ADVANCES IN INTELLIGENT SYSTEMS AND COMPUTING 176

Natarajan Meghanathan
Dhinaharan Nagamalai
Nabendu Chaki (Eds.)

Advances in Computing and Information Technology

Springer

Advances in Intelligent Systems and Computing 176

For further volumes:
http://www.springer.com/series/11156

Natarajan Meghanathan, Dhinaharan Nagamalai,
and Nabendu Chaki (Eds.)

Advances in Computing and Information Technology

Proceedings of the Second International Conference on Advances in Computing and Information Technology (ACITY) July 13–15, 2012, Chennai, India – Volume 1

Springer

Editors

Dr. Natarajan Meghanathan
Department of Computer Science
Jackson State University
Jackson
USA

Dr. Nabendu Chaki
Department of Computer Science &
Engineering
University of Calcutta
Calcutta
India

Dr. Dhinaharan Nagamalai
Wireilla Net Solutions PTY Ltd
Melbourne
VIC
Australia

ISSN 2194-5357 e-ISSN 2194-5365
ISBN 978-3-642-31512-1 e-ISBN 978-3-642-31513-8
DOI 10.1007/978-3-642-31513-8
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012940793

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The Second International Conference on Advances in Computing and Information Technology (ACITY-2012) was held in Chennai, India, during July 13–15, 2012. ACITY attracted many local and international delegates, presenting a balanced mixture of intellect from the East and from the West. The goal of this conference series is to bring together researchers and practitioners from academia and industry and share cutting-edge development in the field. The conference will provide an excellent international forum for sharing knowledge and results in theory, methodology and applications of Computer Science and Information Technology. Authors are invited to contribute to the conference by submitting articles that illustrate research results, projects, survey work and industrial experiences describing significant advances in all areas of Computer Science and Information Technology.

The ACITY-2012 Committees rigorously invited submissions for many months from researchers, scientists, engineers, students and practitioners related to the relevant themes and tracks of the conference. This effort guaranteed submissions from an unparalleled number of internationally recognized top-level researchers. All the submissions underwent a strenuous peer-review process which comprised expert reviewers. These reviewers were selected from a talented pool of Technical Committee members and external reviewers on the basis of their expertise. The papers were then reviewed based on their contributions, technical content, originality and clarity. The entire process, which includes the submission, review and acceptance processes, was done electronically. The overall acceptance rate of ACITY-2012 is less than 20%. Extended versions of selected papers from the conference will be invited for publication in several international journals. All these efforts undertaken by the Organizing and Technical Committees led to an exciting, rich and a high quality technical conference program, which featured high-impact presentations for all attendees to enjoy, appreciate and expand their expertise in the latest developments in various research areas of Computer Science and Information Technology. In closing, ACITY-2012 brought together researchers, scientists, engineers, students and practitioners to exchange and share their experiences, new ideas and research results in all aspects of the main workshop themes and tracks, and to discuss the practical challenges encountered and the solutions adopted. We would like to thank the General and Program Chairs, organization staff, the members of the Technical

Program Committees and external reviewers for their excellent and tireless work. We sincerely wish that all attendees benefited scientifically from the conference and wish them every success in their research.

It is the humble wish of the conference organizers that the professional dialogue among the researchers, scientists, engineers, students and educators continues beyond the event and that the friendships and collaborations forged will linger and prosper for many years to come.

Natarajan Meghanathan
Dhinaharan Nagamalai
Nabendu Chaki

Organization

General Chairs

David C. Wyld	Southeastern Louisiana University, USA
E.V. Krishnamurthy	Australian National University, Australia
Jae Kwang Lee	Hannam University, South Korea
Jan Zizka	SoNet/DI, FBE, Mendel University in Brno, Czech Republic
V.L. Narasimhan	Pentagram R&D Intl. Inc., New Bern, USA
Michal Wozniak	Wroclaw University of Technology, Poland

Steering Committee

Abdul Kadhir Ozcan	Karatay University, Turkey
Brajesh Kumar Kaushik	Indian Institute of Technology-Roorkee, India
Dhinaharan Nagamalai	Wireilla Net Solutions PTY LTD, Australia
Eric Renault	Institut Telecom - Telecom SudParis, Evry, France
Jacques Demerjian	Communication & Systems, France
James Henrydoss	AT&T and University of Colorado, USA
Krzysztof Walkowiak	Wroclaw University of Technology,Poland
Murugan D.	Manonmaniam Sundaranar University, India
Nabendu Chaki	University of Calcutta, India
Natarajan Meghanathan	Jackson State University, USA
Raja Kumar M.	Taylor’s University, Malaysia
Salah Al-Majeed	University of Essex, UK
Selma Boumerdassi	Conservatoire National des Arts Et Metiers (CNAM), France
Sundarapandian Vaidyanathan	VelTech Dr. RR & Dr. SR Technical University, India

Program Committee Members

A.H.T. Mohammad	University of Bradford, UK
A.P. Sathish Kumar	PSG Institute of Advanced Studies, India
AAA. Atayero	Covenant University, Nigeria
Abdul Aziz	University of Central Punjab, Pakistan
Abdul Kadhir Ozcan	Karatay University, Turkey
Abdul Kadir Ozcan	The American University, Cyprus
Abdulbaset Mohammad	University of Bradford, United Kingdom
Ahmad Saad Al-Mogren	King Saud University, Saudi Arabia
Ahmed M. Khedr	Sharjah University, Sharjah, UAE
Ahmed Nada	Al-Quds University, Palestinian
Ajay K. Sharma	Dr. B R Ambedkar National Institute of Technology, India
Alaa Ismail Elnashar	Taif University, KSA
Alejandro Garces	Jaume I University, Spain
Alejandro Regalado Mendez	Universidad del Mar - México, USA
Alfio Lombardo	University of Catania, Italy
Ali El-Rashedy	University of Bridgeport, CT, USA
Ali M.	University of Bradford, United Kingdom
Ali Maqousi	Petra University, Jordan
Alireza Mahini	Islamic Azad University-Gorgan, Iran
Alvin Lim	Auburn University, USA
Amandeep Singh Thethi	Guru Nanak Dev University Amritsar, India
Amit Choudhary	Maharaja Surajmal Institute,India
Anand Sharma	MITS-Rajasthan, India
Anjan K.	RVCE-Bangalore, India
Ankit Thakkar	Nirma University, India
Ankit	BITS, PILANI India
Anthony Atayero	Covenant University, Nigeria
Aravind P.A.	Amrita School of Engineering India
Arun Pujari	Sambalpur University, India
Arunita Jaekel	University of Windsor, Canada
Ashok Kumar Das	IIT Hyderabad, India
Ashok kumar Sharma	YMCA Institute of Engineering, India
Ashutosh Dubey	NRI Institute of Science & Technology, Bhopal
Ashutosh Gupta	MJP Rohilkhand University, Bareilly
Athanasios Vasilakos	University of Western Macedonia, Greece
Azween Bin Abdullah	Universiti Teknologi Petronas, Malaysia
B. Srinivasan	Monash University, Australia
Babak Khosravifar	Concordia University, Canada
Balakannan S.P.	Chonbuk Nat. Univ., Jeonju
Balasubramanian K.	Lefke European University, Cyprus
Balasubramanian Karuppiah	Dr. MGR University, India
Bari A.	University of Western Ontario, Canada

Gomathi Kandasamy	Avinashilingam Deemed University for Women, India
Gopalakrishnan Kaliaperumal	Anna University, Chennai
Govardhan A.	JNTUH College of Engineering, India
Guo Bin	Institute TELECOM SudParis, France
H.V. Ramakrishnan	Dr. MGR University, India
Haider M. Alsabbagh	Basra University, Iraq
Haller Piroska	Petru Maior University-Tirgu Mures, Romania
Hao Shi	Victoria University, Australia
Hao-En Chueh	yuanpei University, Taiwan
Hari Chavan	National Institute of Technology, Jamshedpur, India
Henrique J.A. Holanda	UERN - Universidade do Estado do Rio Grande do Norte, Brasil
Henrique Joao Lopes Domingos	University of Lisbon, Portugal
Hiroyuki Hisamatsu	Osaka Electro-Communication University, Japan
Ho Dac Tu	Waseda University, Japan
Homam Reda El-Taj	Universiti Sains Malaysia, Malaysia
Hong yu	Capitol College, USA
Huosheng Hu	University of Essex, UK
Hussein Al-Bahadili	Petra University, Jordan
Hussein Ismail Khalaf Al-Bahadili	Petra University, Jordan
Hwangjun Song	Pohang University of Science and Technology,South Korea
Ignacio Gonzalez Alonso	University of Oviedo, Europe
Indrajit Bhattacharya	Kalyani Govt. Engg. College, India
Intisar Al-Mejibli	University of Essex, UK
Ioannis Karamitsos	Itokk Communications, Canada
J.K. Mandal	University of Kalyani, India
Jacques Demerjian	Communications & Systems, France
Jae Kwang Lee	Hannam University, South Korea
Jalel Akaichi	University of Tunis, Tunisia
Jan Zizka	SoNet/DI, FBE, Mendel University in Brno, Czech Republic
Jeong-Hyun Park	Electronics Telecommunication Research Institute, South Korea
Jeyanthy N.	VIT University, India
Jifeng Wang	University of Illinois at Urbana Champaign, USA
Johann Groschdl	University of Bristol, UK
Jose Enrique Armendariz-Inigo	Universidad Publica de Navarra, Spain
Juan Li	North Dakota State University, USA
Jyoti Singhai	Electronics and Communication Deptt-MANIT, India
Jyotirmay Gadewadikar	Alcorn State University, USA
Kai Xu	University of Bradford, United Kingdom
Kamalrulnizam Abu Bakar	Universiti Teknologi Malaysia, Malaysia

Karim Konate	University Cheikh Anta DIOP, Dakar
Kaushik Chakraborty	Jadavpur University, India
Kayhan Erciyes	Izmir University, Turkey
Khaled Shuaib	United Arab Emirates University, UAE
Khamish Malhotra	University of Glamorgan, UK
Khoa N. Le	University of Western Sydney, Australia
Krishnamurthy E.V.	ANU College of Engg & Computer Science, Australia
Krzysztof Walkowiak	Wroclaw University of Technology, Poland
Kuribayashi	Seikei University, Japan
L. Nirmala Devi	Osmania University - Hyderabad, India
Laiali Almazaydeh	University of Bridgeport, USA
Lu Yan	University of Hertfordshire, UK
Lus Veiga	Technical University of Lisbon, Portugal
Lylia Abrouk	University of Burgundy, France
M. Aqeel Iqbal	FUIEMS, Pakistan
M. Rajarajan	City University, UK
M. Ali	University of Bradford, UK
Maode Ma	Nanyang Technological University, Singapore
Marco Folli	University of Pavia, Italy
Marco Roccetti	Universty of Bologna, Italy
Massimo Esposito	ICAR-CNR, Italy
Md. Sipon Miah	Islamic University-Kushtia, Bangladesh
Michal Wozniak	Wroclaw University of Technology, Poland
Michel Owayjan	American University of Science & Technology, Lebanon
Miguel A. Wister	Juarez Autonomous University of Tabasco, Mexico
Mohamed Hassan	American University of Sharjah, UAE
Mohammad Ali Jabreil Jamali	Islamic Azad University, Iran
Mohammad Hadi Zahedi	Ferdowsi University of Mashhad, Iran
Mohammad Hajjar	Lebanese University, Lebanon
Mohammad Kaghazgaran	Islamic Azad University, Iran
Mohammad Mehdi Farhangia	Universiti Teknologi Malaysia, Malaysian
Mohammad Momani	University of technology Sydney, Australia
Mohammad Talib	University of Botswana, Botswana
Mohammad Zaidul Karim	Daffodil International University, Bangladesh
Mohammed Feham	University of Tlemcen, Algeria
Mohammed M. Alkhawlani	University of Science and Technology, Yemen
Mohsen Sharifi	Iran University of Science and Technology, Iran
Muhammad Sajjadur Rahim	University of Rajshahi, Bangladesh
Murty	Ch A S, JNTU, Hyderabad
Murugan D.	Manonmaniam Sundaranar University, India
Mydhili Nair	M S Ramaiah Institute of Technology, India
N. Krishnan	Manonmaniam Sundaranar University, India
Nabendu Chaki	University of Calcutta, India

Nadine Akkari	King abdulaziz University, Saudi Arabia
Naohiro Ishii	Aichi Institute of Technology, Japan
Nasrollah M. Charkari	Tarbiat Modares University, Iran
Natarajan Meghanathan	Jackson State University, USA
Nicolas Sklavos	Technological Educational Institute of Patras, Greece
Nidaa Abdual Muhsin Abbas	University of Babylon, Iraq
Nour Eldin Elmadany	Arab Acadmy for Science and Technology, Egypt
Ognjen Kuljaca	Alcorn State University, USA
Olakanmi Oladayo	University of Ibadan, Nigeria
Omar Almomani	Universiti Utara Malaysia, Malaysia
Orhan Dagdeviren	Izmir University, Turkey
Osman B. Ghazali	Universiti Utara Malaysia, Malaysia
Othon Marcelo Nunes Batista	Universidade Salvador, Brazil
Padmalochan Bera	Indian Institute of Technology, Kharagpur, India
Partha Pratim Bhattacharya	Mody Institute of Technology & Science, India
Patricia Marcu	Leibniz Supercomputing Centre, Germany
Patrick Seeling	University of Wisconsin - Stevens Point, USA
R. Thandeeswaran	VIT University, India
Phan Cong Vinh	London South Bank University, UK
Pinaki Sarkar	Jadavpur University, India
Polgar Zsolt Alfred	Technical University of Cluj Napoca, Romania
Ponpit Wongthongtham	Curtin University of Technology, Australia
Quan (Alex) Yuan	University of Wisconsin-Stevens Point, USA
Rafael Timoteo	University of Brasilia - UnB, Brazil
Raied Salman	Virginia Commonwealth University, USA
Rajendra Akerkar	Technomathematics Research Foundation, India
Rajeswari Balasubramaniam	Dr. MGR University, India
Rajkumar Kannan	Bishop Heber College, India
Rakhesh Singh Kshetrimayum	Indian Institute of Technology, Guwahati, India
Raman Maini	Punjabi University, India
Ramayah Thurasamy	Universiti Sains Malaysia, Malaysia
Ramayah	Universiti Sains Malaysia, Malaysia
Ramin karimi	University Technology Malaysia
Razvan Deaconescu	University Politehnica of Bucharest, Romania
Reena Dadhich	Govt. Engineering College Ajmer
Reshmi Maulik	University of Calcutta, India
Reza Ebrahimi Atani	University of Guilan, Iran
Rituparna Chaki	West Bengal University of Technology, India
Robert C. Hsu	Chung Hua University, Taiwan
Roberts Masillamani	Hindustan University, India
Rohitha Goonatilake	Texas A&M International University, USA
Rushed Kanawati	LIPN - Universite Paris 13, France
S. Geetha	Anna University - Tiruchirappalli, India
S. Hariharan	B.S. Abdur Rahman University, India

S. Venkatesan	University of Texas at Dallas - Richardson, USA
S.A.V. Satyamurty	Indira Gandhi Centre for Atomic Research, India
S. Arivazhagan	Mepco Schlenk Engineering College, India
S. Li	Swansea University, UK
S. Senthil Kumar	Universiti Sains Malaysia, Malaysia
Sajid Hussain	Acadia University, Canada
Salah M. Saleh Al-Majeed	University of Essex, United Kingdom
Saleena Ameen	B.S.Abdur Rahman University, India
Salem Nasri	ENIM, Monastir University, Tunisia
Salim Lahmiri	University of Qubec at Montreal, Canada
Salini P.	Pondichery Engineering College, India
Salman Abdul Moiz	Centre for Development of Advanced Computing, India
Samarendra Nath Sur	Sikkim Manipal University, India
Sami Ouali	ENSI, Compus of Manouba, Manouba, Tunisia
Samiran Chattopadhyay	Jadavpur University, India
Samodar reddy	India school of mines , India
Samuel Falaki	Federal University of Technology-Akure, Nigeria
Sanjay Singh	Manipal Institute of Technology, India
Sara Najafzadeh	University Technology Malaysia
Sarada Prasad Dakua	IIT-Bombay, India
Sarmistha Neogy	Jadavpur University, India
Satish Mittal	Punjabi University, India
S.C. SHARMA	IIT - Roorkee, India
Seetha Maddala	CBIT, Hyderabad
Selma Boumerdassi	Cnam/Cedric, France
Sergio Ilarri	University of Zaragoza, Spain
Serguei A. Mokhov	Concordia University, Canada
Shaoen Wu	The University of Southern Mississippi, USA
Sharvani G.S.	RV College of Engineering, Inida
Sherif S. Rashad	Morehead State University, USA
Shin-ichi Kuribayashi	Seikei University, Japan
Shivan Haran	Arizona state University, USA
Shobha Shankar	Vidya vardhaka College of Engineering, India
Shrikant K. Bodhe	Bosh Technologies, India
Shriram Vasudevan	VIT University, India
Shrirang Ambaji Kulkarni	National Institute of Engineering, India
Shubhamoy Dey	Indian Institute of Management Indore, India
Solange Rito Lima	University of Minho, Portugal
Souad Zid	National Engineering School of Tunis, Tunisia
Soumyabrata Saha	Guru Tegh Bahadur Institute of Technology, India
Sridharan	CEG Campus - Anna University, India
Sriman Narayana Iyengar	VIT University, India
Srinivasulu Pamidi	V R Siddhartha Engineering College Vijayawada, India

Sriram Maturi	Osmania University, India
Subhabrata Mukherjee	Jadavpur University, India
Subir Sarkar	Jadavpur University, India
Sundarapandian Vaidyanathan	VelTech Dr. RR & Dr. SR Technical University, India
Sunil Singh	Bharati vidyapeeth's College of Engineering, India
Sunilkumar S. Manvi	REVA Institute of Technology and Management Kattigenhalli, India
SunYoung Han	Konkuk University, South Korea
Susana Sargento	University of Aveiro, Portugal
Swarup Mitra	Jadavpur University, Kolkata, India
T. Ambaji Venkat Narayana Rao	Hyderabad Institution of Technology and Management , India
T.G. Basavaraju	National Institute of Technology Karnataka (NITK), India
Thomas Yang	Embry Riddle Aeronautical University, USA
Tri Kurniawan Wijaya	Technische Universitat Dresden, Germany
Tsung Teng Chen	National Taipei Univ., Taiwan
Utpal Biswas	University of Kalyani, India
V.M. Pandharipande	Dr. Babasaheb Ambedkar Marathwada University, India
Valli Kumari Vatsavayi	AU College of Engineering, India
Vijayalakshmi S.	VIT University, India
Virgil Dobrota	Technical University of Cluj-Napoca, Romania
Vishal Sharma	Metanoia Inc., USA
Wei Jie	University of Manchester, UK
Wichian Sittiprapaporn	Mahasarakham University, Thailand
Wided Oueslati	l'institut Superieur de Gestion de Tunis, Tunisia
William R. Simpson	Institute for Defense Analyses, USA
Wojciech Mazurczyk	Warsaw University of Technology, Poland
Xiaohong Yuan	North Carolina A & T State University, USA
Xin Bai	The City University of New York, USA
Yahya Slimani	Faculty of Sciences of Tunis, Tunisia
Yannick Le Moullec	Aalborg University, Denmark
Yaser M. Khamayseh	Jordan University of Science and Technology, Jordan
Yedehalli Kumara Swamy	Dayanand Sagar College of Engineering, India
Yeong Deok Kim	Woosong University, South Korea
Yogeshwar Kosta	Marwadi Education Foundations Group of Institutions, India
Yuh-Shyan Chen	National Taipei University, Taiwan
Yung-Fa Huang	Chaoyang University of Technology, Taiwan
Zaier Aida	National Engeneering School of GABES, Tunisia
Zakaria Moudam	Université sidi mohammed ben Abdellah, Morocco
Zuqing Zhu	Cisco Systems, USA

