming from the Internet and with a destination of one of the web servers goes through the proxy server.

3. Working of Proxy server

Proxy server is placed between the real server and clients. Clients sends the request to the proxy server, if proxy server is able to fulfill the request of the clients by its stored web pages then it forwards the data from its cache to the clients. If proxy server is not able to fulfill the request of the clients, then it forwards all such request to the real server in World Wide Web take the response of such request and sends the request to the clients. It also protects the internal networks from accessing of unauthorized users. Caching the web document is its main works. Caching refers to store copies of the popular documents in proxy memory and thus used it for future references and reduces the bandwidth requirement.

4. Characteristics of Proxy Server

As such there is no RFC that describes the characteristics of the proxy server. It is up to the designer that how much facility designer is going to provide for the users. Some of most important features are as follows:

- Caching
- Security
- Content filtering
- Does port redirection or forwarding

5. Caching Policies

There are different types of caching policies [7] used today's in proxy server for caching the web documents. Each caching algorithm has its own advantage and disadvantage. Some caching algorithms are based on the some value of the web documents. Some algorithms are used for caching the audio and large video files. Some algorithms are used for caching the partial video files to efficiently utilize the disk bandwidth, disk space of proxy server and internet bandwidth provided to the network. But still these caching policies are not perfect for caching the all types of the web document. Some algorithm suits well for some types of web documents and some suited well for some other type of web documents. A cache (pronounced) is a component that improves performance by transparently storing data such that future requests for that data can be served faster. The data that is stored within a cache might be values that have been computed earlier or duplicates of original values that are stored elsewhere. If requested data is contained in the cache (cache hit), this request can be served by simply reading the cache, which is comparably faster. Otherwise (cache miss), the data has to be recomputed or fetched from its original storage location, which is comparably slower.

Hence, the more requests can be served from the cache the better the overall system performance is.

6. Literature Survey

Related work in the area of caching of proxy server is described in this paper. There are many techniques designed for caching the web objects. All the caching policies are categorized depending on their features. Different categories are described in following section.

6.1 Value-Based Caching

These policies are based on some value that these algorithms decide for the selected objects to be cached. These values are like frequency, recency, expires date etc. Some algorithms combined two or more approaches and make new one. These policies are least recently used (LRU), least frequently used (LFU), LRU-threshold, and LRU-min, Day and first in first out (FIFO).

Drawbacks of these algorithms are, they never does the caching on resource basis means They always considered all objects same. For example it does same things for image, text, html, jpg and video or audio files. Second drawback is that these algorithms never consider the I/O bandwidth and network bandwidth into consideration, it considers only disk space. Third drawback is that it does not tell any thing about organizing the data on the disk storage of the proxy server.

6.2 On disk Caching

These caching algorithms are used for caching the document on disk storage of the proxy Server. On disk caching of web objects give the techniques for caching web objects on secondary storage. There are several difficulties created by the non-uniformity of document sizes when placing them on disk.

Drawbacks of these algorithms are, they never does the caching on resource basis means they always considered all objects same. For example it does same things for image, text, html, jpg and video or audio files. Second drawback is that these algorithms never consider the I/O bandwidth and network bandwidth into consideration, it considers only disk space.

6.3 LRU-Distance and LRU-Segmentation Caching

These are the modified form of LRU that describes a problem Cache Pollution [4] means that a cache contains objects that are not frequently used in the near future. One of the main weaknesses of LRU in WWW applications is that it suffers from "cold cache pollution". The term "cold cache pollution" refers to the unpopular objects that remain in the cache.

The LRU-Distance algorithm places a new object at a distance D from the bottom of the cache stack (as opposed to the traditional LRU that places the new objects on top

of the stack). The LRU-Distance algorithm reduces the cold cache pollution since it takes a shorter time to drop out a new object if it is not accessed again. In the SLRU algorithm, the cache is partitioned into a lower and an upper partition. An object is first placed in the Lower partition, but it is promoted to the upper partition when it is qualified (e.g., when it is accessed k times). We have used trace-driven simulation to evaluate the effectiveness of these two algorithms in relieving cold cache pollution.

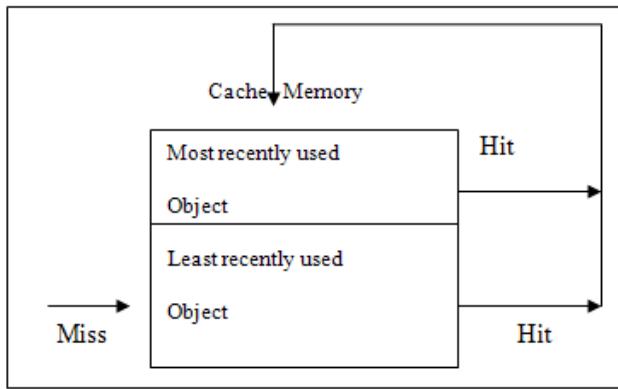


Fig 1: Logical Flow of LRU Distance

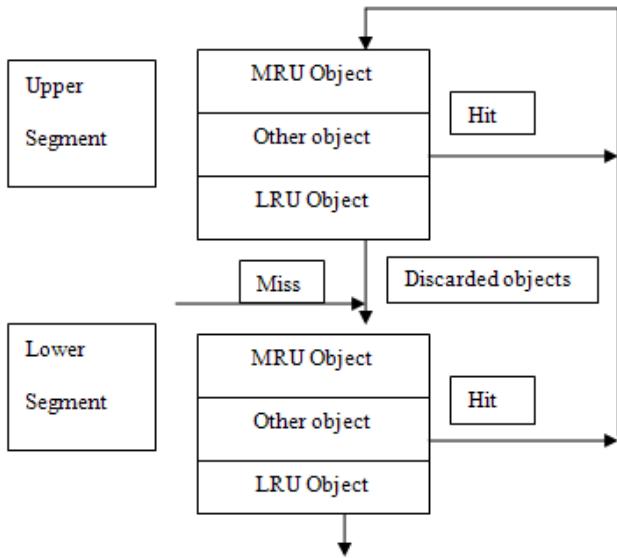


Fig 2: Logical flow of Segmented- LRU

Drawbacks of these algorithms are, they never does the caching on resource basis means they always considered all objects same. For example it does same things for image, text, html, jpg and video or audio files. Second drawback is that these algorithms never consider the I/O bandwidth and network bandwidth into consideration, it considers only disk space. Third drawback is that it does not tell any thing about organizing the data on the disk storage of the proxy server. Fourth drawback is that cold

pollution is not completely removed although it is used for removing the cold cache pollution.

6.4 Cooperative Caching

A group of proxy server is used for cooperative caching in which mesh structure of proxy server is used where each server is cooperating with other server just like the distributed caching.

Drawbacks of these algorithms are, they never does the caching on resource basis means they always considered all objects same. For example it does same things for image, text, html, jpg and video or audio files. Second drawback is that these algorithms never consider the I/O bandwidth and network bandwidth into consideration, it considers only disk space. Third drawback is that it does not tell any thing about organizing the data on the disk storage of the proxy server. . Fourth drawback is that they require the knowledge about the workloads on the proxy server.

6.5 Progressive Caching

Progressive Caching means caching the more part of the document as requirement increases. There are good reasons to allow the proxy server caching partial video instead of the entire video.

Drawbacks of these algorithms are that it is applicable only for the video files not for other resource types. Second drawback is that it does not tell any thing about organizing the data on the disk storage of the proxy server. Third drawback is that it is difficult to implement. Fourth drawback is that they require the knowledge about the workloads on the proxy server.

7. Problem Identification & Methodology

Efficient utilization of proxy resources such as I/O bandwidth and disk space as well as saving the network bandwidth are very important issues to be considered so that proxy server can provides the best facility to the local area network users.

7.1 Drawbacks of Value-Based Caching

These policies are based on some value that these algorithms decide for the selected objects to be cached. These values are like frequency, recency, expires date etc. Some Algorithms combined two or more approaches and make new one.

Drawbacks of these algorithms are, they never does the caching on resource basis means They always considered all objects same.

7.2 Drawbacks of On disk Caching

These caching algorithms are used for caching the document on disk storage of the proxy server. On disk caching of web objects give the techniques for caching

web objects on secondary storage. These algorithms tell about how to organize the data on the disk storage of the proxy server.

Drawbacks of these algorithms are, they never does the caching on resource basis means they always considered all objects same.

7.3 Drawbacks of Resource-Based Caching

The resource based caching algorithm characterizes each object by the bandwidth and space requirement pair and a caching gain. The cache is modeled as a two constraint Knapsack. The algorithm describes heuristics to convert the two constraints knapsack model to multiple single constraint models.

Drawback is that it does not tell any thing about organizing the data on the disk storage of the proxy server.

7.4 Drawbacks of Cooperative Caching

A group of proxy server is used for cooperative caching in which mesh structure of proxy server is used where each server is cooperating with other server just like the distributed caching.

Drawbacks of these algorithms are, they never does the caching on resource basis means they always considered all objects same.

7.5 Drawbacks of LRU-Distance and LRU-Segmentation Caching

The LRU-Distance algorithm places a new object at a distance D from the bottom of the cache stack (as opposed to the traditional LRU that places the new objects on top of the stack). The LRU-Distance algorithm reduces the cold cache pollution since it takes a shorter time to drop out a new object if it is not accessed again. In the SLRU algorithm, the cache is partitioned into a lower and an upper partition. An object is first placed in the Lower partition, but it is promoted to the upper partition when it is qualified (e.g., when it is accessed k times). We have used trace-driven simulation to evaluate the effectiveness of these two algorithms in relieving cold cache pollution. Drawback of these algorithms is that cold pollution is not completely removed although it is used for removing the cold cache pollution.

Problem Definition: To remove the cold cache pollution LRU-Distance places the recently accessed web pages at a distance d from the bottom of the stack. If this web page doesn't access again by any client then it takes D - time to remove the cache. But if it accessed again then it is placed on the top of the stack. If it is not accessed again then it takes time equal to the length of the cache, so again causing cold cache pollution. So this existing mechanism doesn't remove cold cache pollution fully. This paper is proposes a new improved technique to deal with this cold cache pollution.

8. Proposed Solution

As described in the previous description all the existing algorithms used for caching suffers from various disadvantages. This paper is proposing a technique to remove the problem of cold cache pollution which is proved mathematically that it is better than the existing LRU-Distance algorithm. The proposed technique along with its associated advantages and disadvantages is described here.

Main cause of cold cache pollution is due the objects present in the cache that are once accessed and not accessed again. These objects take time equal to the length of the stack. So, when object is first accessed it is placed at a distance D from the bottom of the stack. If it is not accessed again then it takes D time to remove from the cache. If it is accessed again then in spite putting the object on the top of the stack as in case of LRU-Distance algorithm, it is placed in mid of top of the stack and distance D. If it is accessed third time then it is placed in mid of top and previous position the stack. If it accessed again then same process goes on until object reached at the top position. If objects are not popular at any stage then it takes less time to drop from the cache.

An important issue in LRU-Distance is how to determine the value of D. Initially its value is taken as total no of objects divided by two, means middle of the stack. It can be changed time to time if performance is decreased. Generally its value is increased because as value is increased, it is proceeding towards traditional LRU. If D is as big as Cache size then it is equivalent to the traditional LRU.

Suppose total 20 pages can be inserting in cache stack, then size of D will be 10 because

$$D = \text{Total No Of Object Present In Cache Stack} / 2$$

When first object will be inserting in cache stack then it will go to the D distance from the bottom of the cache stack. When Second object will be inserting in cache stack then it will go to the Mid. When third Object comes then it will insert between previous object and top of the cache stack.

Mid = ⌈ Top Of Stack(Index No.) + Distance D from Bottom of the stack (index of where previous object placed) / 2 ⌉

This process will be continuing such type.

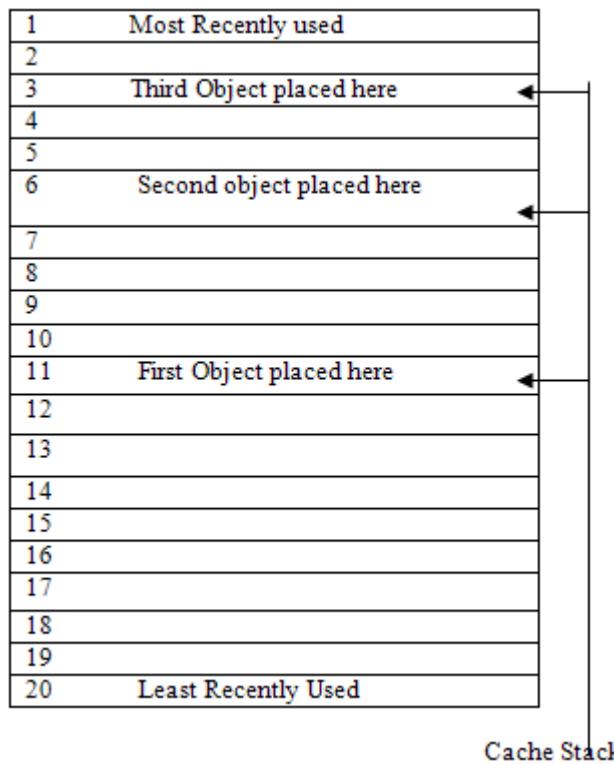


Fig 3: Page Insertion by Modified LRU Distance Algorithm.

9. Proposed Algorithm

Proposed algorithms with the assumptions are described here:

Total size allocated for caching of document is CS.

Size of total information stored at any time is IS.

Total number of web objects stored at any time is TNO.

Size of object to be inserted is OS.

Top of stack is represented by TOS.

Jumping of web document is represented by JB.

Position of web document returned by linear search function is Index.

Significance of D is described in section 4.

Algorithm for storing the new web document:

Step 1. If (OS>(CS-IS)) then do the following

(a). “Delete one or more web documents from bottom of stack, sum of whose size is just greater than or equal to object to be inserted.

(b). Goto step 2.

Step 2. Do the following

(a). “Insert the document at the distance D from bottom of stack”.

(b). IS = IS + OS.

Step 3. If (D != TOS)

JB=1;

Step 4. Exit.

Modified LRU-Distance Algorithm:

Step 1. If (JB == 0) then do the following

“Put the accessed object at the top of Stack”, and go to step 3.

else

pos = Index / 2.

“Put the object at the position pos in the stack”.

Step 2. If (pos == TOS) then do

JB = 0.

Step 3. Exit.

Advantages of Proposed Algorithm:

1. Better than the LRU-Distance algorithm for removing cold cache pollution.
2. Doesn't require previous knowledge about workloads.
3. Easier to implement than other advanced algorithms.
4. Can handle all types of data type.

Disadvantages of Proposed Algorithm:

1. Not efficiently works for Video data types.
2. Doesn't consider Network bandwidth and disk I/O bandwidth of proxy server.
3. It considers only disk space of the proxy server.
4. Doesn't tell any thing about the organization of data on the disk space.
5. It considers all the data of same types.

10. Conclusions and Future Work

10.1 Conclusions

LRU is a very simple and widely used cache replacement algorithm that suffers from cache pollution in proxy caching. In the recent years, several more efficient replacement algorithms have been suggested. But, these advanced algorithms require more knowledge about the workloads and are generally more difficult to implement. The main attraction of LRU is its simplicity. Then two modified LRU algorithms, LRU-Distance and SLRU proposed, to reduce cold cache pollution. These two algorithms are also simple to implement. But cold cache pollution is partially removed by these two modified form of the LRU. This paper is proposing a new and better technique by modifying the LRU-Distance algorithm, to reduce the cache pollution further.

10.2 Future Work

No strategy can be hundred percent perfect and this proposed algorithm also suffers from some inherent disadvantages. The proposed technique doesn't consider I/O bandwidth and network bandwidth in storing or replacing the web pages. This technique doesn't tell any thing that how to organize the data on the secondary storage, what should be the file structure. It does consider all the resources of same types.

Acknowledgement

This is not possible for a person to write a paper in a day and night program, this really is a long journey. In this long journey we meet many renewed personality of Computer Science field. This is possible only due to their blessing and love to complete a paper.

First and foremost We would like to thank Mr. Daulat singh chouhan, Managing director, ITM University, Gwalior (M.P) and Ms. Pooja Rai, Chairperson, SR group of Institutions, Jhansi (U.P) in developing the idea for this paper, for their patient reading, many valuable suggestions and criticism in producing this paper in a good shape.

We would like to thanks Prof. (Dr.) Anand Tripathi, Director, Iswarchand Vidya Sagar Institute of Technology (IVSIT) Mathura (U.P) for the inspirations in the development of this material. We also wish to show my gratitude to Prof. Sanjay Gupta, Head, School of computer application, Jiwaji University, Gwalior (M.P) for his constant encouragements in shaping this paper.

We wish to express my sincere gratitude to Dr. P.K. Sharma, Principal, IASCA, ITM University Gwalior(M.P) and Dr. B.M. Singhal, Director, School of computer application, ITM University Gwalior(M.P).

At last We would like from the bottom of our heart to thanks to our parents who believe in our to complete this paper is the power of success.

References

- [1] W. Ma and D. H.C. Du, "Reducing bandwidth requirement for delivering video over wide area networks with proxy server," *IEEE Trans. Multimedia*, vol. 4, pp. 539–550, Dec. 2002.
- [2] S. Sen, J. Rexford, and D. Towsley, "Proxy prefix caching for multimedia streams," presented at the *INFOCOM '99*, Mar. 1999.
- [3] R. Tewari, H. Vin, A. Dan, and D. Sitaram, "Resource-based caching for web servers," presented at the *Proc. of SPIE/ACM Conference on Multimedia Computing and Networking*, San Jose, Jan. 1998.
- [4] R. Ayani, Yong Meng, "Cache Pollution in Web Proxy Servers," *IEEE proceedings of the IPDPS*, 2003.
- [5] Yong Woon Park and Ki Dong Chung , "A Caching Policy for Continuous Media Objects Based on Logical Caches and Object partitioning " *IEEE Trans. On Multimedia*, 2001.
- [6] MAJ Richard Howard and MAJ Bernard J . Jansen "A Proxy Server Experiment: an Indication of the Changing Nature of the Web" *IEEE Trans.* , 1998.
- [7] Wei-hsiu Ma and David H. C. Du, " Design a Progressive Video Caching Policy for Video Proxy Servers" *IEEE Trans. on Multimedia*, VOL. 6, NO. 4, August 2004.

Jitendra Singh Kushwah He did M.Tech. from RGPV Bhopal (M.P and)MCA from Jiwaji University, Gwalior (M.P) after B.Sc. Mathematics from Govt. P.G.college Morena (M.P). He has seven years teaching experience

from academic and two years industry experience. He published a book entitled "Digital Principles" in Shree sai Publications, Meerut (U.P) and also published many research papers in international, national journals and conferences. He organized many national level workshops and conferences and also presented papers in conferences. He is also life member of Computer Society of India (CSI).

Jitendra Kumar Gupta He did M.Tech from MANIT Bhopal (M.P) after B.E. from IET Lucknow (U.P) and also did diploma in polytechnic engineering from Jhansi (U.P). He has three years teaching experience from academic. He organized many national level workshops and conferences and also presented papers in conferences. Currently he is coordinator of M.Tech. and different events. He is also member of Computer Society of India (CSI).

Mahendra singh Yadav He did MCA from RGPV, Bhopal (M.P) after B.Sc. Mathematics from Govt. Science college Gwalior (M.P). He has four years teaching experience from academic. He published many research papers in international, national journals and conferences. He organized many national level workshops and conferences and also presented papers in conferences.

Hitesh Madan He did MCA from RGPV, Bhopal (M.P) after B.Sc. (Electronics) from Jiwaji University Gwalior(M.P). He has two years teaching experience from academic. He published many research papers in international, national journals and conferences. He organized many national level workshops and conferences and also presented papers in conferences.

Classification of speech for Clinical Data using Artificial Neural Network

C.R.Bharathi¹,Dr.V. Shanthi²

¹. Research Scholar, Department of ECE, Sathyabama University,
AP, Vel Tech University, Chennai, India.

² Professor, Department of MCA, St. Joseph's College of Engineering,
Chennai, India.

Abstract

A wide range of researches are carried out in speech signal processing for denoising, enhancement and more. Besides the other, stress management is important to improve disabled children speech. In order to provide proper speech practice for the disabled children, their speech is analyzed. Initially, the normal and pathological subjects speech are obtained with the same set of words. In this paper, classification of normal and pathological subjects speech is discussed. Initially Feature Extraction is implemented using well known Mel Frequency Cepstrum Coefficients (MFCC) for both words of normal and pathological subjects' speech. Dimensionality reduction of features extracted is implemented using Principal Component Analysis (PCA). Finally the features are trained using Artificial Neural Network (ANN) for classification.

Keywords: speech signal, stress management, Mel Frequency Cepstrum Coefficients (MFCC), Principal Component Analysis (PCA), trained

1. Introduction

Terminology for Mental Retardation/Intellectual Disability (MR/ID) has been particularly challenging as the term *mentally retarded* carries significant social and emotional stigma. Developmental delay is often used inappropriately as synonymous with MR/ID. Developmental delay is an overlay inclusive term and should generally be used for infants and young children in which the diagnosis is difficult, such as those too young for formal testing.

Speech is the most basic means of communication between humans. Effective determination of quantitative speech disability remains one of the challenges in medical

profession. However millions of people are suffering from disabilities associated with speech. Speech synthesis among the disabled children is the first step for making speech corrective tool kit.

Approximately 10% of children have some learning impairment, while as many as 3% manifest some degree of MR/ID. In every verbal communication, the quality and precision of speech are given greater importance [2]. Due to the presence of deleterious properties in the acoustic environment such as multipath distortion (reverberation) and ambient noise, the performance of speech and speaker recognizers are often degraded [3]. Speech communication applications such as voice-controlled devices, hearing aids, and hands-free telephones mostly suffer from poor speech quality because of background noise and room echo [4]. Most of the time, particularly during travel, we meet noisy environment. Signal processing techniques removes the noise from the signal-noise mixture and provides an almost noise-free sound for enhanced communication [2]. Signal processing techniques are exploited in several applications such as speech acquisition, acoustic imaging and communications purposes [5].

In the modern period, there is a great interest for developing techniques for both speech (and character/word sequences) recognition and synthesis [6]. Generally, the speech recognition has two stages: feature extraction and classification [8].

In this paper, we concentrate on the pathological subject's speech and to analyse the speech of them by comparing it with the normal children. This work is very useful for the speech practitioners in the way in which position they have to

improve the speech of the abnormal person. Initially, the samples of the normal as well as the pathological subject's speech have been obtained and with the aid of these samples, the further process has to be carried out. Initially, the MFCC of both the speeches are extracted and the PCA is applied to the MFCC to reduce the dimensionality of the speeches. After that the parameters that are extracted from this MFCC of both speeches are then sent as an input to generate the ANN. The abnormal and normal features are used to train the network for classification.

Feature extraction is a key issue for efficient speaker recognition. Additionally, a reduced feature set would allow more robust estimates of the model parameters, and less computational resources would be required. Best features are those that help to discriminate among speakers. A small amount of data is enough to estimate good models.

State-of-the-art systems use the same short-term spectrum features (Mel-Frequency Cepstral Coefficients, MFCC) for speech and speaker recognition, because MFCC convey not only the frequency distribution identifying sounds, but also the glottal source and the vocal tract shape and length, which are speaker specific features. Additionally, it has been shown that dynamic information improves significantly the performance of recognizers, so MFCC are commonly used as features.

Principal Component Analysis (PCA), an technique of multivariate statistical analysis [1], consists of computing the eigenvectors of the $D \times D$ covariance matrix X , then sorting them according to the corresponding eigenvalues, in descending order, and finally building the projection matrix A (called Karhunen-Loeve Transform, KLT) with the largest K eigenvectors (i.e. the K directions of greatest variance). Each feature vector X is then pre-processed according to the expression $Y = A(X - \mu)$, where μ represents the mean feature vector. KLT decorrelates the features and provides the smallest possible reconstruction error among all linear transforms, i.e. the smallest possible mean-square error between the data vectors in the original D -feature space and the data vectors in the projection K -feature space.

The study of Neural Networks, was initially inspired by neurobiology, but it has since become a very interdisciplinary field, spanning computer science, electrical engineering, mathematics, physics, psychology, and linguistics as well. Some researchers are still studying the neurophysiology of the human brain, but much attention is now being focused on the general properties of neural computation, using simplified neural models. These properties include:

- **Trainability:** Networks can be taught to form associations between any input and output patterns. This can be used, for example, to teach the network to classify speech patterns into phoneme categories.

- **Generalization:** Networks don't just memorize the training data; rather, they learn the underlying patterns, so they can generalize from the training data to new examples. This is essential in speech recognition, because acoustical patterns are never exactly the same.

- **Nonlinearity:** Networks can compute nonlinear, nonparametric functions of their input, enabling them to perform arbitrarily complex transformations of data. This is useful since speech is a highly nonlinear process.

- **Robustness:** Networks are tolerant of both physical damage and noisy data; in fact noisy data can help the networks to form better generalizations. This is a valuable feature, because speech patterns are notoriously noisy.

- **Uniformity:** Networks offer a uniform computational paradigm which can easily integrate constraints from different types of inputs. This makes it easy to use both basic and differential speech inputs, for example, or to combine acoustic and visual cues in a multimodal system.

- **Parallelism:** Networks are highly parallel in nature, so they are well-suited to implementations on massively parallel computers. This will ultimately permit very fast processing of speech or other data.

An neural network consists of a potentially large number of simple processing elements (called *units*, *nodes*, or *neurons*), which influence each other's behavior via a network of excitatory or inhibitory weights. Each unit simply computes a nonlinear weighted sum of its inputs, and broadcasts the result over its outgoing connections to other units. A training set consists of patterns of values that are assigned to designated input and/or output units. As patterns are presented from the training set, a learning rule modifies the strengths of the weights so that the network gradually learns the training set. This basic paradigm can be fleshed out in any different ways, so that different types of networks can learn to compute implicit functions from input to output vectors, or automatically cluster input data, or generate compact representations of data, or provide content-addressable memory and perform pattern completion.

Neural networks are usually used to perform static pattern recognition, that is, to statically map complex inputs to simple outputs, such as an N-ary classification of the input patterns.

Moreover, the most common way to train a neural network for this task is via a procedure called *backpropagation*, whereby the network's weights are modified in proportion to their contribution to the observed error in the output unit activations (relative to desired outputs). To date, there have been many successful applications of neural networks trained by backpropagation.

The rest of the paper is organized as follows. The Feature Extraction and Classification using ANN are described in Section 2. The experimental setup is described in Section 3, including the speech database used to train. Experiments Results are presented and discussed in Section 4: (1) Feature Extraction using MFCC and PCA and (2) Training and Classification using ANN. Finally, Section 5 summarizes our approach.

2. Methodology

2.1 Feature Extraction

In this work, MR children speech of mild degree is taken as Pathological children speech dataset. For improving the abnormal speech initially Feature extraction, dimensionality reduction and Classification is done. Normal and pathological subject's speech dataset is obtained and then the MFCC is extracted from both the speeches. This speech dataset is developed through in which both the databases are same set of scripts. Subsequently, with the aid of the PCA the dimensionality is reduced for MFCC features extracted. The step by step processes are explained in detail in the following subsections.

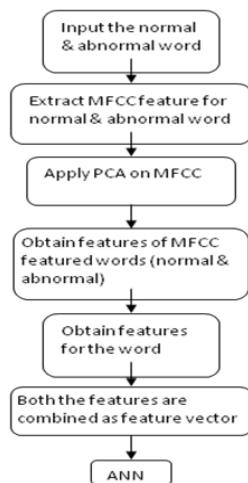


Fig.1 Data Flow for Feature Extraction

2.1.1 MFCC

In this segment, the speech samples are extracted from the normal and pathological subjects with the aid of the audio synthesizer. Let D_a and D_b are the abnormal children, normal children speech datasets respectively and from these datasets the MFCC feature is extracted.

$$D_a = \{w_1, w_2, w_3 \dots w_{N_w-1}\} \quad (1)$$

$$D_b = \{\omega_1, \omega_2, \omega_3 \dots \omega_{N_w-1}\} \quad (2)$$

MFCC's are based on the known variation of the human ear's critical bandwidths with frequency. The MFCC technique makes use of two types of filter, namely, linearly spaced filters and logarithmically spaced filters. To capture the phonetically important characteristics of speech, signal is expressed in the Mel frequency scale. This scale has a linear frequency spacing below 1000 Hz and a logarithmic spacing above 1000 Hz. Normal speech waveform may vary from time to time depending on the physical condition of speakers' vocal cord. Rather than the speech waveforms themselves, MFCCs are less susceptible to the said variations [13]. The following steps are involved in extracting the MFCC feature

- Fourier transform is taken for the signal
- With the aid of triangular overlapping windows, the power spectrums are compared with mel scale
- At each of the mel frequency the logs of powers are taken
- Discrete Cosine Transform (DCT) is taken for the mel log powers
- Obtaining the resulting spectrums is the amplitude of MFCCs.

With the utilization of above steps, the MFCC features are obtained from the normal as well as abnormal datasets which is referred as M_a and M_b

2.1.2 Principal Component Analysis (PCA)

PCA is used abundantly in all forms of analysis - from neuroscience to computer graphics - because it is a simple, non-parametric method of extracting relevant information from confusing data sets. With minimal additional effort PCA provides a roadmap for how to reduce a complex data set to a lower dimension to reveal the sometimes hidden, simplified dynamics that often underlie it.

$$N = w - \mu \quad (3)$$

$$cv = \frac{N \times N^T}{n-1} \quad (4)$$

$$Y = N * E^T \quad (5)$$

$$\frac{Y}{E^T} * \mu = w \quad (6)$$

N – Mean Deviation

w - window

μ - mean

cv - covariance vector

E – Eigenvector

The aforesaid equations are utilized to obtain the PCA of both M_a and M_b , where given equations are the general sets of equations to generate PCA. In PCA the data are processed by window by window. After PCA the inverse PCA also applied to obtain the dimensionality reduced original information again. After this process completed, the following parameters are obtained from the MFCC featured vectors M_a , M_b . The parameters are mean, standard deviation, maximum amplitude value and its id, minimum amplitude value and its id, MFCC length are extracted for the MFCC featured word and as well as for the original word also extracted and hence for each word we have 14 inputs.

2.2 Classification through Neural Network (NN)

The Artificial neural network (ANN) used in the model was a multilayer perceptron (MLP) with two layers of neurons. The number of neurons in the hidden layer is dependent on the size of the input vector [13]. The output layer has one neuron. Both the words of normal and pathological subjects are inputted to the ANN to be trained and to identify the abnormal word. The 14 features extracted for each word is given as input to neural network.

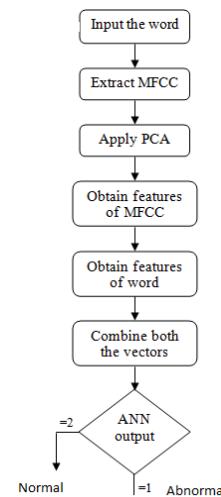


Fig. 2 Data Flow of Classification

NN is utilized for the classification purpose in order to identify the normal speech and the abnormal speech. In order to identify this, initially the MFCC features are obtained from both the normal and abnormal words. After that, both the words are inputted to the NN to identify the abnormal word. A artificial neural network is developed with a systematic step-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.

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.

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 3.

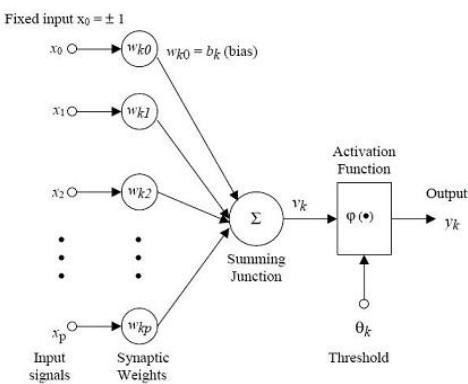


Fig. 3 ANN Model

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, y_k , would therefore be the outcome of some activation function on the value of v_k .

3. Experimental Setup

3.1 The speech database

In this with the aid of the **Free Audio Editor** we generate the dataset with the normal and abnormal female children within the age limit 6-10. For normal data we utilized 2 female children and for abnormal data we utilized a female child for our system and their normal frequency range is from 20 - 4 kHz. The proposed technique is tested with the database of 100 words with two normal children and an abnormal child each.

3.2 Classification using NN

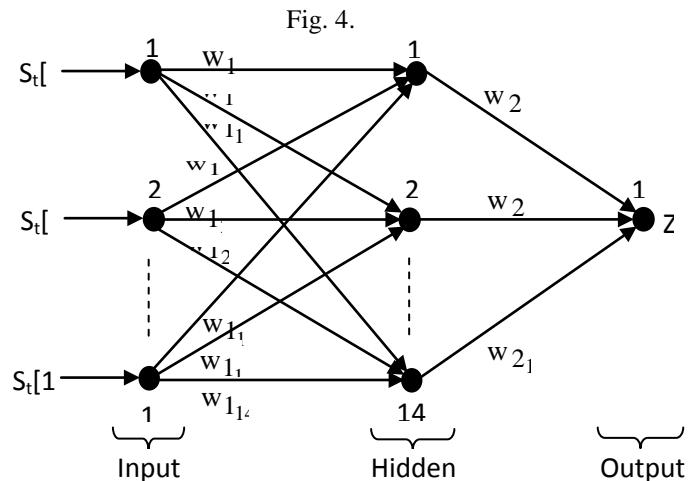
NN is used for the classification purpose in order to separate the normal speech and for identifying the abnormal speech. The following steps details the training process of ANN

Step 1: As the first step, set up the input weights to every neuron, apart from the neurons of the input layer.

Step 2: This neural network has 14 input layers which are the parameters of a word as said in the section 2.2. N_h hidden layers and one output layer to identify the word inputted is either normal or abnormal. The value at the output layer is

either true or false depending on whether the input word is abnormal or normal. In this neural network, 14 input neurons and a bias neuron, N_h hidden neurons and a bias neuron and one output neuron are present.

Step 3: The weights are added to the designed network N, also it is biased. The developed N is shown in Fig.4.



Step 4: The basis function and the activation function which are chosen for the designed N is given below.

$$Z_i = \alpha + \sum_{j=1}^{N_h} (w_{ij} S_t[1] + w_{ij} S_t[2] + w_{ij} S_t[3] + w_{ij} S_t[4] + \dots + w_{ij} S_t[14]) \quad (7)$$

$$h(Z_i) = \frac{1}{1 + e^{-Z_i}} \quad (8)$$

$$h(Z_i) = Z_i \quad (9)$$

Eq. (7) is the basis function for the input layer, where $S_t[1] - S_t[14]$ are the parameters for the word which are mean, standard deviation, maximum amplitude value and its id, minimum amplitude value and its id, MFCC length for MFCC featured word and for the original word we extract the same word for the MFCC length here we extract the word length. Here w_{ij} is the weight of the neuron and α is the bias. The sigmoid function for the hidden layer is given in Eq.(7) and the activation function for the output layer is given in Eq.(8). The basis function given in Eq. (7) is commonly used in all the remaining layers (hidden and output layer, but with the number of hidden and output neurons, respectively). The output of the ANN is obtained by giving the region vector as its input.

Step 5: The learning error is determined for the network N as follows

$$Er = \frac{1}{N_h} \sum_{o=0}^{N_h-1} D_o - Z_o z \quad (10)$$

Here, Er is the error in the FF-ANN, D_o is the desired output and Z_o is the actual output.

3.2.1 Error minimization by BP algorithm

The steps involved in the training of BP algorithm based NN is given below.

(1) Randomly generate weights in the interval $[0,1]$ and assign it to the neurons of the hidden layer and the output layer. But all neurons of the input layer have a constant weight of unity.

(2) Determine the BP error using Eq. (9), and give the training gene data sequence as input to the N Eq. (7), Eq. (38) and Eq. (9) show the basis function and transfer function.

(3) Adjust the weights of all the neurons when the BP error is determined as follows,

$$w_{ij} = w_{ij} + \delta w_{ij} \quad (11)$$

The change in weight δw_{ij} given in Eq. (7) can be determined as $\delta w_{ij} = \gamma \cdot Z_{ij} \cdot Er$, where Er is the BP error and γ is the learning rate and it normally ranges from 0.2 to 0.5.

(4) After adjusting the weights, repeat steps until the BP error gets minimized. Normally, it is repeated till the criterion, $E < 0.1$ is satisfied.

Once the error gets minimized to a minimum value it is concluded that the designed FF-ANN is well trained for its further testing phase and the BP algorithm is terminated. Thus the neural network is trained using the parameters for each word.

4. Experimental Results

4.1 Feature Extraction

Initially, the words are extracted from the both normal and abnormal children and then the MFCC feature has been

extracted from it. Subsequently, the PCA is applied to reduce the dimensionality of the words. Few speech samples which are sent as input to MFCC Feature Extraction.

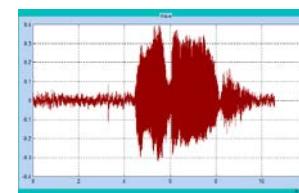


Fig. 5 Normal speech 1
 “wild animals”

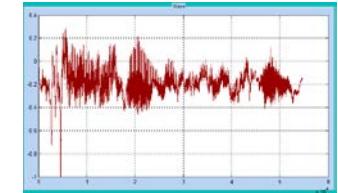


Fig. 6 Normal speech 2 – “wild animals”

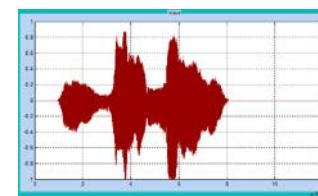


Fig. 7 Abnormal speech – “wild animals”

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	-2.46E-15	2.853887	0.615143	2	-17.9884	318	613	-2.64E-06	0.001638	0.061798	26206	-0.06305	26702	49558
2	-2.90E-15	3.621083	3.73356	2	-14.1841	78	361	0.00052	0.056483	0.325134	16083	-0.26419	16712	99425
3	-1.40E-15	2.415081	4.694251	215	-13.0707	123	544	0.00078	0.071582	0.418274	24326	-0.31506	24213	44081
4	1.55E-15	7.594511	9.461903	209	-15.3411	278	599	0.00068	0.106159	0.487091	10647	-0.37173	8448	48476
5	-3.44E-15	2.805239	9.911978	132	-16.9131	283	1013	0.000489	0.011632	0.249678	10717	-0.73849	10377	81585
6	-2.95E-15	4.703249	1.463655	2	-18.8383	202	291	3.61E-05	0.002095	0.054871	15063	-0.0545	15386	23808
7	6.39E-15	3.230252	1.033956	2	-17.7546	448	577	8.30E-06	0.006413	0.076124	34500	-0.10092	33941	46700
8	-1.60E-14	2.979884	0.783359	1	-19.5246	1190	2612	6.76E-05	0.005292	0.091888	80831	-0.07892	81574	209475
9	1.27E-15	4.920547	1.830258	2	-18.4647	52	324	8.95E-05	0.07539	0.078125	16030	-0.07117	17941	26473
10	2.31E-15	5.644728	6.367292	1	-14.6163	87	357	-0.00051	0.049236	0.231201	12750	-0.24442	3162	29066
11	-9.34E-16	7.616504	3.632384	2	-25.1537	318	564	-0.0001	0.009292	0.105042	24115	-0.14551	23902	45648
12	1.01E-15	3.531965	4.534858	130	-17.9956	36	204	0.001253	0.130464	0.542511	5747	-0.67053	6376	17602
13	1.55E-15	7.594511	9.461903	209	-15.3411	278	599	0.00068	0.106159	0.487091	10647	-0.37173	8448	48476
14	-8.21E-16	2.6914	3.196227	1	-8.0118	23	185	0.000527	0.065959	0.296417	5354	-0.33234	5371	15296
15	1.58E-16	2.372321	3.511143	56	-5.8508	1	73	-0.00162	0.081005	0.297546	867	-0.25574	1869	6344
16	5.48E-16	5.310417	3.168171	1	-16.0582	139	486	-7.05E-05	0.011762	0.111603	9146	-0.08908	10580	39426
17	7.13E-15	4.637277	5.039176	481	-17.1554	236	682	0.000685	0.037202	0.326416	39150	-0.29849	38890	55078
18	1.07E-14	8.656101	10.17948	496	-17.355	726	968	0.000662	0.050097	0.476685	40681	-0.16469	42249	79180
19	1.73E-15	9.102771	7.729523	1	-17.6176	125	360	-0.00047	0.026006	0.176223	5388	-0.16824	5595	24500
20	-1.94E-14	4.896758	3.171191	1	-19.8483	111	438	0.00041	0.037635	0.239727	13120	-0.16919	15554	35571
21	-1.27E-14	8.402864	6.621747	396	-26.2784	97	875	4.02E-05	0.038873	0.307251	32248	-0.27298	30045	70560

Fig. 8 Feature Extraction output of Normal Speech of Child 1

4.2 Classification using ANN

Subsequently, for MFCC features the PCA is applied to reduce the dimensionality of the words and then they are inputted to the neural network to identify the abnormal and the normal word. The target is fixed as 1 and 2 for abnormal and normal speech respectively. Fig.9 is the generated neural network structure.

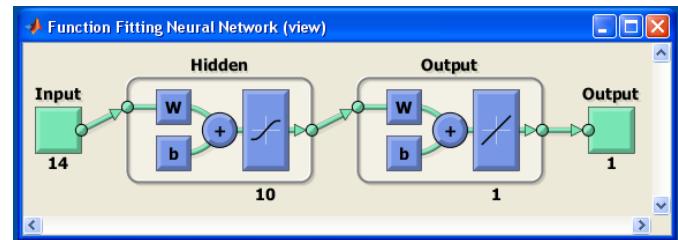


Fig. 9 Generated ANN for classification

5. CONCLUSIONS AND FUTURE WORK

The proposed system was implemented in the working platform of **MATLAB (version 7.11)**. In this with the aid of the **Free Audio Editor** we generate the dataset with the normal and abnormal female children within the age limit 6-10. For 100 normal data (words) we utilized 2 female children each and for 100 abnormal data we utilized a female child for our system and their normal frequency range is from 20 – 4khz. The speech input is recorded at a sampling rate of 44.1kHz. To develop an effective system to identify the abnormal word and the spot, where the speech has to be improved, initially the MFCC is obtained from both the normal and abnormal words and then PCA dimensionality reduction is done. Then classification of normal and abnormal words are done using ANN. The result of this work is discussed with outputs. After that, the work will be extended by (i) Testing Phase (ii) acute Spotting aberration in speech of pathological subject respectively.

References

- [1] Shirbahadurkar and Bormane, "Speech Synthesizer Using Concatenative Synthesis Strategy for Marathi language (Spoken in Maharashtra, India)", International Journal of Recent Trends in Engineering, Vol. 2, No. 4, pp. 80-82, November 2009
- [2] Singaram, Guru Raghavendran, Shivaramakrishnan and Srinivasan, "Real Time Speech Enhancement using Blackfin Processor BF533", J. Instrument Society of India, Vol. 37, No. 2, pp. 67-79, 2009
- [3] Qiguang Lin Ea-Ee Jan and James Flanagan, "Microphone Arrays and Speaker Identification", IEEE Transactions on Speech and Audio Processing, Vol. 2, No. 4, pp. 622-629, October 1994
- [4] Thomas Lotter, Christian Benien and Peter Vary, "Multichannel Direction-Independent Speech Enhancement Using Spectral Amplitude Estimation", EURASIP Journal on Applied Signal Processing, Vol. 2003, No. 11, pp. 1147-1156, 2003
- [5] Sven Nordholm, Thushara Abhayapala, Simon Doclo, Sharon Gannot, Patrick Naylor and Ivan Tashev, "Microphone Array Speech Processing", EURASIP Journal on Advances in Signal Processing, Vol. 2010, pp. 1-3, 2010
- [6] Marius Crisan, "Chaos and Natural Language Processing", Acta Polytechnica Hungarica, Vol. 4, No. 3, pp. 61-74, 2007
- [7] Aida-Zade, Ardin and Rustamov, "Investigation of Combined use of MFCC and LPC Features in Speech Recognition Systems", World Academy of Science, Engineering and Technology, Vol. 3, No. 2, pp. 74-80, Spring 2007
- [8] Chulhee Lee, Donghoon Hyun, Euisun Choi, Jinwook Go and Chungyong Lee, "Optimizing Feature Extraction for Speech Recognition", IEEE Transactions on Speech and Audio Processing, Vol. 11, No. 1, pp. 80-87, January 2003
- [9] Rashad, Hazem M. El-Bakry and Islam R. Ismail, "Diphone Speech Synthesis System for Arabic Using MARY TTS ", International journal of computer science & information Technology (IJCSIT), Vol. 2, No. 4, pp. 18-26, August 2010
- [10] Stelzle, Ugrinovic, Knipfer, Bocklet, Noth, Schuster, Eitner, Seiss and Nkenke, "Automatic, computer-based speech assessment on edentulous patients with and without complete dentures - preliminary results", Journal of Oral Rehabilitation, Vol. 37, No. 3, pp. 209-216, March 2010
- [11] Sumathi and SanthaKumaran, "Pre-Diagnosis of Hypertension Using Artificial Neural Network", Global Journal of Computer Science and Technology, Vol.11, No.2, pp.43-47, February 2011
- [12] El-Shafie, Mukhlisin, Najah and Taha, "Performance of artificial neural network and regression techniques for rainfall-runoff prediction", International Journal of the Physical Sciences, Vol.6, No.8, pp.1997-2003, April 2011
- [13] Rashidul Hasan, Mustafa Jamil, Golam Rabbani and Saifur Rahman, "Speaker Identification Using Mel Frequency Cepstral Coefficients", In proceedings of 3rd International Conference on Electrical and Computer Engineering, Dhaka, Bangladesh, December 2004
- C.R.Bharathi** received AMIE (ECE) and M.E.(Applied Electronics) in 2001 and 2005. Since June 2002, she has been a Lecturer in an Engineering College and at present working in Vel Tech University, Avadi, Chennai. She is an Research scholar at Sathyabama University.
- Dr. V. Shanthi** working as Professor in St.Joseph's College of Engineering,Chennai. Her research interest includes Artificial Intelligence, cloud computing. Dr. V. Shanthi is co-author.

Efficient Retrieval of Text for Biomedical Domain using Expectation Maximization Algorithm

Sumit Vashishtha¹, Dr. Yogendra Kumar Jain²

¹ MTECH Scholar, Computer Science Department
Samrat Ashok Technological Institute, Vidisha, M.P, INDIA

² Head, Computer Science Department,
Samrat Ashok Technological Institute Vidisha, MP, INDIA

Abstract

Data mining, a branch of computer science [1], is the process of extracting patterns from large data sets by combining methods from statistics and artificial intelligence with database management. Data mining is seen as an increasingly important tool by modern business to transform data into business intelligence giving an informational advantage. Biomedical text retrieval refers to text retrieval techniques applied to biomedical resources and literature available of the biomedical and molecular biology domain. The volume of published biomedical research, and therefore the underlying biomedical knowledge base, is expanding at an increasing rate. Biomedical text retrieval is a way to aid researchers in coping with information overload. By discovering predictive relationships between different pieces of extracted data, data-mining algorithms can be used to improve the accuracy of information extraction. However, textual variation due to typos, abbreviations, and other sources can prevent the productive discovery and utilization of hard-matching rules. Recent methods of soft clustering can exploit predictive relationships in textual data. This paper presents a technique for using soft clustering data mining algorithm to increase the accuracy of biomedical text extraction. Experimental results demonstrate that this approach improves text extraction more effectively than hard keyword matching rules.

Keywords-: Data mining, Biomedical text extraction, Biomedical text mining,Maximization algorithm

1. Introduction

This paper aims to use data mining techniques to extract text from biomedical literature with reasonably high recall and precision. In recent years, along with development of bioinformatics and information technology, biomedical technology grows rapidly. With the growth of the biomedical technology, enormous biomedical databases are produced. It creates a need and challenge for data mining. Data mining is a process of the knowledge discovery in databases and the goal is to find out the hidden and interesting information[3]. The technology includes association rules, classification, clustering, and evolution analysis etc. Clustering algorithms are used as the essential tools to group analogous patterns and separate outliers according to its principles that elements in the same cluster are more homogenous while elements in the different

ones are more dissimilar [2]. Furthermore, data mining algorithms do not need to rely on the pre-defined classes and the training examples while classifying the classes and can produce the good quality of clustering, so they fit to extract the biomedical text better. A major challenge for information retrieval in the life science domain is coping with its complex and inconsistent terminology. In this paper we try to devise an algorithm which makes word-based retrieval more robust. We will investigate how data mining algorithms based on keywords affects retrieval effectiveness in the biomedical domain. We will try to answer the following research question in this paper “How can the effectiveness of word-based biomedical information retrieval be improved using data mining algorithm?”

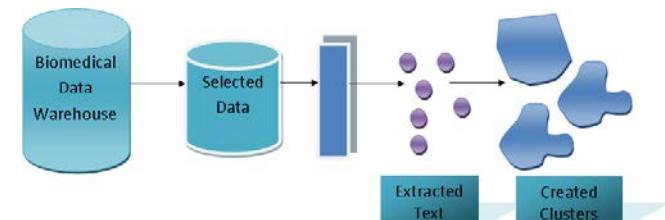


Figure 1: Text extraction from Biomedical literature base

2. Background

Biomedical text extraction refers to text mining applied to texts and literature of the biomedical and molecular biology domain. It is a rather recent research field on the edge of natural language processing, bioinformatics, medical informatics and computational linguistics.

There is an increasing interest in text mining and information extraction strategies applied to the biomedical and molecular biology literature due to the increasing number of electronically available publications stored in databases.

The main developments in this area have been related to the identification of biological entities (named entity recognition), such as protein and gene names in free text, the association of gene clusters obtained by microarray experiments with the

biological context provided by the corresponding literature, automatic extraction of protein interactions and associations of proteins to functional concepts (e.g. gene ontology terms). Even the extraction of kinetic parameters from text or the subcellular location of proteins have been addressed by information extraction and text mining technology.

The optimal retrieval of a literature search in biomedicine depends on the appropriate use of Medical Subject Headings, descriptors and keywords among authors and indexers. We hypothesized that authors, investigators and indexers in four biomedical databases are not consistent in their use of terminology in Complementary and Alternative Medicine.

The increasing research in Complementary and Alternative Medicine and the importance placed on practicing evidence-based medicine require ready access to the biomedical scientific literature. The optimal retrieval of a literature search in biomedicine depends on the appropriate use of Medical Subject Headings, descriptors and keywords among authors, indexers, and investigators [4]. It has been recognized that available online databases for biomedical domain differed in their thesaurus construction and indexing procedures, making effective and efficient searching difficult [5].