External Reviewers

A. Kannan	K.L.N. College of Engineering, India
Martin	Sri Manakula Vinayagar Engineering College, India
Abhishek Samanta	Jadavpur University, Kolkata, India
Ayman Khalil	Institute of Electronics and Telecommunications of Rennes, France
Cauvery Giri	RVCE, India
Ch. V. Rama Rao	Gudlavalleru Engineering College, India
Chandra Mohan	Bapatla Engineering College, India
E.P. Ephzibah	VIT University-Vellore, India
Hameem Shanavas	Vivekananda Institute of Technolgy, India
Kota Sunitha	G. Narayanamma Institute of Technology and Science, Hyderabad
Kunjal B. Mankad	ISTAR, Gujarat, India
Lakshmi Rajamani	Osmania University, India
Lavanya	Blekinge Institute of Technology, Sweden
M.P. Singh	National Institute of Technology, Patna
M. Tariq Banday	University of Kashmir, India
M.M.A. Hashem	Khulna University of Engineering and Technology, Bangladesh
Mahalinga V. Mandi	Dr. Ambedkar Institute of Technology, Bangalore, Karnataka, India
Mahesh Goyani	G H Patel College of Engineering and Technology, India
Maragathavalli P.	Pondicherry Engineering College, India
M.P. Singh	National Institute of Technology, Patna
M. Tariq Banday	University of Kashmir, India
M.M.A. Hashem	Khulna University of Engineering and Technology, Bangladesh
Mahalinga V. Mandi	Dr. Ambedkar Institute of Technology, India
Monika Verma	Punjab Technical University, India
Moses Ekpenyong	University of Uyo, Nigeria
Mini Patel	Malwa Institute of Technology, India
N. Kaliammal	NPR College of Engg &Tech, India
N. Adhikari	Biju Pattnaik University of Technology, India
N.K. Choudhari	Bhagwati Chaturvedi College of Engineering, India
Naga Prasad Bandaru	PVP Siddartha Institute of Technology, India
Nagamanjula Prasad	Padmasri Institute of Technology, India
Nagaraj Aitha	I.T, Kamala Institute of Tech & Science, India
Nana Patil	NIT Surat, Gujrat
Nitiket N. Mhala	B.D. College of Engineering - Sewagram, India
P. Ashok Babu	Narsimhareddy Engineering College, India
P. Sheik Abdul Khader	B.S. Abdur Rahman University, India

Pesn Krishna Prasad	Aditya Engineering College, India
Partha Pratim Bhattacharya	Mody Institute of Technology & Science, India
Pappa Rajan	Anna University, India
Pooja Jain	JUIT Waknaghat, India
Prabu Dorairaj	NetApp Inc, India
Pradeepini Gera	Jawaharlal Nehru Technological University, India
Pravin P. Karde	HVPM's College of Engineering & Technology - Amravati, India
Premanand K. Kadbe	Vidya Pratishthan's College of Engineering, India
Priti Sajja	Sardar Patel University, India
R. Baskaran	Anna University - Chennai, India
R. Murali	Dr. Ambedkar Institute of Technology, Bangalore
R.M. Suresh	Mysore University
Rabindranath Bera	Sikkim Manipal Inst. of Technol., India
Rajashree Biradar	Ballari Institute of Technology and Management, India
Rajesh Kumar Krishnan	Bannari Amman Inst. of Technol., India
Rajesh Kumar P.	The Best International, Australia
Rajeshwari Hegde	BMS College of Engineering, India
Rajagopal Palsonkennedy	Dr. MGR University, India
Selvakumar Ramachandran	Blekinge Institute of Technology, Sweden

Contents

Network Security and Applications

Intelligent Network-Based Intrusion Detection System (iNIDS) 1
P.R. Mahalingam

Mutual Authentication for Wireless Communication Using Elliptic Curve Digital Signature Based on Pre-known Password 11
Tumpa Roy, Poonam Sisodia, Divye Upadhyay, Kamlesh Dutta

Securing Multi-agent Systems: A Survey 23
S.V. Nagaraj

Personal Secret Information Based Authentication towards Preventing Phishing Attacks 31
Gaurav Varshney, Ramesh Chandra Joshi, Anjali Sardana

Key Distribution Schemes in Wireless Sensor Networks: Novel Classification and Analysis 43
Premraj Mahajan, Anjali Sardana

An Integrated Intrusion Detection System for Credit Card Fraud Detection 55
M. Sasirekha, I. Sumaiya Thaseen, J. Saira Banu

Analysis and Performance Evaluation of Application Specific Processors for Network-Based Intrusion Detection Systems 61
Majid Nezakatolhoseini, Sam Jabbehdari, Mohammad Ali Pourmina

ECDLP Based Proxy Multi-signature Scheme 71
Ramanuj Chouksey, R. Sivashankari, Piyush Singhai

SVIP-Enhanced Security Mechanism for SIP Based VoIP Systems and Its Issues . . . 81
D. Chandramohan, D. Veeraiah, M. Shanmugam, N. Balaji, G. Sambasivam, Shailesh Khapre

Host-Based Bot Detection Using Destination White-Lists for User's Profile . . . 87
B. Soniya, M. Wilscy

Effective Implementation and Evaluation of AES in Matlab . . . 95
Amish Kumar, Namita Tiwari

Low Overhead Handoff Based Secure Checkpointing for Mobile Hosts . . . 103
Priyanka Dey, Suparna Biswas

An Obfuscated Implementation of RC4 . . . 113
Roger Zahno, Amr M. Youssef

A Statistical Pattern Mining Approach for Identifying Wireless Network Intruders . . . 131
Nur Al Hasan Haldar, Muhammad Abulaish, Syed Asim Pasha

SLA for a Pervasive Healthcare Environment . . . 141
J. Valarmathi, K. Lakshmi, R.S. Menaga, K.V. Abirami, V. Rhymend Uthariaraj

Composing Signatures for Misuse Intrusion Detection System Using Genetic Algorithm in an Offline Environment . . . 151
Mayank Kumar Goyal, Alok Aggarwal

SRSnF: A Strategy for Secured Routing in Spray and Focus Routing Protocol for DTN . . . 159
Sujoy Saha, Rohit Verma, Satadal Sengupta, Vineet Mishra, Subrata Nandi

Multi Tree View of Complex Attack – Stuxnet . . . 171
Shivani Mishra, Krishna Kant, R.S. Yadav

Secure Peer-Link Establishment in Wireless Mesh Networks . . . 189
Swathi Bhumireddy, Somanath Tripathy, Rakesh Matam

Secret Image Embedded Authentication of Song Signal through Wavelet Transform (IAWT) . . . 199
Uttam Kr. Mondal, Jyotsna Kumar Mandal

Specification Based IDS for Power Enhancement Related Vulnerabilities in AODV . . . 209
Chaitali Biswas Dutta, Utpal Biswas

English to Hindi Machine Translator Using GMT and RBMT Approach . . . 219
Ambrish Srivastav, Nitin Hambir

Plugging DHCP Security Holes Using S-DHCP 227
Amit Kumar Srivastava, Arun Kumar Misra

Efficient Cryptography Technique on Perturbed Data in Distributed Environment 239
Nishant Goswami, Tarulata Chauhan, Nishant Doshi

A Novel Triangle Centroid Authentication Protocol for Cloud Environment 245
K. Anitha Kumari, G. Sudha Sadasivam, Bhandari Chetna, S. Rubika

Security and Availability of Data in the Cloud 255
Ahtesham Akhtar Patel, S. Jaya Nirmala, S. Mary Saira Bhanu

A Novel Power Balanced Encryption Scheme for Secure Information Exchange in Wireless Sensor Networks 263
Shanta Mandal, Rituparna Chaki

A Cryptographic Approach towards Black Hole Attack Detection 273
Pratima Sarkar, Rituparna Chaki

Connecting Entropy-Based Detection Methods and Entropy to Detect Covert Timing Channels 279
Bukke Devendra Naik, Sarath Chandra Boddukolu, Pothula Sujatha, P. Dhavachelvan

Route and Load Aware Channel Assignment Algorithm for Multichannel and Multi Radio Vehicular Ad-Hoc Networks 289
Jagadeesh Kakarla, S. Siva Sathya

Performance Analysis of TCP & UDP in Co-located Variable Bandwidth Environment Sharing Same Transmission Links 299
Mayank Kumar Goyal, Yatendra Kumar Verma, Paras Bassi, Paurush Kumar Misra

Personalised High Quality Search with in a Web Site: No User Profiling 307
L.K. Joshila Grace, V. Maheswari, Dhinaharan Nagamalai

Usability Evaluation Using Specialized Heuristics with Qualitative Indicators for Intrusion Detection System 317
Tulsidas Patil, Ganesh Bhutkar, Noshir Tarapore

Networks and Communications

Analysis and Synchronization of the Hyperchaotic Yujun Systems via Sliding Mode Control 329
Sundarapandian Vaidyanathan

String Matching Technique Based on Hardware: A Comparative Analysis 339
Aakanksha Pandey, Nilay Khare

Impact of Blackhole and Rushing Attack on the Location-Based Routing Protocol for Wireless Sensor Networks 349
R. Shyamala, S. Valli

Analysis of Fractional Frequency Reuse (FFR) over Classical Reuse Scheme in 4G (LTE) Cellular Network 361
Chandra Thapa, C. Chandrasekhar

Temporary Parallel Route Recovery for Frequent Link Failure in VANET 369
B. Siva Kumar Reddy, M. Sakthi Ganesh, P. Venkata Krishna

Analysis of MIMO Channel Characteristics in Indoor Environment Using Ray Tracing Simulator 375
Manjusha Karkhanis, Achala Deshmukh

Efficient Target Recovery in Wireless Sensor Network 385
Shailaja Patil, Ashish Gupta, Mukesh Zaveri

Generic Middleware Architecture Supporting Heterogeneous Sensors Management for Any Smart System 395
Soma Bandyopadhyay, Abhijan Bhattacharyya

Improving TCP Performance in Hybrid Networks 405
O.S. GnanaPrakasi, P. Varalakshmi

Synchronization in Distributed Systems 417
Amritha Sampath, C. Tripti

NAAS: Negotiation Automation Architecture with Buyer's Behavior Pattern Prediction Component 425
Debajyoti Mukhopadhyay, Sheetal Vij, Suyog Tasare

Quantum DOT Sensor for Image Capturing and Routing Based on Temporal Power and Critical Factor 435
S. Indu Vadhani, G. Vithya, B. Vinayagasundaram

Checkpointing and Recovery Using Node Mobility among Clusters in Mobile Ad Hoc Network 447
Suparna Biswas, Sarmistha Neogy

Design of Broadband Optical Sources for OCDMA/WDMA Applications . . . 457
Lakshmi Priya, M. Meenakshi, G. Geetha

Localization in Wireless Sensor Network: A Distributed Approach 467
Shailaja Patil, Mukesh Zaveri

Web Mining and Security in E-commerce 477
Shaikh Mohammed Atiq, Dayanand Ingle, B.B. Meshram

Performance Improvement in MIMO Systems Using Rotating Codebooks 489
J. Julia, M. Meenakshi

Efficient Techniques for the Implementation of AES SubByte and MixColumn Transformations 497
K. Rahimunnisa, M. Priya Zach, S. Suresh Kumar, J. Jayakumar

Feature Selection for Detection of Ad Hoc Flooding Attacks 507
Sevil Sen, Zeynep Dogmus

Performance Analysis of SCTP Compared to TCP and UDP 515
Nagesha, S.S. Manvi

Wireless and Mobile Networks

MSRCC – Mitigation of Security Risks in Cloud Computing 525
D. Chandramohan, T. Vengattaraman, M.S.S. Basha, P. Dhavachelvan

Topology Construction of 3D Wireless Sensor Network 533
Sarbani Roy, Nandini Mukherjee

Design and Analysis of Dual Capacitively Loaded C-PIFA 543
Kirti Dhwaj, Rachit Garg, Gaurav Mishra, Neetesh Purohit

Jong Nang 3-Input NOR Channel 551
Moon Ho Lee, Md. Hashem Ali Khan, Daechul Park

Multiuser Transmitter Preprocessing Aided Downlink Communications in Correlated Frequency-Selective Channels 563
Nithin Srinivasan, Sriram Ravichandran, Shruthi Ravichandran, Prabagarane Nagaradjane

Multi-dimensional Performance Characterization of Directional Antennas for Applications in Energy Efficient Ad-Hoc Network 575
C.H. Sandhya, Deepali R. Borade, Rinki Sharma, Govind R. Kadambi

Reliability Enhanced Routing Protocol for Wireless Mesh Networks 587
Rakesh Matam, Somanath Tripathy

DARIH: Distributed Adaptive Routing via Information Highway in Sensor Network 597
Monomita Mazumdar, Srimanta Halder, Prasenjit Chanak, Indrajit Banerjee

A Comparative Study of Cache Replacement Policies in Wireless Mobile Networks 609
Preetha Theresa Joy, K. Polouse Jacob

FTLBS: Fault Tolerant Load Balancing Scheme in Wireless Sensor Network 621
Srimanta Halder, Monomita Mazumdar, Prasenjit Chanak, Indrajit Banerjee

Effective Resource Allocation Technique for Wireless Cellular System 633
Banda Sreenivas, S. Ramesh Babu, S. Prabhakar, D. Karunakar Reddy

Performance Analysis of AODV and DSDV Protocols Using RPGM Model for Application in Co-operative Ad-Hoc Mobile Robots 641
Rajesh P. Barnwal, Arnab Thakur

Enhanced SAFER+ Algorithm for Bluetooth to Withstand Against Key Pairing Attack 651
Payal Chaudhari, Hiteishi Diwanji

Super Peer Deployment in Unstructured Peer-to-Peer Networks 661
R. Venkadeshan, M. Jegatha

Efficient Path Selection to Propagate Data Message for Optimizing the Energy Dissipation in WSN 671
Subrata Dutta, Nandini Mukherjee, Monideepa Roy, Sarmistha Neogy

Peak to Average Power Ratio Reduction in OFDM System over PAM, QAM and QPSK Modulation 685
Gaurav Sikri, Rajni

MERCC: Multiple Events Routing with Congestion Control for WSN 691
Ayan Kumar Das, Rituparna Chaki

Peer-to-Peer Networks and Trust Management

Awareness Based Approach against E-Mail Attacks 699
Gaurav Kumar Tak, Gaurav Ojha

On the Fly File Dereplication Mechanism 709
Paras Gupta, Manu Vardhan, Akhil Goel, Abhinav Verma, Dharmender Singh Kushwaha

Security Service Level Agreements Based Authentication and Authorization Model for Accessing Cloud Services 719
Durgesh Bajpai, Manu Vardhan, Sachin Gupta, Ravinder Kumar, Dharmender Singh Kushwaha

Load Balancing in Cluster Using BLCR Checkpoint/Restart 729
Hemant Hariyale, Manu Vardhan, Ankit Pandey, Ankit Mishra, Dharmender Singh Kushwaha

Adaptive Region Based Huffman Compression Technique with Selective Code Interchanging 739
Utpal Nandi, Jyotsna Kumar Mandal

Human Tracking in Video Surveillance 749
Helly Patel, Mahesh P. Wankhade

Performance Evaluation of V2VCommunication by Implementing Security Algorithm in VANET 757
Manpreet Kaur, Rajni, Parminder Singh

An Efficient Approach to Secure Routing in MANET 765
Dipayan Bose, Arnab Banerjee, Aniruddha Bhattacharyya, Himadri Nath Saha, Debika Bhattacharyya, P.K. Banerjee

Trust Oriented Secured AODV Routing Protocol against Rushing Attack . . . 777
Swarnali Hazra, S.K. Setua

SEPastry: Security Enhanced Pastry 789
Madhumita Mishra, Somanath Tripathy, Sathya Peri

Hybrid Scenario Based Analysis of the Effect of Variable Node Speed on the Performance of DSDV and DSR 797
Koushik Majumder, Sudhabindu Ray, Subir Kumar Sarkar

Author Index 807

Intelligent Network-Based Intrusion Detection System (iNIDS)

P.R. Mahalingam

Department of Computer Science
Rajagiri School of Engineering & Technology, Rajagiri valley, Cochin, India
prmahalingam@gmail.com

Abstract. Networks are regarded as one of the biggest advancements in the field of computer science. But they enable outsiders to “intrude” into our information. Intrusions can be in the form of simple eavesdropping, or gaining access to the host itself. Here, intruders are identified using two main methods – signature analysis and anomaly analysis. The proposed method is such that the signature analysis is strengthened by anomaly analysis, which in turn uses some level of intelligence based on the traffic parameters, obtained and processed using neural networks. The initial intelligence is obtained using the KDDCUP99 dataset, which trains a neural network. The neural network will take care of further detections, and it strengthens itself during the run itself. The result obtained suggests that even with minimal initial intelligence, iNIDS can reach accuracy levels of over 70%, and by increasing the initial set a little more, it reaches accuracy levels exceeding 80%.

Keywords: Intrusion detection, neural networks, intelligence, anomaly analysis, signature analysis, KDDCUP99, JpCap.

1 Introduction

Advancements and increased usage of computer networks paved way for increase in variety and complexity of security threats. The scenario is getting worse in the sense even single firewall strategies are insufficient to counter security threats[1]. Nowadays people are aware of the risks involved in securing a computer network. So a system which is capable of detecting network security threats is developed [2]. Here, an Network based Intrusion Detection System (NIDS) is proposed that uses real time internet traffic for analysis. Also, the system uses Artificial Intelligence for improving the performance and speed of detection.

Real time packets in the network are captured online i.e. from the internet as and when they reach the interface of the network, using suitable Java[3]-based packages. iNIDS is designed to provide the basic detection techniques so as to secure the systems inside a computer network that are directly or indirectly connected to the Internet.

Network intrusion[4] can be defined as any deliberate attempt to enter or gain unauthorized access to a network and thereby break the security of the network and thus gaining access to confidential information present in the computers inside the

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 1–9.
springerlink.com

network. An IDS[4] captures and inspects all traffic, regardless of whether it's permitted or not. Based on the contents of the packet, their flow, length etc, at either the IP or application level, an alert is generated.

An *intrusion signature*[4] can be defined as a special TCP state set such as [SYN|RST] in one packet, special bytes in the IP header, or a special byte stream in the payload of a packet that will provide a pattern which can be used for packet analysis for identifying threats.

The primary goals of the whole proposal can be summarized to the following.

1) Detect Network intrusions[8][9]
2) Use of Artificial Intelligence to improve detection[5][6]
3) Use Network traffic for analysis and detection[5][7].

2 NIDS and ANN

Dr. Dorothy E Denning proposed an Intrusion detection system in 1987 which became a benchmark in the research in this area[2]. Many researches have been conducted based on this paper and currently researchers are more interested in developing intrusion detection systems based on Artificial Neural Networks. Artificial Neural Networks possess features like generalization, flexibility etc. Wang Zhenqi and Wang Xinyu proposed a Netflow[1] based intrusion detection system, which can resist network attacks and intrusions. It was found to be cost effective and does not affect the performance of backbone network.[1][13]

Usually, sampled data from Kddcup99 dataset[14], an attack or intrusion database is the standard for evaluating the security detection mechanisms. This dataset is used for signature analysis, for training neural network for anomaly analysis and for testing the IDS itself. The advantage of using Backpropagation algorithm is that it can train (learn) data at a faster rate and it provides efficient generalization and flexibility when compared to other existing Neural Network technologies[13]. But, the performance of a Neural Network depends mainly on the amount of training data given[15][16][17].