In this paper we try to employ an algorithm that extracts the biomedical texts from the biomedical database based on the some data mining algorithm. Our approach first identifies the keywords contained in the biomedical database and then clustering these keywords to group all the text that fall into the category of the given keyword i.e. if that keyword is being used for searching the returned cluster for that particular keyword will contain all the text corresponding to that keyword.

3. Method

The main goal of the proposed system is to find valuable information of biomedical domain from the web. In the next step these information is exploited to mining user navigation pattern based on Expectation Maximization (EM) algorithm.

We adopted the usage mining system to exploit user navigation pattern based on the EM algorithm. According to different function, the system is partitioned into two main modules; Data pretreatment and navigation pattern mining.

Data pretreatment in a web usage mining model, aims to reformat the original web logs to identify all web access sessions. There are several tasks in this module of the system. The Web server usually registers all users' access activities of the website as Web server logs. Due to different server setting parameters, there are many types of web logs, but typically the log files share the same basic information, such as: client IP address, request time, requested URL, HTTP status code, referrer, etc. Data cleaning and filtering methods are the next step in this module. Not every access to the content should be taken into consideration. We need to remove accesses to irrelevant items, redundant items, accesses by Web crawlers, and failed requests. Data cleaning also identifies Web robots

and removes their request. Web robots (also called spiders) are software tools that scan a web site to extract its content. In the web usage mining algorithm, to get knowledge about each user's identity is not necessary. However, a mechanism to distinguish different users is still required for analyzing user access behavior. A user session is a delimited set of pages visited by the same user within the duration of one particular visit to a Web site. Session identification is carried out using the assumption that if a certain predefined period of time between two accesses is exceeded, a new session starts at that point. Sessions can have some missing parts. This is due to the browser's own caching mechanism and also because of the intermediate proxy-caches. The missing parts can be inferred from the site's structure. Web usage data is prepared for applying navigation patterns mining algorithms by doing these pretreatment tasks.

Mining of navigation patterns is the main task of the proposed system. The main objective of this module is mining of user's navigation pattern. A user navigation pattern is common browsing characteristics among a group of users. Since many users may have common interests up to a point during their navigation, navigation patterns should capture the overlapping interests or the information needs of these users. In addition, navigation patterns should also be capable to distinguish among web pages based on their different significance to each pattern. In the web usage mining systems, the large majority of methods that have been used for navigation pattern mining from Web data are clustering methods. Clustering aims to divide a data set into groups that are very different from each other and whose members are very similar to each other. In this paper, a partitioning method is used for clustering of user's navigation patterns. Expectation maximization (EM) is a clustering algorithm that works based on partitioning methods. The EM is a memory efficient and easy to implement algorithm, with a profound probabilistic background.

Expectation maximization (EM) is a well-known algorithm used for clustering in the context of mixture models. This method estimates missing parameters of probabilistic models. Generally, this is an optimization approach, which had given some initial approximation of the cluster parameters, iteratively performs two steps: first, the expectation step computes the values expected for the cluster probabilities, and second, the maximization step computes the distribution parameters and their likelihood given the data. It iterates until the parameters being optimized reach a fix point or until the log-likelihood function, which measures the quality of clustering, reaches its maximum. To simplify the discussion we first briefly describe the EM algorithm. The algorithm is similar to the K-means procedure in that a set of parameters are re-computed until a desired convergence value is achieved. The parameters are re-computed until a desired convergence value is achieved. The finite mixtures model assumes all attributes to be independent random variables. A mixture is a set of N probability distributions where each distribution

represents a cluster. An individual instance is assigned a probability that it would have a certain set of attribute values given it was a member of a specific cluster. In the simplest case N=2, the probability distributes are assumed to be normal and data instances consist of a single real-valued attribute. Using the scenario, the job of the algorithm is to determine the value of five parameters, specifically:

1. The mean and standard deviation for cluster 1
2. The mean and standard deviation for cluster 2
3. The sampling probability P for cluster 1 (the probability for cluster 2 is 1-P)

And the general procedure states as follow:

1. Guess initial values for the five parameters.
2. Use the probability density function for a normal distribution to compute the cluster probability for each instance. In the case of a single independent variable with mean μ and standard deviation σ , the formula is:

$$f(x) = \frac{1}{(\sqrt{2\pi}\sigma)e^{-\frac{(x-\mu)^2}{2\sigma^2}}}$$

In the two-cluster case, we will have the two probability distribution formulas each having differing mean and standard deviation values.

3. Use the probability scores to re-estimate the five parameters.
4. Return to Step 2.

The algorithm terminates when a formula that measures cluster quality no longer shows significant increases.

4. Proposed Model

Clustering is the process of organizing objects into groups whose members are similar in some way. It can be considered the most important unsupervised learning problem which deals with finding a structure in a collection of unlabeled data. A cluster is therefore a collection of objects which are “similar” between them and are “dissimilar” to the objects belonging to other clusters. Hard clustering is the techniques in which any pattern can be in only one cluster at any time.

Soft clustering is the technique which permits patterns to be in more than one cluster at any time. There are various clustering approaches that can be applied to cluster the biomedical keywords extracted from full text articles, some of them are k-means, k-median, Hierarchical Clustering

Algorithm, Nearest Neighbor Algorithm etc. Here we are using modified fuzzy C mean clustering algorithm.

Here the proposed algorithm is responsible for extracting keywords present in the full text biomedical article store these keywords in a relation. Then the actual work of algorithm begins, it starts clustering of keywords.

The algorithm initially picks some keywords that are extracted. It groups the full text articles based on these keywords. It means each cluster contains only those articles which contain that keyword as their part. Then it starts using fuzzy C mean clustering to combine the clusters together on some similarity measure. Here we combine two clusters if their similarity measure is greater than or equal to a specified threshold value.

The proposed Algorithm repeats this process until no more changes are made to the clusters. Finally the proposed algorithm stores all the clusters in an xml file. Here our motive to extract all the full text articles which may be relevant for the user providing the search string, for this out of all clusters the cluster with largest number of articles is our target.

5. Proposed Algorithm

The proposed algorithm will take a complete list of all the biomedical articles and the output will be the XML files containing the clusters created using fuzzy c mean algorithm on keywords.

Input: List of full text biomedical articles.

Output: XML files containing the created clusters.

Algorithm

1. Read the next article in the list of biomedical text
2. Read the full text article
3. Extract the keywords from the article using KEA algorithm
4. Refer to the biomedical lexicon and discard the irrelevant keywords
5. Put the data in following relation so that the full text can be retrieved later using keywords only.

Article UID	Article Name	Keywords	Full text	Source
----------------	-----------------	----------	--------------	--------

6. Go to step 1 and repeat till all the articles in the list of biomedical articles are processed.
7. Apply navigation pattern mining using Expectation maximization.
8. Store a text file of all clusters formed.
9. Determine the cluster having maximum item count.
10. Cache all the URLs found in the cluster containing the maximum number of weblog entries.

Note: The relation created step 10 will be used at the time of retrieval. Whenever the biomedical database is searched for any word the cluster containing the matching keywords is returned. The respective full text and other details corresponding to the returned cluster can be retrieved using this relation.

6. Result

The experiments were performed on the test application developed in .Net 2.0. The database contains all the article entries populated manually from the web resources like "<http://www.medilexicon.com>" and few more, starting with letter 'A'.

The search was performed using the traditional keyword based search algorithm and compared with the proposed algorithm. The snapshot for asset of search results is shown in Table 1.

Given the same data for text extraction, the proposed algorithm seems to be retrieving approximately 69% more relevant search results than the keyword based searching. Figure 2 illustrates the improvement achieved using the proposed algorithm.

Table 1. Comparison of results using traditional and proposed algorithm

Search Keyword	List of matching articles found	
	Keyword based search	Proposed algorithm
abarognosis	42	71
abasia	23	39
abasia-astasia	34	57
abasic	32	54
abatement	42	71
abatic	5	8
abaxial	53	90
Abbé	43	73
Abbé condenser	44	74

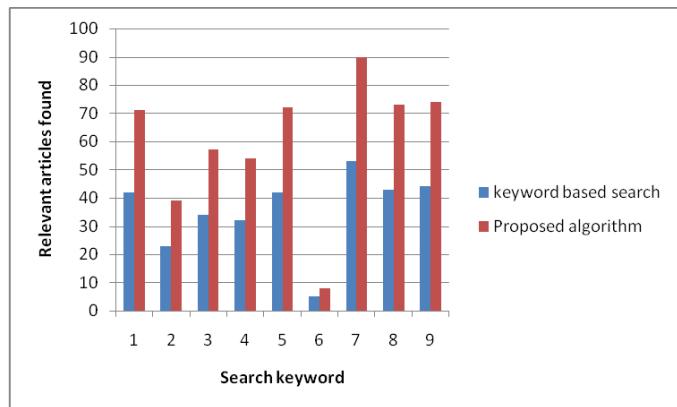


Figure 2: Improved text extraction using proposed algorithm

7. Conclusion

Extraction of text from biomedical literature is an essential operation. Given that there have been many text extraction methods developed; this paper presents a novel technique that employs keyword based article clustering to further enhance the text extraction process. The development of the proposed algorithm is of practical significance; however it is challenging to design a unified approach of text extraction that retrieves the relevant text articles more efficiently. The proposed algorithm, using data mining algorithm, seems to extract the text with contextual completeness in overall, individual and collective forms, making it able to significantly enhance the text extraction process from biomedical literature.

ACKNOWLEDGMENT

This research is supported by the Computer Science and Engineering department, SATI, Vidisha.

REFERENCES

- [1] Clifton, Christopher (2010). "Encyclopedia Britannica: Definition of Data Mining". Retrieved 2010-12-09.
- [2] Berkhin, P., Survey of Clustering Data Mining Techniques. Technical Report, Accrue Software, San Jose, CA, 2002.
- [3] Han, J., & Kamber, M., Data Mining Concepts and Techniques. CA : Morgan Kaufmann, 2001.
- [4] Badgett RG: How to search for and evaluate medical evidence. Seminars in Medical Practice 1999, 2:8-14, 28.
- [5] Richardson J: Building CAM databases: the challenges ahead. J Altern Complement Med 2002, 8:7-8.
- [6] Miller, H. and Han, J., (eds.), 2001, Geographic Data Mining and Knowledge Discovery, (London: Taylor & Francis).
- [7] Manu Aery, Naveen Ramamurthy, and Y. Alp Aslandogan. Topic identification of textual data. Technical report, The University of Texas at Arlington, 2003.
- [8] Pavel Berkhin. Survey of clustering data mining techniques. Technical report, Accrue Software, San Jose, CA, 2002.
- [9] Cecil Chua, Roger H.L. Chiang, and Ee-Peng Lim. An integrated data mining system to automate discovery of

- measures of association. In Proceedings of the 33rd Hawaii International Conference on System Sciences, 2000.
- [10] George Forman. An extensive empirical study of feature selection metrics for text classification. *J. Mach. Learn. Res.*, 3:1289-1305, 2003.
- [11] Rayid Ghani. Combining labeled and unlabeled data for text classification with a large number of categories. In IEEE Conference on Data Mining, 2001.
- [12] George Karypis and Eui-Hong Han. Concept indexing: A fast dimensionality reduction algorithm with applications to document retrieval and categorization. Technical report TR-00-0016, University of Minnesota, 2000.
- [13] Jerome Moore, Eui-Hong Han, Daniel Boley, Maria Gini, Robert Gross, Kyle Hastings, George Karypis, Vipin Kumar, and Bamshad Mobasher. Web page categorization and feature selection using association rule and principal component clustering. In 7th Workshop on Information Technologies and Systems, 1997.
- [14] Sam Scott and Sam Matwin. Text classification using wordnet hypernyms. In Proceedings of the COLING/ACL Workshop on Usage of WordNet in Natural Language Processing Systems, Montreal, 1998.
- [15] Michael Steinbach, George Karypis, and Vipin Kumar. A comparison of document clustering techniques. In KDD Workshop on Text Mining, 2000.
- [16] Andreas Weingessel, Martin Natter, and Kurt Hornik. Using independent component analysis for feature extraction and multivariate data projection, 1998.
- [17] Robert Nisbet (2006) Data Mining Tools: Which One is Best for CRM? Part 1, *Information Management Special Reports*, January 2006.
- [18] Dominique Haughton, Joel Deichmann, Abdolreza Eshghi, Selin Sayek, Nicholas Teebagy, & Heikki Topi (2003) A Review of Software Packages for Data Mining, *The American Statistician*, Vol. 57, No. 4, pp. 290-309.
- [19] R. Agrawal et al., Fast discovery of association rules, in Advances in knowledge discovery and data mining pp. 307-328, MIT Press, 1996.

Authors Profile

Sumit Vashishta 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 CSE (HONS) from Rajiv Gandhi Technical University, Bhopal (M.P.) India in 2006.



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 40 Research papers in various Journals-Conferences, which include 15 research papers in International Journals.

Symmetrical Dispersion Compensation For High Speed Optical Links

Ojuswini Arora¹, Dr.Amit kumar Garg², Savita Punia³

¹Electronics & Communication Department

M.M.E.C, M.M University, Mullana, Ambala, 133203 Haryana, India

²Electronics & Communication Department

M.M.E.C, M.M.University, Mullana, Ambala, 133203 Haryana, India

³Electrical Department

M.M.E.C, M.M University, Mullana, Ambala, 133203 Haryana, India

Abstract

In this paper, the performance of high speed optical fiber based network is analysed by using dispersion compensating module (DCM). The optimal operating condition of the DCM is obtained by considering dispersion management configurations for the symmetrical system i.e Pre-compensation & Post-compensation. The dispersion compensating fiber (DCF) is tested for a single span, single channel system operating at a speed of 10 Gb/s with a transmitting wavelength of 1550 nm, over 120 km single mode fibre by using the compensating fiber for 24 km, 30km and 35Km. So far, most of the investigations for single mode fiber (SMF) transmission at high amplifier spacings in the order of 90 km to 120 km is focused on conventional Non Return to Zero(NRZ) format. The simulation results are validated by analysing the Q-factor and Bit error rate (BER) in the numerical simulator OptSim.

Keywords: Dispersion, Dispersion Compensation, Symmetrical compensation, Bit error rate (BER), Dispersion compensating fiber (DCF), single mode fiber(SMF).

1. Introduction & Previous Work

It is seen that attenuation in the single-mode fiber is lowest in the 1.55- mm wavelength region (0.2 dB/km). However, the intramodal dispersion in this wavelength region is so severe that its effect is intolerable for very high *BL* systems. Because of the high dispersion the operating speed of the network reduces to 2.5 Gb/s. In order to take the advantage of lowest attenuation at this wavelength window, it was realized that the optical fibers need to be modified to have low dispersion in this window. In this technique, fibres of negative dispersion coefficient are made to alternate along the length of the optical link. Negative dispersion fibers (NDF) have a large dispersion in comparison to standard SMF, thus a relatively short NDF can compensate for dispersion accumulated over long links of SMFs. NDFs are easy to

install and require little modification to an already existing system. The major disadvantage of NDF is that it exhibits a large attenuation in signal power, which results into using more optical amplifiers to be employed in the system. This in turn will overcome the other limitations in the system because the non-linear attributes of this fibre are considerably higher. So, which has been validated by using Symmetric Compensation (Pre, Post compensation).Also the results have been validated by numerical simulations with the optical simulator OptSim. Nutys et al.[1] investigated theoretically and experimentally the transmission performance of a 10 Gb/s repeater transmission system using DCF. The system configuration that was considered is a 360 km standard (1300 nm zero-dispersion) fiber transmission system with an optical repeater including DCFs located every 120 km (or every 2100 ps/nm dispersion). The transmitter was a DFB laser externally modulated by a zero-chirp LiNbO₃ modulator with NRZ (non-return to zero), 440 PRBS data. The results of this investigation clearly demonstrate that the use of DCFs is an extremely effective method to overcome the chromatic dispersion in high-speed transmission systems. Weinert et al. [2] investigated the possibilities of 40 and 440 Gb/s time division multiplexing/wavelength division multiplexing (TDM/WDM) return-to-zero (RZ) transmission over embedded standard single-mode fibers (SMF) at a transmission wavelength of 1.55μ m both experimentally and theoretically. Dispersion of the SMF was compensated by a dispersion compensating fiber (DCF). Transmission over a span of 150 km of SMF in the single channel case and of 100 km SMF in the multichannel case is reported. It was shown numerically that improvement was achieved by employing the newest type DCF which also compensates the dispersion

slope of the SMF. Mob et al.[3] theoretically and experimentally analyses advantages of nonlinear RZ over NRZ on 10 Gb/s single-span links. In India, during the past few years' fiber dispersion and nonlinearities have been studied and their impact on the system performance has been investigated. Sharma et al. [4] reviewed the various fiber dispersion compensation methods and investigated techniques for compensation of dispersion by differential delay method including the impact of higher order dispersion terms Kaler et al. [5] discussed the limitations due to Group velocity dispersion (GVD) on transmission distance, bit rate and laser line width including the higher order dispersion effects. The power penalty analysis for different realistic weight functions for combating the pulse broadening effects of group-velocity dispersion in a fiber-optic communication link using differential time delay method with higher-order dispersion terms [6] was discussed. Further, the propagation of signal and noise in the transmission medium to observe the validity of higher order dispersion terms [7] was described. The comparison of pre-compensation, post-compensation and symmetrical-dispersion compensation schemes for 10 Gb/s NRZ links using standard and dispersion compensated fibers was also investigated[8].

Section 2 discusses the proposed work regarding the DCF system configuration and its Symmetric Compensation taking into account the pre-compensation & post-compensation. Results for the simulation are validated and discussed in section 3, followed by the concluding remarks related to the dispersion compensation of various configurations in the Section 4.

2. Proposed Work

- The fiber based method employs the dispersion compensation through a small section of fiber length. There are various techniques such as dispersion compensation fiber (DCF), reverse dispersion fiber, negative dispersion fiber to compensate the dispersion of the system.
- Dispersion compensating fiber (DCF) is the predominant technology for dispersion compensation. It consists of an optical fiber that has a special design such as providing a large negative dispersion coefficient while the dispersion of the transport fiber is positive. A proper length of DCF allows the compensation of the chromatic dispersion accumulated over a given length of the transport fiber, although standard modules with predetermined dispersion values (with a typical granularity corresponding to the dispersion of 20 km of SSMF) are commercially available.
- The main advantage of this technology is the fact that it provides a broadband operation with a smooth dispersion

property and good optical characteristics. In the first generation of DCF, only about 60% of the SSMF dispersion slope was allowed to be compensated. Now, 100% slope-matching for both SSMF and E-LEAF is commercially available.

- However, dispersion compensation modules based on a first-generation DCF are largely deployed and their associated slope-mismatch is a problem we have to live with. DCF also presents a quite large insertion loss although improvements have been reported recently. Dispersion compensation modules based on DCF are also bulky and again, size reduction is expected in the future as bend loss reduction could allow a significant improvement in the compactness. The disadvantages of the fiber-based methods are the extra fiber loss, high nonlinearities and an additional cost of the DCF is imposed. The maximum dispersion of such DCF is about -100 ps/nm/km, which is limited by the mismatching of the glass properties between the core and the cladding.
- Very long lengths of dispersion compensating fibers are required to compensate for the dispersion of even modest lengths of transmission. So, it is proposed by using the Symmetrical optimization for the optical fibers using Pre & Post compensation both. Dispersion management can be achieved with various combinations of fibre layout. The widely implemented configurations are Pre-compensation, Post-Compensation and Symmetrical compensation which uses both the techniques in one link as shown in Fig 1.

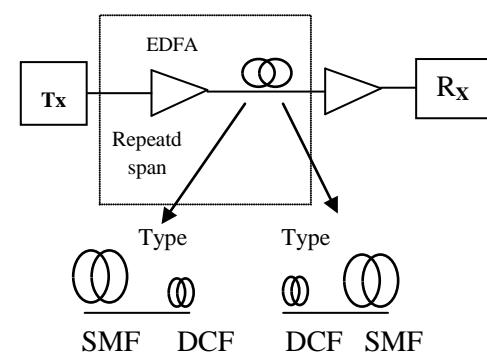


Fig.1 Dispersion management using Symmetric Compensation.

- To achieve optimum dispersion compensation, a dispersion management scheme is modelled using a transmission line consists of equal numbers of 120 km SMF and DCF sections of 24 km, 30 km and 35km for Pre compensation & Post compensation both.

- By varying the lengths of the Post compensating & the Pre compensating fibers , the fiber link is then tested out for an optimum compensation scheme, at which a minimum amount of dispersion , good eye opening,a.high quality factor (Q- Factor) and a low BER is observed

Table1: Simulation Parameters used for the symmetrical compensation

Simulation Parameters	SMF	DCF pre-compensation	DCF post-compensation
Length(km)	120	24	24
Dispersion [ps/km/nm]	16	-80	-80
Loss(dB/km)	0.2	0.6	0.6
Nonlinear Coefficient $(Wkm)^{-1}$	1.26677	1.8	1.8
Transmitter & Receiver Parameters	NRZ	-	-
Bit rate [Gb/s]	10	-	-

- The transmitter section consists of data source, electrical driver (NRZ), laser source (CW Lorentzian) and amplitude modulator (\sin^2 MZ). The fiber parameters for SMF and DCF are listed in Table 1 and the simulated model is shown in fig.2. The figure comprises of equal sections of SMF & DCF (Pre compensation & Post Compensation).
- In this technique, a partial compensation of second-order dispersion by employing DCF units is employed. In the zero path-average dispersion in all schemes or NRZ modulation formats, the transmitter emits chirp-free modulated pulses.
- Dispersion Compensation is employed to increase the transmission distance in systems operating at higher bit rates.Furthermore, Dispersion compensating devices are required to have a sufficiently large bandwidth in order to achieve simultaneous across all the channels.

- Several dispersion and dispersion slope compensating devices are also available including single mode and higher order mode dispersion compensating devices. Although they have great potential for dispersion compensation but DCF are widely employed

- However fiber nonlinearity coefficient for DCF is 1.8 and for SMF is 1.266. At the receiver, the signal is electrically filtered by using Bessel filter(low pass), bandwidth,8dB. As a measure of system performance Q factor and BER are evaluated from the simulations in standard fiber transmissions operating at 10Gb/s at high amplifier spacings of 120km, the impact of fiber nonlinearity is diminished by symmetrical ordering of dispersion compensating fibres.

3. Results & Discussions

By taking into account the dispersion management scheme which is employed using a transmission line consisting of equal numbers of 120 km SMF and DCF sections of 24 km,30km and 35km for Pre compensation & Post compensation. The observations made are as shown in Table 2.