The strategy[18] dictates that NIDS uses a hybrid detection engine i.e. a combination of Signature detection and Anomaly detection capabilities. "Rule –based" detection technique is used for signature analysis and "Pattern matching" is used for anomaly analysis.

3 Signature Analysis and Anomaly Analysis

Firewalls cannot or do not analyze packets once they are inside the network and it only analyze them while it enters the network.[19] So, if some anomalies or activities happen from inside the network, then firewalls won't respond to those activities. But, NIDS analyze network packets internally as well as while it enters or leaves the network. With the explosive growth of networking and data sharing, NIDS have become the most popular form of Intrusion Detection[19]. A NIDS is capable of detecting network security threats. Many different NIDSs have been developed and each of them has its own advantages and disadvantages.

Signature Analysis: An NIDS use signature based detection, based on known traffic data to analyze network traffic. This type of detection is very fast, and easy to configure[20]. However, an attacker can slightly modify an attack to render it undetectable by a signature based IDS. Still, signature-based detection, although limited in its detection capability, can be very accurate. A common strategy for IDS in detecting intrusions is to memorize signatures of known attacks. These signatures are written based on data collected from known and previous attacks, and this unfortunately ensures that these signatures “will always be a step behind the latest underground exploits” [20][21].

Anomaly Analysis: Anomaly analysis[17][21] is an efficient way to detect intrusions and thus forms a vital part of the next generation Intrusion Detection Systems. The most efficient Anomaly analysis technique is the pattern based anomaly detection. In pattern based anomaly analysis, the Intrusion Detection System is given a pre-defined set of intrusion patterns. Network packets are collected for a specified period of time or till a specified number of packets. These packets are considered as a block for analysis. The predefined patterns are then matched with this packet block and if any patterns match, an alert is given. During anomaly analysis, a normal behavior model is used as the base for analyzing incoming traffic and any deviation or variation from the normal behavior model is considered as an intrusion or threat[22]. But this can produce a rather high degree of false alarms.

The following attacks are stressed upon in iNIDS.

- Denial of Service[23] - UDP Flooding[24], TCP SYN Attack[26], Smurf attacks[28]
- User to Root (U2R)[29] - Type signatures
- Remote to Local (R2L)[8]
- Probing Attack[30] - Portsweep, Satan, Nmap, etc.

Each possesses its own signature, and attack characteristics, which make it easier to detect and handle. They can sometimes be identified directly from the signature, or by using the anomaly detection methods.

4 KDDCUP99 Dataset

The “KDD CUP’99” dataset [14], which derived from the DARPA dataset, was used for the KDD (Knowledge Discovery and Data Mining Tools Conference) Cup 99 Competition. The complete dataset has around 5 million input patterns and each record represents a TCP/IP connection that is composed of 41 features. The dataset used in this study is a smaller subset (10% of the original training set), it has 494 021 instances (patterns) and it was employed as the training set in the original competition. Each record of the KDD Cup 99 dataset captures various features of the connections, as for example, the source and destination bytes of a TCP connection, the number of failed login attempts or the duration of a connection. Complex relationships exist between the features, which are difficult for human experts to discover.

An NIDS must therefore reduce the amount of data to be processed so as to maintain an accurate and real-time detection. Some input data may not be useful to the Network based IDS and thus can be eliminated before processing. In complex classification systems, the features may contain false correlations, which block the process of detecting intrusions/attacks. Furthermore, some features may be redundant since the information they add is contained in other features.[14][31]

KDD Cup 99 dataset feature selection[32] consists of detecting the relevant features and discarding the irrelevant features. Relevant features are features that can be used for analysis with ease and that can deliver relevant information as well as can work without any performance degradation.

KDDCUP99 attributes[33] can be categorized into four. They are: *Intrinsic Attributes*, *Content Attributes*, *Traffic Attributes*, and *Class Attributes*.

5 JPCAP

Jpcap(Java Packet Capturer)[34] is a Java library for sniffing, capturing and sending network packets, from an available network interface. It also facilitates visualization, creation and analysis of network packets by appropriate coding in Java. The Java language gives it the capability to work in multiple platforms (Operating systems). Jpcap has been tested on Microsoft Windows (98/2000/XP/Vista), Linux (Fedora, Mandriva, Ubuntu), Mac OS X (Darwin), FreeBSD, and Solaris and was found to be working successfully. Jpcap can capture Ethernet, IPv4, IPv6, ARP/RARP, TCP, UDP, and ICMPv4 packets[34]. It is open source, and is licensed under GNU LGPL.

6 Artificial Neural Networks

Artificial Neural Networks, also known as “Artificial Neural Nets”, “neural nets”, or ANN for short, is a computational tool modeled based on the interconnection of the biological neurons in the nervous systems of the human body. ANN can be trained / taught to solve certain type of problems using a training method and data to train. [35] By following this method, Artificial Neural Networks made can be used to perform different tasks depending on the amount training given. A properly trained ANN is capable of generalization, the ability to recognize similarities among many different number and type of input patterns, especially patterns that have been corrupted by noise which has a wide variety of applications.

Artificial Neural Networks have a number of features and characteristics that make them an attractive alternative to traditional problem-solving techniques. We can design a complex network of neurons using multiple layers[35][36], the first of which was called the *Multi-layered Perceptron*. There are a variety of thresholds[37] available for ANN, an example of which is the *sigmoid*.

No matter what organization is used, the ANN has to be trained (or it learns)[38]. It can be *supervised* or *unsupervised*. Here, Backpropagation is preferred due to its speed, which is essential in networked environments. It works on a supervised training model. So, the initial training data, extracted fron KDDCUP99 dataset is important.

7 Implementation

During development, an existing packet capture program based of JpCap was used. The anomaly analysis module was added on top of that by analyzing the parameters of the packets on the fly, and reporting any threats before the attacker has a chance to enter.

The overall system looks like below.

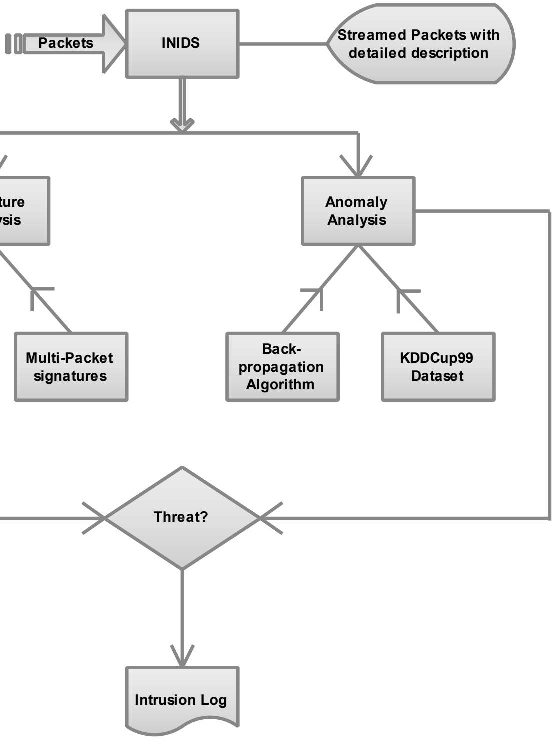


Fig. 1. Implementation

When newer threats are identified by the neural network, it implicitly trains itself to identify that pattern, and it is in turn added to the library. So, further analysis will check for this pattern also. Also, since we use neural networks, it doesn't need an exact match. Any similar pattern of that genre can be caught. So, iNIDS improves on the go, by learning newer ways in which attackers can threaten the integrity and security of the data.

8 Results

The performance was as expected during the design phase. The initial training data was taken such that there was only one sample for each type of intrusion expected.

So, the initial performance was very bad, at about 10%. But eventually, it improved by adding more patterns to the arsenal, and finally, it was able to reach accuracy of more than 70%, in just 21 runs.

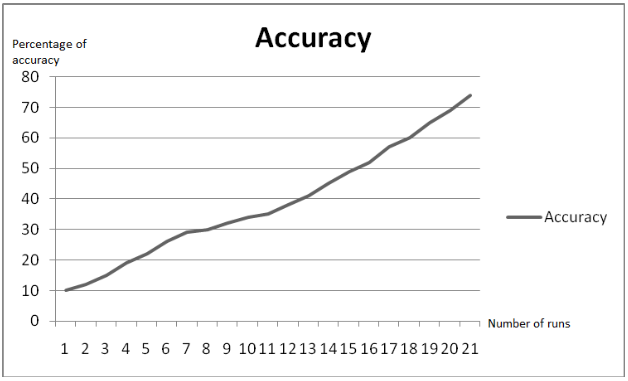


Fig. 2. Accuracy plot with smaller training data, over 21 runs

On a later run, the initial training data itself was improved, with multiple patterns per type of attack. In this case, the initial performance itself went up to 40%, and later, it improved to more than 85% in the same number of runs.

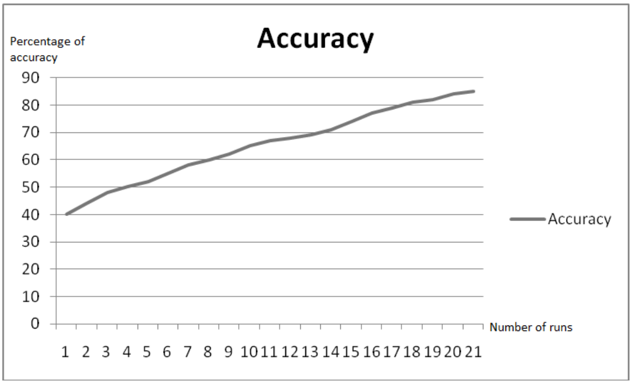


Fig. 3. Accuracy plot with more comprehensive training data, over 21 runs

9 Conclusion

iNIDS initially kicked off as an abstract deign that can improve the current packet capture and analysis routines in intrusion detection, by adding a level of knowledge and intelligence to the analysis. In this case, the threat detection is reinforced by past knowledge of similar attacks, and we are able to achieve levels of accuracy much higher than conventional systems that use the same level of knowledge.

10 Future Work

Even though the research proposes a new technique, it can't be claimed as the best possible technique. The field of network security is subject to continuous research in search for new and better techniques that can ensure better security capabilities. Applying Artificial Intelligence to an NIDS gives it the capability to detect unknown threats which is an important characteristic.

The intelligence thus given to the NIDS can be improved by the use of a better Neural Network architecture as well as a much better training algorithm than Backpropagation which is used in this implementation. A better Neural Network architecture will give better input processing capabilities such as improved speed, pattern matching etc. This will thus enable the ANN to detect patterns much more efficiently with very little data. Better training algorithms can provide reduced training time thereby allowing more training data to be used in very limited time. This will thus improve the training speed and also the training data thereby resulting in an improved performance. So a thorough research is recommended in the field of NIDSs based on Artificial Intelligence.

References

[1] Wang, Z., Wang, X.: NetFlow Based Intrusion Detection System. In: International Conference on MultiMedia and Information Technology, MMIT 2008, December 30-31, pp. 825–828 (2008)

[2] Denning, D.E.: An Intrusion-Detection Model. IEEE Transactions on Software Engineering SE-13(2), 222–232 (1987)

[3] The JavaTM Tutorials,
`http://download.oracle.com/javase/tutorial/`
(accessed August 20, 2011)

[4] Garuba, M., Liu, C., Fraites, D.: Intrusion Techniques: Comparative Study of Network Intrusion Detection Systems. In: Fifth International Conference on Information Technology: New Generations, ITNG 2008, April 7-9, pp. 592–598 (2008)

[5] Shun, J., Malki, H.A.: Network Intrusion Detection System Using Neural Networks. In: Fourth International Conference on Natural Computation, ICNC 2008, October 18-20, vol. 5, pp. 242–246 (2008)

[6] Wang, Y., Huang, G.X., Peng, D.G.: Model of Network Intrusion Detection System based on BP Algorithm. In: 2006 1st IEEE Conference on Industrial Electronics and Applications, May 24-26, pp. 1–4 (2006)

[7] Liu, B., Lin, C., Ruan, D., Peng, X.: Netfiow Based Flow Analysis and Monitor. In: International Conference on Communication Technology, ICCT 2006, November 27-30, pp. 1–4 (2006)

[8] Ahmad, I., Abdullah, A.B., Alghamdi, A.S.: Remote to Local attack detection using supervised neural network. In: 2010 International Conference for Internet Technology and Secured Transactions (ICITST), November 8-11, pp. 1–6 (2010)

[9] Stolfo, S.J., Fan, W., Lee, W., Prodromidis, A., Chan, P.K.: Cost-based modeling for fraud and intrusion detection: results from the JAM project. In: Proceedings of DARPA Information Survivability Conference and Exposition, DISCEX 2000, vol. 2, pp. 130–144 (2000)

[10] Zihao, S., Hui, W.: Network Data Packet Capture and Protocol Analysis on Jpcap-Based. In: 2009 International Conference on Information Management, Innovation Management and Industrial Engineering, December 26-27, vol. 3, pp. 329–332 (2009)

[11] Al-Shaer, E.: Managing firewall and network-edge security policies. In: IEEE/IFIP Network Operations and Management Symposium, NOMS 2004, April 23-23, vol. 1, p. 926 (2004)

[12] Yang, Y., Mi, J.: Design and implementation of distributed intrusion detection system based on honeypot. In: 2010 2nd International Conference on Computer Engineering and Technology (ICCET), April 16-18, vol. 6, pp. V6-260–V6-263 (2010)

[13] Ahmad, I., Ansari, M.A., Mohsin, S.: Performance Comparison between Backpropagation Algorithms Applied to Intrusion Detection in Computer Network Systems. In: 9th WSEAS International Conference on Neural Networks, May 2-4, pp. 47–52 (2008)

[14] KDD Cup 1999 Data (1999), http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html (accessed August 13, 2011)

[15] Lee, S.M., Kim, D.S., Park, J.S.: A Hybrid Approach for Real-Time Network Intrusion Detection Systems. In: 2007 International Conference on Computational Intelligence and Security, December 15-19, pp. 712–715 (2007)

[16] Abdel-Azim, M., Abdel-Fatah, A.I., Awad, M.: Performance analysis of artificial neural network intrusion detection systems. In: International Conference on Electrical and Electronics Engineering, ELECO 2009, November 5-8, pp. II-385–II-389 (2009)

[17] Yu, X.: A new model of intelligent hybrid network intrusion detection system. In: 2010 International Conference on Bioinformatics and Biomedical Technology (ICBBT), April 16-18, pp. 386–389 (2010)

[18] Successful software development - Google Books, http://books.google.co.uk/books?id=lrix5MNRiu4C&pg=PA184&dq=software+development+life+cycle+preliminary+design+detailed+design&hl=en&ei=0h1SToP2GdKwhAf9rNHaBg&sa=X&oi=book_result&ct=result&resnum=1&ved=0CDAQ6AEwAA#v=onepage&q=software%20development%20life%20cycle%20preliminary%20design%20detailed%20design&f=false (accessed August 22, 2011)

[19] Network-Based IDS (NIDS) overview | IDStutorial, http://idstutorial.com/network-based-ids.php (accessed August 13, 2011)

[20] Wu, T.M.: Intrusion Detection Systems, September 25 (2009), http://iac.dtic.mil/iatac/download/intrusion_detection.pdf (accessed August 12, 2011)

[21] SANS institute, Host- vs. Network-Based Intrusion Detection Systems (2005), http://www.giac.org, http://www.giac.org/paper/gsec/1377/host-vs-network-based-intrusion-detection-systems/102574 (accessed August 12, 2011)

[22] Zhang, W., Yang, Q., Geng, Y.: A Survey of Anomaly Detection Methods in Networks. In: International Symposium on Computer Network and Multimedia Technology, CNMT 2009, January 18-20, pp. 1–3 (2009)

[23] Gill, K., Yang, S.-H.: A scheme for preventing denial of service attacks on wireless sensor networks. In: 35th Annual Conference of IEEE Industrial Electronics, IECON 2009, November 3-5, pp. 2603–2609 (2009)

[24] Chang, R.K.C.: Defending against flooding-based distributed denialof-service attacks: a tutorial. IEEE Communications Magazine 40(10), 42–51 (2002)

[25] IntelliGuard I.T. - Eliminate DDoS and Flash crowd problems, `http://www.intelliguardit.net/library_attackscenarios.html` (accessed August 19, 2011)

[26] Wang, H., Zhang, D., Shin, K.G.: Detecting SYN flooding attacks. In: Proceedings of the Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies, p.1530 (June 2002)

[27] ASA/PIX 7.x and Later: Mitigating the Network Attacks - Cisco Systems, `http://www.cisco.com/en/US/products/ps6120/products_tech_note09186a00809763ea.shtml` (accessed August 19, 2011)

[28] Mirkovic, J., Reiher, P.: A taxonomy of DDoS attack and DDoS defense mechanisms. ACM SIGCOMM Computer Communication Review, 39–53 (2004)

[29] Ahmad, I., Abdullah, A.B., Alghamdi, A.S.: Applying neural network to U2R attacks. In: 2010 IEEE Symposium on Industrial Electronics & Applications (ISIEA), October 3-5, pp. 295–299 (2010)

[30] Ahmad, I., Abdullah, A.B., Alghamdi, A.S.: Application of artificial neural network in detection of probing attacks. In: IEEE Symposium on Industrial Electronics & Applications, ISIEA 2009, October 4-6, vol. 2, pp. 557–562 (2009)

[31] MIT Lincoln Laboratory: Communication Systems and Cyber Security: Cyber Systems and Technology: DARPA Intrusion Detection Evaluation, `http://www.ll.mit.edu/mission/communications/ist/corpora/ideval/data/index.html` (accessed August 14, 2011)

[32] Bolon-Canedo, V., Sanchez-Maroo, N., Alonso-Betanzos, A.: A combination of discretization and filter methods for improving classification performance in KDD Cup 99 dataset. In: International Joint Conference on Neural Networks, IJCNN 2009, June 14-19, pp. 359–366 (2009)

[33] Index of /acwaldap/gureKddcup, `http://www.sc.ehu.es/acwaldap/gureKddcup/README.pdf` (accessed August 14, 2011)

[34] Jpcap - a Java library for capturing and sending network packets, `http://netresearch.ics.uci.edu/kfujii/Jpcap/doc/index.html` (accessed August 14, 2011)

[35] Artificial Neural Networks/Neural Network Basics - Wikibooks, open books for an open world, `http://en.wikibooks.org/wiki/Artificial_Neural_Networks/Neural_Network_Basics` (accessed August 14, 2011)

[36] Advances in Data Analytical Techniques, `http://www.iasri.res.in/ebook/EBADAT/5-Modeling%20and%20Forecasting%20Techniques%20in%20Agriculture/5-ANN_GKJHA_2007.pdf` (accessed August 14, 2011)

[37] Neural Networks: Tutorials: Paras Chopra, `http://paraschopra.com/tutorials/nn/index.php` (accessed August 14, 2011)

[38] Basic Concepts for Neural Networks, `http://www.cheshireeng.com/Neuralyst/nnbg.html` (accessed August 14, 2011)

Mutual Authentication for Wireless Communication Using Elliptic Curve Digital Signature Based on Pre-known Password

Tumpa Roy[1], Poonam Sisodia[1], Divye Upadhyay[1], and Kamlesh Dutta[2]

[1] GLNA Institute of Technology, Mathure- 281406
tumpa.nit@gmail.com,
poonamsisodi22@gmail.com,
divs08divyaa@gmail.com
[2] National Institute Of Technology Hamirpur
Hamirpur, Himachal Pradesh-177005
India
kdnith@gmail.com

Abstract. The appearance of public access wireless networks enables ever-present Internet services, whereas it inducing more challenges of security due to open air mediums. As one of the most widely used security mechanisms, authentication is provide for secure communications by preventing unauthorized usage and negotiating credentials for verification. In the intervening time, it generates heavy overhead and delay to communications, further deteriorating overall system performance. First, a system model based on challenge/response authentication mechanism by using the elliptic curve cryptographic digital signature is introduced, which is wide applied in wireless environment to reduce the computational cost, communication bandwidth and the server overload . Then, the concept of security levels is proposed to describe the protection of communications with regard to the nature of security.

Keywords: Elliptic curve cryptography (ECC), security, wireless communication, Public key cryptography (PKC), Authentication, verification.

1 Introduction

You Wireless communications is advancing rapidly in recent years. After 2G (e.g. GSM) widely deployed in the world, 3G mobile communication systems are spreading step by step in many areas. At present, some countries have already launched investigations beyond 3G (B3G) and 4G. Along with the wireless communications' rapid development, the secure access authentication of the users within wireless networks is becoming very critical, and so, more and more attention is focused on it. As the wireless industry explodes, it faces a growing need for security. Applications in sectors of the economy such as healthcare, financial services, and government depend on the underlying security already available in the wired computing environment. Both for secure (authenticated, private) Web transactions and for secure (signed, encrypted) messaging, a full and efficient public key

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 11–22.
springerlink.com

infrastructure is needed. Three basic choices for public key systems are available for these applications:

- RSA
- Diffie-Hellman (DH) or Digital Signature Algorithm (DSA) modulo a prime p
- Elliptic Curve Diffie-Hellman (ECDH) or Elliptic Curve Digital Signature Algorithm (ECDSA).

RSA is a system that was published in 1978 by Rivest, Shamir, and Adleman, based on the difficulty of factoring large integers. Whitfield Diffie and Martin Hellman proposed the public key system now called Diffie-Hellman Key Exchange in 1976. DH is key agreement and DSA is signature, and they are not directly interchangeable, although they can be combined to do authenticate key agreement. Both the key exchange and digital signature algorithm are based on the difficulty of solving the discrete logarithm problem [15] in the multiplicative group of integers modulo a prime p. Elliptic curve groups were proposed in 1985 as a substitute for the multiplicative groups modulo p in either the DH or DSA protocols. For the same level of security per best currently known attacks, elliptic curve based systems [7,10] can be implemented with much smaller parameters, leading to significant performance advantages. Such performance improvements are particularly important in the wireless arena where computing power, memory, and battery life of devices are more constrained. In this article we will highlight the performance advantages of elliptic curve systems [8] by comparing their performance with RSA in the context of protocols from different standards.

Authentication is the act of establishing or confirming something as authentic, that is, that claims made by or about the subject are true. There are several methods concerning strong authentication. The main difference consists whether secret-key or public-key cryptography is used. In secret-key cryptography the signer and the verifier must share a secret where the problem of the key exchange must be solved. The main difference consists whether secret-key or public-key cryptography is used. In secret-key cryptography the signer and the verifier must share a secret where a public key is distributed for signature verification. The method using public-key cryptography is known as a digital signature. The protocols used for authentication consists of zero-knowledge protocols and challenge-response protocols. The Diffie-Hellman protocol [9] is used in wireless communication.

Deffie-hellman algorithm has five parts:

1. Global Public Elements
2. User A Key Generation
3. User B Key Generation
4. Generation of Secret Key by User A
5. Generation of Secret Key by User B

Global Public Elements:

q is a Prime number
$\alpha, \alpha < q$ and α is a primitive root of q

The global public elements are also sometimes called the domain parameters.

User A Key Generation:

Select private X_A, where $X_A < q$
Calculate public Y_A ,where $Y_A = \alpha .X_A \bmod q$

User B Key Generation:

Select private X_B, where $X_B < q$
Calculate public Y_B, where $Y_B = \alpha. X_B \bmod q$

Generation of Secret Key by User A:

$K = (Y_B).X_A \bmod q$

Generation of Secret Key by User B:

$K = (Y_A).X_B \bmod q$

If user A and user B are genuine then they can communicate to each other. The ECC version of algorithm is used in wireless communication for authentication proof.

2 Preliminaries

2.1 Elliptic Curve Cryptography

Elliptic curves [11] take the general form of the equation:

$$Y_2 + axy + by = x_3 + cx_2 + dx +e$$

where a, b, c, d and e are real numbers satisfy some conditions which depends on the field it belongs to, such as real number or finite field. Finite field may be F(p) or $F(2^m)$

The F(p) Field:

The elements of Fp [13] should be represented by the set of integers: {0, 1,. . . p – 1} With addition and multiplication defined as follows:

Addition: If a, b $\in$ F(p), then $a + b = r$ where r is the remainder of the division of a + b by p and $0< r < p-1$. This operation is called addition modulo p.

Multiplication: if a, b $\in$ F(p), then $a . b = s$ where s is the remainder of the division of a . b by p and $0< s < p-1$. This operation is called multiplication modulo p.

The $F(2^m)$ Field:

The elements of $F(2^m)$ should be represented by the set of binary polynomials of degree m – 1 or less: $a = \alpha_{m-1}x^{m-1} + \ldots + \alpha_1 x + \alpha_0$ with addition and multiplication defined as follows:

Addition: $a + b = c = \{_{cm-1},..c_1,c_0\}$ where $c_i = (a_i + b_i) \bmod 2$. $c \in F(2^m)$.

Multiplication: $a . b = c = \{c_{m-1},..c_1,c_0\}$ where c is the remainder of the division of the polynomial a(x) . b(x) by an irreducible polynomial of degree m. $c \in F(2^m)$.

There is a point 0 called the point at infinity or the zero point [12]. The basic operation of elliptic curve is addition. The addition of two distinct points on elliptic curve can be illustrated by the following figure [3] (figure 1):

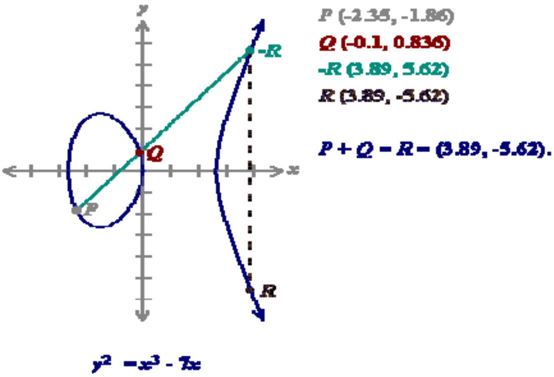


Fig. 1.

Elliptic Curve over F(p):

Let F(p) be a finite field, p > 3, and let a, b ∈ F(p) are constant such that

$$4a^3 + 27b_2 \equiv 0 \text{ (mod p)}.$$

An elliptic curve, E(a,b)(F(p)), is defined as the set of points (x,y) ∈ F(p) * F(p) which satisfy the equation

$$y^2 \equiv x^3 + ax + b \text{ (mod p)}$$

together with a special point, O, called the point at infinity.

Elliptic Curve over F(2m) for some m ≥ 1. :

Elliptic curve E(a,b)(F(2m)) [14] is defined to be the set of points (x,y) ∈ F(2m) * F(2m) which satisfy the equation

$$y^2 + xy = x^3 + ax^2 + b;$$

where a, b ∈ F(2^m) and b≠0, together with the point on the curve at infinity, O.The points on an elliptic curve form an abelian group under a well defined group operation. The identity of the group operation is the point O.

P and Q be two points on E(a,b)(F(p)) or F(2m) and O is the point at infinity.

P+O = O+P = P

If P = (x_1,y_1) then -P = (x_1 ,-y_1) and P + (-P) = O.

If P = (x_1,y_1) and Q = (x_2,y_2), and P and Q are not O.

Then P +Q = (x_3 ,y_3) where

$x_3 = \lambda_2 - x_1 - x_2$,

$y_3 = \lambda(x_1 - x_3) - y_1$ and

$\lambda = (y_2-y_1)/(x_2-x_1)$ if $P \neq Q$; $\lambda = (3x_1^2+a)/ 2y_1$ if $P = Q$

2.2 Elliptic Curve Digital Signature Algorithm

The private key in DSA is a number X. It is known only to the signer. The public key in DSA consists of four numbers:

P = a prime number, between 512 and 1024 bits long
Q = a 160-bit prime factor of P-1.
G = h(P - 1)/Q, where H< P -1 and G mod Q> 1.
Y = G X mod P, which is a 160-bit number.

A signature on a document's hash value H consists of two numbers R and S:

R = (G K mod P) mod Q, where K is a randomly-chosen number <Q.
S = (K -1 (H + XR)) mod Q

To verify the signature, a recipient must compute a value V from the known information:

W = S -1 mod Q
U1 = HW mod Q
U2 = RW mod Q
V = ((G U1 Y U2) mod P) mod Q

If V = R, then document was signed by the person with the public key (P, Q, G, Y). The security of DSA is based on the computational infeasibility of finding a solution for the equation S = (K -1 (H + XR)) mod Q, when X is not known.

3 Proposed Protocol

Choosing a finite field Fq. An elliptic curve E defined over Fq with large group order and a point P of large order n is selected and made public, where n is a prime number. Zn is a class of modulo n, where n is the order of p over E(Fq). Given r,t ∈ Zn , where r+t = o mod n, r is called the additive inverse of t and denoted as r =-t mod n. the server and client share a secret password S and a secret key K. the server and client individually compute two integers t and r. t is derived from Sand (n-1) in any predetermined way and it yields a unique value. The whole protocol divided into two phases:

Key establishment phase,
Verification phase.

3.1 Key Establishment Phase

The steps of the key establishment phase are explain bellow:

e.1 the client choose a random integer r_c which is belongs in between 1 to n-1 ie. $r_c \in (1, n\text{-}1)$. And compute $Q_c = (r_c + t)P$. the client send Q_c to the server.

e.2 The server then select a random integer r_s which is belongs in between 1 to n-1 ie. $r_s \in (1, n\text{-}1)$. And compute $Q_s = (r_c + t)P$. the server send Q_s to the client.

e.3 client compute $X = Q_s + rP$

$=(r_s+t)P+rP$
$= r_sP+tP+(-t)P$
$=r_sP$

And compute the session key $K_c=r_cX=r_cr_sP$

e.4. Server compute $Y=Q_c+rP$
$=(r_c+t)P+rP$
$= r_cP+tP+(-t)P$
$=r_cP$

And compute the session key $K_s=r_sY=r_cr_sP$

The session key computed by the server and client individually are same ie. Kc=Ks.

The figure 2 show the key establishment procedure between the client and server.

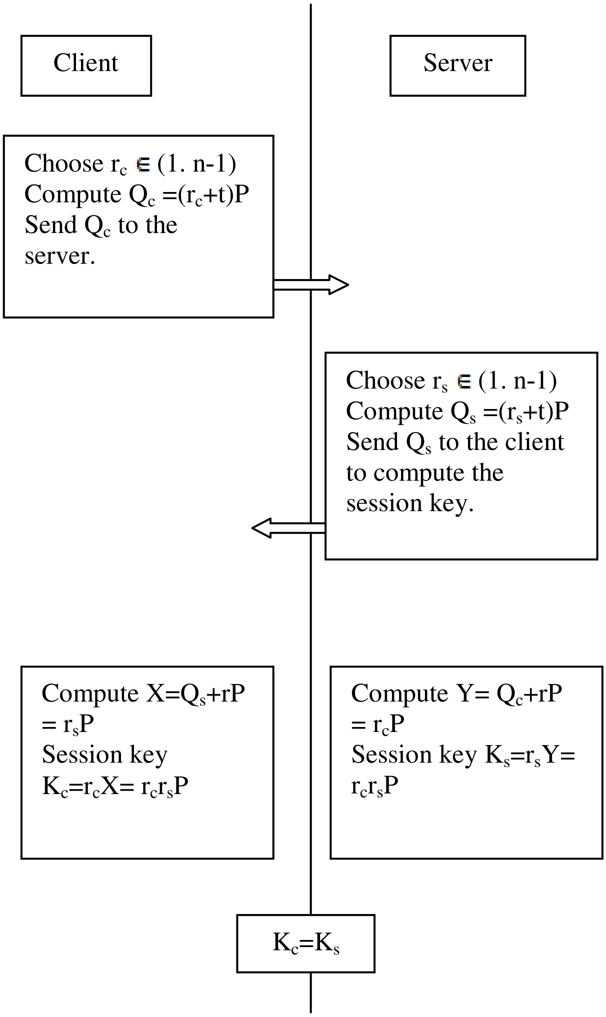

Client
Server
Choose $r_c \in (1. n-1)$
Compute $Q_c =(r_c+t)P$
Send Q_c to the server.
Choose $r_s \in (1. n-1)$
Compute $Q_s =(r_s+t)P$
Send Q_s to the client to compute the session key.
Compute $X=Q_s+rP$
$= r_sP$
Session key
$K_c=r_cX= r_cr_sP$
Compute $Y= Q_c+rP$
$= r_cP$
Session key $K_s=r_sY= r_cr_sP$
$K_c=K_s$

Fig. 2. Key Establishment Phase

3.2 Verification Phase

v.1 The client compute $K*K_c=K*r_c *r_s*P$ where K is the secret key which is known by the server and client. Client send the $K*K_c$ to the server to proof its validation.

v.2 server checks whether $K*K_c=K*K_s$ hold or not. if it dose server believes that it and the client have obtain the same session key i.e $K_c=K_s$ and the client is not duplicate because it knows the secret key which is only known by the server and the clinent. Since the server knows r_s , it believes it has obtain the accurate Q_C. Since client knows r_c , server believes client obtain the correct Q_c ie server condensed that the K_s is valid and the server compute $K*Q_c$ and send it to the client.

v.3 client checks $K*Q_c$. If $K*Q_c$ is correct, client believes that B has obtain the correct Q_c . since only server knows the the secret key K which is shared between the server and client and t is known by the server. So the server is not duplicate. The server knows the t beside client. Client believes that it has obtain the correct Qs and they have obtain the same session key $K_c=K_s$. Client convinced that the K_c is valid.

After the verification procedure has been completed by both sides, the client and the server are now ready to use the session key.

The figure 2 show the Verification procedure between the client and server.

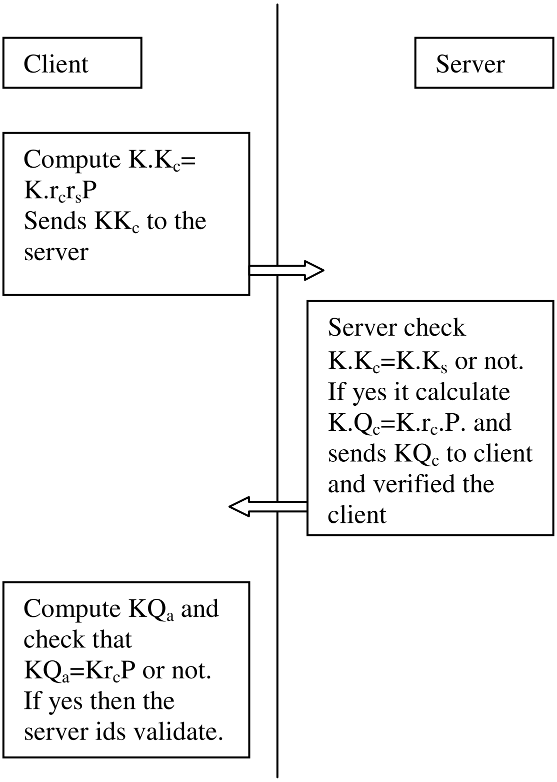

Client
Server
Compute $K.K_c$= $K.r_cr_sP$
Sends KK_c to the server
Server check $K.K_c=K.K_s$ or not.
If yes it calculate $K.Q_c=K.r_c.P$. and sends KQ_c to client and verified the client
Compute KQ_a and check that $KQ_a=Kr_cP$ or not.
If yes then the server ids validate.

Fig. 3. Validation Phase

4 Analysis of the Proposed Protocol

4.1 Security Analysis

In this section, we scrutinize our proposed key agreement protocol in detail so as to ensure that our protocol is able to achieve the desired security attributes of a key agreement protocol and also able to resist against the known cryptographic attacks.

Known session key security (KSK-S). As shown in our protocol description, the session key is derived from the ephemeral keys (r_c, r_s) of the specific session and the long term keys (S,K) of the protocol entities. This would result in distinct independent session key in each protocol execution. On top of that, a one-way collision-resistant cryptographic function is used to derive the session key. Thus, obtaining any other session keys would not benefit the adversary in mounting a successful attack against a protocol run without the information set (r_c ,r_s),(t,r) which is required in the computation of the shared secret K. Therefore, we claim that the knowledge of some previous session keys would not allow the adversary to gain any advantage in deriving any future and other previous session keys.

Weak Perfect Forward secrecy (wPFS). Suppose that both client and server's long term secret key and password S and K have been exposed. However, the adversary, with the eavesdropped information of any particular session, would not be able to recover the respective established session key since the adversary does not know the involved ephemeral private key r_c or r_s which are needed in the computation of the shared secret K_c and K_s . And also, the intractability of ECCDHP has significantly thwarted the adversary's attempt in computing K_c and K_s by using S and K . Hence, we claim that our enhanced protocol enjoys weak perfect forward secrecy.

Key-Compromise Impersonation Resilience (KCI-R). Suppose that client's and server's long term private key S, K has been compromised and instead of directly impersonating client, the adversary now wishes to impersonate server in order to establish a session with client. However, the adversary is unable to compute the shared secret K_c with the available information (S, r_s, K) since the required information set is (r_c,S, K). Hence, the adversary is significantly prevented from launching a successful KCI attack against our protocol. Generally, the same situation will result when the long term key S is compromised (the adversary would impersonate client in this case and her effort will be foiled in computing K_c as our key agreement protocol is symmetric. As a result, we claim that this protocol is able to withstand the KCI attack under all circumstances.

Key Replicating Resilience (KR-R). The key replicating attack was first introduced by Krawczyk [1] where the illustration of it involves oracle queries described in Bellare and Rogaway's random oracle model [2,3]. This attack, if successfully carried out by the adversary, would force the establishment of a session, K (other than the Test session or its matching session) to agree on a same session key as the Test session, by means of intercepting and altering the message from both communicating parties during transmission. Since the Test session and K are non-matching, the adversary may issue a Reveal query to the oracle associated with K and she can then distinguish whether the Test session key is real or random. Notice that the message integrity of Qc and Qs has been guaranteed by having each party to calculate K_c and K_s which will be bound to X and Y respectively. Since the adversary has no idea in forging X or Y along with Q_c or Q_b she would not be able to force the establishment of non matching sessions to possess a common session key. As a result, if the adversary reveals client's session key, she would not be able to guess server's session key correctly with non-negligible probability and vice versa. Therefore, we claim that our protocol is secure against the key replicating attack.