Table2: Simulation Results for the symmetrical compensation for various lengths of Pre & Post compensation

Pre compensating Fiber length(Km)	Post compensating Fiber length(Km)	Q Value (dB)	BER
24	24	30.76	1e-040
30	24	20.53	1.80072e-0.26
30	30	15.814	3.46214 e-010
35	35	7.1682	.0111766

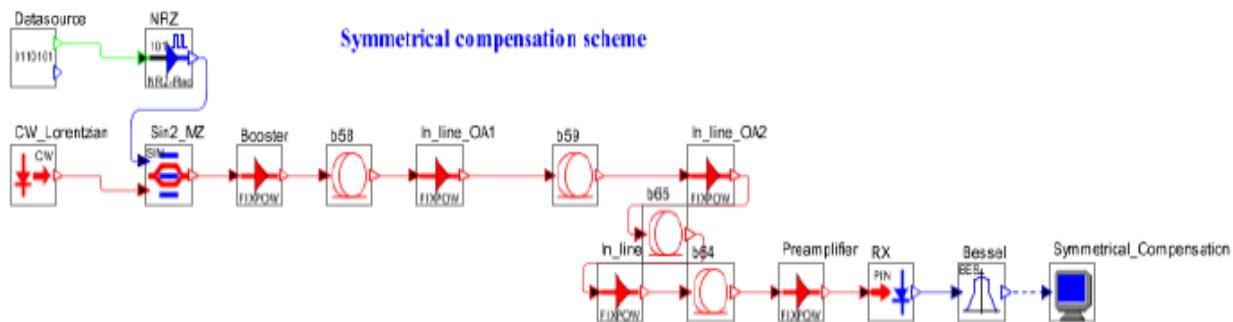


Fig.2 Simulated model for the Symmetrical Dispersion Compensation

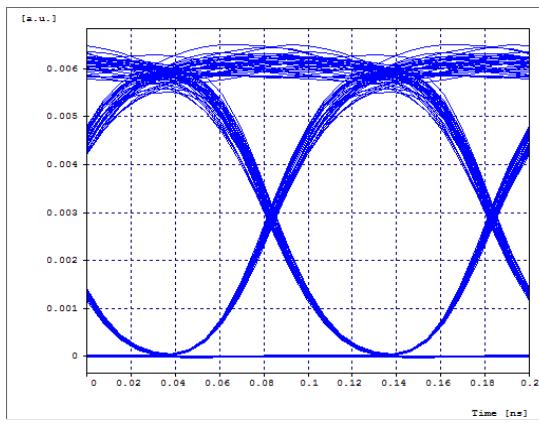


Fig.3 Pre-compensation fiber length=24km, SMF=120Km, Post-Compensation fiber length=24km,Q Value(dB)=30.76, BER=1e-0404

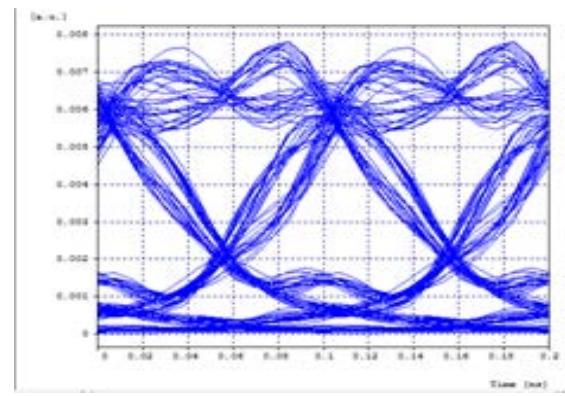


Fig.5 Pre-compensation fiber length=30km, SMF=120Km, Post-Compensation fiber length=30km.Q Value(dB)=15.814, BER=3.47e-01

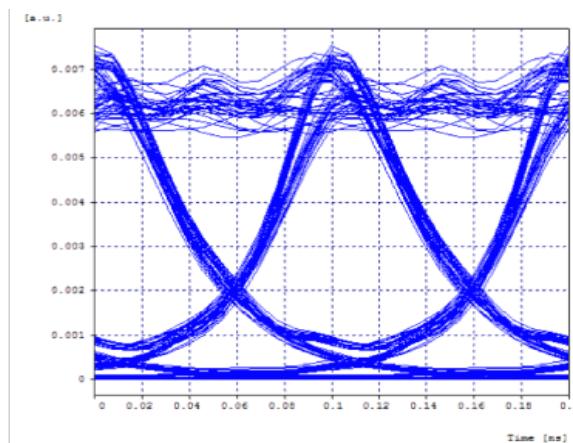


Fig.4 Pre-compensation fiber length=30km,SMF=120Km, Post-Compensation fiber length=24km, Q Value(dB)=20.53, BER=1.8e-0.26

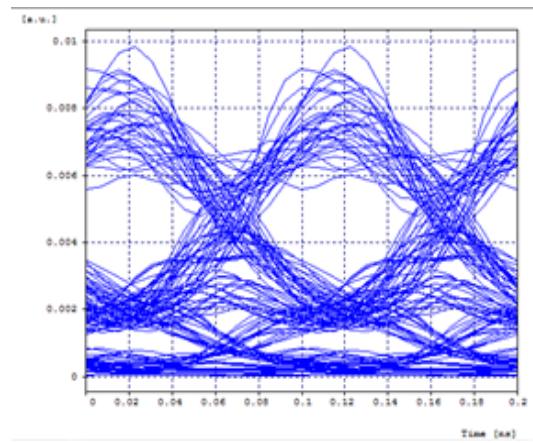


Fig.6 Pre- compensation fiber length=35km, SMF=120Km, Post- Compensation fiber length=35km,Q Value(dB) = 7.1682, BER=.01117

In the proposed methodology, a comparison is made between the various configurations taken, for pre-compensation & post-compensation. It is observed that the effect of dispersion compensating fibers using 30Km fiber for pre compensation, SMF of 120Km and 24Km fiber for the Post compensation, the estimated values of the Q Factor=20.5346dB, BER=1.80072 e 0.26 and Jitter=0.223014ns [fig.4]. Analysing other configurations of Pre-compensating fiber length=30km, SMF=120Km, Post-Compensation fiber length=30km [fig.5] and DCF Pre compensation fiber length=35km, SMF=120Km, Post-Compensation fiber length = 35km [fig.6]. However, if the DCF length for the Pre-compensation & Post compensation fiber = 24Km, SMF=120Km;the estimated values of the Q Factor=30.7602dB, BER=1e-040 and Jitter=0.024919ns [fig3], this is the optimum compensation scheme which has been obtained.

In this work, it is found that the system performance gradually improved as the total dispersion of the transmission fiber tend towards that of the DCF and in a similar fashion, the system performance decreases as the total dispersion of fiber exceeded that of the DCF. Results show that symmetrical compensation for 24Km fiber of DCF is much better than the results obtained for the compensation obtained for the 30Km and 35Km length of the fiber. Furthermore, analysis of the Q-factor revealed that system performance has exceeded by an amount of 10dB which shows a significant increase..

4. Conclusions

The work has emphasised on the dispersion compensation techniques at 10Gb/s. A simulation model is presented for the dispersion compensation in optical fibers for a single channel single span fiber. It is observed that the length of dispersion compensating fiber (DCF) effects the dispersion compensation for the schemes Pre-compensation & Post- compensation. The results show that the optimum compensation scheme is achieved when a DCF of 24km for pre & post compensation is used. It has been found that there is a considerable improvement in terms of Q factor by an amount of 10dB when the results are being compared with the other schemes used.

References

- [1] Roeland J. Nuyts, Yong Kwan Park , and Philippe Gallion, "Dispersion equalization of a 10 Gb/s Repeatered Transmission System Using Dispersion compensating Fibers", Journal of Lightwave Technology, Vol.15, Issue 1,1997. pp.31-42
- [2] C. M. Weinert, R. Ludwig, W. Pieper, H. G. Weber, D. Breuer, K. Petermann, and F. Kuppers," 40 and 4 40 Gb/s time TDM/WDM standard transmission fiber", Journal of Lightwave Technology, Vol. 17, No.11, Nov. 1999.
- [3] G. Mohs, C. Furst, H. Geiger, and G. Fischer, "Advantages of nonlinear RZ and NRZ on 10 Gb/s single-span links," Optical Fiber Communication Conference (OFC), 2000, pp.35-37
- [4] Ajay K. Sharma, R.K. Sinha, R.A.Agrawal, "Improved Analysis of Dispersion Compensation using Differential time delay for high speed long span optical links," Fiber and Integrated Optics (USA), Vol.16, No.4,1997, pp.415-426.
- [5] Ajay K. Sharma, R.K. Sinha, R.A.Agrawal, "Higher Order Dispersion Compensation by Differential Time Delay," Optical Fiber Technology USA, Vol. 4, 1998, pp. 135-143
- [6] Ajay K. Sharma, R.K. Sinha, R.A.Agrawal, "Wavelength Division Multiplexing Systems and Networks," Journal of IETE (Tech Review) Vol 15, 1998, pp. 235-250.
- [7] R.S.Kaler, T.S.Kamal and Ajay K. Sharma, Sandeep K. Arya, R.A.Aggarwala,"Large Signal Analysis of FM-AM Conversion in Dispersive Optical Fibers for PCM systems including Second Order Dispersion," Fiber and Integrated Optics Incorporating International Journal on Optoelectronics, Vol.21, No.3, 2002, pp 193-203.
- [8] R.S.Kaler, Ajay K. Sharma and T.S.Kamal, "Power Penalty Analysis for Realistic Weight Functions using Differential Time Delay with Higher Order Dispersion,"International Journal on Optical Fiber Technology, Vol.8, No.3, 2002, pp 197-207.
- [9] R.S.Kaler, Ajay K. Sharma and T.S.Kamal, "Approximate and Exact Small Signal Analysis for Single Mode Fiber Near Zero Dispersion Wavelength with Higher Order Dispersion," Fiber and Integrated Optics Incorporating International Journal on Optoelectronics, Vol.21, No.5, 2002 , pp 391-415.
- [10] R.S.Kaler, Ajay K. Sharma and T.S.Kamal, "Comparison of Pre-, Post- and Symmetrical-Dispersion Compensation Schemes for 10 Gb/s NRZ Links Using Standard and Dispersion Compensated Fibers," International Journal of Optics Communication,Elsevier Science, vol. 209/1-3, 2002, pp 107-123,
- [11] Mr.Ramesh Pawase, Mrs.R.P.Labade, Dr.S.B.Deosarkar "Dispersion Post- Compensation

Using DCF at 10Gbps” *Global journal of computer science & Technology* Volume XI, Issue III, Version I, 2011.

- [12] D.Breuer, K.Peterman, A. Mattheus, S.K.Turitsyn , “Combatting Fibre Nonlinearity In Symmetrical Compensation Schemes Using RZ Modulation Format At 120 km Amplifier Spacing Over Standard Fibre”, IEEE, 1997.
- [13] M. I. Hayee and A. E. Willner, Senior Member, IEEE “Pre and Post-Compensation of Dispersion and linearities in 10-Gb/s WDM”, IEEE Photonics technology letters,Vol.9, No 9, 1997.
- [14] C. Caspar, H.-M. Foisel, A. Gladisch, N. Hanik, F. K“ppers, R. Ludwig A. Mattheus, W. Pieper, B. Strelbel, and H. G. Weber “RZ Versus NRZ Modulation Format for Dispersion Compensated SMF-Based 10-Gb/s Transmission with More Than 100-km Amplifier Spacing” IEEE Photonics technology letters, Vol. 11,No. 4,April 1999.
- [15] G.P. Agrawal, “Fibre-Optic Communication Systems,” 2nd edition, John Wiley & Sons, Inc ,New York 1997
- [16] G. Keiser, “Optical fibre communication system,” second edition, John Wiley & Sons, Inc, New York 1997

Ojuswini Arora B.Tech in Electronics & Instrumentation from Kurukshetra University, Pursuing M.TECH in Electronics & Communication from M.M. university. She has published research articles in International Journals. Currently employed as a lecturer in Electrical Departmentnet, M.M.E.C, M.M. University, Mullana, INDIA

Dr. Amit Kumar Garg Received his B.E. and M.E. degrees in Electronics & Communication Engineering with distinction from Gulbarga University and Panjab University (India) respectively., Ph.D from Thapar University, Patiala. he is presently working as Professor & Head in Electronics & Communication Engineering Department at M.M.E.C, M.M. University, He has published several scholarly articles in various international journals of high repute & is also the reviewer of various esteemed journals.

Savita Punia M.tech pursuing form Electrical Department, M.M..University, Mullana, .Currently employed in Electrical Department M.M.E.C, M.M.University, Mullana.

Improvement the Flatness, Gain and Bandwidth of Cascaded Raman Amplifiers for Long-Haul UW-WDM Optical Communications Systems

Fathy M. Mustafa¹, Ashraf A. Khalaf² and F. A. El-Geldawy³

¹ Electronics and Communications Engineering Department, Bani-suef University, Egypt

² Electronics and Communications Engineering Department, Mina University, Egypt

³ Electronics and Communications Engineering Department, Mina University, Egypt

Abstract

In the present paper, the problem of multi-pumping all Raman amplifier has been investigated to obtain a maximum flatness gain and increases the bandwidth of the cascaded Raman amplifiers over a wide range of optical signal wavelengths for long-haul ultra-wide wavelength division multiplexing (UW-WDM) transmission systems due to multi-pumping wavelengths and pumping power. Three cases are analyzed where, five, eight and ten Raman pumping of special pumping powers are launched in the forward direction. We obtained a different bandwidth and different gain with flatness [1] but in the present paper we modified the flatness of the gain. The model equations are numerically handled and processed via specially cast software (Matlab). The gain is computed over the spectral optical wavelengths ($1.45\mu\text{m} \leq \lambda_{\text{signal}} \leq 1.65\mu\text{m}$).

Keywords: Raman amplifier, Distributed multi-pumping Raman amplifier (DMRA), Raman gain, pumping power and wavelength, ultra wide-wavelength division multiplexing (UW-WDM).

1. Introduction

Optical amplifiers have played a critical role in the telecommunication revolution that has begun two decades ago. Raman amplification has enabled a dramatic increase in the distance and capacity of light wave systems [2]. Two approaches to Raman amplification have received the most attention. These are designated as single-order and dual-order Raman amplification, where the pump and signal laser are separated by a single Raman-Stokes shift [3].

In the mid-1990s, the development of suitable high power pumps sparked a renewed interest. Researchers were quick to demonstrate some of the advantages that Raman amplifiers have over erbium doped fiber amplifiers (EDFAs), particularly when the transmission fiber itself is turned into a Raman amplifier. This, in turn, shows the exponential

increase since 1994 in the capacity-distance product of transmission experiments reported in literature [4].

There are mainly three reasons for the interest in Raman amplifier. First its capability to provide distributed amplification second is the possibility to provide gain at any wavelength by selecting appropriate pump wavelengths, and the third is the fact that the amplification bandwidth may be broadened simply by adding more pump wavelengths. An important feature of the Raman amplification process is that amplification is achievable at any wavelength by choosing the pump wavelength in accordance with the signal wavelength [5].

The term distributed amplification refers to the method of cancellation of the intrinsic fiber loss. The loss in distributed amplifiers is counterbalanced at every point along the transmission fiber in an ideal distributed amplifier [5].

In the late eighties, Raman amplification was perceived as the way to overcome attenuation in optical fibers and research on long haul transmission was carried out demonstrating transmission over several thousand kilometers using distributed Raman amplification. However, with the development and commercialization of erbium-doped fiber amplifiers through the early nineties, work on distributed Raman amplifiers was abandoned because of its poor pump power efficiency when compared to erbium-doped fiber amplifiers (EDFAs). In the mid-nineties, high-power pump lasers became available and in the years following, several system experiments demonstrated the benefits of distributed Raman amplification including repeater-less undersea experiments, high-capacity terrestrial as well as submarine systems transmission experiments, shorter span single-channel systems including 320 Gbit/s pseudo linear transmissions, and in soliton systems [5].

distributed Raman amplifiers improved noise performance because of amplification at any wavelength controlled simply by selecting the appropriate pump wavelength, extended bandwidth achieved by using multiple pumps when compared to amplification using EDFA, and finally control of the spectral shape of the gain and the noise figure, which may be adjusted by combining and controlling the wavelength and power among multiple pumps [5].

The use of distributed Raman amplification has already been demonstrated in ultra-high-capacity optical communication systems as the enabling method to transmit 40Gbit/s per channel in a wavelength-division-multiplexed transmission system [5].

The most important feature of Raman-gain spectrum is that the peak-gain wavelength only depends on the pump wavelength. The peak-gain wavelength for each pump still exists although the total gain spectrum of a multi-pumped fiber Raman amplifier (FRA) is the comprehensive result of all pumps [6].

Two critical merits of distributed Raman amplifier (DRA) are the low noise and the arbitrary gain band. Experiments show that 2.5 Gb/s system could be upgraded to 10Gb/s by only adding a Raman amplifier [7].

Raman amplifiers pumped at multiple wavelengths draw significant attention in high-speed long-haul WDM transmission, for example, because of their wideband flat-gain profile (100nm with 12 channel-WDM pumping) and superior signal-to-noise ratio (SNR) performance. However, they require numbers of high power pump lasers to achieve high-gain and high bandwidth which makes it very expensive at the initial deployment stage where the WDM bandwidth is not in full use. While modular band-by-band and high upgrade like EDFA-based WDM systems reduces system introduction cost very much, in which either C or L-band EDFA can be added later when a new bandwidth becomes needed. However, such modular addition of amplifiers is not possible for a DRA in which a transmission fiber is shared as common-gain medium. Neglecting nonlinear pump interaction or saturation WDM-pumped Raman amplifier gain can be approximated as the linear superposition of Raman gains induced by each pump laser [8].

Currently, RFAs are the only silica-fiber based technology that can extend the amplification bandwidth to the S band while providing performance and reliability comparable with those of EDFA. However, the noise figure remains high compared to that of the C and L bands [9].

In this paper, Raman gain coefficient and Raman differential gain are processed through a numerical solution of the mathematical model.

2. Mathematical Model

In the present section, we cast the basic model and the governing equation to process N-Raman amplifiers in a cascaded form of special pumping powers $P_{r1}, P_{r2}, P_{r3}, P_{r4}, \dots, P_{rN}$ and corresponding pumping wavelengths $\lambda_{r1}, \lambda_{r2}, \lambda_{r3}, \lambda_{r4}, \dots, \lambda_{rN}$. The map of δ -g is as shown in Fig. 1, where δ is the Raman shift and g is the Raman differential gain coefficient; both were cast based on [10-14] as:

$$\delta = \frac{\lambda_s - \lambda_r}{\lambda_s \lambda_r} \times 10^4, \text{ cm}^{-1} \quad (1)$$

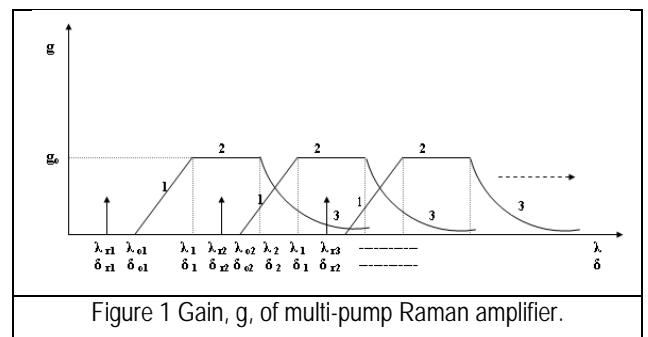


Figure 1 Gain, g, of multi-pump Raman amplifier.

The map of δ -g shown in Fig. 1 describes the basic model. This section depends on the position of the gain of each amplifier with wavelength, where the gain of each amplifier consists of three parts (three equations). A special software program is used to indicate the position of $\delta_{o,i}$ or $\lambda_{o,i}$ and studying the total gain of the amplifiers. In this case, the basic model depends on using more than one amplifier which is put in a cascaded form to increase the bandwidth of the amplifier to multiplexing more signals in the transmission system. The overall amplifier bandwidth increases and the gain flatness improved depend on the position of each amplifier corresponding to other amplifiers. This is achieved by more trials of changing of $\delta_{o,i}$ or $\lambda_{o,i}$ for each amplifier.

The general equations representing the Raman gain in the three regions are respectively [14].

$$g_{1,i} = g_o \frac{\delta}{440}, \quad 0 \leq \delta \leq 440 \quad (2)$$

Where

$$\delta = \frac{\lambda_s - \lambda_r}{\lambda_s \lambda_r} \times 10^4, \text{ cm}^{-1} \quad (3)$$

$$g_{2,i} = g_o, \quad \delta_{1,i} \leq \delta \leq \delta_{2,i} \text{ and } \lambda_1 \leq \lambda \leq \lambda_2 \quad (4)$$

Where $g_o = 7.4 \times 10^{-14} \text{ m/W}$ and

$$\delta_{1,i} = \frac{\lambda_{1,i} - \lambda_{r,i}}{\lambda_{1,i} \lambda_{r,i}} \times 10^4, \text{ cm}^{-1} \quad (5)$$

$$\delta_{2,i} = \frac{\lambda_{2,i} - \lambda_{r,i}}{\lambda_{2,i} \lambda_{r,i}} \times 10^4, \text{ cm}^{-1} \quad (6)$$

And

$$g_{3,i} = g_o e^{-0.005(\delta-440)}, \quad \delta \geq 440 \quad (7)$$

$$\Delta\lambda = \lambda_2 - \lambda_1 = 15 \text{ nm} = (\text{fixed value})$$

$$\lambda_1 = \frac{\lambda_{r1}}{1 - 0.044\lambda_{r1}} \times 10^4, \mu\text{m} \quad (8)$$

$$g_{1,i} = g_o \frac{\delta - \delta_{o,i}}{440} \quad (9)$$

Where

$$\delta_{o,i} \leq \delta \geq \delta_{1,i}, \quad 0 \leq \delta - \delta_{o,i} \leq 440 \quad (10)$$

With

$$\delta_{o,i} = \frac{\lambda_{o,i} - \lambda_{r,i}}{\lambda_{o,i} \lambda_{r,i}} \times 10^4, \text{ cm}^{-1} \quad (11)$$

With $1 \text{ cm}^{-1} = 30 \text{ GHz}$ [15], where $\lambda_{o,i}$ indicates the offset wavelength and $\lambda_{r,i}$ indicates the pumping wavelength of each amplifier. These wavelengths are then used to indicate $\delta_{o,i}$ for each amplifier.

$$g_{2,i} = g_o, \quad \delta_{1,i} \leq \delta \leq \delta_{2,i} \quad (12)$$

Where, $g_o = 7.4 \times 10^{-14} \text{ m/W}$ is the differential Raman gain constant (of pure SiO_2 at $\lambda = 1.34 \mu\text{m}$), and

$$\delta_{1,i} = \frac{\lambda_{1,i} - \lambda_{r,i}}{\lambda_{1,i} \lambda_{r,i}} \times 10^4, \text{ cm}^{-1} \quad (13)$$

$$\delta_{2,i} = \frac{\lambda_{2,i} - \lambda_{r,i}}{\lambda_{2,i} \lambda_{r,i}} \times 10^4, \text{ cm}^{-1} \quad (14)$$

And

$$g_{3,i} = g_o e^{-0.025(\delta - \delta_{2,i})}, \quad \delta \geq \delta_{2,i} \quad (15)$$

$$\Delta\lambda = \lambda_2 - \lambda_1 = 16 \text{ nm} = (\text{fixed value})$$

$$\lambda_1 = \frac{\lambda_{o1}}{1 - 0.044\lambda_{o1}} \times 10^4, \mu\text{m} \quad (16)$$

Where, λ_r is Raman pump wavelength and $\lambda_o \geq 1.35 \mu\text{m}$.

The shift $\delta_{o,i}$ is the Raman shift that indicates the position of each amplifier. By changing this position, the total bandwidth and the flatness of the amplifier are changed. We are interested in obtaining a large bandwidth with flatness by more trials of changing $\delta_{o,i}$ or $\lambda_{o,i}$. In this case, one uses $\delta > \delta_r$ or $\lambda > \lambda_r$ and $\delta_o \geq \delta_r$ or $\lambda_o \geq \lambda_r$, where λ_r is Raman pump wavelength. Raman differential gain constant, g , and the effective core area, A , are defined as [11]:

$$g = 1.34 \times 10^{-6} \times g_o \frac{1 + 80\Delta}{\lambda_r} \quad (17)$$

$$A = \frac{\pi}{2} (W_s^2 + W_r^2), \quad (18)$$

Where

$$W = \frac{0.21\lambda}{\sqrt{\Delta}}, \quad (19)$$

Where, λ_r is the pump wavelength, W_s and W_r are the mode field radii of two light waves coupled with each other with $W=W_s$ at $\lambda=\lambda_s$ and $W=W_r$ at $\lambda=\lambda_r$ and Δ is the relative refractive index difference.

Neglecting the cross coupling among the signal channels, one has the differential equation governing the signal propagation for N-channels Raman pumping [12]:

$$\frac{ds_i}{dz} + \sigma_{si}s_i = \left(\sum_{i=1}^{i=N} \sum_{j=1}^{j=M} \frac{g_{ij}}{A_{ij}} P_{Rj} \right) s_i, \quad (20)$$

Where, $i = 1, 2, 3, \dots, N$, M is the number of pumps, s_i is signal power and P_{Rj} is the pump power. Assume the R.H.S of equation (20) equals g_{ti} , as:

$$g_{ti} = \left(\sum_{i=1}^{i=N} \sum_{j=1}^{j=M} \frac{g_{ij}}{A_{ij}} P_{Rj} \right), \quad (21)$$

The total gain coefficient in m^{-1} which represents the total gain coefficient of the i^{th} signal due to the N -pumping. It is clear that g_{ti} is a function of the set of variables {signal wavelength, fiber radius, Raman wavelength, relative refractive index difference, Raman power}. This term can be written in the form:

$$g_{ti} = \left(\sum_{i=1}^{i=N} \sum_{j=1}^{j=M} g_{di} P_{Rj} \right), \quad (22)$$

Define g_{ci} , the total gain coefficient per watt, as

$$g_{ci} = \left(\sum_{i=1}^{i=N} \sum_{j=1}^{j=M} \frac{g_{dij}}{A_{ij}} \right), m^{-1} W^{-1} \quad (23)$$

Then, the total differential gain, g_{di} , is:

$$g_{di} = \left(\sum_{i=1}^{i=N} \sum_{j=1}^{j=M} g_{ij} \right), mW^{-1} \quad (24)$$

The three gain coefficients g_{di} , g_{ci} and g_{ti} are also functions of the propagation distance.

3. Simulation Results and Discussions

The bandwidth for cascaded multi-pumping Raman amplifier is optimized. Optimal results show that the amplifier bandwidth, $\Delta\lambda_r$, can be evidently broadened by means of increasing the number of pumps and according to the position of each amplifier. It is found that $\Delta\lambda_r$ decreases with the increase of Raman gain and with the improvement of flatness. The hybrid EDFA and DMRA can availablely overcome the weakness of pure DMRA. In this paper,

we discuss three different models; namely, five, eight and the ten Raman pumping optical wavelengths and pumping powers are shown in the tables I, II, and III, respectively, where the sum of pumping powers is one watt. The three gain coefficients g_{di} , g_{ci} , and g_{ti} are displayed for each case.

Case I

Table I Number of amplifiers = 5

$$\lambda_1 = \frac{\lambda_o}{1 - 0.044\lambda_o} \times 10^4, \mu\text{m}$$

$\lambda_1 - \lambda_o$ =fixed value (0.096294798) and $\lambda_2 - \lambda_1 = 16$ nm

λ_r	λ_o	λ_1	λ_2	$P_p(\text{W})$
1.4	1.432	1.528294799	1.544294799	0.17
1.44	1.477	1.573294799	1.589294799	0.25
1.46	1.500	1.596294799	1.612294799	0.18
1.48	1.523	1.619294799	1.635294799	0.24
1.5	1.532	1.628294799	1.644294799	0.16

Differential gain

Figure 2 displays the differential Raman gain, g , with wavelength, λ , at different values of the relative refractive index difference. If relative index difference increases, Raman gains increases.

We note that Raman gain starts to increase from the first pumping wavelength to reach to peak value at $1.59 \mu\text{m}$, then the gain is start to decrease exponentially tended to zero at $1.65 \mu\text{m}$. Because of optical amplifiers and optical signals are operated in range $1.45 \mu\text{m}$ to $1.65 \mu\text{m}$.

In this case we obtained, total bandwidth = 110nm , where $\lambda_{1t} = 1.51 \mu\text{m}$ (for all amplifiers) and $\lambda_{2t} = 1.62 \mu\text{m}$ (for all amplifiers).