Replay Resilience (R-R). In any protocol run, an adversary may attempt to deceive a legitimate participant through retransmitting the eavesdropped information of the impersonated entity from a previous protocol execution. Although the adversary might be unable to compute the session key at the end of the protocol run, her attack is still considered successful if she manage to trick the protocol entity to complete a session with her, believing that the adversary is indeed the impersonated party. In this replay analysis, we reasonably assume that the prime order n of point P is arbitrarily large such that the probability of a protocol entity selecting the same ephemeral key (r_c , $r_s \in [1, n - 1]$) in two different sessions is negligible. Consider a situation where the adversary (masquerading as A) replays A's first message from a previous protocol run between client and server. After server has sent her a fresh (Q_s, Y) in the second message flow, the adversary would abort since she could not produce (by means of forging or replaying) X corresponding to Q_s . Notice that the same replay prevention works in the reverse situation where server's message is replayed. The adversary would fail eventually in generating Y corresponding to the fresh Q_c. Hence, we claim that message replay in our protocol execution can always be detected by both client and server.

Identity authentication. On the one hand, assuming Eve can impersonate B. When Eve receives Q_c, E →A: $Q_e = (r_e + t)P$. But Eve does not know t and re, and she cannot make the validation message KrcrsP, thus the key validation fails. On the other hand, with (v.2) and (v.3), A and B believe that only knowing t can generate the valid validation messages.

Man-in-the-middle attacks. In the original Diffie-Hellman protocol, Eve can alter the public values such as g_a mod n or g_b mod n with her own values. Thus Eve can share session keys with client or server. In our protocol, when Eve receives $Qc= (r_c+ t)P$, she cannot guess rc and t. If she still tries to eavesdrop, she mus t generate $r_eP = (r^{c'} + t)P$ and send it to server; server will obtain a wrong value rc'rsP, which is impossible for Eve to know. Thus Eve cannot share a session key with server or client. Based on ECDH algorithm [4], our protocol with pre-shared password is proposed. It makes use of the difficulty of computing discrete logarithms over elliptic curves. It provides identity authentication, key validation and perfect forward secrecy, and it can foil man-in-the-middle attacks.

4.2 Performance Analysis

Efficiency Analysis

Atay et. al. have conducted detailed studies on Computational Cost Analysis of Elliptic Curve Arithmetic [5]. They have reported the point addition arithmetic is applied on two and three dimensional coordinate systems. The computational cost of each arithmetic operation should be taken into consideration in order to compare the efficiency of algorithms in different coordinate systems. The efficiency is measured as the computational cost, which is in terms of elapsed time. The measured units in Fig. 4 [10] are as follows:

1. Inversion (I) is the multiplicative inverse in modular arithmetic. It has the highest computational cost and one inversion is approximately equals nineteen times of the cost of multiplication cost and denoted as 1I =19M .
2. Multiplication (M) has a lower cost than inversion; therefore all inversions should be converted either to multiplication or to addition.
3. Addition (A) and subtraction (S) have the lowest cost, therefore omitted.

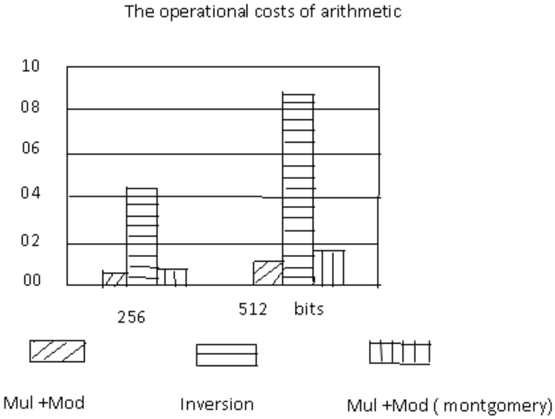


Fig. 4. The operational cost of Arithmetic operation

Computational Cost Analysis
The major advantage of ECC over RSA is ECC needs less computation than RSA but still can achieve the same or even higher level of security. Table 1[6] gives cost-equivalent key sizes. It gives the size, in bits, for equivalent keys. The time to break is computed assuming a machine can break a 56-bit DES key in 100 seconds, and then scaling accordingly.

Table 1.

ECC key	RAS key	Time to break	Machines	Memory
112	430	<5 minutes	105	`Trivial
106	760	600 months	4300	4Gb
192	1020	3 million years	114	170 Gb
256	1620	10^16 years	16	120Gb

5 Conclusion

Attack that monitor side-channel information, Key Replicating Resilience (KR-R).the key replicating attack was first introduced by Krawczyk have recently been receiving much attention in wireless communication. The result presented in this paper conform that the key replacing attack quite powerful and need to be addressed. Any addition to memory or processing capacity increases the cost of each card. ECC needs less computation power, thus it is more suitable than RSA. We have described an authentication and key agreement protocol for wireless communication based on elliptic curve cryptographic techniques. The proposed protocol is a public key type with the feature of signature generation procedure. The new protocols are based on previous classic authentication protocols, including the protection of integrity and session key exchange. This can be used to provide the integrity of the data being transferred during the authentication process in order to prevent from active attacks. The smaller key sizes result in smaller system parameters, smaller public key signatures, bandwidth savings, faster implementations, and smaller hardware processors.

References

1. Krawczyk, H.: HMQV: A High-Performance Secure Diffie-Hellman Protocol. In: Shoup, V. (ed.) CRYPTO 2005. LNCS, vol. 3621, pp. 546–566. Springer, Heidelberg (2005)
2. Bellare, M., Rogaway, P.: Provably Secure Session Key Distribution: The Three Party Case. In: 27th ACM Symposium on the Theory of Computing - ACM STOC, pp. 57–66 (1995)
3. Blake-Wilson, S., Johnson, D., Menezes, A.: Key Agreement Protocols and their Security Analysis. In: Darnell, M.J. (ed.) Cryptography and Coding 1997. LNCS, vol. 1355, pp. 30–45. Springer, Heidelberg (1997)
4. Miller, V.S.: Use of Elliptic Curves in Cryptography. In: Williams, H.C. (ed.) CRYPTO 1985. LNCS, vol. 218, pp. 417–426. Springer, Heidelberg (1986)
5. Atay, S., Koltuksuz, A., Hışıl, H., Eren, S.: Computational Cost Analysis of Elliptic Curve Aurithmetic. In: International Conference on Brid Information Technology, ICHIT 2006, vol. 1, pp. 578–582 (2006)
6. Chatterjee, K., Gupta, D.: Secure access of smart cards using Elliptic curves Cryptography. In: 5th International Conferences on Wireless Communications, Networking and Mobile Computing, WiCom 2009, pp. 1–4 (2004)
7. Blake, I.F., Seroussi, G., Smart, N.P.: Elliptic Curves in Cryptography. London Math. Soc. Lecture Note Series, vol. 265. Cambridge Univ. Press (2000)
8. Aydos, M., Sunar, B., Koç, Ç.K.: An Elliptic Curve Cryptography Based Authentication and Key Agreement Protocol for Wireless Communication. In: 2nd Int. Workshop Discrete Algorithms and Methods for Mobility, DIALM 1998, Dallas, TX (1998)
9. Diffie, W., Hellman, M.: New Directions in Cryptography. IEEE Transactions on Information Theory 22(6), 644–654 (1976)
10. Strangio, M.A.: Efficient Diffie-Hellmann Two-Party Key Agreement Protocols based on Elliptic Curves. In: Proceedings of the 2005 ACM Symposium on Applied Computing, pp. 324–331 (2005)

11. Miller, V.S.: Use of Elliptic Curves in Cryptography. In: Williams, H.C. (ed.) CRYPTO 1985. LNCS, vol. 218, pp. 417–426. Springer, Heidelberg (1986)
12. Blake, I.F., Seroussi, G., Smart, N.P.: Elliptic curves in cryptography. Cambridge University Press, New York (1999)
13. Chen, L., Yanpu, C., Zhengzhong, B.: An implementation of fast algorithm for Elliptic Curve Cryptosystem over GF(p). Journal of Electronics 21(4), 346–352 (2004)
14. Morales-Sandoval, M., Feregrino-Uribe, C., Cumplido, R., Algredo-Badillo, I.: An area/performance trade-off analysis of a GF(2m) multiplier architecture for Elliptic Curve Cryptography. Computers and Engineering 35, 54–58 (2009)
15. Smart, N.P.: The discrete logarithm problem on elliptic curves of trace one. Journal of Cryptology 12(3), 193–196 (1999)
16. Lin, X., Wong, J.W., Kou, W.: Performance analysis of secure web server based on SSL, pp. 249–261. Springer, Heidelberg (2000)

Securing Multi-agent Systems: A Survey

S.V. Nagaraj

Dept. of Computer Science
RMK Engineering College
RSM Nagar, Kavaraipettai 601206, India
http://www.rmkec.ac.in

Abstract. We look at various security aspects of multi-agent systems that are often overlooked by developers designing such systems. We look at some of the key security challenges in multi-agent systems. We focus on techniques that help to ensure that security of multi-agent systems is not compromised.

Keywords: Security, Multi-agent systems, Design, Implementation, Applications.

1 Introduction

A multi-agent system is a system made up of multiple interacting agents that are intelligent (see [33], [69]). Agent technologies have rapidly moved from the research laboratory to industrial application over the last few years [62]. Mobile agents are agents with the ability to migrate from one host to another where they can resume their execution. Security issues pertaining to mobile agents have been studied in [9], [21], [25], [32], [36], [37], [41], [45], [46], [64], [71]. Many of these issues also apply to multi-agents systems. The development of secure protocols for mobile agent computation against static, semi-honest or malicious adversaries without relying on any third party or trusting any specific participant in the system is quite challenging. There are few results in this direction [71]. Security issues related to the protection of host resources as well as the agents themselves form a major obstacle in the application of the agent paradigm [36]. It is important to protect servers from malicious agents and likewise agent data from tampering by malicious servers. Multi-agent systems are quite different from systems that make use of stand alone agents. There seems to be a lack of trust in multi-agent systems that are being developed or that have been already deployed mainly due to the security issues involving them not being addressed properly. Hence, it becomes necessary to establish the confidence of users in multi-agent systems. We look at those aspects pertaining to the security of multi-agent systems that are often overlooked by developers of such systems.

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 23–30.
springerlink.com

2 Security Aspects often Overlooked When Designing Multi-agent Systems

Numerous multi-agent based systems are being developed with practical applications such as the multi-agent based marketplace [34]. However, in such systems often the developers tend to overlook important security features. This only leads to loss of confidence in such systems. Standard mechanisms for specifying security in multi-agent systems must be developed [52]. It is often said that agent-oriented software engineering methodologies have not integrated security concerns throughout their development phase (see [48]).

The integration of security concerns during the whole range of the development stages will be beneficial for the development of more secure multi-agent systems [49]. However, such an integration is no easy task (see [48]). Methodologies such as secure Tropos have been proposed for this integration to take place (see [48]). Here too modeling can play a vital role. It is often stated that security requirements of any system (and multi-agent systems in particular) must also be addressed in the early stages of system development. They must also be addressed throughout the development of the system [35]. Analysis of the security requirements of multi-agent systems is often neglected (see [10], [11]). The usual approach towards the inclusion of security within a system is to identify security requirements after the definition of the system. This has aggravated the growth of computer systems impaired with security vulnerabilities. An analysis of the security requirements helps in identifying security bottlenecks.

In many real-life applications such as a multi-agent based e-learning environment [68] and health-care applications such as the one discussed in [4] security is paramount. Agents themselves need to be protected from each other as well from the systems in which they will be deployed. Agents must be protected from the vulnerabilities of the systems in which they will be deployed. It is important to secure agents and systems from malicious agents. Sandboxing is a technique often used in providing a secure execution environment for agents. However, sandboxing has certain limitations too. Often, encryption is not employed for securing agent communication. Models for secure communication between agents have been studied in [15], [50]. Such models can provide guarantees for code, data and execution integrity, data privacy and prevention from malicious routing. Models for agent coordination, authentication and authorization have been studied by [13]. Privacy and integrity of information is not often provided when it is needed.

It is very important to protect agents from the misbehavior of other agents. Such unacceptable behaviors may include denial-of-service by malicious agents as well as spying by such agents. Many systems that are otherwise good do not have the provision for strong authentication and authorization mechanisms [66]. This greatly affects their usability. Very often in multi-agent systems simple access control models are employed. Such models may be unsatisfactory.

The issue of trust amongst agents is often not clearly established in many multi-agent systems (see [53]). It is necessary to establish clearly which agents can be trusted and which agents cannot be trusted. Very often in multi-agent systems there is a strong reliance on the security mechanisms of the underlying operating system. Often multi-agent systems seem to rely heavily on the security policies of the underlying operating system: for example, for file access control. If the security provided by the underlying operating system is weak then the multi-agent system also provides weak security.

In some situations, the confidentiality of communication between agents may need to be ensured by employing encryption. However, this is often not done. Some multi-agent systems may require non-repudiation to be in built. However, the designers may not be offering non-repudiation. Digital signatures need to be employed in e-commerce multi-agent system applications. Secure transport should be ensured by using appropriate protocols. X.509 certificates need to be used for security purposes. A secure execution environment for agents needs to be established. Fault tolerance mechanisms should be included in safety-critical applications. In some applications, there may be a need for public key infrastructures. This is especially true in case of e-commerce applications. Role-based access control mechanisms can be set up so that access is controlled to the extent it is required. Privilege management is an often overlooked aspect in many software applications and this also applies to those involving multi-agent systems. The principle of least privilege is often not used by developers in order to ensure better security. Protocols that are often employed to provide security include IPSec, SSL, and TLS.

It is expected that unauthorized access to information must be prevented. Unauthorized alteration of data should also be prevented. In some multi-agent based systems, man-in-the-middle attacks should be overcome. The use of cryptography should be considered in multi-agent systems when confidentially becomes critical. Some multi-agent systems may require single sign-on since users may not prefer to login too many times. Single sign-on in case of Web services based applications is not easy to achieve in terms of implementation as there are many complexities involved. Traditionally, security is provided by making use of firewalls, proxies, intrusion detection systems, and intrusion prevention systems. In e-health applications such as those discussed in [4] and [70], the confidentiality of health records must be maintained.

In many applications it is necessary to maintain the confidentiality of agent interactions [6]. Delegation is frequently not done in a proper fashion. Insecure delegation leads to security breaches. Accountability should be ensured within multi-agent systems. Modal logic has been employed in [40] for characterizing the relationship among trust, information acquisition and trust in multi-agent systems. Public key infrastructure (PKI) may be used as in [31] for access control and delegation purposes. Identity certificates may be used as explained in [31] for authentication of agents whereas authorization certificates may be used for authorization of agents. PKI may also be useful for authentication purposes (see [23]).

Table 1. Security aspects often overlooked when designing multi-agent systems

No.	Aspect
1	Authentication
2	Authorization
3	Confidentiality
4	Non-repudiation
5	Agent Integrity
6	Message Integrity
7	Host Integrity
8	Message Privacy
9	Trust
10	Availability
11	Delegation
12	Integration of security concerns during the development phase
13	Analysis of security requirements
14	Protection from the vulnerabilities of the underlying system
15	Secure execution environment for agents
16	Spying by malicious agents
17	Fault tolerance
18	Least privilege
19	Reliance on the security mechanisms of the underlying operating system

3 Challenges in Multi-agent Systems

Malicious hosts pose a major problem for agents (refer [8], [28], [29], [45]). Many experts believe that this problem has no easy solution. A malicious host can observe code, data, and control flow. Malicious hosts may also manipulate code, data and control flow. They may also alter the routes of agents. Malicious hosts can cause incorrect execution of code and sometimes re-execution of code. Such hosts may also deny execution by agents. This may be in entirety or perhaps partially. Malicious hosts may also pretend as if they are some other hosts. Communication between agents may also be observed by malicious hosts. Communication between agents is also vulnerable to manipulation by malicious hosts.

There are many threats in multi-agent systems. Some of the common threats include: man-in-the-middle attacks, modification of data, replay attacks, breaking crypto-systems by deriving private key data from public key data, and denial of service attacks. So it becomes important to develop countermeasures to deal with such threats (refer [14], [16]). Attack trees may be used for identifying possible attacks. Ensuring privacy in multi-agent applications handling sensitive personal data is a key challenge [22].

4 Conclusion

We have seen various security aspects of multi-agent systems that are often overlooked by developers. It is important to ensure the security of multi-agent systems. Often the success or failure of multi-agent systems depends on the robustness of the security provided by them.

Acknowledgment. The author is thankful to his institution for encouragement.

References

1. Special issue on multi-agent and distributed information security. IET Information Security 4(4), 185–421 (December 2010) ISSN 1751-8709
2. Beautement, P., Allsopp, D.N., Greaves, M., Goldsmith, S., Spires, S.V., Thompson, S.G., Janicke, H.: Autonomous Agents and Multi –agent Systems (AAMAS) for the Military - Issues and Challenges. In: Thompson, S.G., Ghanea-Hercock, R. (eds.) DAMAS 2005. LNCS (LNAI), vol. 3890, pp. 1–13. Springer, Heidelberg (2006)
3. Becerra, G.: A security pattern for multi-agent systems. In: Proc. of ATS 2003, pp. 142–153 (2003)
4. Bergenti, F., Poggi, A.: Multi-agent systems for e-health: recent projects and initiatives. In: 10th Int. Workshop on Objects and Agents (2009)
5. Bibu, G.D.: Security in the context of multi-agent systems (Extended Abstract). In: Yolum, Tumer, Stone, Sonenberg (eds.) Proc. of 10th Int. Conf. on Autonomous Agents and Multiagent Systems, AAMAS 2011, Taipei, Taiwan, May 26, pp. 1339–1340 (2011)
6. Biskup, J., Kern-Isberner, G., Thimm, M.: Towards Enforcement of Confidentiality in Agent Interactions. In: Proceedings of the Twelfth International Workshop on Non-Monotonic Reasoning (2008)
7. Boella, G., van der Torre, L.W.N.: Permission and Authorization in Policies for Virtual Communities of Agents. In: Moro, G., Bergamaschi, S., Aberer, K. (eds.) AP2PC 2004. LNCS (LNAI), vol. 3601, pp. 86–97. Springer, Heidelberg (2005)
8. Borelius, N.: Multi-agent system security for mobile communication, PhD Thesis, Royal Holloway, University of London (2003)
9. Borelius, N.: Mobile agent security. IET Electronics and Communication Engineering Journal 14(5), 211–218 (2002)
10. Bresciani, P., Giorgini, P., Mouratidis, H.: On security requirements analysis for multi-agent systems. In: Second Int. Workshop on Software Engineering for Large-Scale Multi-Agent Systems, SELMAS, Portland, Oregon (2003)
11. Bresciani, P., Giorgini, P., Mouratidis, H., Manson, G.: Multi-agent Systems and Security Requirements Analysis. In: Lucena, C., Garcia, A., Romanovsky, A., Castro, J., Alencar, P.S.C. (eds.) SELMAS 2003. LNCS, vol. 2940, pp. 35–48. Springer, Heidelberg (2004)
12. Chi Wong, H., Sycara, K.: Adding security and trust to multi-agent systems. Applied Artificial Intelligence 14(9), 927–941 (2000)
13. Cremonini, M., Omicini, A., Zambonelli, F.: Multi-Agent Systems on the Internet: Extending the Scope of Coordination towards Security and Topology. In: Garijo, F.J., Boman, M. (eds.) MAAMAW 1999. LNCS, vol. 1647, pp. 77–88. Springer, Heidelberg (1999)
14. Endusuleit, R., Calmet, J.: A security analysis on JADE(-S) V. 3.2. In: Proceedings of NORDSEC, pp. 20–28 (2005)
15. Endusuleit, R., Mie, T.: Secure multi-agent computations. In: Proc. of Int. Conf. on Security and Management, vol. 1, pp. 149–155 (2003)

16. Endusuleit, R., Wagner, A.: Possible Attacks on and Countermeasures for Secure Multi-Agent Computation. In: Proceedings of the International Conference on Security and Management, SAM 2004, pp. 221–227, Las Vegas, Nevada, USA. CSREA Press (2004) ISBN 1-932415-37-8
17. Fasli, M.: On agent technology for e-commerce: trust, security, and legal issues. Knowl. Eng. Rev. 22, 3–35 (2007)
18. Fatih, T.: Developing a security mechanism for software agents, M.Sc Thesis, Izmir Institute of Technology (2006)
19. Ferber, J., Gutknecht, O.: A meta-model for the analysis and design of organizations in multi-agent systems. In: Proc. International Conference on Multi Agent Systems, pp. 128–135 (1998)
20. Ferber, J., Gutknecht, O., Michel, F.: From Agents to Organizations: An Organizational View of Multi-agent Systems. In: Giorgini, P., Müller, J.P., Odell, J.J. (eds.) AOSE 2003. LNCS, vol. 2935, pp. 214–230. Springer, Heidelberg (2004)
21. Fernandes, D.L., Saboia, V.F.S., De Castro, M.F., De Souza, J.N.: A secure mobile agents platform based on a peer-to-peer infrastructure. In: Int. Conf. on Networking, Systems and Mobile Communications and Learning Technologies, pp. 186–189 (2006)
22. Foner, L.N.: A security architecture for multi-agent matchmaking. In: Proc. of Second International Conference on Multi-Agent System, Mario Tokoro (1996)
23. Fugkeaw, S., Manpanpanich, P., Juntapremjitt, S.: Multiapplication authentication based on multi-agent system. IAENG Int. J. Comput. Sci. 33(2), 37–42 (2007)
24. Fugkeaw, S., Manpanpanich, P., Juntapremjitt, S.: Achieving DRBAC Authorization in Multi-trust Domains with MAS Architecture and PMI. In: Ghose, A., Governatori, G., Sadananda, R. (eds.) PRIMA 2007. LNCS, vol. 5044, pp. 339–348. Springer, Heidelberg (2009)
25. Garrigues, C., Robles, S., Borrell, J., Navarro-Arribas, G.: Promoting the development of secure mobile agent applications. J. Syst. Softw. 83(6), 959–971 (2010)
26. Gerhard, W. (ed.): Multiagent Systems, A Modern Approach to Distributed Artificial Intelligence. MIT Press (1999) ISBN 0-262-23203-0
27. Gokce, B., Laleci, D.A., Olduz, M., Tasyurt, I., Yuksel, M., Okcan, A.: SAPHIRE: a multi-agent system for remote healthcare monitoring through computerized clini- cal guidelines. In: Series, W. (ed.) Agent Technology and e-Health 2008. Agent Technology and e-Health Whilestan Series in Software Agent Technologies. Springer and Autonomic Computing, pp. 25–44 (2008)
28. Gray, R.S.: Agent Tcl: A flexible and secure mobile-agent system. PhD Thesis, Dartmouth College (June 1997)
29. Gray, R.S., Kotz, D., Cybenko, G., Rus, D.: D'Agents: Security in a Multiple-Language, Mobile-Agent System. In: Vigna, G. (ed.) Mobile Agents and Security. LNCS, vol. 1419, pp. 154–187. Springer, Heidelberg (1998)
30. Gunupudi, V., Tate, S.R.: SAgent: a security framework for JADE. In: Proc. Fifth Int. Joint Conf. on Autonomous Agents and Multiagent Systems, pp. 1116–1118 (2006)
31. Hu, Y.-J., Tang, C.-W.: Agent Oriented Public Key Infrastructure for Multi-Agent E-service. In: Palade, V., Howlett, R.J., Jain, L. (eds.) KES 2003. LNCS, vol. 2773, pp. 1215–1221. Springer, Heidelberg (2003)
32. Ismail, L.: A secure mobile agents platform. J. Commun. 3(2), 12 (2008)
33. Jacques, F.: Multi-Agent Systems: An Introduction to Artificial Intelligence. Addison-Wesley (1999) ISBN 0-201-36048-9
34. Jaiswal, A., Kim, Y., Gini, M.: Design and Implementation of a Secure Multi-Agent Marketplace. Electronic Commerce Research and Applications 3(4), 355–368 (2004)
35. Janicke, H.T.: The development of secure multi-agent systems. PhD Thesis. De Montfort University (2007)

36. Karnick, N.M., Tripathi, A.R.: Security in the Ajanta Mobile Agent System. Software Practice and Experience 31(4), 301–329 (2001)
37. Karygiannis, T., Jansen, W.: Mobile agent security. Technical Report NIST SP 800-19, National Institute of Standards and Technology (1999)
38. Karygiannis, A., Antonakakis, E.: Security and Privacy Issues in Agent-Based Location-Aware Mobile Commerce. In: Barley, M., Mouratidis, H., Unruh, A., Spears, D., Scerri, P., Massacci, F. (eds.) SASEMAS 2004-2006. LNCS, vol. 4324, pp. 308–329. Springer, Heidelberg (2009)
39. Laleci, G.B., Dogac, A., Olduz, M., Tasyurt, I., Yusel, M., Okcan, A.: SAPHIRE: A Multi-Agent System for Remote Healthcare Monitoring through Computerized Clinical Guidelines, Project Deliverable. METU, Turkey
40. Liau, C.-J.: Belief, information acquisition and trust in multi-agent systems - A model logic formulation. Artificial Intelligence, 31–60 (2003)
41. Malik, N.S., Kupzog, F., Sonntag, M.: An Approach to Secure Mobile Agents in Automatic Meter Reading. In: Proc. of International Conference on Cyberworlds, pp. 187–193 (2010)
42. Mamadou, T.K., Shimazu, A., Nakajima, T.: The State of the Art in Agent Communication Languages (ACL). Knowledge and Information Systems Journal (KAIS) 2(2), 1–26 (2000)
43. Maña, A., Muñoz, A., Serrano, D.: Towards Secure Agent Computing for Ubiquitous Computing and Ambient Intelligence. In: Indulska, J., Ma, J., Yang, L.T., Ungerer, T., Cao, J. (eds.) UIC 2007. LNCS, vol. 4611, pp. 1201–1212. Springer, Heidelberg (2007)
44. Martínez-García, C., Navarro-Arribas, G., Borrell, J., Martín-Campillo, A.: An Access Control Scheme for Multi-agent Systems over Multi-Domain Environments. In: Demazeau, Y., Pavón, J., Corchado, J.M., Bajo, J. (eds.) PAAMS 2009. AISC, vol. 55, pp. 401–410. Springer, Heidelberg (2009)
45. Marques, P., Silva, L., Silva, J.: Security mechanisms for using mobile agents in electronic commerce. In: 18th IEEE Symp. on Reliable Distributed Systems, Lausanne, Switzerland (1999)
46. McDonald, J.T., Yasinsac, A.: Trust in Mobile Agent Systems, Florida State University. Tech. Rep. (2005)
47. Moradian, E., Håkansson, A.: Approach to Solving Security Problems Using Meta-Agents in Multi Agent System. In: Nguyen, N.T., Jo, G.-S., Howlett, R.J., Jain, L.C. (eds.) KES-AMSTA 2008. LNCS (LNAI), vol. 4953, pp. 122–131. Springer, Heidelberg (2008)
48. Mouratidis, H., Giorgini, P.: Enhancing Secure Tropos to Effectively Deal with Security Requirements in the Development of Multiagent Systems. In: Barley, M., Mouratidis, H., Unruh, A., Spears, D., Scerri, P., Massacci, F. (eds.) SASEMAS 2004-2006. LNCS, vol. 4324, pp. 8–26. Springer, Heidelberg (2009)
49. Mouratidis, H., Giorgini, P., Mason, G.: Modelling secure multiagent systems. In: Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems, pp. 859–866 (2003)
50. Novák, P., Rollo, M., Hodík, J., Vlcek, T.: Communication Security in Multi-agent Systems. In: Mařík, V., Müller, J.P., Pěchouček, M. (eds.) CEEMAS 2003. LNCS (LNAI), vol. 2691, pp. 454–463. Springer, Heidelberg (2003)
51. Pechoucek, M., Marik, V.: Industrial deployment of multi-agent technologies: review and selected case studies. Auton. Agent Multi-Agent Syst. 17, 397–431 (2008)
52. Poslad, S., Charlton, P., Calisti, M.: Specifying Standard Security Mechanisms in Multi-agent Systems. In: Falcone, R., Barber, S.K., Korba, L., Singh, M.P. (eds.) AAMAS 2002. LNCS (LNAI), vol. 2631, pp. 163–176. Springer, Heidelberg (2003)

53. Ramchurn, S.D., Dong, H., Jennings, N.R.: Trust in multiagent systems. The Knowledge Engineering Review 19(01), 1–25 (2004)
54. Ramchurn, S.D., Jennings, N.R.: Trust in agent-based software. Cyber Trust and Crime Prevention Project, Review
55. Ramchurn, S.D.: Multi-agent negotiation using trust and persuasion. PhD Thesis, University of Southampton (2004)
56. Rashvand, H.F., Salah, K., Calero, J.M.A., Harn, L.: Distributed security for multi-agent systems - review and applications. IET Inf. Secur. 4(4), 188–201 (2010)
57. Rindebck, C.: Designing and maintaining trustworthy online services. Blekinge Institute of Technology Licentiate Dissertation Series No 2007:08, Sweden (2007) ISSN 1650-2140, ISBN 978-91-7295-120-4
58. Shoham, Y., Leyton-Brown, K.: Multiagent Systems: Algorithmic, Game-Theoretic and Logical Foundations. Cambridge University Press (2008) ISBN 9780521899437
59. Sulaiman, R., Sharma, D., Ma, W., Tran, D.: A Multi-agent Security Framework for e-Health Services. In: Apolloni, B., Howlett, R.J., Jain, L. (eds.) KES 2007, Part II. LNCS (LNAI), vol. 4693, pp. 547–554. Springer, Heidelberg (2007)
60. Takahashi, K., Mitsuyuki, Y., Mine, T., Sakurai, K., Amamiya, M.: Design and Implementation of Security Mechanisms for a Hierarchical Community-Based Multi-Agent System. In: Ghose, A., Governatori, G., Sadananda, R. (eds.) PRIMA 2007. LNCS, vol. 5044, pp. 134–145. Springer, Heidelberg (2009)
61. Tweedalea, J., Ichalkaranjeb, N., Sioutisb, C., Jarvisb, B., Consolib, A., Phillips-Wrenc, G.: Innovations in multi-agent systems. Comput. Appl. 30, 1089–1115 (2007)
62. Van Dyke Parunak, H.: Agents in Overalls: Experiences and Issues in the Development and Deployment of Industrial Agent-Based Systems. Expanded version of invited talk at PAAM (1999)
63. Van Dyke Parunak, H.: A Practitioners' Review of Industrial Agent Applications. Autonomous Agents and Multi-agent Systems 3(4), 389–407 (2000)
64. van 't Noordende, G.J., Brazier, F.M.T., Tanenbaum, A.S., Security in a Mobile Agent System. In: IEEE First Symposium on Multi-Agent Security and Survivability, pp. 35–45 (2004)
65. Vigna, G. (ed.): Mobile Agents and Security. LNCS, vol. 1419, pp. 3–540. Springer, Heidelberg (1998)
66. Vila, X., Schuster, A., Riera, A.: Security for a multi-agent system based on JADE. Comput. Sec. 26, 391–400 (2007)
67. Vitabile, S., Conti, V., Militello, C., Sorbello, F.: An extended JADE-S based framework for developing secure multiagent systems. Comput. Stand. Interfaces 31, 913–930 (2009)
68. Webber, C.G., De Fatima, M., Lima, W.P., Casa, M.E., Ribiero, A.M.: Towards securing e-learning applications: A multiagent platform. Journal of Software 2(1), 60–69 (2007)
69. Woolridge, M.: An Introduction to Multi-agent Systems. John Wiley and Sons (2002) ISBN 0-471-49691-X
70. Xiao, L., Peet, A., Lewis, P., et al.: An adaptive security model for multi-agent systems and application to a clinical trials environment. In: 31st IEEE Annual Int. Computer Software and Applications Conf., COMPSAC 2007 (2007)
71. Xu, K.: Mobile agent security through multi-agent cryptographic protocols. PhD Thesis, University of North Texas (2004)
72. Zhao, S., Liu, H., Sun, Z.: Scalable trust in multi-agent e-commerce system. In: Int. Symp. on Electronic Commerce and Security, pp. 990–993 (2008)

Personal Secret Information Based Authentication towards Preventing Phishing Attacks

Gaurav Varshney, Ramesh Chandra Joshi, and Anjali Sardana

Electronics and Computer Engineering Department, Indian Institute of Technology, Roorkee, India
{gauravdtsi,rcjosfec,dr.anjalisardana}@gmail.com

Abstract. Phishing is a well-known technique used by internet fraudsters for acquiring sensitive and personal information from users by impersonating a real identity. A Phishing attack involves various deceptions & advanced cybercrime techniques, some of them includes email spoofing, exploiting browser side vulnerabilities, fraudulent emails and Phished websites creation techniques using scripting languages and technologies. Phishing causes identity, goodwill and money loss to companies and individuals. One of the major problems we identified is the reduced usage and reliability on the email Infrastructure as a communication medium between customers and companies. Previous schemes for phishing prevention such as those which use browser extension, Quick Response code, Extended Authentication server & device and smart card based techniques are complex and difficult to make use in real world scenario. We present an architecture that can be used by companies for preventing phishing attacks by sharing a piece of secret information with every customer and using it as an authentication mechanism to prove their originality when a customer login to their websites using links provided in their emails. The unavailability of secret information which is securely shared between customer and the company will prevent a phisher in creating deception and hence will prevent phishing attacks which occur due to malicious links in phished emails. This will increase the reliability of email service as an authentic communication medium. The efficacy of this technique does not rely on results of any spam or phishing prevention scheme provided at email service provider side.

Keywords: Phishing, phisher, authentication.

1 Introduction

Phishing was known to people in the year 1996. It can be defined as an art of deceiving people on the internet, so as to steal the personal information secret to them such as user names, passwords, bank account numbers, credit card details etc. The concept was termed as phishing as the fraudsters are using emails as a medium to “Phish” user information such as usernames and passwords in the sea of internet users. The name resembles the word fishing; ‘ph’ is used instead of ‘f’ for two reasons:

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 31–42.
springerlink.com

1. To make it a different word
2. The letter ‘ph’ is derived from the word “phreaking” which is known to be the earliest form of hacking of telephone lines.

Phishing come first time into the knowledge of people as a severe attack in 1996 when cyber criminals stole American Online Passwords by deceiving the AOL users through phishing [11].

Phishing is a deception technique used by attackers (Commonly known as Phishers) for gaining personal information from end users, with the help of fraudulent and spoofed emails, Phished Websites and various deception techniques. The aim of the phisher lies in obtaining personal information or credentials from an end user such as bank account numbers their passwords, credit card details etc. They use this information in doing mischievous and fraudulent activities such as accessing important information and secrets, withdrawing money of an individual on web.

Phishing starts when an attacker uses a mass-mailer for sending fraudulent and spoofed emails by impersonating themselves as an authenticated bank, financial or social institution to a large population of end users. Phishing generally starts with a mass mailing activity to increase the population of end users that will eventually fall for Phishing. Phishers also use a phished website that looks exactly same as that of the original one he is targeting to phish, except for the domain name and the DNS entry it will use. The attack scenario starts when the attacker sends a phished email using spoofing and advanced email creation techniques such as those used in email newsletters, with other fraudulent techniques to fulfill their specific needs. The end user or the victim opens the email and because of deception techniques used inside it, trusts on the originality of its contents and its sender and clicks on the URL specified in it. The URL looks normal but it will take to a phished Web site.

The phished website is created in a way to look like an original highly trusted site that a phisher is targeting. As an example a phishing website can be of a highly trusted bank having the same text the same logo and animations as it is on the original bank website. When a user reaches a phished website which he can’t identify as phished one, he enters information asked by the website such as user id’s password, credit card numbers etc. which eventually get stored in the servers of the phisher.

Phishing is termed as a deception technique as it creates an illusion to the receiver of an email that, it is from an entity on which user’s trust, but behind the scenes it is not as expected. Email phishing is carried out with the help of many other tricky techniques which are used for internet fraud in today’s internet world, one of which is Email spoofing. Email spoofing technique allow an attacker to send email using other’s identity which causes a severe problem, because now he can send anything such as wrong information, malicious codes etc. and held others responsible for his wrongdoings. Spoofing creates two problems: one is of creating wrong trust in the mind of end user, hence gaining confidence, so that he will do what is required and second is of wrong backtracking because an innocent user or group will be held responsible for the problems created. Email spoofing plays an important role in carrying out email phishing as it makes the user to believe on the illusion of reality, created by a Phisher.

The statistics as obtained from Avira shown in Table 1. are of February 2011 which shows that phishing attacks are more Top level domain and business centric. The most phishing attacks are on the .com top level domain and the companies which gets most affected are those which involve some kind of electronic money transfers and social networking. Hence Phishing is from one of the most important threats in the internet world that is to be taken care of. The damages that it causes include loss of money, information and good will.

Table 1. Statistics of Phishing Attacks

#	Top Level Domain	%	#	Brand Name	%
1	.com	51.56	1	PayPal	53.5 9
2	Others	15.82	2	Others	20.0 3
3	.org	6.20	3	HSBC Bank	5.07
4	.net	5.94	4	Chase Bank	4.43
5	.uk	3.69	5	Facebook	4.09
6	IP address	3.22	6	EBay	3.48
7	.br	2.44	7	Bank Of America	3.16
8	.tk	2.18	8	Visa	2.19
9	.ru	2.01	9	Lloyds	2.07
10	.tl	1.23	10	Banco Satander	1.88

In this paper we propose an architecture that will solve the problem of phishing that is launched through Phished website links in emails. The problem created by this attack is of bad trust in email service as an authentic communication medium and loss of credentials of users.

The next section will describe the previous schemes for phishing prevention. Section 3 will describe our proposed scheme with section 4 giving the description of overall benefits. Section 5 ends the discussion with conclusions and future work.

2 Previous Work

A detailed description of previous schemes proposed for preventing phishing attacks with their assumptions, advantages and disadvantages are in a Table shown on the next page. From the study of previous techniques we concluded that there are some common shortcomings in them which are as follows:

1. Various Schemes based on Secret Images shared between website and user and which are revealed during pre-logging when the user enters the secret key are annoying and vulnerable. As they ask users to enter a new watermark image and its position each time a user logs out, also a Phisher can obtain the secret key from the user by creating fake login pages and can obtain the Watermark image and its location from the original web page by using the secret key obtained from the user.
2. Use of Advanced Technologies such as Radio frequency Identification Technology (RFID) for authentication requires that a user must carry the RFID reader and Tokens for Login.