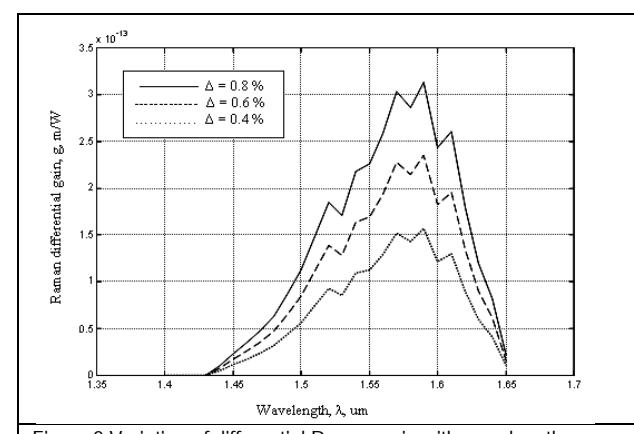


Figure 2 Variation of differential Raman gain with wavelength.

Fig.3 depicts the relation between Raman gain, g m/w and pumping wavelength. This figure is plotted

at different values of relative index difference, where pumping wavelengths for optical signals in range from 1.4 to 1.55 μm , this range is suitable for Raman amplifier to avoid noise and losses.

Then any source has pumping wavelength and pumping power must be suitable for choice design to obtained suitable gain and bandwidth

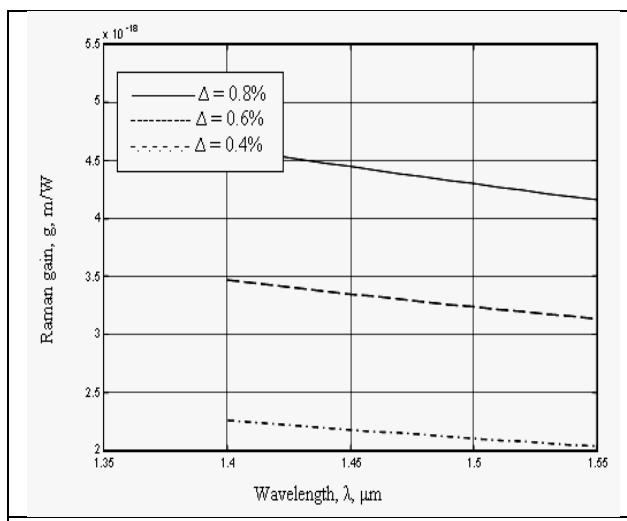


Figure 3 Raman gain against pumping wavelength.

Figure 4 displays Raman gain, g , against the relative refractive index difference. This figure is plotted for special pumping wavelengths.

So relative index difference of the materials must be take in account in design for optical amplifiers.

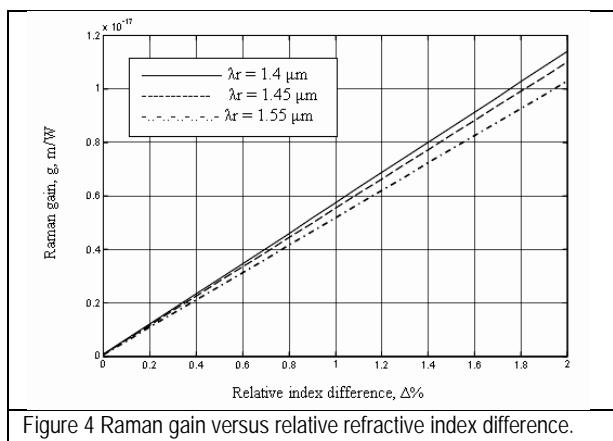


Figure 4 Raman gain versus relative refractive index difference.

Gain coefficient per unit watt

The gain coefficient/unit watt, $\sum g_i / A_i, \text{m}^{-1} \text{W}^{-1}$ against wavelength is shown in Fig. 5 at different values of relative refractive index difference.

We note that gain coefficient/unit watt is starts to increase from the first pumping wavelength to reach to peak value at 1.59 μm , then the gain is start to

decrease exponentially tended to zero at 1.65 μm . In this case more than one parameter can be control in gain coefficient per unit watt such that effective core area, relative refractive index difference, pumping wavelengths and pumping powers. So these parameters take in account for any design. Where each parameter can effected in design.

In this case we obtained, total bandwidth =110nm, where $\lambda_{1t} = 1.51 \mu\text{m}$ (for all amplifiers) and $\lambda_{2t} = 1.62 \mu\text{m}$ (for all amplifiers).

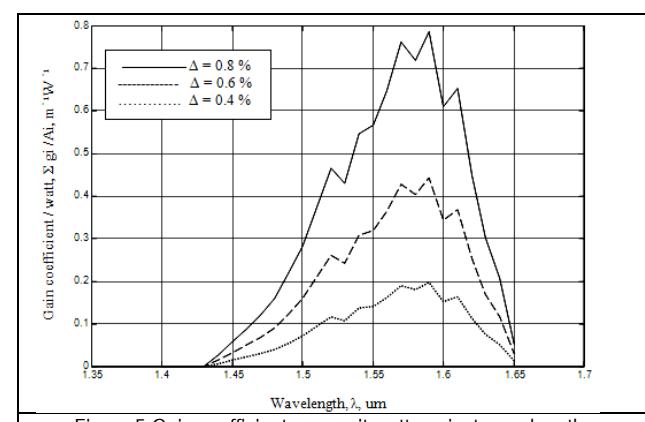


Figure 5 Gain coefficients per unit watt against wavelength.

Total gain coefficient

Figure 6 displays the variation of the total gain coefficient with wavelength. In this case, a bandwidth of 110 nm is obtained. By similar we note the total gain coefficient is start to increase from the first pumping wavelength to reach to peak value at 1.59 μm , the gain is start to decrease exponentially tended to zero at 1.65 μm . Gain in this case is affected by pumping powers, effective core area and relative index difference.

Where pumping powers increase the total gain is increase so Raman amplifiers is used to sources with high pumping powers.

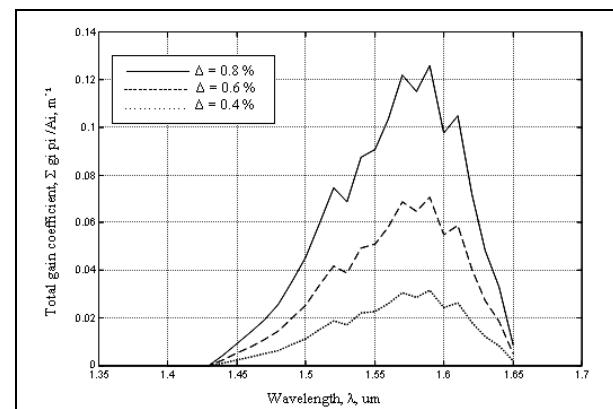


Figure 6 Variation of total gain coefficient with wavelength.

Case II

Table II Number of amplifiers = 8

$\lambda_1 - \lambda_0$ =fixed value (0.093262399) and $\lambda_2 - \lambda_1 = 16$ nm

λ_r	λ_0	λ_1	λ_2	$P_p(W)$
1.41	1.41	1.503262399	1.519262399	0.14
1.44	1.444	1.537262399	1.553262399	0.12
1.45	1.455	1.548262399	1.564262399	0.14
1.46	1.466	1.559262399	1.575262399	0.10
1.47	1.478	1.571262399	1.587262399	0.14
1.48	1.489	1.582262399	1.598262399	0.12
1.49	1.501	1.594262399	1.610262399	0.11
1.5	1.512	1.605262399	1.621262399	0.13

The differential Raman gain and the gain coefficient per unit watt are displayed, respectively, in Figs. 7 and 8, while the total gain is displayed in Fig. 9. In this case, a 110 nm bandwidth is obtained. Peak value in this case at 1.55 μm for the different gain.

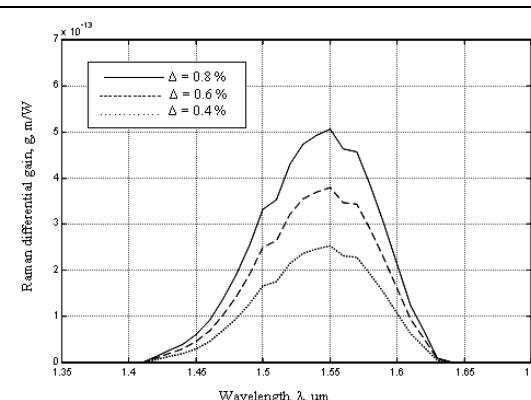


Figure 7 Variation of differential Raman gain with wavelength.

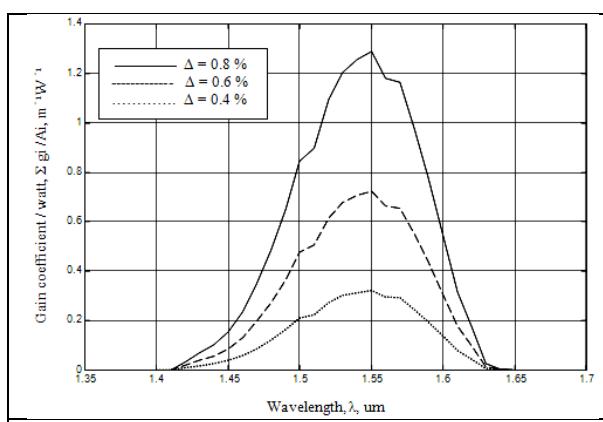


Figure 8 Gain coefficients per unit watt against wavelength.

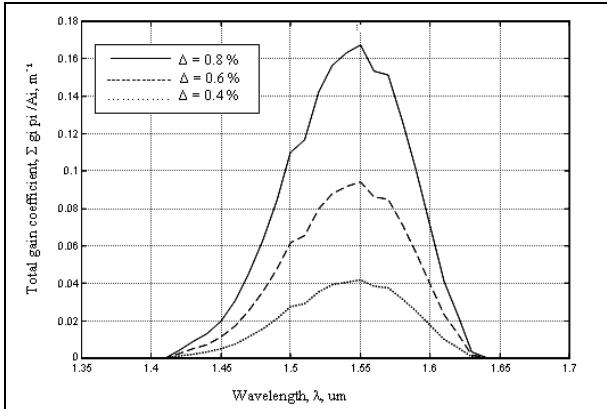


Figure 9 Variation of total gain coefficient with wavelength.

Case III

Table III Number of amplifiers = 10

$\lambda_1 - \lambda_0$ =fixed value (0.093262399) and $\lambda_2 - \lambda_1 = 16$ nm

λ_r	λ_0	λ_1	λ_2	$P_p(W)$
1.41	1.41	1.503262399	1.519262399	0.08
1.42	1.421	1.514262399	1.530262399	0.08
1.43	1.432	1.525262399	1.541262399	0.12
1.44	1.444	1.537262399	1.553262399	0.13
1.45	1.455	1.548262399	1.564262399	0.10
1.46	1.466	1.559262399	1.575262399	0.10
1.47	1.478	1.571262399	1.587262399	0.12
1.48	1.489	1.582262399	1.598262399	0.11
1.49	1.501	1.594262399	1.610262399	0.08
1.5	1.512	1.605262399	1.621262399	0.08

The differential Raman gain and the gain coefficient per unit watt are displayed, respectively, in Figs. 10 and 11, while the total gain is displayed in Fig. 12. In this case, a 110 nm bandwidth is obtained. Peak value in this case at 1.52 μm for the different gain.

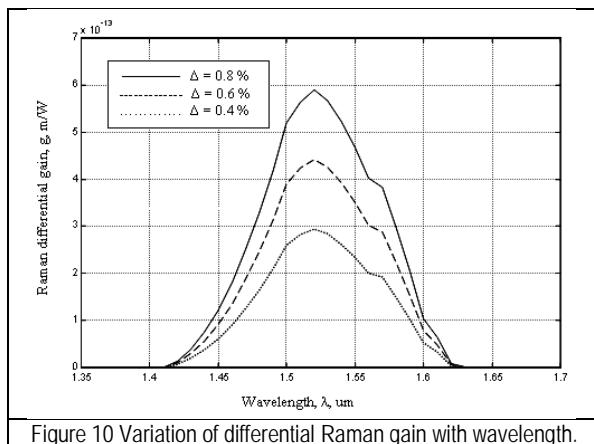


Figure 10 Variation of differential Raman gain with wavelength.

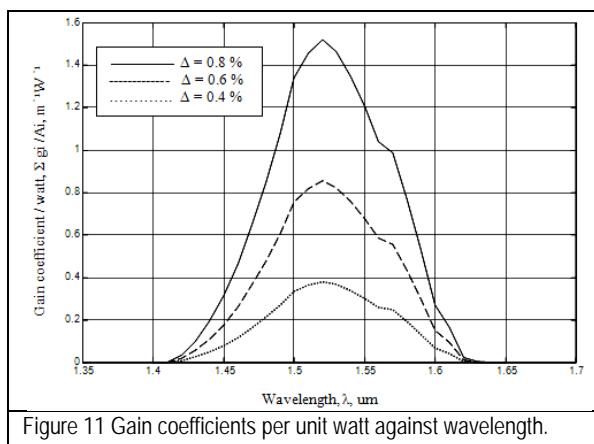


Figure 11 Gain coefficients per unit watt against wavelength.

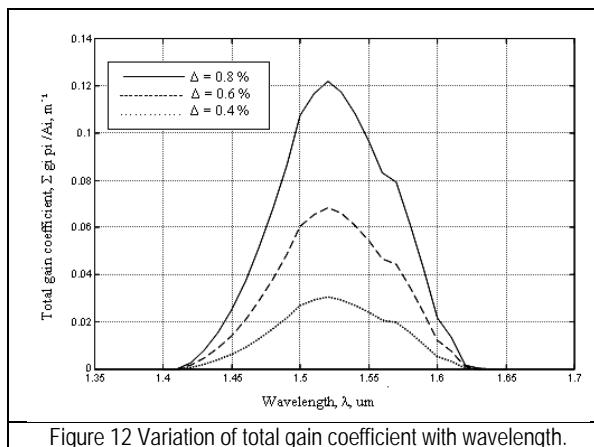


Figure 12 Variation of total gain coefficient with wavelength.

A summary of the obtained results, in different cases, is found in the following comparison table, where one can note that the maximum gain increases with the relative refractive index difference increase. And bandwidth is change according to the change of the position of optical amplifiers.

From table IV we conclude that:

- 1- If number of optical amplifiers increases, Raman gains increase.
- 2- If relative index difference increase then we gets Raman gain is increase.
- 3- Also, for each case only if relative index difference increases, Raman gain is increase.
- 4- Bandwidth and flatness of the gain depends on the position of amplifiers corresponding to each other's and number of amplifiers, where in case 1 the flatness of the gain is better than in case 2 of [1], then for case 1 BW = 110 nm for different number of optical amplifiers but the flatness of the gain for N = 8 is better than in case of N= 5, also for case 2 [1] the bandwidth is change according to the number of optical amplifiers and the position of each amplifier corresponding to each others, for N = 4 , BW = 100 nm and for N = 6, BW = 110 nm and 130 nm.

The we get case 1 is better than case 2 [1], where the flatness of the gain is improved.

Table IV Maximum gain and bandwidth for different number of amplifiers.

Case 1			
No of optical amplifiers	g_{\max} (m/W)	$\Delta \%$	BW(nm)
5	3.1308×10^{-13}	0.8	
	2.3481×10^{-13}	0.6	
	1.5654×10^{-13}	0.4	110
8	5.0577×10^{-13}	0.8	
	3.7933×10^{-13}	0.6	
	2.5289×10^{-13}	0.4	110
10	5.8978×10^{-13}	0.8	
	4.4234×10^{-13}	0.6	
	2.9489×10^{-13}	0.4	110
Case 2 [1]			
No of optical amplifiers	g_{\max} (m/W)	$\Delta \%$	BW(nm)
4	2.4326×10^{-13}	0.8	
	1.8245×10^{-13}	0.6	
	1.2163×10^{-13}	0.4	100
6	3.6239×10^{-13}	0.8	
	2.7179×10^{-13}	0.6	
	1.8119×10^{-13}	0.4	110
6	3.2401×10^{-13}	0.8	
	2.4301×10^{-13}	0.6	
	1.6201×10^{-13}	0.4	130

4. Conclusions

The bandwidth of cascaded multi-pumping Raman amplifier is investigated, where N Raman pumping signals are injected in a parallel processing at different pumping powers wavelengths. The differential gain of each pumping is according to the straight line-exponential model of a small maximum constant gain of 7.4×10^{-14} m/W over an optical wavelength interval of 16 nm. The processed gains are functions of the set of variables $\{\lambda_s, \lambda_r, \Delta\}$ and the locations of the maximum constant gain interval}. We have obtained bandwidth of about, 110 nm at different value of Δ % for use 5, 8 and 10 optical Raman amplifiers, respectively.

5. References

- [1] F. M. Mustafa ""Distributed Multi-Raman Amplifier for Long-Haul UW-WDM Optical Communications Systems" Bulletin of Journal of Al Azher University Engineering Sector, Nasr City 11371, Cairo, Egypt, 2011.
- [2] M. N. Islam, "Raman Amplifiers for Telecommunications," IEEE J. Selected Topics in Quantum Electron., Vol. 8, No. 3, pp.548-559, 2002.
- [3] M. D. Mermelstein, K. Brar, and C. Headly, "RIN Transfer Measurement and Modeling in Dual-Order Raman Fiber amplifiers," J. Lightwave Technol., Vol. 21, No. 6, p. 1518, 2003.
- [4] J. Bromage, "Raman Amplification for Fiber Communications Systems," J. Lightwave Technol., Vol. 22, No. 1, pp. 79-93, 2004.
- [5] C. Headley, G. P. Agrawal "Raman Amplification in Fiber Optical Communication Systems", Elsevier Inc. 2005.
- [6] P. Xiao, O. Zeng, J. Huang, and J. Liu, "A New Optimal Algorithm for Multi-Pumping Sources of Distributed Fiber Raman Amplifier," IEEE Photonics Technol. Lett., Vol. 15, No.2, pp. 206-208, 2003.
- [7] X. Liu and B. Lee, "A Fast Stable Method for Raman Amplifier Propagation Equation," Optics Express, Vol. 11, No. 18, pp. 2163-2176, 2003.
- [8] N. Kikuchi, "Novel In-Service Wavelength-Based Upgrade Scheme for Fiber Raman Amplifier," IEEE Photonics Technol. Lett., Vol. 15, No. 1, pp. 27-29, 2003.
- [9] Y. Cao and M. Raja, "Gain-Flattened Ultra-Wideband Fiber Amplifiers," Opt. Eng., Vol. 42, No. 12, pp. 4447-4451, 2003.
- [10] M. S. Kao and J. Wu, "Signal Light Amplification by Stimulated Raman Scattering in an N-Channel WDM Optical Communication System", J. Lightwave Technol., Vol.7, No. 9, pp. 1290-1299, 1989.
- [11] T. Nakashima, S. Seikai, N. Nakazawa, and Y. Negishi, "Theoretical Limit of Repeater Spacing in Optical Transmission Line Utilizing Raman Amplification," J. Lightwave Technol., Vol. LT-4, No. 8, pp. 1267-1272, 1986.
- [12] Y. Aoki, "Properties of Fiber Raman Amplifiers and Their Applicability to Digital Optical Communication Systems," J. Lightwave Technol., Vol. 6 No. 7, pp. 1227-1239, 1988.
- [13] W. Jiang and P. Ye., "Crosstalk in Raman Amplification for WDM Systems," J. Lightwave Technol., Vol. 7, No. 9, pp. 1407-1411, 1989.
- [14] A. A. Mohammed, "All Broadband Raman Amplifiers for Long-Haul UW-WDM Optical Communication Systems," Bulletin of Faculty of Electronic Engineering, Menouf, 32951, Egypt, 2004.
- [15] A. Yariv, Optical Electronics in Modern Communications, 5th ed., Oxford Univ. Press, 1997.

Fathy M. Mustafa received the B.Sc. degree in Electronics and communications department with honors from the Faculty of Engineering, Fayoum University, Fayoum, Egypt, in 2003. He is currently working a research assistant in Electronics and communications department at Bani-suef University. He is earned the M.Sc degree in Electronics and communication engineering in 2007 from Arab Academy for Science and Technology & Maritime Transport College of Engineering and Technology, Alexandria, Egypt. He is joined the PhD program in Mina university in 2011. Her areas of interest include optical communications, optical amplifiers.

Ashraf A. Khalaf: received his B.Sc. and M.Sc. degrees in electrical engineering from Minia university, Egypt, in 1989 and 1994 respectively. He received his Ph.D in electrical engineering from Graduate School of Natural Science and Technology, Kanazawa university, Japan in 2000. He works at electronics and communications engineering Department, Minia University. He is a member of IEEE since 12 years.

F. A. El-Geldawy: is a professor Electric Engineering Dept., faculty of engineering, Minia, Egypt.

Fuzzy Priority CPU Scheduling Algorithm

Bashir Alam¹, M.N. Doja¹, R. Biswas³, M. Alam⁴

¹**Department of Computer Engineering, Jamia Millia Islamia**
New Delhi-110025, India

²**Department of Computer Engineering, Jamia Millia Islamia**
New Delhi-110025, India

³**Department of Computer Science , Jamia Hamdard**
New Delhi-110062, India

⁴**Department of computer Science, Jamia Millia Islamia,**
New Delhi-110025, India

Abstract

There are several CPU scheduling algorithms like FCFS, SRTN,RR , priority etc. Scheduling decision of these algorithms are based on parameters which are assumed to be crisp. However, in many circumstances these parameters are vague. The vagueness of these parameters suggests that scheduler should use fuzzy logic in scheduling the jobs. A fuzzy priority CPU scheduling algorithm has been proposed. This proposed algorithm improves the priority based CPU scheduling algorithm as obvious from simulation results.

Keywords:- FIS, Priority CPU Scheduling, Fuzzy Logic

1. Introduction

When a computer is multiprogrammed , it frequently has multiple processes competing for the CPU at the same time. When more than one process is in the ready state and there is only one CPU available, the operating system must decide which process to run first. The part of operating system that makes the choice is called short term scheduler or CPU scheduler. The algorithm that it uses is called scheduling algorithm. There are several scheduling algorithms. Different scheduling algorithms have different properties and the choice of a particular algorithm may favor one class of processes over another. Many criteria have been suggested for

comparing CPU scheduling algorithms and deciding which one is the best algorithm[1]. Some of the criteria include (i)Fairness(ii)CPU utilization(iii)Throughput(iv)Turnaround time(v)Waiting time(vi)Response time

It is desirable to maximize CPU utilization and throughput, to minimize turnaround time, waiting time and response time and to avoid starvation of any process.[1,2] Some of the scheduling algorithms are briefly described here. FCFS: In First come First serve scheduling algorithm the process that request first is scheduled for execution[1,2,3]SJF: In shortest Job first scheduling algorithm the process with the minimum burst time is scheduled for execution.[1,2]SRTN: In shortest Remaining time next scheduling algorithm , the process with shortest remaining time is scheduled for execution.[3]Priority: in Priority Scheduling algorithm the process with highest priority is scheduled for execution. Round-robin: In this the CPU scheduler goes around the ready queue allocating the CPU to each process for a time interval of up to one time quantum. [1,2,3]Multilevel queue scheduling: In this the ready queue is partitioned into several separate queue. The processes are permanently assigned to one queue generally based on some property of the process such as memory size, process priority or process type. Each queue has its own scheduling algorithm. There is scheduling among the queues, which is commonly implemented as fixed-priority preemptive scheduling . Each queue has absolute priority over low priority queues.[1]Multilevel feedback-queue scheduling:-

This allows a process to move between queues.[1] Fair share Scheduling: Fair share scheduler considers the execution history of a related group of processes, along with the individual execution history of each process in making scheduling decision. The user community is divided into a fair-share groups. Each group is allocated a fraction of CPU time. Scheduling is done on the basis of priority of the process, Its recent processor usage and the recent processor usages of the group to which the process belongs. Each process is assigned a base priority. The priority of a process drops as the process uses the processor and as the group to which process belongs uses the processor.[3]Guaranteed scheduling:-In this a ratio of actual CPU time a process had and its entitled CPU time is calculated. The process with this lowest ratio is scheduled[2] Lottery Scheduling:-The basic idea is to give processes lottery tickets for CPU time. Whenever a scheduling decision has to be made, a lottery ticket is chosen at random and the process holding the ticket gets the CPU.[2] HRRN :-In this response ration is calculated for each process. The process with the highest ratio is scheduled for execution. [3] In all the these scheduling algorithms the parameters used are crisp. However, in many circumstances these parameters are vague. To exploit these vagueness we have used fuzzy logic in our proposed scheduling algorithm.

2. Fuzzy Inference Systems and Fuzzy Logic

A fuzzy inference system (FIS) tries to derive answers from a knowledgebase by using a fuzzy inference engine. The inference engine which is considered to be the brain of the expert systems provides the methodologies for reasoning around the information in the knowledgebase and formulating the results. Fuzzy logic is an extension of Boolean logic dealing with the concept of partial truth that denotes the extent to which a proposition is true. Whereas classical logic holds that everything can be expressed in binary terms (0 or 1, black or white, yes or no), fuzzy logic replaces Boolean truth values with the degree of truth. Degree of truth is often employed to capture the imprecise modes of reasoning that play an essential role in the human ability to make decisions in an environment of uncertainty and imprecision. The membership function of a fuzzy set corresponds to the indicator function of the classical sets. It can be expressed in the form of a curve that defines how each point in the input space is mapped to a membership value or a degree of truth between 0 and 1. The most common shape of a membership function is triangular, although trapezoidal and bell

curves are also used. The input space is sometimes referred to as the universe of discourse [4]. Fuzzy Inference Systems are conceptually very simple. An FIS consists of an input stage, a processing stage, and an output stage. The input stage maps the inputs, such as deadline, execution time, and so on, to the appropriate membership functions and truth values. The processing stage invokes each appropriate rule and generates a result for each. It then combines the results of the rules. Finally, the output stage converts the combined result back into a specific output value [4]. The processing stage, which is called the inference engine, is based on a collection of logic rules in the form of IF-THEN statements, where the IF part is called the "antecedent" and the THEN part is called the "consequent". Typical fuzzy inference subsystems have dozens of rules. These rules are stored in a knowledgebase. An example of fuzzy IF-THEN rules is: IF *Remaining Time* is *Extremely short* then *priority* is *very high*, in which *Remaining Time* and *priority* are linguistics variables and *Extremely short* and *very high* are linguistics terms. The five steps toward a fuzzy inference are (i) fuzzifying inputs(ii) applying fuzzy operators(iii) applying implication methods(iv) aggregating outputs(v)defuzzifying results

Bellow is a quick review of these steps. However, a detailed study is not in the scope of this paper. Fuzzifying the inputs is the act of determining the degree to which they belong to each of the appropriate fuzzy sets via membership functions. once the inputs have been fuzzified, the degree to which each part of the antecedent has been satisfied for each rule is known. If the antecedent of a given rule has more than one part, the fuzzy operator is applied to obtain one value that represents the result of the antecedent for that rule. The implication function then modifies that output fuzzy set to the degree specified by the antecedent. Since decisions are based on the testing of all of the rules in the Fuzzy Inference Subsystem (FIS), the results from each rule must be combined in order to make the final decision. Aggregation is the process by which the fuzzy sets that represent the outputs of each rule are processes into a single fuzzy set. The input for the defuzzification process is the aggregated output fuzzy set and the output is then a single crisp value [4]. This can be summarized as follows: mapping input characteristics to input membership functions, input membership function to rules, rules to a set of output characteristics, output characteristics to output membership functions, and the output membership function to a single crisp valued output. There are two common inference methods [4]. The first one is called Mamdani's fuzzy inference method proposed in 1975 by Ebrahim Mamdani [5] and the second one

is Takagi-Sugeno-Kang, or simply Sugeno, method of fuzzy inference introduced in 1985 [6]. These two methods are the same in many respects, such as the procedure of fuzzifying the inputs and fuzzy operators. The main difference between Mamdani and Sugeno is that the Sugeno's output membership functions are either linear or constant but Mamdani's inference expects the output membership functions to be fuzzy sets. Sugeno's method has three advantages. Firstly, it is computationally efficient, which is an essential benefit to real-time systems. Secondly, it works well with optimization and adaptive techniques. These adaptive techniques provide a method for the fuzzy modeling procedure to extract proper knowledge about a data set, in order to compute the membership function parameters that best allow the associated fuzzy inference system to track the given input/output data. The third, advantage of Sugeno type inference is that it is well-suited to mathematical analysis.

3. The Proposed Model

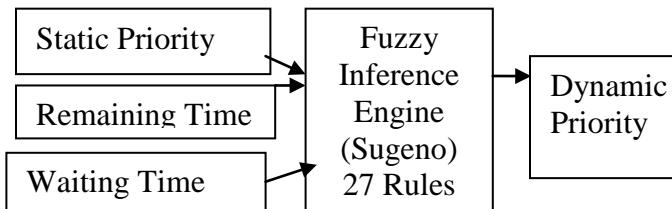


Figure 1: Block diagram of Fuzzy Inference System

The block diagram of the proposed fuzzy inference system is given in figure1. In the proposed model, the input stage consists of three linguistic variables. The first one is the static priority that is assigned to the process before its execution. The second is the expected remaining time of the process. The third input is the waiting time of the process. The output stage consists of one linguistic variable called Dynamic priority.

The input and out variables are mapped into fuzzy sets using appropriate membership functions. Membership functions are given below

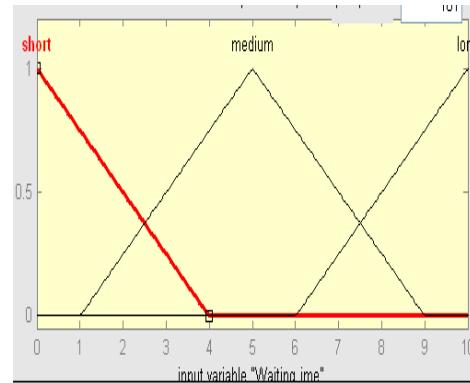


Figure 2: Membership Function for Waiting Time

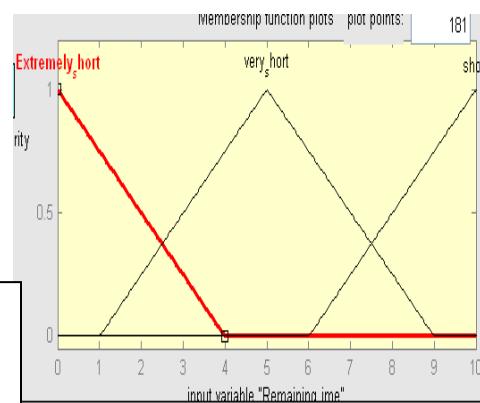


Figure 3 : Membership Function for remaining Time

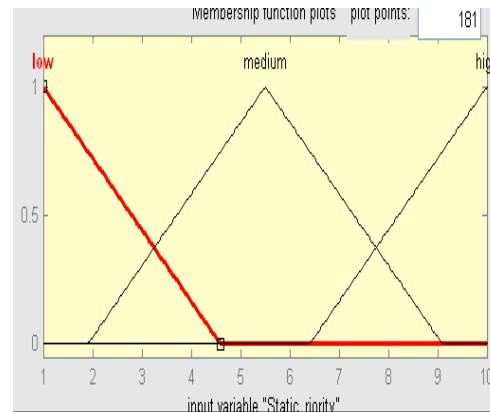


Figure 4: Membership Function for Static Priority

The shape of the membership function for each linguistic term is determined by the expert. Adjusting these membership functions in an optimal mode is very difficult. However, there are some techniques for adjusting membership functions [6, 8].

Twenty seven rules are formulated and a Sugeno type fuzzy Inference system is built. Some of the rules are

listed here : if the static priority is ‘low’ and remaining time is ‘extremely short’ and waiting time is ‘long’ then the dynamic priority is ‘very high’. if the static priority is ‘low’ and remaining time is ‘short’ and waiting time is ‘short’ then the dynamic priority is ‘very low’. if the static priority is ‘medium’ and remaining time is ‘extremely short’ and waiting time is ‘long’ then the dynamic priority is ‘very high’. if the static priority is ‘medium’ and remaining time is ‘short’ and waiting time is ‘short’ then the dynamic priority is ‘medium’.

Table 1: Rule base for Fuzzy Inference System

S.No.	Static Priority	Remaining Time	Waiting Time	Dynamic Priority
1.	Low	Extremely short	Short	Very High
2.	Low	Extremely short	Medium	Very High
3.	Low	Extremely short	Long	Very High
4.	Low	Very short	Short	Very low
5.	Low	Very short	Medium	Low
6.	Low	Very short	Long	High
7.	Low	Short	Short	Very low
8.	Low	Short	Medium	Low
9.	Low	Short	Long	High
10.	Medium	Extremely short	Short	Very high
11.	Medium	Extremely short	Medium	Very high
12.	Medium	Extremely short	Long	Very high
13.	Medium	Very short	Short	Medium
14.	Medium	Very short	Medium	Medium
15.	Medium	Very short	Long	Very High
16.	Medium	Short	Short	Medium
17.	Medium	Short	Medium	Medium
18.	Medium	Short	Long	High
19.	High	Extremely short	Short	Very high
20.	High	Extremely short	Medium	Very high
21.	High	Extremely short	Long	Very high
22.	High	Very short	Short	High
23.	High	Very short	Medium	High
24.	High	Very short	Long	Very High
25.	High	Short	Short	High
26.	High	Short	Medium	High
27.	High	Short	Long	Very High

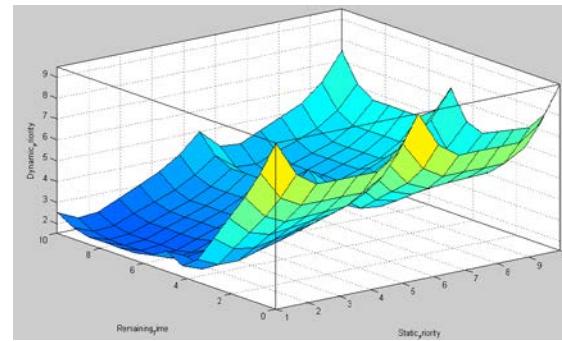


Figure 5: Surface view of FIS for Priority scheduling

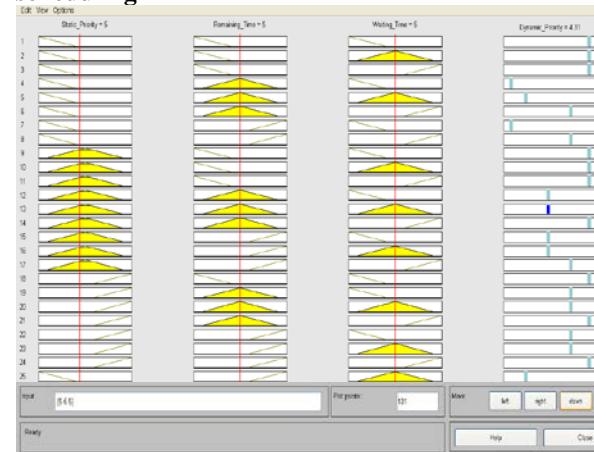


Figure 6: Rule View of FIS for Priority Scheduling

4. Proposed Algorithm

The parameters of process are stored in table called Process Control Block (PCB). Each process has its own PCB. The structure of the Process Control Block is given in Table II.

Table 2: Structure of Process control Block for priority scheduling

Process
Process name=“bash”
Process identifier=100
State=Ready
CPU reserve
{CPU Burst Time=
CPU Remaining Time=
Priority=
Waiting time=
Arrival Time=
Start time=
...
....
}

The parameters remaining time Rt_i , static priority sp_i , dynamic priority dp_i and waiting time wt_i of process P_i are stored in Process Control Block PCB_i . The proposed algorithm is as follows:

- i. For each process P_i in ready queue fetch its parameters Rt_i , sp_i , and wt_i from PCB_i and give them as input to FIS and then set dp_i to the output of FIS
- ii. Schedule the process P_i with the highest value of dp_i for execution.
- iii. If the scheduled process finishes and no new request arrives go to step ii
- iv. If new request arrives go to step first .

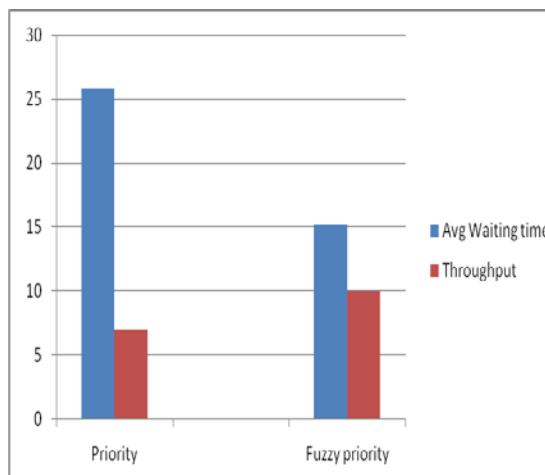


Figure 7 : Performance of Priority and Fuzzy Priority Scheduling Algorithms

4. Performance

For comparing the performance of Priority CPU Scheduling Algorithm and Fuzzy Priority CPU Scheduling algorithm, we did simulation on 1000 processes in groups of ten each. We assumed random burst time of processes and random arrivals. Max burst time of a process should not exceed 10 ms. Throughput and average waiting time of the processes in a group was computed and then average was taken over all groups to give average throughput and average waiting time. The performance is shown in the column chart given in figure 7 above.

6. Conclusion

We have proposed fuzzy priority CPU scheduling algorithm based on FIS. This algorithm is having benefits of shortest remaining time next (SRTN) as well as Priority scheduling algorithm and is capable of removing the starvation problem of priority scheduling algorithm. This proposed algorithm also

improves the system performance by not pre empting the process that requires extremely less fraction of CPU time. Fuzzy priority scheduling algorithm gives better throughput and lesser waiting time than traditional Priority Scheduling as obvious from the figure 7.

References

- [1] Silberschatz, A., Peterson, J. L., and Galvin, B., *Operating System Concepts*, Addison Wesley, 7th Edition, 2006.
- [2] Andrew S. Tanenbaum , and Albert S. Woodhull , Operating Systems Design and Implementation, Second Edition, 2005
- [3] William Stallings, Operating Systems Internal and Design Principles, 5th Edition ,2006
- [4] Wang Lie-Xin, A course in fuzzy systems and control, Prentice Hall, August 1996.
- [5] Mamdani E.H., Assilian S., An experiment in linguistic synthesis with a fuzzy logic controller, International Journal of Man-Machine Studies, Vol.7, No. 1, 1975.
- [6] Sugeno, M., Industrial applications of fuzzy control, Elsevier Science Inc., New York, NY, 1985.
- [7] Jang, J.-S. R., ANFIS: Adaptive-Network-based Fuzzy Inference Systems, IEEE Transactions on Systems, Man, and Cybernetics, Vol. 23(3), 685, May 1993.
- [8] Simon D, Training fuzzy systems with the extended Kalman filter, Fuzzy Sets and Systems, Vol. 132, No. 2, 1, December 2002.

Dynamic Gesture Recognition Using Hidden Markov Model in Static Background

Malvika Bansal¹, Shivin Saxena², Devendra Desale³ and Dnyaneshwar Jadhav⁴

¹ Department of Computer Engineering, MES College Of Engineering
Pune University, Maharashtra, India

² Department of Computer Engineering, MES College Of Engineering
Pune University, Maharashtra, India

³ Department of Computer Engineering, MES College Of Engineering
Pune University, Maharashtra, India

⁴ Department of Computer Engineering, MES College Of Engineering
Pune University, Maharashtra, India

Abstract

Human Computer Interaction is a challenging endeavour. Being able to communicate with your computer (or robot) just as we humans interact with one another has been the prime objective of HCI research since the last two decades. A number of devices have been invented, each bringing with it a new aspect of interaction. Much work has gone into Speech and Gesture Recognition to develop an approach that would allow users to interact with their system by simple using their voice or simple intuitive gestures as against sitting in front of the computer and using a mouse or keyboard. Natural Interaction must be fast, convenient and reliable. In our project, we intend to develop one such natural interaction interface, one that can recognize hand gesture movements in real time using HMM but by using Computer Vision instead of sensory gloves.

Keywords: HCI, Dynamic, Gesture Recognition, Skin colour detection, Computer Vision, HMM

1. Introduction

Technology has come a long way from computers running on vacuum tubes to semi-conductor chips and finally, the super computers of today's era. In the earlier stages of computing inventions, a keyboard was the only means of communication to interact with the computer (after magnetic tapes and punch cards became obsolete). The mouse brought with it a revolution and an entirely new dimension to Human Computer Interaction, or popularly abbreviated as HCI. Many inventions followed such as the light pen, tablets, digitizers and more recently the "Space Mouse" each bringing with it an innovative style of interacting with the computer and opening new possibilities and dimensions of interaction.

Thus, Human Computer Interaction is not just a necessity but a challenging endeavour to push current boundaries of interaction. Speech Recognition and Synthesis has been a prominent domain of research over the last decade. Even more prominent has been interaction by the use of simple intuitive Hand Gestures. Sign language is a common form of communication between auditory handicapped people. It can further be employed to communicate with a robot or any computer. Imagine sitting on the couch and operation your computer from a distance with your voice or just simple day-to-day hand movements. It would not only eliminate the need to actually physically touch your mouse and keyboard unless absolutely necessary, but will also be so much more convenient and quick. This is the power of Human Computer Interaction.

Both static and dynamic hand gesture recognition is a challenging aspect of HCI. Gesture Recognition can be implemented by one of the two methods, one, by wearing sensory gloves on one's hand or second, by with the aid of Computer Vision (CV). For glove based techniques, it mainly utilizes sensory gloves to measure the angles and spatial positions of a hand and fingers. Referring to one of the papers based on this field, we came across an approach in gesture recognition that used a sensing glove with 6 embedded accelerometers. It could recognize 28 static hand gestures and the computation time was about 1 characters/second. However, the proposed algorithm was not efficient enough to be applied in real time. Although the former is a powerful technique it's not really a natural way to interact with the computer because one has to continually wear the gloves. Natural HCI should be glove free, fast, reliable and convenient. A sensory glove would not be a convenient option for daily usage.

For Computer Vision based techniques, one or a set of cameras are used to capture images for hand gesture recognition. Using backward reference from a paper based on the use of Haar Wavelet for recognizing gestures, we came across another proposed algorithm that would first separate the hand region from the complex background images by measuring entropy from adjacent frames. Hand gestures would then be recognized by the approach of improved centroidal profile. However, mis-recognitions can be caused by hand gestures with similar spatial features and therefore the number of gestures that could be recognized was limited.

Referring to [1], the paper based on the use of Haar Wavelet Representation for gesture recognition, we came across another approach for an effective computer vision system. The proposed approach was based on skin-colour detection. Hands were extracted by detecting the skin colour. The problem of hand orientation in the image is also solved by utilizing the idea of axis of elongation. It helped immensely in keeping the database small by standardizing the hand gestures in the database using fixed orientations. To facilitate the searching process, a codeword scheme mechanism was utilized. To further improve the success rate the use of a measurement metric with a penalty score during recognition was proposed. Experimental results showed a good hit rate for recognition of gestures.

Referring to [2], we then came across an entirely different approach to gesture recognition that used Hidden Markov Models. In this paper, a Graphic Editor that could recognize five static and twelve dynamic gestures, was being developed. Recognition was facilitated by the using a structural analysis for static gestures and HMM for dynamic ones. According to [2], Hidden Markov models have intrinsic properties which make them an attractive option for gesture recognition, and also explicit segmentation is not necessary for either training or recognition.

Further, referring to [3], Jinli and Tianding propose a thesis for hand trajectory recognition based on HMM, that can model spatio-temporal information in a natural way. In order to be able to differentiate undefined gestures, a modified threshold model was proposed. The hand was separated from its background by the use of skin colour detection by first converting the RGB based pixels to YCbCr colour model and then defining a suitable range to recognize skin colour.

In our final year project, we intend to propose an approach for dynamic hand gesture recognition using HMM model against a static background. The concept is

to develop an application that would recognize some defined gestures and perform associated tasks, such as opening and closing some application, zooming in and out of an image, rotating it and even printing the image. Following are the gestures that we will be working with:

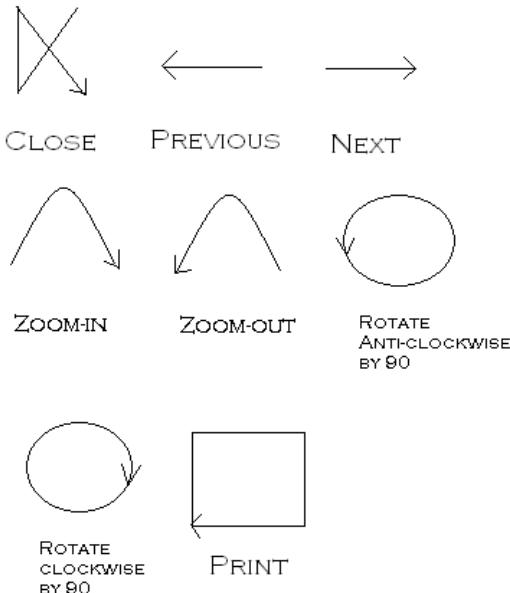


Fig. 1 Various Gestures used in the System

In addition to the HMM model, which will be used to represent the gestures in an adjacency matrix, we will also be making use of the idea of principal axis through the centroid of the hand to standardize the gestures, thereby reducing the size of the image dataset. We will be working with the HSV colour model.

The structure of this paper is as follows: Section II gives the System Overview, Section III describes Segmentation based on Skin Colour Detection, Section IV demonstrated use of HMM for hand gestures, Section V is the implementation and finally Section VI will be the Conclusion.

2. System Overview

The following block diagram summarizes the approach that we will be using to implement our system:-

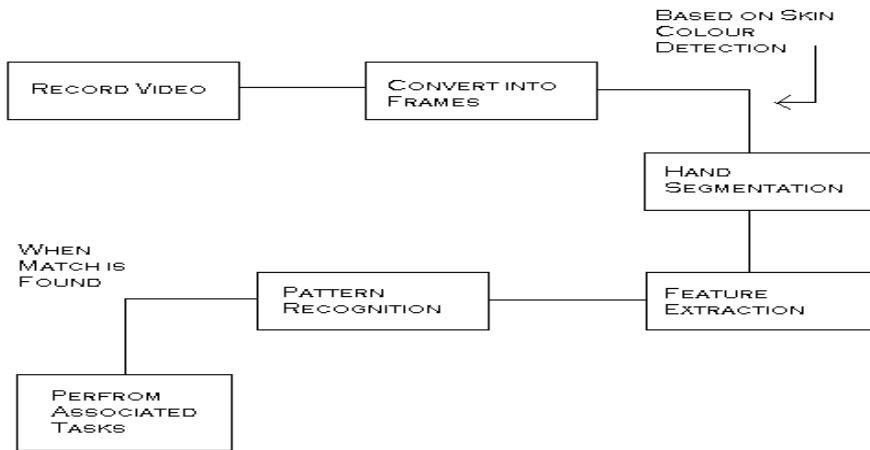


Fig. 2 Block Diagram for the System

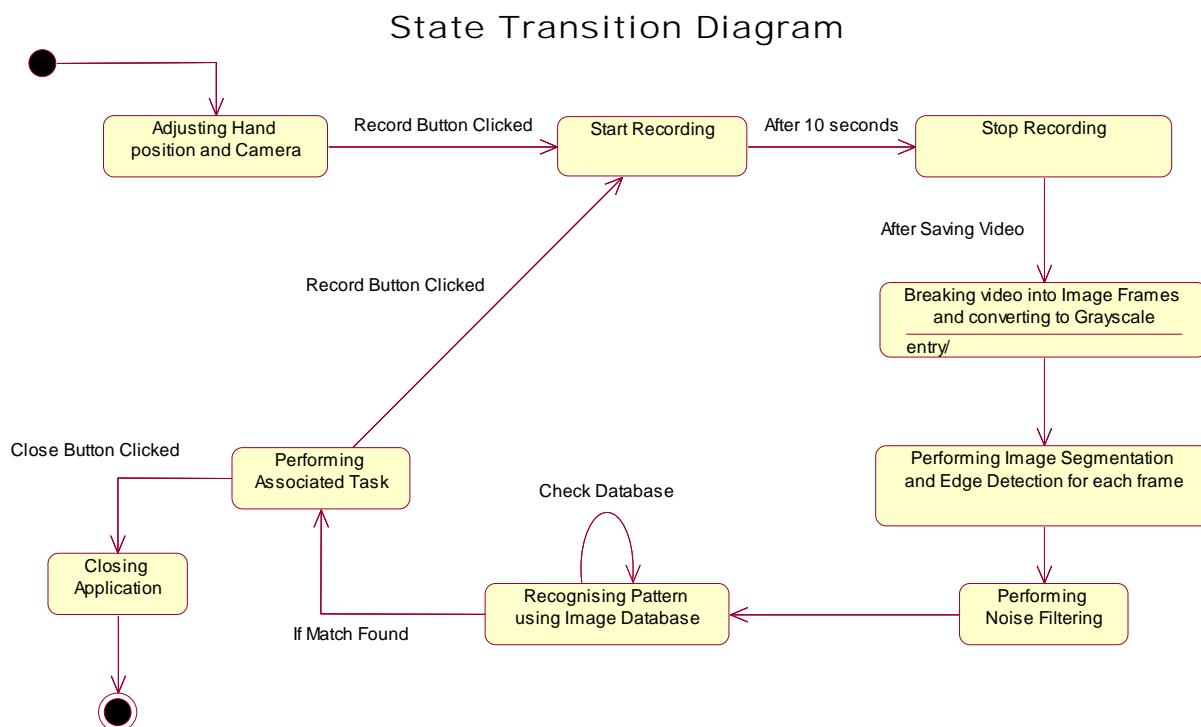


Fig. 3 State Transition Diagram for the System

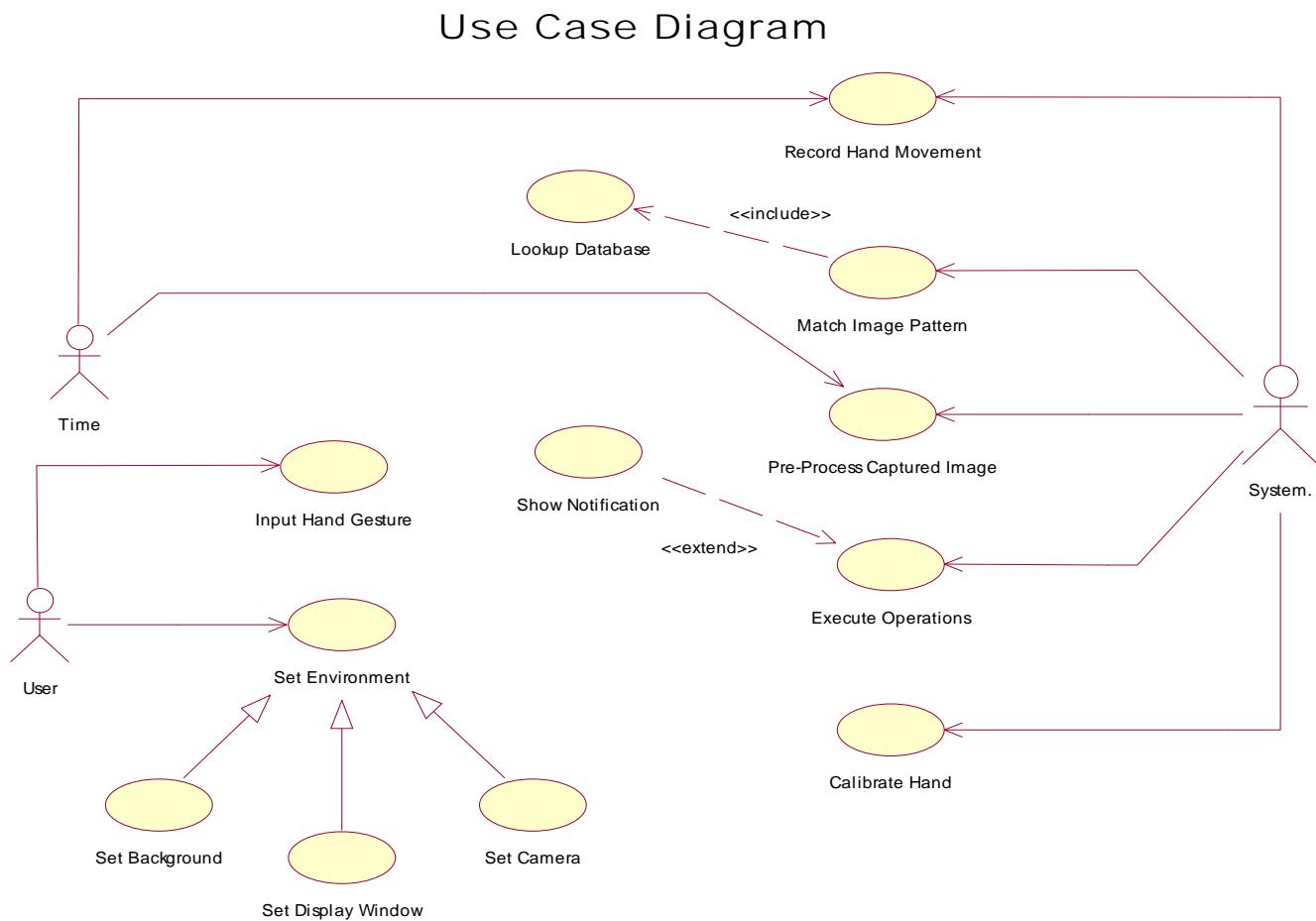


Fig. 4 Use Case Diagram for the System

3. Hand Segmentation using Skin Detection

The RGB colour model includes the information of both colour and brightness, which is vulnerable to the changes in background illumination and environment. To ensure these doesn't affect the skin colour, we will be converting the RGB colour model to HSV colour model, because the latter is more related to human colour perception. The skin in channel H is characterised by values between 0 to 50, in the channel S between 0.23 to 0.68 for Asian and Caucasian skin. The figure below shows the original image in RGB colour format.