3. External authenticating device usage in prevention of phishing attacks add on to the cost and complexity. Complexity is increased because the user interaction will be secure but complex as now it requires external device communicating with the browser which will then eventually contact to the target server. Also it will always require an external device for secure access.
4. Those solutions that require client server support require changes in the underlying frameworks and architectures also their maintenance and proper synchronization is a requirement. Also real time implementation, deployment and performance are of great concern.
5. Email Authentication and verification schemes require key management activities which will increase an extra burden on the system as now system has to take care of keys for each and every user.
6. Techniques implemented for avoiding key logging and improving password schemes are complex with respect to a normal user as he certainly doesn't want to annoy on every time he logs on to the website by entering keys through keypads and hence require initial user training.
7. Those schemes that implement security of user passwords at client's side with browser extensions are difficult to implement.
8. Client server interaction for authentication during every secure transaction increases the communication and computation cost at both client and server side.
9. Schemes based on short time passwords and certificates require special systems such as offline card readers (FINREAD reader) and Smart cards for their generation which add up extra cost and complexity to the underlying system.

Abbreviation	**Proposed Scheme**	**Paper name**	**Assumptions**	**Advantages**	**Disadvantages**
Watermarking Based [1]	Author proposed an anti-phishing approach based on Dynamic watermarking technique. According to this approach user will be asked for some additional information like watermark image, its fixing position and secret key at the time of user's registration and these credentials of particular user will be changed at per login. During each login phase a user will verify the authentic watermark with its position and decide the authenticity of website.	**Detection and Prevention of Phishing Attack Using Dynamic Watermarking** A.P. Singh et.al **2011**	User will select a watermark image and its position at website while logging out. User Account database that will store secret key & Watermark Image with regular credentials.	Doesn't require any external mobile device. Feasible to implement with minimal changes.	During every logout user is asked for reentering new watermark image and its position which is annoying. A phisher can obtain the initial secret key through phishing and can obtain the watermark position and its location.
RFID Based [2]	Authors proposed a RFID (Radio Frequency Identification Technology) Factor Authentication Application (RFAA) techniques; an enhanced technique from SofToken scheme that acts as a technique for two-factor authentication.	**A Sophisticated RFID Application on Multi-Factor Authentication** J.C. Liou et.al 2011	RFID tags and RFID reader and changed Login Infrastructure.	RFAA is a two factor authentication scheme for more secure identification. RFAA can be used for both online transactions and computer system access as opposed to the SofToken application that primary addresses to online transaction security	RFID reader and Tokens will be required each time user login.

External Authentication Device based [3]	Author proposed techniques based on external authenticator. They proposed that user's must be authenticated by an external authenticator that they cannot reveal to malicious parties. Scheme uses additional authenticator on a trusted device, which can be a cell phone or a PDA, such that the attacker will have to compromise the device to obtain user password and to obtain user account.	**Phoolproof Phishing Prevention** B. Parno et.al 2005	User can establish a secure connection between their cellphone and their browser and the cellphone itself has not been compromised.	Prevent active man in middle attacks. Use of cellphones allows us to minimize the effect of hijacked browser windows and facilitates user convenience since it can be used at multiple machines.	Requires the usage of external authenticating device to solve the purpose which will add complexity in the way a user account will be accessed.
Post Phishing Rescue based [4]	Authors proposed post Phishing Rescue technique. Here client identifies whether user have entered valid credentials on a faked website and Server capture this information from various clients. if there is some phishing going on server transfer the information to target domain for immediate attention.	**Password Rescue: A New Approach to Phishing Prevention** D. Florˆencio and C. Herley 2006	Assuming that a white list and a black list are maintained and updated regularly and a notion of trusted client and server ends who will cooperate.	The scheme doesn't protect the user from information leakage but rather try to detect and then rescue the user from bad trust decisions.	Complex and require client and server deployment and synchronization. Also require to maintain white list and blacklist of sites. Real time implementation considerations
Email spoofing detection based [5]	Authors proposed a novel key distribution architecture and identity based digital signature for making email trustworthy and hence detecting & mitigating spam mails by detecting email spoofing	**Fighting phishing attacks lightweight trust architecture for detecting phished mails.** B. Adida et.al 2005	Upgraded email client and at least one key server.	The scheme is lightweight neither pre-established public key infrastructure nor cooperation between email domains is required. all legitimate uses of email remain fully functional after the changes required by the scheme	Real time implementation considerations. Requires noticeable changes in the email service provider's side
Picture passwords Based [6]	Author has shown the usability of Picture passwords and shown how picture keypads can be used for entering credentials instead of typing through keyboard. A number of features of keypad are personalized to the user such as background color border design of keypad which differ from other users, and selected from the user's stored account record by means of the user's username. This provides protection against phishing, by alerting the user when any changes to their familiar keypad 'look-and-feel' occur, which is unknown to the phisher.	**The usability of picture passwords** N. Fraser	Set of pictures from which a subset of pictures will be issued as password to a particular user	Avoid logging by key loggers, also it is impossible for a user to disclose their password on a randomly generated phisher keypad as it is hard for a phisher to randomly generate a keypad that contains all picture necessary for entering the password by a user.	It will be complex from user's perspective to enter the password time during login by picture keypad and will require user training.
Dynamic security skin based [7]	Authors proposed two interaction techniques to prevent spoofing. 1. Browser extension provides a trusted Window in the browser for username and password entry. A photographic image for creating a trusted path between the user and the window so as to prevent spoofing of the Window and text entry fields. 2. The scheme allows the remote server to generate a Unique abstract image for individual user for each transaction. The image will create a "skin" that will automatically customize the Window or the user interface elements in the content of a remote web page. The extension will allow the browser to inde-	**The Battle Against Phishing: Dynamic Security Skins** R. Dhamija, J.D. Tygar 2005	Configured remote server and browser extension.	To authenticate, the user has to recognize only one image and remember one low entropy password, no matter how many Servers he wishes to interact with. To authenticate content from an authenticated server, the user only needs to perform one visual matching operation to compare two images.	Increases the complexity of user interface, require initial user training requires client server interaction each time a transaction is performed. Extended browser window, increases the complexity of user interaction.

	pendently compute the image it expects to receive from the remote server. To authenticate content from the server, the user can visually verify that the images match.				
Browser Extension Based [8]	Authors described a browser extension, PwdHash that transparently produces a different password for each site, which improves web password security and defends against phishing and other attacks. Browser extension apply a cryptographic hash function to a combination of the plaintext password entered by the user, data associated with the web site, and (optionally) a private salt which is stored on the client machine.	**Stronger Password Authentication Using Browser Extensions** B. Ross et.al 2005	Browser Extension, a good cryptographic hash function.	The scheme requires no changes on the server side. Theft of password received at one of the website will not reveal the password that will be used at another website.	Implementing this password method securely and transparently in a web browser extension turns out to be quite difficult
Smart card based [9]	Authors proposed two solutions: 1. Short-time password solution. This authentication scheme uses an offline card reader and a smart card to produce short-lived passwords on demand. 2. Certificate-based solution. This authentication scheme uses a secure online card reader, the FINREAD card reader, and a smart card to sign SSL/TLS challenges on demand	**Secure Internet Banking Authentication,** A. Hiltgen, et.al 2006	Java Applet Websites that can detect FINREAD card reader, Card Readers.	The user's credentials are stored on the smart card and can only be accessed via an offline smartcard reader, so malicious software can't get the user's symmetric cryptographic key or related functionality. Scheme effectively thwarts both offline credential-stealing attacks as well as online channel-breaking attacks.	Necessity of mobile equipment's. Require major changes in underlying system. Complex.
QR Code based [10]	Author proposed an anti-phishing single sign-on (SSO) authentication model using QR code. This scheme is secure against phishing attack and even on the distrusted computer environment. Scheme consists of three phases: login request phase, QR code generation phase, and verification phase.	**A mobile based anti-phishing authentication scheme using QR code**, K. Choi et.al 2011	QR (Quick Response) Code reader, extended authentication server, External Mobile device.	User can access the web sites in online Environment of distrust local computer and web server using mobile device. Even if the user's sensitive information is exposed, attacker cannot obtain the mobile information because user data is encrypted by the mobile device. Users can check the web server whether this server is the phishing server through the extended authentication server	Complex Scheme for deployment Requires an extended authentication server, whose reliability is a concern. Requires a secure external mobile device. Feasible, efficient and transparent deployment in real world is a question.

10. Extended authentication schemes for prevention of Phishing requires configuration, management and security consideration for extended authentication servers which are used to authenticate the web-servers a user is communicating with. Also it increases the communication time and cost for each user login.

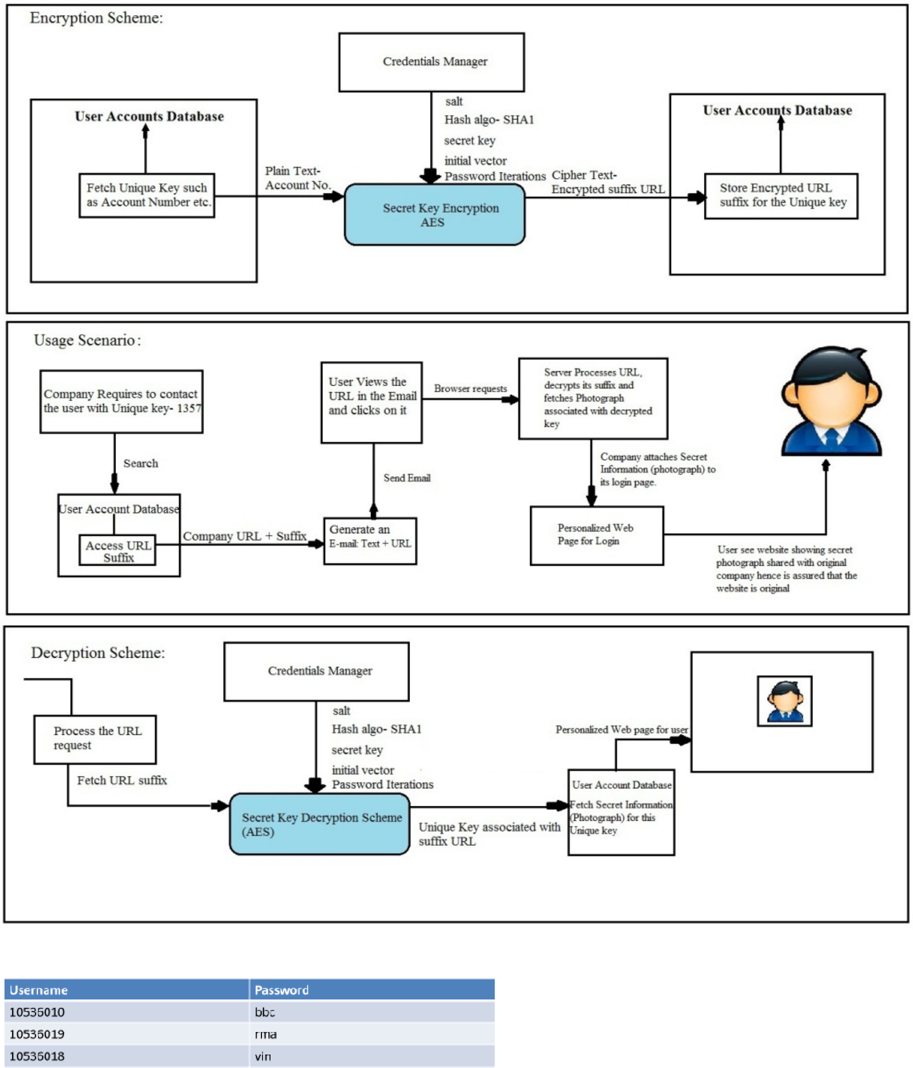
3 Personal Secret Information Based Authentication towards Preventing Phishing Attacks

Our proposed sheme is for prevention of phishing attacks by providing users a way to verify the originality of the website they are logging with while they click on a link in the e-mail that comes to their E-mailbox. This is achieved by using a piece of secret information that can be a photograph or a key which is shared between the user and the website and is provided by the user at the time of online account creation. In our implementation we will use a photograph as a piece of secret information. The overall architecture proposed is shown in fig1.

The architecture proposed require minimal changes to the underlying database that is used by websites for storing user credentials. Generally websites of online banking, social networking and others store userid and passwords as secret credential for a unique user and the verification includes checking these credentials when a user login with them. Our proposed scheme requires that websites should include some more user secret information to prevent phishing and use them in a way that will help the user to discriminate between original and phished websites. By secret information we mean userid, password as usual with the additional use of a user's photograph or a secret phrase.

The overall scenario and the underlying scheme we propose is based on the assumption that if a user can see the login pages of their websites personlaized then there will be low chances that they will fall for phishing as a phisher cannot provide such a personalization on a phished website as he is unaware of the secret information shared between user and the original company website and which is being used as a way to provide personalized experience to each unique user.

The personalization for each user while they click on the links sent to them through email from original companies is achieved by storing URL suffix for each user which is encrypted userid and will be used with the compaanies URL whenever company wants any communication. This will result in URL suffixes for each individual user which are stored with the user credentials in the database. these will be then sent to the user whenever company wants to contact the specific customer. when the user click on such a link the url suffix will be retrieved and processed to get the user id associated from where he can extract the secretv information shared with the website and display them at appropriate places during login to provide user a kind of personlization. The Encryption scheme for converting unique user information such as user id's to URL suffixes with the decryption scheme showing the server side processing of URL to obtain the userid for providing user personlization is shown in Fig.1 also explains the usage scenario in which the scheme will be deployed and used.


Encryption Scheme:
Credentials Manager
salt
Hash algo- SHA1
secret key
initial vector
Password Iterations
User Accounts Database
Fetch Unique Key such as Account Number etc.
Plain Text- Account No.
Secret Key Encryption AES
Cipher Text- Encrypted suffix URL
User Accounts Database
Store Encrypted URL suffix for the Unique key
Usage Scenario :
Company Requires to contact the user with Unique key- 1357
Search
User Account Database
Access URL Suffix
Company URL + Suffix
Generate an E-mail: Text + URL
Send Email
User Views the URL in the Email and clicks on it
Browser requests
Server Processes URL, decrypts its suffix and fetches Photograph associated with decrypted key
Company attaches Secret Information (photograph) to its login page.
Personalized Web Page for Login
User see website showing secret photograph shared with original company hence is assured that the website is original
Decryption Scheme:
Process the URL request
Fetch URL suffix
Credentials Manager
salt
Hash algo- SHA1
secret key
initial vector
Password Iterations
Secret Key Decryption Scheme (AES)
Unique Key associated with suffix URL
User Account Database
Fetch Secret Information (Photograph) for this Unique key
Personalized Web page for user

Username	Password
10535010	bbc
10535019	rma
10535018	vin

Fig. 1. Proposed Architecture

3.1 Encryption Scheme

When a user supplies the user credentials in the form of userid, password and secret information(photograph) the information will be stored as usual in the companies database. From there the userid (unique key) is extracted and is encrypted by a symmetric encryption scheme (AES) to develop a URL suffix that will be stored in the user database with that user id and is then sent as URL suffix with the companies URL link that company will sent to the user for logging in emails.

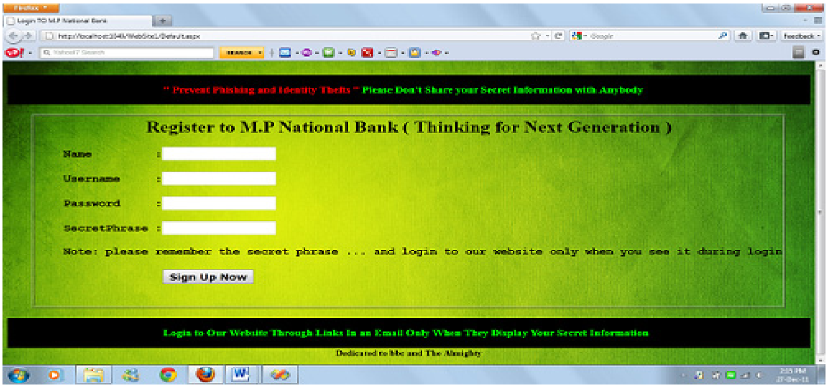
3.2 Decryption Scheme

Whenever a user will click on the link in the mail from the original companies website the link is processed at the server side. The processing includes fetching the URL suffix associated with the URL and then decrypting it with the symmetric decryption scheme(AES) to obatin the userid which was encrypted. the user id is then used to extract the secret information which is photograph associated with the user and then eventually displaying it on the login window. This will make sure the user that the page with which is he logging with is original as phisher has no knowledge of the secret photograph he shared with the company during account creation.

Both Encryption and decryption schemes require the usage of a trusted component at the server side that will store the secrets for encryption and decryption scheme. we call it in our scheme as Credentials manager. for encryption and decryption we have proposed Advanced Encryption standard (AES). The initial study shows that AES is a good symmetric enryption scheme as the only way of breaking it is through brute force attacks and those kind of attacks on hug key sizes as provided by AES are proven to be difficult and is also used in [10]. The credential manager will store the information for the AES encryption scheme and are as follows:

1. Salt which act as second secret password
2. Hash Algorithm can be SHA-1 or MD-5.
3. Secret key used for encryption and decryption
4. Initial vector which is an collection of 16 ASCII characters
5. Password iteration that defines the no of times the algorithm is run on the plain text.

The screenshots of our prototype implementation are shown below:



" Prevent Phishing and Identity Thefts " Please Don't Share your Secret Information with Anybody

Register to M.P National Bank (Thinking for Next Generation)

Name :

Username :

Password :

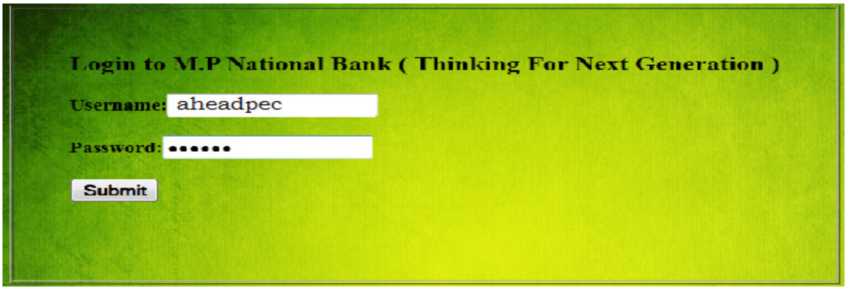
SecretPhrase :

Note: please remember the secret phrase ... and login to our website only when you see it during login

Login to Our Website Through Links In an Email Only When They Display Your Secret Information

Dedicated to Me and The Almighty

Fig. 2. Sign up page of a website as per proposed scheme

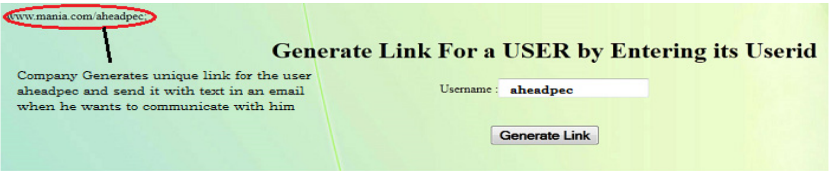


Login to M.P National Bank (Thinking For Next Generation)

Username: aheadpec

Password: ••••••

Fig. 3. Login page



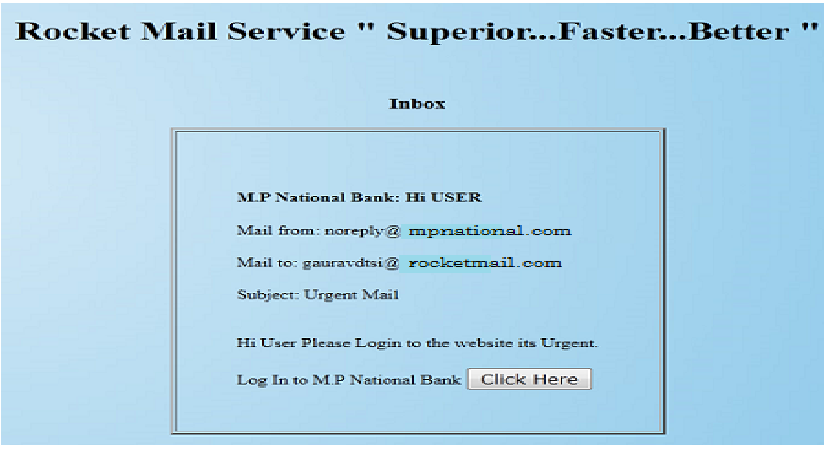
www.mania.com/aheadpec;

Company Generates unique link for the user aheadpec and send it with text in an email when he wants to communicate with him

Generate Link For a USER by Entering its Userid

Username : aheadpec

Fig. 4. Comapny Generating URL for user aheadpec



Rocket Mail Service " Superior...Faster...Better "

Inbox

M.P National Bank: Hi USER

Mail from: noreply@ mpnational.com

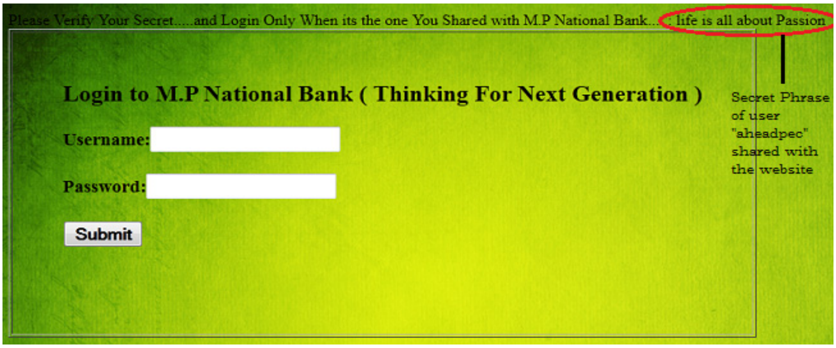
Mail to: gauravdtsi@ rocketmail.com

Subject: Urgent Mail

Hi User Please Login to the website its Urgent.