Fig. 5 Original Image

After changing the pixel colour values, using RGB to HSV conversion, we get the following image:



Fig. 6 Image after conversion from RGB to HSV

The next image is an intermediate image which is obtained immediately after setting all pixels that fall in our skin colour range to 255(white) and the rest non-skin pixels to 0 (black). However, this image has some noises in it i.e. some pixels that are actually not part of the skin but still have fallen in the given range and must be eliminated. This is accomplished by the help of morphological filters.

The final image obtained after noise removal is shown as under:



Fig. 7 Final Image after Noise Removal

Finally only the skin regions are represented as white pixels. To convert from RGB to HSV (assuming normalized RGB values) first find the maximum and minimum values from the RGB triplet.

Saturation, S, is then given by:

$$S = \frac{\text{max} - \text{min}}{\text{Max}} \quad \text{Eq. (1)}$$

and Value, V, is given by:

$$V = \text{max}$$

The Hue, H, is then calculated as follows. First calculate R'G'B':

$$R' = (\text{max}-R) / (\text{max-min})$$

$$G' = (\text{max}-G) / (\text{max-min})$$

$$B' = \text{max}-B / (\text{max-min}) \quad \text{Eq. (2)}$$

If saturation, S, is 0 (zero) then hue is undefined (i.e. the colour has no hue therefore it is monochrome) otherwise: then,

$$\text{if } R = \text{max} \text{ and } G = \text{min}$$

$$H = 5 + B' \quad \text{Eq. (3)}$$

$$\text{else if } R = \text{max} \text{ and } G = \text{min}$$

$$H = 1 - G' \quad \text{Eq. (4)}$$

$$\text{else if } G = \text{max} \text{ and } B = \text{min}$$

$$H = R' + 1 \quad \text{Eq. (5)}$$

$$\text{else if } G = \text{max} \text{ and } B = \text{min}$$

$$H = 3 - B' \quad \text{Eq. (6)}$$

$$\text{else if } R = \text{max}$$

$$H = 3 + G' \quad \text{Eq. (7)}$$

$$\text{otherwise}$$

$$H = 5 - R' \quad \text{Eq. (8)}$$

Hue, H, is then converted to degrees by multiplying by 60 giving HSV with S and V between 0 and 1 and H between 0 and 360.

4. The Markov Model

We shall first give a brief introduction to the Hidden Markov Model. Consider a person who is sitting inside a

room and has three coins. He is tossing these coins in sequence known only to him. We are positioned outside this room and are shown by means of a display (also placed outside the room), the outcomes of the man flipping the coins, for example, HTHTHTTHHTHHTHH. This is referred called the Observation Sequence. We do not know the sequence in which the coins are being tossed and neither do we know the bias of the individual coins. To understand the significance of the impact of the bias of the coins on the outcome, imagine that the third coin is highly biased to generate Tails. Now, if all the coins are tossed with equal probability then it would be naturally expected that the output will have more Tails than Heads. Further, consider that the probability of moving from the first or second coin (state) to the third coin (state) is zero. Now if we had started the tossing with the first and second coins then the output sequence will generate more tails because of the transition probability between the coins/states and also, the initial state. These three sets, namely, the set of individual bias of the three coins, the set of transition probabilities from one coin to the next and the set of initial probabilities characterize what is called as the Hidden Markov Model.

The HMM Model is specified by:

- The set of states $S = \{s_1, s_2, \dots, s_N\}$, (corresponding to the N possible gesture conditions above),

And a set of parameters: $\lambda = \{\Pi, A, B\}$

- Transition probabilities
 $A = \{a_{ij} = P(q_j \text{ at } t+1 | q_i \text{ at } t)\}$,
- where $P(a | b)$ is the conditional probability of a given b , $t = 1, \dots, T$ is time, and q_i in \mathbf{Q} . Informally, A is the probability that the next state is q_j given that the current state is q_i .
- Observations (symbols) $\mathbf{O} = \{O_k\}, k = 1, \dots, M$.
- Emission probabilities B is:
 $B = \{b_{ik} = b_i(O_k) = P(O_k | q_i)\}$, where o_k in \mathbf{O} . Informally, B is the probability that the output is o_k given that the current state is q_i .
- Initial state probabilities $\Pi = \{p_i = P(q_i \text{ at } t=1)\}$.

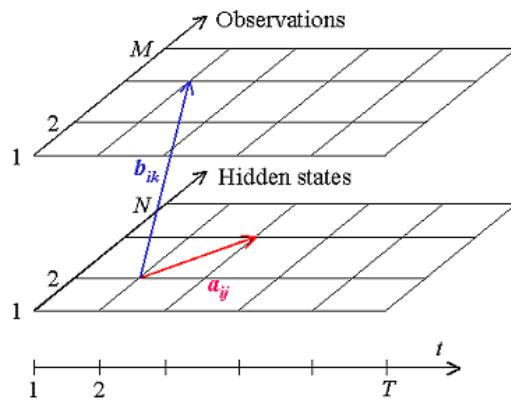


Fig. 8 Observed and Hidden Sequences

A HMM allowing for transitions from any emitting state to any other emitting state is called an Ergodic HMM. The other extreme, a HMM where the transitions only go from one state to itself or to a unique follower is called a left-right HMM.

There are 3 canonical problems to solve with HMMs:

1. Given the model parameters, compute the probability of a particular output sequence. This problem is solved by the Forward and Backward algorithms (described below).
2. Given the model parameters, find the most likely sequence of (hidden) states which could have generated a given output sequence. Solved by the Viterbi algorithm and Posterior decoding.
3. Given an output sequence, find the most likely set of state transition and output probabilities. Solved by the Baum-Welch algorithm.

4.1 Forward Algorithm

Let $\alpha_t(i)$ be the probability of the partial observation sequence $O_t = \{o(1), o(2), \dots, o(t)\}$ to be produced by all possible state sequences that end at the i -th state.

$$\alpha_t(i) = P(o(1), o(2), \dots, o(t) | q(t) = q_i) \quad \text{Eq. (9)}$$

Then the unconditional probability of the partial observation sequence is the sum of $\alpha_t(i)$ over all N states.

The Forward Algorithm is a recursive algorithm for calculating $\alpha_t(i)$ for the observation sequence of increasing length t . First, the probabilities for the single-symbol sequence are calculated as a product of initial i -th state probability and emission probability of the given symbol $o(1)$ in the i -th state. Then the recursive formula is applied. Assume we have calculated $\alpha_t(i)$ for some t . To calculate $\alpha_{t+1}(j)$, we multiply every $\alpha_t(i)$ by the corresponding transition probability from the i -th state to the j -th state, sum the products over all states, and then multiply the result by the emission probability of the symbol $o(t+1)$. Iterating the process, we can eventually calculate $\alpha_T(i)$, and then summing them over all states, we can obtain the required probability.

In a similar manner, we can introduce a symmetrical backward variable $\beta_t(i)$ as the conditional probability of the partial observation sequence from $o(t+1)$ to the end to be produced by all state sequences that start at i -th state (3.13).

$$\beta_t(i) = P(o(t+1), o(t+2), \dots, o(T) | q(t) = q_i) \quad \text{Eq. (10)}$$

The Backward Algorithm calculates recursively backward variables going backward along the observation sequence. The Forward Algorithm is typically used for calculating the probability of an observation sequence to be emitted by an HMM, but, as we shall see later, both procedures are heavily used for finding the optimal state sequence and estimating the HMM parameters.

4.2 Viterbi algorithm

The Viterbi algorithm chooses the best state sequence that maximizes the likelihood of the state sequence for the given observation sequence.

Let $\delta_t(i)$ be the maximal probability of state sequences of the length t that end in state i and produce the t first observations for the given model.

$$\delta_t(i) = \max \{P(q(1), q(2), \dots, q(t-1); o(1), o(2), \dots, o(t) | q(t) = q_i)\} \quad \text{Eq. (11)}$$

The Viterbi algorithm is a dynamic programming algorithm that uses the same schema as the Forward algorithm except for e

1. It uses maximization in place of summation at the recursion and termination steps.
2. It keeps track of the arguments that maximize $\delta_t(i)$ for each t and i , storing them in the N by T matrix ψ . This matrix is used to retrieve the optimal state sequence at the backtracking step.

4.3 Baum-Welch Algorithm

An important aspect of the HMM training is that the model should decode the observation sequence in such a way that if an observation sequence having many characteristics similar to the given one be encountered later it should be able to identify it. There are two methods that can be used for such identification:

- The Segmental K-means Algorithm
- The Baum-Welch Re-estimation Formulas

(We will be using and discussing the second method)

Let us define $\xi_t(i, j)$, the joint probability of being in state q_i at time t and state q_j at time $t+1$, given the model and the observed sequence:

$$\xi_t(i, j) = P(q(t) = q_i, q(t+1) = q_j | O, \Lambda) \quad \text{Eq. (12)}$$

Therefore we get

$$\xi_t(i, j) = \frac{\alpha_t(i) a_{ij} b_j(o(t+1)) \beta_{t+1}(j)}{P(O | \Lambda)} \quad \text{Eq. (13)}$$

The probability of output sequence can be expressed as

$$P(O | \Lambda) = \sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) a_{ij} b_j(o(t+1)) \beta_{t+1}(j) = \sum_{i=1}^N \alpha_t(i) \beta_t(i) \quad \text{Eq. (14)}$$

The probability of being in state q_i at time t :

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j) = \frac{\alpha_t(i) \beta_t(i)}{P(O | \Lambda)} \quad \text{Eq. (15)}$$

5. Conclusion

The Hidden Markov Model serves as an indispensable tool for the recognition of dynamic gestures in real time. Also, standardising the axis through the centroid will greatly reduce the database size. Based on observations from other references the accuracy of our proposed algorithm is expected to be high.

In the future, this approach can be further improved upon by taking into account the effect of speed of movement of hand and also, implement the system for both hands and to able to recognize and make use of the entire forearm.

5.1 Advantages

- The system can be used conveniently to communicate with the computer at a distance and since we are making use of HMM, the accuracy rate is also increased with increased training of the system.
- The idea of standardizing the orientation of images in the database greatly reduces its size and thus facilitates a faster searching process.

5.2 Disadvantages

- There might be miss-recognitions in case the background has elements that resemble the human skin.

- A number of other factors such as velocity of movement, orientation and low background illumination might take a toll on the system's accuracy.

5.3 Acknowledgment

All the authors would like to extend their sincere gratitude and thanks to their project guides who in spite of their hectic schedule were a constant source of guidance and motivation and helped them explore the depths of image processing and computer vision.

6. References

- [1] Wing Kwong Chung, Xinyu Wu, Yangsheng Xu, "A realtime hand gesture recognition based on Haar wavelet representation", Robotics and Biomimetics, 2008. ROBIO 2008. IEEE International Conference, pp. 336 – 341, 22-25 Feb. 2009.
- [2] Byung-Woo Min, Ho-Sub Yoon, Jung Soh, Yun-Mo Yang, Tosakiaki Ejima, "Hand Gesture Recognition Using Hidden Markov Model", pp. 305-333.
- [3] Jinli Zhao and Tianding Chen, "An Approach to Dynamic Gesture Recognition for Real-time Interaction", ISNN 2009, pp. 369-377.
- [4] Shuying Zhao, Wenjun Tan, Chengdong Wu Chunjiang Liu, Shiguang Wen, "A novel interactive method of virtual reality system based on hand gesture recognition", Control and Decision Conference, 2009. CCDC '09. Chinese, pp. 5879 – 5882, 17-19 June 2009.
- [5] Rokade, Doye, Kokare, "Digital Image Processing, 2009 International Conference", pp. 288-291, 7-9 March 2009.
- [6] Rokade, Doye, Kokare, "Digital Image Processing, 2009 International Conference", pp. 284-287, 7-9 March 2009.

Image Compression Using Partitioning Around Medoids Clustering Algorithm

¹V.B. Nikam ²Vinod J. Kadam, , ³B. B. Meshram

¹Research Scholar, Department of Computer Technology,

²Assistant Professor, Department of Information Technology,
Dr.Babasaheb Ambedkar Technological University, Lonere,(Maharashtra)

³Professor, Department of computer Technology

^{2,3}Veermata Jijabai Technological Institute, Mumbai, (Maharashtra).

Abstract

Clustering is a unsupervised learning technique. This paper presents a clustering based technique that may be applied to Image compression. The proposed technique clusters all the pixels into predetermined number of groups and produces a representative color for each group. Finally for each pixel only clusters number is stored during compression. This technique can be obtained in machine learning which one of the best methods for clustering is. The k -medoids algorithm is a clustering algorithm related to the K-means algorithm and the medoidshift algorithm.

Keywords: Data Mining, K -medoids clustering, Machine learning, Image compression,

1. Introduction

Clustering in data mining is a discovery process that groups similar objects into the same cluster[1]. Various clustering algorithms have been designed to fit various requirements and constraints of application[1]. Machine learning has been applied to many problems, and has demonstrated their superiority over classical methods when dealing with noisy or incomplete data. One such application is for data compression. A useful tool for determining k is the silhouette. Silhouette refers to a method of interpretation and validation of clusters of data. The technique provides a succinct graphical representation of how well each object lies within its cluster. It was first described by Peter J. Rousseeuw in 1986[2]. It is more robust to noise and outliers as compared to K-means because it minimizes a sum of dissimilarities instead of a sum of squared Euclidean distances. This clustering seem to be well suited to this particular function, as they have an ability to preprocess input patterns to produce simpler patterns with fewer components . This compressed information (stored in a hidden layer) preserves the full

information obtained from the external environment. The compressed features may then exit the network into the external environment in their original uncompressed form.

For data compression, the image or data is broken down into smaller vectors for use as input. For each input vector presented, the Euclidean distances to all the output points are computed. The weights of the point with the minimum distance, along with its neighboring points are adjusted. This ensures that the outputs of these points are slightly enhanced. This process is repeated until some criterion for termination is reached. After a sufficient number of input vectors have been presented, each output point becomes sensitive to a group of similar input vectors, and can therefore be used to represent characteristics of the input data. This means that for a very large number of input vectors passed into the model,(uncompressed image or data), the compressed form will be the data exiting from the output points of the model (considerably smaller number). This compressed data may then be further decompressed by another model.

2. The Proposed Method

The basic concept of this method is as follows:

1. Clustering all the pixels in to predetermined number of groups.
2. Producing a representative color for each group.
3. For each pixels storing only cluster number during compression.
4. During decompression restoring cluster number and storing representative color of that cluster.

A brief description of the method is shown in Fig.1.

3. Analysis

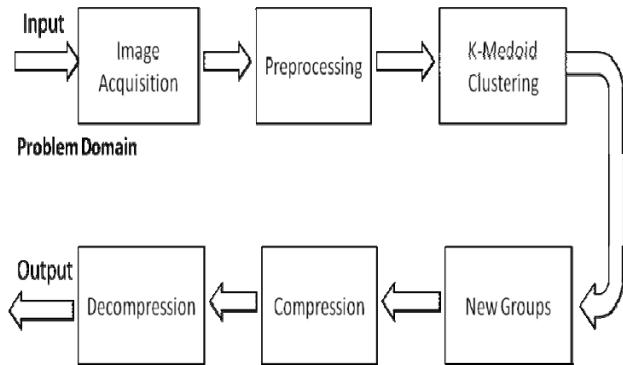


Fig.1 Design Steps for Image Compression

Step1: Image Acquisition and Preprocessing : These are preliminary steps required to feed input to the network. Problem domain is set of images received from user as a part of image acquisition. Preprocessing is required to remove noise from images using low pass filter.

Step2: Clustering : Since the basic colors are Red, Green and Blue, each vector have three parts as Red , Green and Blue quantity. Total number of feature in input pattern will be exactly three. For particular pixel input feature one will accept amount of Red quantity, feature two will accept amount of Green quantity and , feature three will accept amount of Blue quantity .The number of groups in the model will vary based on total number of clusters eg., if number of cluster required is 256 then we will have 256 (k^*) representatives.

1. Initialize: randomly select (k^*) of the *total* data points as the medoids
2. Associate each data point to the closest medoid. ("closest" here is defined using any valid distance metric, most commonly Euclidean distance, Manhattan distance or Minkowski distance). We here used Euclidean Distance.
3. For each medoid m
 1. For each non-medoid data point o
 1. Swap m and o and compute the total cost of the configuration
4. Select the configuration with the lowest cost.
5. repeat steps 2 to 5 until there is no change in the medoid.

Step3: Compression : During this process firstly medoid details obtained in previous step are stored ie., cluster number of closest medoid is stored.

In mathematics, the Euclidean distance or Euclidean metric is the "ordinary" distance between two points that one would measure with a ruler, and is given by the Pythagorean formula[3]. By using this formula as distance, Euclidean space (or even any inner product space) becomes a metric space. The associated norm is called the Euclidean norm. In one dimension, the distance between two points on the real line is the absolute value of their numerical difference[3]. Thus if x and y are two points on the real line, then the distance between them is computed as

$$\sqrt{(x - y)^2} = |x - y|.$$

In one dimension, there is a single homogeneous, translation-invariant metric (in other words, a distance that is induced by a norm), up to a scale factor of length, which is the Euclidean distance. In higher dimensions there are other possible norms. [3]In three-dimensional Euclidean space, the distance is

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_3 - q_3)^2}.$$

In our example, Here P1 red feature, P2 green feature and P3 is Blue feature. The representative who will give minimum $d(p,q)$ will be the winner. A medoid can be defined as the object of a cluster, whose average dissimilarity to all the objects in the cluster is minimal i.e. it is a most centrally located point in the given data set[4].

If j is the medoid of Cluster C , the average distance of all objects of C to j is calculated using the formula

$$\frac{\sum_{i \in C} d_i}{N_j}$$

where N is the number of object *minus* other than j . PAM is a partitioning method which operates on a distance matrix. The core objective of a learner is to generalize from its experience.[5] Medoids are representative objects of a data set or a cluster with a data set whose average dissimilarity to all the objects in the cluster is minimal. Medoids are similar in concept to means or centroids, but medoids are always members of the data set. Medoids are most commonly used on data when a mean or centroid cannot be defined such as 3-D trajectories or in the gene expression context[6]. Note that a medoid is not equivalent to a median or a geometric median. A median is only

defined on 1-dimensional data, and it only minimizes dissimilarity to other points for a specific distance metric (Manhattan norm). A geometric median is not necessarily a point from within the original dataset[6]. This algorithm basically works as follows. First, a set of medoids is chosen at random. Second, the distances to the other points are computed. Third, data are clustered according to the medoid they are most similar to. Fourth, the medoid set is optimized via an iterative process.



Fig.2 Output of Algorithm with PAM K=12

4. Conclusions

In this paper we have proposed a method for Image Compression using a K-medoids clustering. A superior training time and compression could be achieved with our method. It is more robust, because it minimizes a sum of dissimilarities instead of a sum of squared Euclidean distances.

References

- [1] Shu-Chuan Chu, John F. Roddick and J. S. Pan, "An Efficient K -Medoids-Based Algorithm Using Previous Medoid Index, Triangular Inequality Elimination Criteria, and Partial Distance Search", Data Warehousing and Knowledge Discovery, Lecture Notes in Computer Science, 2002, Volume 2454/2002, 301-311, DOI: 10.1007/3-540-46145-0_7.
- [2] Peter J. Rousseeuw, "Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis", Computational and Applied Mathematics **20**: 53–65, (1987).
- [3] Elena Deza, Michel Marie Deza, "Encyclopedia of Distances", page 94, (2009), Springer.
- [4] Sergios Theodoridis, Konstantinos Koutroumbas, Pattern Recognition 3rd ed.. (2006), pp. 635.
- [5] Christopher M. Bishop, Pattern Recognition and Machine Learning, Springer ISBN 0-387-31073-8, (2006).
- [6] Van Der Lann, Mark J; Pollard, Katherine S; Bryan, Jennifer; E, "A New Partitioning Around Medoids Algorithm". Journal of Statistical Computation and

Simulation (Taylor & Francis Group) **73** (8): 575–584, (2003).

Vinod J Kadam is a faculty in Information Technology Department at Dr. Babasaheb Ambedkar Technological University, Lonere, Raigad District, Maharashtra, India. He Holds BE, and MTech in Information Technology. He has seven years of academic experience. His research interests are in the domain of Data Mining, Image Processing.

Valmik B Nikam is a Ph.D. candidate in the Department of Computer Technology at Veermata Jijabai Technological Institute, Matunga, Mumbai. He holds B.E. in Computer Science and Engineering from Government College of Engineering Aurangabad and M.E. in Computer Technology from Veermata Jijabai Technological Institute, Mumbai. He worked at Dr. Babasaheb Ambedkar Technological University, Lonere for several years before joining PhD program. His research interests are data mining, scalable computing, image processing. He is a member of CSI, ACM and IEEE.

B.B.Meshram is a Professor and Head of Department of Computer Technology Department of Veermata Jijabai Technological Institute, Matunga, Mumbai. His current research includes database technologies, data mining, securities, forensic analysis, video processing, distributed computing. His has authored over 150 research publications. He has given numerous invited talks at various conferences, workshops, training programs and also served as chair/co-chair for many conferences/workshops in the area of computer science and engineering.

Lanes and Road Signs Recognition for Driver Assistance System

Ahmed Hechri^{1,2}, Abdellatif Mtibaa^{1,2}

¹ Laboratory E&E, Faculty of Sciences of Monastir, University of Monastir.
² National School of Engineering of Monastir, Av Ibn ElJazzar 5019 Monastir.

Abstract

Driver assistance systems become one of the most important features of the modern vehicles to ensure driver safety and decrease vehicles accidents on roads. According to their type and functionality, they intervene in different levels of the control processes involved in driving. In this paper, multitasks driver assistance system has been proposed. First, the system provides the driver with real time information from lanes markers and road signs, which consist the most important and challenging tasks. Secondly, generate an acoustic warning to the driver in advance of any danger. This warning then allows the driver to take appropriate corrective actions in order to mitigate or completely avoid the event. The proposed system was tested on real road scene captured from moving vehicle. From the experimental results, the system has demonstrated a robust performance for detecting the road lanes and signs under different conditions.

Keywords: Driver assistance systems, multitasks, vehicle, Lanes detection, road signs recognition.

1. Introduction

The rapid growing volume of traffic has a critical affect on economic development but causes large amount of traffic accidents. According to official statistics, 1.2 million people die every year on the roads worldwide and between 20 and 50 million suffer nonfatal injuries. Therefore, several security measures have been taken to reduce the number of victims of road accidents. Among these solutions, we find the development of a driver-assistance system (see Fig1) that catches the attention of the driver to avoid dangerous traffic situations. Apparently, among the complex and challenging tasks of these systems is road lanes detection and road signs recognition.

The objective of an automatic road signs recognition system is to detect and classify one or more road signs from within live colour images captured by a camera. It

attempts to develop an on-board warning system to alert the driver of the warning signs. Developing a reliable road

signs recognition system is considered a challenging task in this work.

Lanes detection plays an important role in driver assistance system (DAS), as it can help estimate the geometry of the road ahead, as well as the lateral position of the vehicle on the road.

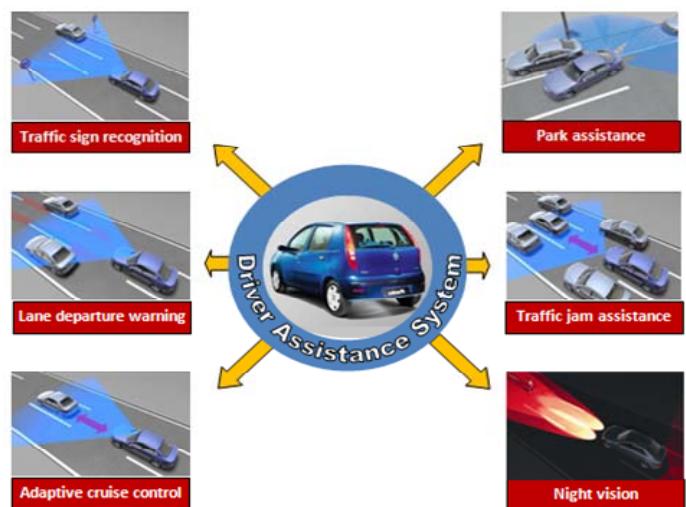


Fig. 1 Driver assistance systems functionality.

This module needs to be robust to a variety of roads conditions. By constantly monitoring the position of a car within a lane, collisions due to unintentional lane departure caused by driver distractions, fatigue, or driving under the influence of a controlled substance could be avoided.

The main objective of this paper is to develop a warning assistance system based on computer vision. First, the system provides the driver with real time information from

lanes markers and road signs, which consist the most important and challenging tasks. Secondly, generate an acoustic warning to the driver in advance of any danger. This warning then allows the driver to take appropriate corrective actions in order to mitigate or completely avoid the event.

The next section presents an overview of previous works in the field of lane detection and road signs recognition. Section 3 presents the proposed warning assistance system including the approaches used to develop each modules of the system. In Section 4, a discussion on the obtained results and performance of our approach is presented. Finally, in Section 5, we conclude our work.

2. Related works

2.1 Lanes detection

Lane detection is one important process in the vision-based driver assistance system and can be used for vehicle navigation, lateral control, collision prevention, or lane departure warning system. Various road conditions make this problem become very challenging including different type of lanes (straight or curvilinear), occlusions caused by obstacles, shadows, lighting changes (like night time), and so on. Therefore, in the literature and recently, there have been many approaches proposed for solving the above problems in lane detection. For example, in [1], He et al. proposed a color-based vision system to detect lanes from urban traffic scenes. Cheng et al. [2] used the color feature to detect lane lines and utilized the size, shape and motion for false lane region elimination. In [3], Yim and Oh combined three features including the starting position, intensity, and direction for lane detection. In addition to the feature-based scheme, the model-based scheme is more robust in lane detection when different lane types with occlusions or shadows are handled. Kang and Jung [4] proposed a searching framework to group edges with similar directions as a road lane. However, when complicated roads were handled, their method tended to detect false candidates of lane. In [6], a deformable template model of lane structure is used to locate lane boundaries by using intensity gradient information. In [7], an approach that combines an edge distribution function and the Hough transform with linear parabolic model is developed for lane detection and lane tracking. The Hough transform is very powerful because it can detect road boundaries successfully even in an extremely snowy environment [8].

2.2. Road sign recognition

Most of the work published in the area of road sign recognition (RSR) separates the detection and classification of RS as two different problems. The detection of road signs from outdoor images is the most complex step in a RSR system. There are many ways in which the characteristics of road signs (RS) (e.g. well established shapes and colours of the signs) could be exploited. However, the necessity to analyze the images in real-time is a limiting factor as to how much information available within the image should be extracted and analyzed. Most approaches to the problem of RS detection use either colour information or shape information

Color information is usually exploited by performing color-based segmentation of the image. Such segmentation is difficult to perform in RGB space. RGB colors are very sensitive to illumination changes and traffic scenes tend to have varying illumination. Some authors [9, 10] try to overcome this by devising simple formulas relating red, green and blue components and experimenting with appropriate thresholds. Others work in HSI [11, 12] or L*a*b [13] color spaces.

Shape information can be obtained by various strategies: Hough transform, fast radial transform, corner detection, pattern matching, genetic algorithms etc. Hough transform is used to locate lines or circles corresponding to a sign [11]. Shape is sometimes determined by using corner information of candidate regions [11, 14]. Some researchers use pattern matching with simple shape templates [15]. A technique based on genetic algorithms was used for detection of circular traffic signs in [16]. Shape detection often fails in cases of insufficient edge contrast, so most researchers choose to augment it with color information.

In the classification stage a pixel-based approach is often adopted and the class of the detected sign is determined by the cross-correlation template matching [17] or neural network [18]. Feature-based approach is used for instance in [19]. Support vector machines also have been reported as a good method to achieve this main target due to their ability to provide good accuracy as well as being sparse methods. In [12], a technique based on genetic algorithms has been proposed to recognize circular road signs by using only the brightness of an input image, which is obtained in the form of a binary image with the help of a smoothing and a Laplacian filter.

3. Proposed system

The proposed system is enclosed in dotted as shown in figure 2.

Driver assistance systems, according to their type and functionality, intervene in different levels of the control processes involved in driving.

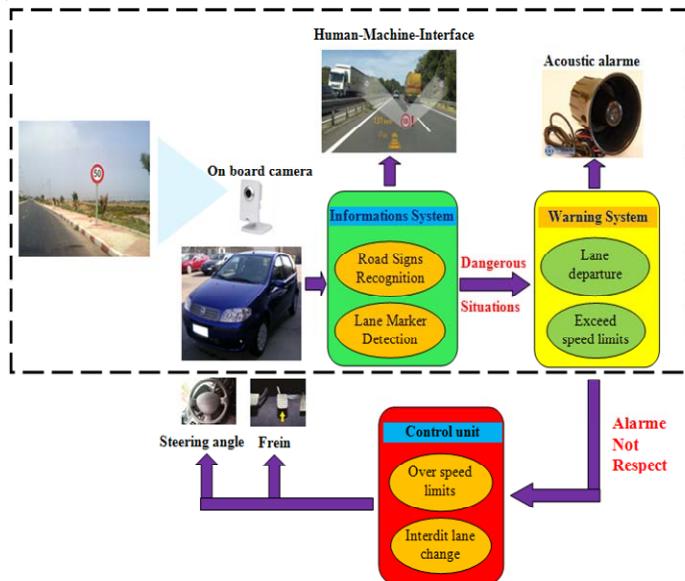


Fig. 2. Proposed system

At the most basic level, they present drivers with information which enables them to make more informed driving decisions, for example information about pedestrians not visible to the driver during night driving. At the next level, they can give the driver warnings of an imminent and possibly perilous situation to give them more time for decision making and reaction. The third level of intervention involves the system not only warning the driver but also advising or guiding them through the situation. At the highest level of intervention, driver assistance systems either take action independently or override the action of the driver.

The proposed system consists of three modules: lane detection, road signs recognition and warning system. The functionality of each module will be explained in details in the following subsections.

3.1. Lanes detection module

The goal of the image processing is to extract information about the position of the vehicle with respect to the road from the video image. Two major processes are implemented: the pre-processing process and then the lane detection process. The goal of pre-processing is to remove image noise and make the images sharper. The goal of lane detection is to detect the desired lane of the vehicle in order to obtain the look ahead distance and the lane angle.

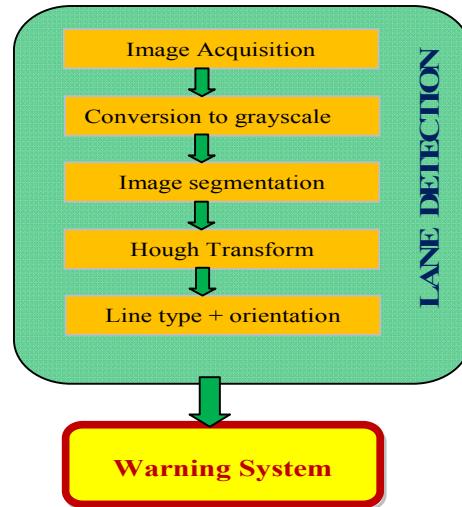


Fig. 3. Lane detection system

The first step of the lane detection algorithm is the conversion of the color image into grayscale. Then image segmentation and noise filter is carried out.

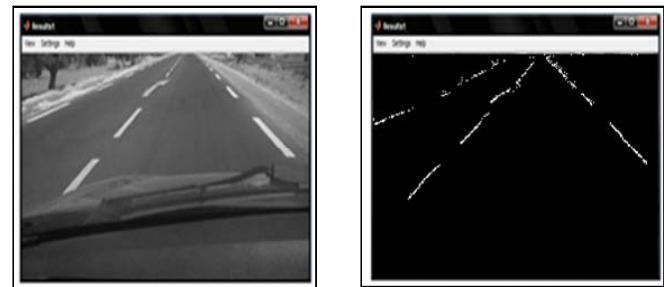


Fig. 4. (a) Original image (b) Image after segmentation

After image segmentation, we apply the standard Hough transform. The main advantage of this technique is relatively unaffected by image noise and present a reducing processing time for finding lines.

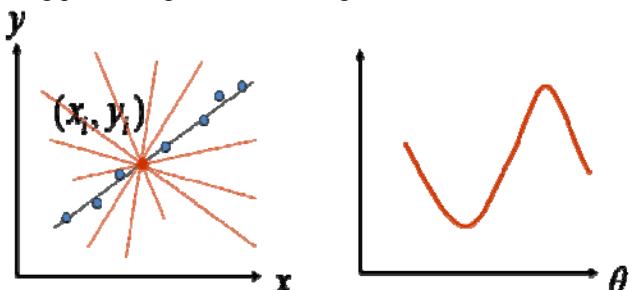


Fig. 5. (a) Image space (b) Hough space

Conversion from Cartesian coordinates $[x, y]$ into polar coordinates $[\rho, \theta]$ is proceeded according to the following equation 1:

$$\rho = x \cos \theta + y \sin \theta \quad (1)$$

Where ρ is the line connecting the polar coordinate to the origin where the x-axis intersects the y-axis, and where θ is the angle between the x axis and ρ .

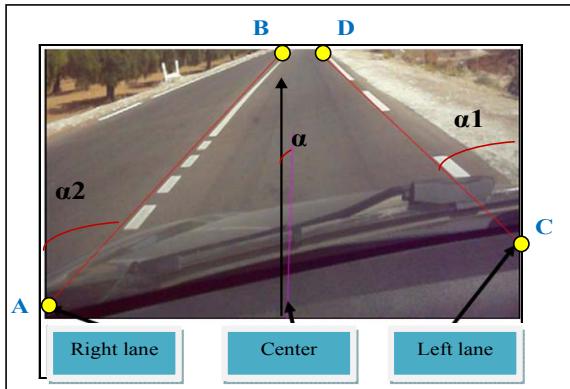


Fig. 5. Example of lane detection using Hough transforms

The steering angle α of the vehicle is given according to the equation 1:

$$\alpha = \frac{\alpha_1 + \alpha_2}{2} \quad (2)$$

$$\text{With } \begin{cases} \alpha_1 = \arctg\left(\frac{y_D - y_c}{x_D - x_c}\right) \\ \alpha_2 = \arctg\left(\frac{y_A - y_B}{x_B - x_A}\right) \end{cases} \quad (3)$$

3.2. Road signs recognition module

In this section, the proposed road sign recognition system (see Fig. 6.) is described. The system consists of two modules: detection and classification. Captured images are fed into the detection module. The regions of the image containing potential road sign patterns are extracted, and then forwarded to the classification module. The classification module further processes the extracted road-sign patterns and identify the type of road signs they represent.

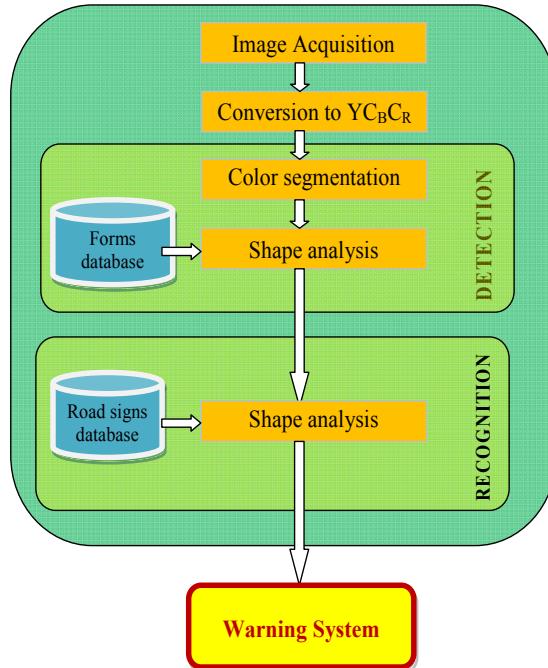


Fig. 6. Road signs recognition system

The illumination variation degrades considerably the performance of the detection module. Thus, in order to reduce the influence of luminance, acquired images were converted from RGB color model to the YCbCr color model according to the following equation:

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & 0.5 \\ 0.5 & -0.419 & -0.081 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (4)$$

First part of proposed algorithm is sign detection. The main task of the detection module is to segment the input image and extract out the areas that contain road sign patterns, and then forward them to the classification module for identification. After color segmentation, some morphological operations have been applied such as dilation, erosion and whole filling. Then we have applied template-matching algorithm to filter out the extracted non road sign objects. Template matching involves comparing a given template with windows of the same size in an image and identifying the window that is most similar to the template. In this work, the mean square error (MSE) shown in (5) is used for evaluating the similarity index.

$$MSE = \frac{1}{MN} \sum_{x=1}^M \sum_{y=1}^N [T(x, y) - I(x, y)]^2 \quad (5)$$

Where $T(x,y)$ is the intensity value of the pattern image at the position (x,y) , and $I(x,y)$ is the intensity of the input image at position (x,y) . M and N are the width and height of the image respectively.

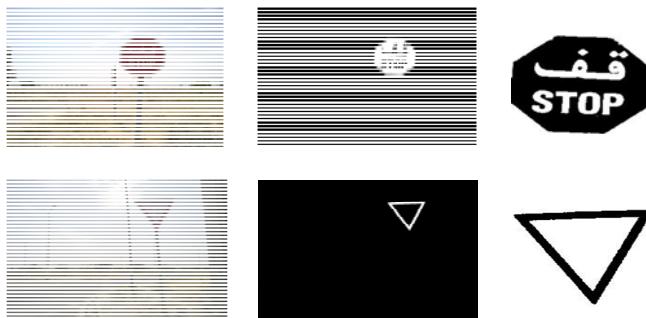


Fig. 7. Detection results

Once one or more candidates to be a road sign are found by the detection algorithm described in the previous section, a recognition stage must validate and possibly classify each candidate according to its similarity to one of the signs of a suitable database. The output of the detection algorithm is an image of road sign with nearly constant orientation, but the unknown factors, such as camera viewing direction and their relative position, will complicate the input image. Therefore, we need to use as many rotated, translated, or scaled templates as possible for identification (see figure 8).

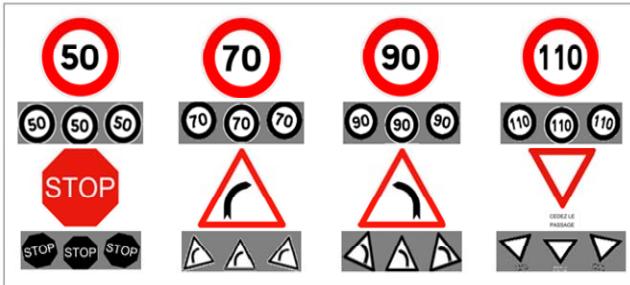


Fig. 8. :Road signs database for recognition system

In this work, a recognition scheme based on a normalized cross correlation between the window of the unknown road sign and the objects of the database work has been adopted. The cross-correlation coefficient $S(I,J)$ between two image I and J is defined as :

$$S(I, J) = \frac{\sum_i (I^i - \bar{I})(J^i - \bar{J})}{\sqrt{\sum_i (I^i - \bar{I})^2} \sqrt{\sum_i (J^i - \bar{J})^2}} \quad (6)$$

Where \bar{I} and \bar{J} denote the corresponding mean of the image intensities.

4. Experimental results

To evaluate the lanes and traffic signs recognition system, experiments were performed on the real data collected from Tunisian roads. Sample video sequences were acquired from an Axis IP video camera mounted in our vehicle were used for the input to our system. Each video sequence was down-sampled to a frame rate of 5Hz before being submitted to the system.

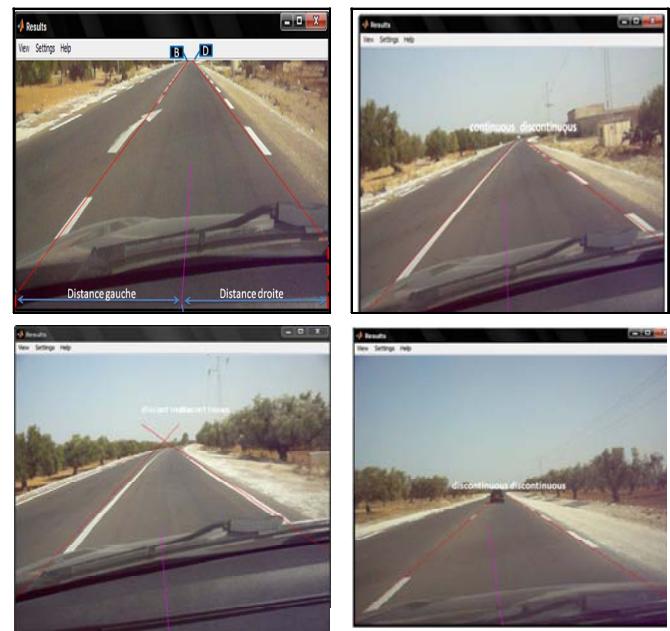


Fig. 9. Some results of the lane detection module

Some sample lane detection results are also provided in Fig8. It can be seen that the performance of the detection algorithm is quite robust in detecting lanes on urban streets with varying conditions. The proposed approach achieves good results for detecting all the visible lane boundaries especially in clear conditions. However we get some false positives due to stop lines at cross streets, at cross walks, near passing cars.

Captured video for the road sign recognition module content depicts the total of 204 signs and includes urban, countryside, and motorway scenes in natural daytime lightning, with numerous signs appearing in shade and in cluttered background. Table 1 illustrates road sign recognition results

Experimental results for the road sign recognition module demonstrate the superiority of the proposed approach which has achieved an average accuracy of 90% on completely novel test images. Our tests have demonstrated the effectiveness of our approach handling

TABLE I

DETECTION AND CLASSIFICATIONS RESULTS

Sign shape	Detection rate	Classification rate	Global
Octagon	100%	100%	100%
Circle	89%	82%	85,5%
Triangle	83%	86%	84,5%

occlusion in real traffic images including high noisy, bad illumination conditions and occlusions.



Fig. 10. Some results of the road sign recognition module

In a few false alarms across all test sequences were reported. The majority of detection failures were caused by the insufficient contrast between a sign's boundary and the background, especially for the signs appearing in shade or seen against sunlight.

5. Conclusion

In this paper, warning assistance system has been presented. Two driving tasks are considered: lane detection and road sign recognition. Firstly, a robust lane detection method has been developed which is based on Hough transform. This approach for safe lane systems has developed a safety system for avoiding lane departures for a large and complex set of traffic scenarios. Then, efficient approach for road sign recognition system was developed

and evaluated extensively through various test sets. The results prove the efficiency of the presented algorithms and show that our approach could have good prospects for application on board of intelligent vehicles.

Future work will include the full development of a HW/SW demonstrator to be mounted in a commercial car to verify the effectiveness of the concepts and the quality of the implemented strategies.

References

- [1] Y. He, H. Wang, and B. Zhang, "Color-Based Road Detection in Urban Traffic Scenes". IEEE Transactions on ITS, vol. 5, pp. 309-318, 2004.
- [2] H.-Y. Cheng, et al., "Lane detection with moving vehicle in the traffic scenes". IEEE Transactions on ITS, Vol. 7, pp. 571-582, 2006.
- [3] Y. U. Yim and S.-Y. Oh, "Three-Feature Base Automatic Lane Detection Algorithm (TFALDA) for Autonomous Driving". IEEE Transaction on ITS, vol. pp. 219-225, 2003.
- [4] D.-J. Kang, M.-H. Jung "Road lane segmentation using dynamic programming for active safety vehicles". Pattern Recognition Letters, Vol. 24, Issue 16, pp. 3177-3185, 2003.
- [5] K. Kluge, "Extracting road curvature and orientation from image edge points without perceptual grouping into features". IEEE Intelligent Vehicle Symposium, pp.109-114, Oct. 1994.
- [6] K. Kluge and S. Lakshmanan, "A deformable template approach to lane detection". IEEE Intelligent Vehicle Symposium, pp.54-59, Sept. 1995.
- [7] C.R. Jung and C.R. Kelber, "A robust linear-parabolic model for lane following". The XVII Brazilian Symposium on Computer Graphics and Image Processing, 2004.
- [8] Ahmed Hechri, Faycal Hamdaoui, Anis Ladgham and Mtibaa abdellatif, "Using fuzzy logic path tracking for an autonomous robot". In International Review of Automatic Control -Theory and Applications(IREACO), Praise Worthy Prize Publishers (Italy), (ISSN: 1974 – 6059) Volume 4 (Issue 1).
- [9] L.D. Lopez and O. Fuentes, "Color-based road sign detection and tracking". Image Analysis and Recognition(ICIAR), Montreal, CA, Agust 2007.
- [10] D. Krumbiegel, K.-F. Kraiss, S. Schrieber, "A connectionist traffic sign recognition system for onboard driver information". Proceedings of the Fifth IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design and Evaluation of Man-Machine Systems, 1992, pp. 201–206.
- [11] Fatmehsan, Y.R. Ghahari, A. Zoroofi, R.A. "Gabor wavelet for road sign detection and recognition using a hybrid classifier". International Conference on Multimedia Computing and Information Technology (MCIT), 2010.
- [12] Aoyagi Y, Asakura T, "A study on traffic sign recognition in scene image using genetic algorithms and neural networks". Proceedings of the 22nd international conference on industrial electronics, control, and instrumentation, Taipeis, pp 1838–1843, 1996.

- [13] S.H. Hsu, C.L. Huang, "Road sign detection and recognition using matching pursuit method". *Image and Vision Computing* 19 (2001) 119–129.
- [14] M. Benallal, J. Meunier, "Real-time color segmentation of road signs". Proceedings of the IEEE Canadian Conference on Electrical and Computer Engineering (CCGEI), 2003.
- [15] H.-M. Yang, C.-L. Liu, K.-H. Liu, and S.-M. Huang, "Traffic sign recognition in disturbing environments". Proceedings of the 14th International Symposium on Methodologies for Intelligent Systems (ISMIS '03), vol. 2871 of Lecture Notes in Computer Science, pp. 252–261, Maebashi City, Japan, October 2003.
- [16] J. Miura, T. Kanda, Y. Shirai, "An active vision system for real-time traffic sign recognition ". Proceedings of the IEEE Intelligent Transportation Systems, 2000, pp. 52–57.
- [17] Piccioli, G., De Micheli, E., Parodi, P., and Campani, M. A robust method for road sign detection and recognition. *Image and Vision Computing*, 14(3):209-223 (1996).
- [18] Escalera, A., Moreno, L. E., Salichs, M. A., and Armengol, J. M. Road trafficsign detection and classification. *IEEE Trans. on Industrial Electronics*, 44(6):848-859 (1997).
- [19] Paclik, P., Novovicova, J., Pudil, P., and Somol, P. Road Sign Classification using the Laplace Kernel Classifier. *Pattern Recognition Letters*, 21(13-14):1165-1173 (2000).



Ahmed Hechri received the degree in Electrical Engineering from Sfax University, Tunisia in 2006. In 2008 he received his M.S degree in Electrical engineering National School of Engineering of Monastir. He is currently preparing his PhD degree in the Electronics and Microelectronics Laboratory (E&E) at the Faculty of Sciences of Monastir.

His current research interests include modeling and control of mobile robots, image and video processing.



Abdellatif Mtibaa received his PhD degree in Electrical Engineering at the National School of Engineering of Tunis. Since 1990 he has been an assistant professor in Micro-Electronics and Hardware Design with Electrical Department at the National School of Engineering of Monastir. His research interests include high level synthesis, rapid prototyping and reconfigurable architecture for real-time multimedia applications, processing, embedded systems and pattern recognition.

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 2 – March 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 January 2012

Notification: 29th February 2012

Revision: 10th March 2012

Publication: 31st March 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>

© IJCSI PUBLICATION 2011

www.IJCSI.org



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