Log In to M.P National Bank

Fig. 5. Mail Sent from company to user aheadpec



Please Verify Your Secret.....and Login Only When its the one You Shared with M.P National Bank... : life is all about Passion

Login to M.P National Bank (Thinking For Next Generation)

Secret Phrase of user "aheadpec" shared with the website

Fig. 6. Login page showing secret phrase aheadpec shared with website during sign up

4 Advantages of Proposed Scheme

The advantages of our scheme wll provide compared to other techniques are as follows:

1. Our Proposed scheme does not require any kind of support from spam and phished mail filters provided by email service providers and also don't rely on their accuracy in detecting phished email.
2. This scheme can be implemented in real time by companies with minimal changes to their server side processing.
3. Key management is not a task as no sharing of key is done anywhere in the scheme. However the credentials managers have to be designed in a way so that the credentials can be protected from attackers reach. Also credentials manager can refresh the scheme by changing the values of the credentials stored after a certain period of time.
4. The scheme can rely on a single unique key used for encryption and decryption of all user ids. But generally credentials manager can implement numerous other schemes in which he can allot certain group of user id's a different set of credentials for encryption and decryption and the other group a different one that will eventually increase the security.
5. Our scheme require no changes in the browser or the at the client machine.
6. There is no requirement for any external authenticating device.
7. It requires no user training and is not annoying compared to other techniques. Also user doesn't have to remember the position, color, shape and sizes of any browser window or watermark Image.
8. Our Scheme doesn't require any special external card readers or tokens as used in some techniques for phishing preventions and hence our solution doesn't add cost and complexity to the underlying system.
9. There is no requirement for extended authentication server which cuts off extra server maintenance and configuration with reduced time per login.

5 Conclusion and Future Work

We have proposed a novel scheme based on personal secret information for authentication towards preventing Phishing attacks which are launched through Phished website links in emails. The Scheme is easy to deploy in real world scenario with minimal changes and better efficiency. A working prototype of the proposed scheme is developed and its accessment on various measures such as communication cost, time and efficiency is under study.

In future we will try to apply this technique in a way to prevent Web Phishing which occurs when a user reaches a Phished website through typing mistakes on browsers etc. and not from links in Phished emails.

References

1. Singh., A.P., et al.: Detection and Prevention of Phishing Attack Using Dynamic Watermarking. Information Technology and Mobile Communication Communications in Computer and Information Science, Part 1 147, 132–137 (2011), doi:10.1007/978-3-642-20573-6_212011
2. Liou, J., et al.: A Sophisticated RFID Application on Multi-Factor Authentication. In: 2011 Eighth International Conference Information Technology: New Generations (ITNG), Las Vegas, pp. 180–185 (2011), doi:10.1109/ITNG.2011.38
3. Parno, B., Kuo, C., Perrig, A.: Phoolproof Phishing Prevention. In: Di Crescenzo, G., Rubin, A. (eds.) FC 2006. LNCS, vol. 4107, pp. 1–19. Springer, Heidelberg (2006)
4. Florencio, D., Herley, C.: Password Rescue: A New Approach to Phishing Prevention. In: Proceedings of the 1st USENIX Workshop on Hot Topics in Security, HOTSEC (2006)
5. Adida., B., et al.: Fighting Phishing Attacks: A Lightweight Trust Architecture for Detecting Spoofed Emails. In: DIMACS Workshop on Theft in E-Commerce (2005)
6. Fraser, N.: The usability of picture password (unpublished)
7. Dhamija, R., Tygar, J.D.: The Battle Against Phishing: Dynamic Security Skins. In: Proceedings of the 2005 symposium on Usable privacy and security, SOUPS (2005)
8. Ross, B., et al.: Stronger Password Authentication Using Browser Extensions. In: Security 2005 Technical Program (2005)
9. Hiltgen, A., et al.: Secure Internet banking authentication. IEEE Security & Privacy 4(2), 21–29 (2006), doi:10.1109/MSP.2006.50
10. Kyeongwon, C., et al.: A mobile based anti-phishing authentication scheme using QR code. In: 2011 International Conference on Mobile IT Convergence (ICMIC), September 26-28, pp. 109–113 (2011)
11. APWG.: Origins of the Word "Phishing", `http://www.antiphishing.org/word_phish.html`

Key Distribution Schemes in Wireless Sensor Networks: Novel Classification and Analysis

Premraj Mahajan and Anjali Sardana

Electronics and Computer Science Department,
IIT Roorkee, India
{prem228434,dr.anjalisardana}@gmail.com

Abstract. Security is one of the important and challenging aspects in wireless sensor network owing to their wireless nature combined with limited memory, energy, and computation. We can classify security issue of the wireless sensor network into five broad categories as cryptography techniques, key management, routing protocols, intrusion detection and data aggregation. Since the key management forms an underlying factor for efficient routing protocol and cryptography in wireless sensor network, we focus on key management issue. This paper outlines the constraints, security requirements and attacks, which are related to the key management and routing. Further novel classification of key distribution schemes have been proposed. The proposed novel classification and comparison distinctly brings to the fore gaps in the existing solutions of research which can be put to use by researchers in the area to identify current challenges for designing efficient key distribution scheme. The paper concludes with possible future research directions on key distribution in WSNs.

Keywords: Key distribution schemes, Security, Sensor network.

1 Introduction

Wireless Sensor Network contains hundreds or thousands of sensor nodes and these sensor nodes have the ability to communicate either amongst each other or directly to an external base station (BS). Figure 1 shows a schematic diagram of sensor node components. Basically, sensor node comprises of sensing, processing, transmission, mobilizer, position finding system, and power units. The same figure shows the communication architecture of a wireless sensor network (WSN) [1, 2].

These types of the sensor nodes are deployed into the field for the purpose of sensing some specific information. But these sensor nodes are resource constraints. Sensor nodes have limitations like computational power, storage, battery etc. So possibility of the attacks like hello flood on sensor node is more. Hence it is important to utilize available resources effectively with fulfilling the basic requirements like encryption, authentication etc. These (encryption, authentication) services are based on operations which involves the different [3]keys like encryption-decryption keys, cluster key, key which is used in hash function etc. So energy efficient key distribution in sensor nodes plays vital role in security of WSNs Section 2 of this paper presents constraints of the wireless sensor network along with security requirements. Section 3

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 43–53.
springerlink.com

presents attacks related to the key management and routing. In section 4, a novel classification of the key management schemes is presented. Section 5 discusses about conclusions and future work to be done.

2 Constraints in Wireless Sensor Network

Sensor nodes have limited processing power, storage capacity and transmission range because of the energy and the physical size.

Energy: Energy in sensor network is conserved for many purposes like sensing, ADC, computation, communication. So for long lasting working of the sensor, all these operations should be performed efficiently.

Computation: Embedded processors in sensor nodes are not so powerful that they can perform the complex cryptographic functions. Typically 8bit, 4-12 MHz[4].

Memory: Memory includes flash memory and RAM. Flash memory is used for storing downloaded application code and RAM is used for sensed data, intermediate computation. In SmartDust project, tiny OS code space is 3500bytes, and only 4500bytes [4] are there for the security application.

Transmission range: Again range is also dependent on the energy limitation. It also depends on the environment factors like whether and terrain.

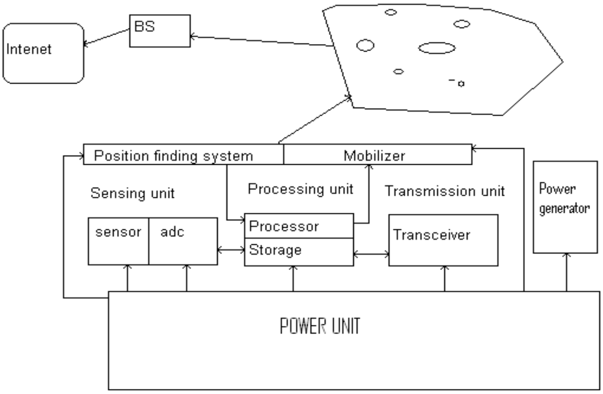


Fig. 1. The component of sensor network [3]

Security requirement: To protect the information and resource from attacks, security services are provided in WSNs. These security requirements include:

- **Authentication:** It ensures that communicating nodes are genuine and no any malicious node can inject or spoof the message.
- **Availability:** It ensures that message is made available to the destination node even in presence of the intermediate node capture or Denial-of-service attack.

- **Authorization:** It ensures that only authorized nodes can be involved in providing information to network services.
- **Confidentiality:** It ensures that the given information cannot be understood by the attacker or any unauthorized person.
- **Integrity:** It ensures that information cannot be altered by any intermediate malicious node.

3 Attacks Related to Key Management and Routing

Wireless Sensor Network is vulnerable to various types of attacks. In following section the attacks which are related to key management and routing are considered.

Spoofing, altering and replaying attack: In presence of the spoof and replay attack, the network traffic can be extensively corrupted. Continuous alteration in the message transmits the incorrect message and source node has to retransmit the packets. It reduces the battery life in large extend due to power exhaustion. In replay attack, malicious node may capture the any of the network message and replay that message, and hence damaging the network performance.[4, 5]

Selective forwarding attack: Normally sensor nodes are multi-hop systems and the assumption in such network is that intermediate nodes faithfully forward the received message. In this type of attack the malicious node may refuse or simply drop some part of message [4-6]. Such type of attack is most effective when attacker is explicitly included on the path of data flow.

Sybil attack: The Sybil attack is a case in which malicious node shows multiple identities. Malicious node behaves as it is a large number of the nodes for example impersonating other node or simply claiming false identities. In worst case, an attacker may generate an arbitrary number of additional node identities, using single device [4, 7].

Sinkhole attack: The attacker tries to pass nearly all the traffic from a particular area through a particular/malicious node. An attacker makes a compromised node look more attractive to the surrounding nodes by forging routing information and ultimately surrounding nodes will choose next node to route the information through the compromised node giving access to all data. Many attacks can be initiated [4, 5] through the sinkhole attack ex. Wormhole, selective forwarding or eavesdropping.

Wormhole attack: A wormhole is low-latency link between two portions of the network over which attacker replays the network messages [5, 8]. An attacker receives the packets at one portion of network and tunnels them to another portion, and then replays them into the network. These tunneled packets arrive sooner than the other packets transmitted over normal multi-hope route because these tunneled distances are longer than the normal wireless transmission range of a single hop. The wormhole attack is possible even if the attacker has not compromised any hosts and even if all communication provides authenticity and confidentiality.

Hello flood attack: Many of protocol use HELLO packets for getting the list of the neighboring nodes and assume that replied nodes are within their transmission range and are therefore neighbors. But an attacker may use high-powered transmitter to track maximum areas[5] so that, other nodes will believe that they are neighbor. If the attacker falsely broadcast a superior route to the base station then all nodes will pass the information through those attacking nodes even that node is out of range.

Acknowledgement spoofing attack: Acknowledgements are sometimes required in the sensor networks. An attacker node can spoof acknowledgements. Goal of the spoofing the acknowledgement is that attacker can convince[5] the sender node by giving false information like a weak link as strong or dead node as alive.

Tampering attack: Tampering is a physical layer attack. Given physical access to node, attacker can extract sensitive information such as cryptographic keys and some other data on a node [9] and may create false identity.

Table 1. Different types of attacks and their defense mechanism along with related issue

TYPES OF ATTACK	DEFENSE MECHANISM	ISSUE
Tampering	a. Tamper-proofing b. Hiding	High cost
Spoofed, Altered, or Replayed Routing Information	a. MAC b. Monitoring c. Lightweight Authentication d. SPINS protocol	Computation power Computation power
Selective Forwarding	a. Multi path routing b. Probing	Computation power
Sinkhole	a. Authentication b. Geographical routing c. Redundancy d. Monitoring	Computation power, key distribution Energy consumption
Sybil	a. Use of symmetric keys b. Probing	Computation power, key distribution Energy Consuming
Wormhole	a. Authentication b. Time synchronization c. Packet leashing by geographical and temporal information	Computation power, key distribution Infeasible
Hello flood Attack	a. Authentication b. Verify the bidirectional link	Computation power, key distribution
Ack. Spoofing	a. Authentication	Computation power, key distribution
Node replication attack	a. Localized voting system b. Key renewing	Replication attacks

4 Key Distribution Schemes

In wireless sensor network, to provide the basic security requirement like encryption, decryption, authentication etc. we have to perform some operations involving the different types of keys. With considering the constraints of the sensor node, we have to distribute these keys to all the sensor nodes. This key distribution operation must be energy efficient so as to increase the life-time of sensor node. An open research problem is how to set-up secrete keys among the communicating nodes. There are different schemes are proposed for key distribution among the sensor nodes. These schemes are categorized with the following properties [3, 10, 11]:

- **Pre-distribution/Post-distribution:** In pre-distribution schemes the keys are stored into nodes before deployment into the field and in post-distribution schemes the keys are distributed after the deployment into the field with the help of trusted server or self-enforcing property.
- **Homogeneous/Heterogeneous:** In homogeneous sensor network, all the nodes are identical and having the same computational power, storage capacity and energy level whereas in case of the heterogeneous sensor network, small number of sensor nodes are more powerful in terms of the energy, storage and computational power than other large number of the sensor nodes.
- **With deployment knowledge/without deployment knowledge:** Sensor network which knows that where and how the sensor nodes are deployed into the network that comes under deployment knowledge category. And other sensor networks, which don't have information about the deployment knowledge, that comes under without deployment knowledge category.

Different types of key distribution schemes are classified as shown in the figure 2:

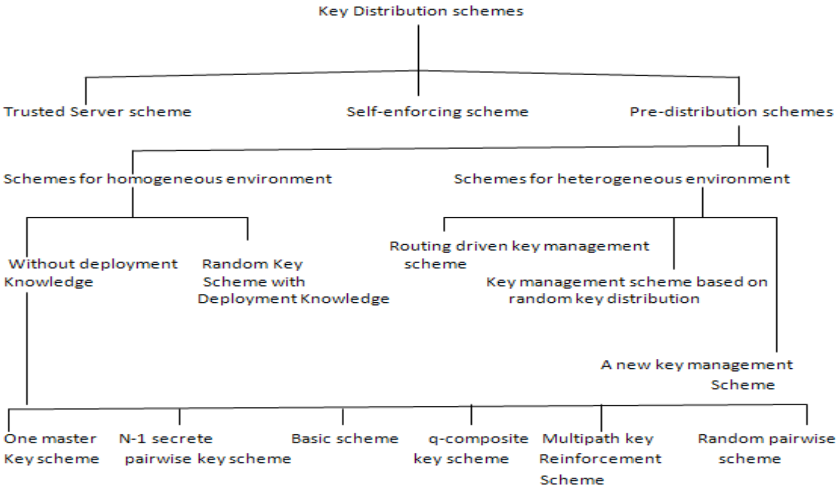

Key Distribution schemes
Trusted Server scheme
Self-enforcing scheme
Pre-distribution schemes
Schemes for homogeneous environment
Schemes for heterogeneous environment
Without deployment Knowledge
Random Key Scheme with Deployment Knowledge
Routing driven key management scheme
Key management scheme based on random key distribution
A new key management Scheme
One master Key scheme
N-1 secrete pairwise key scheme
Basic scheme
q-composite key scheme
Multipath key Reinforcement Scheme
Random pairwise scheme

Fig. 2. Classification of key distribution schemes

4.1 Trusted Server Scheme

Trusted server scheme depends on the trusted server for key agreement between two different nodes, e.g. Kerberos. Such a third party key distribution requires infrastructure which is impractical [11, 12]for sensor network.

4.2 Self-enforcing Scheme

Self-enforcing scheme depends on asymmetric cryptography. it is very good solution for key management and distribution in WSN but sensor nodes have lot of limitations like memory and processing power. Limited computation and energy resources of the sensor nodes often make it undesirable to use public key algorithms, such as Diffie-Hellman key agreement or RSA. Several works shows[13] that lightweight versions of the public key algorithms can be utilized in the sensor networks.

4.3 Pre-distribution Schemes

In pre-distribution schemes the keys are stored into nodes before deployment into the field. These can be further divided into schemes for homogeneous and heterogeneous environment.

4.3.1 Schemes for Homogeneous Network

In homogeneous sensor network, all the nodes are identical and having the same computational power, storage capacity and energy level.

4.3.1.1 Without Deployment Knowledge. In these schemes, deployment knowledge is not considered.

4.3.1.1.1 One Master Secrete Key Scheme [11]: In one master secrete key scheme, each node carry one master key, pre-distributed before deployment. This master key is used to achieve the key agreement and obtain a new pair wise key. Because of one master key, this scheme doesn't exhibit desirable network resilience. If any node is compromised then the entire sensor network will be compromised. In this scheme giving the temper proofing mechanism, will increase the cost as well as energy consumption of each node.

4.3.1.1.2 N-1 Secrete Pair-Wise Key[11]: In N-1 secrete pair-wise key scheme, if there are N nodes then each node should have to carry n-1 secrete pair-wise keys. Each of which is known to this sensor and one of the other to n-1 sensor node. Resilience is perfect as compared to other scheme because if any of the nodes is compromised then that node does not affect the security of communications of other nodes. But this system has main two drawbacks. It is not practical because of extremely limited amount of memory. As the network grows (N), memory required for storing keys also increases. Second one is, adding new node to pre-existing network is complex because the existing nodes do not have the keys of the new sensor node.

4.3.1.1.3 Basic Scheme [14]: It consists of three phases. Key pre-distribution, shared key discovery and path key establishment. First phase store small number of the keys into nodes key ring, taken from generated pool of keys to ensure that two node share at least one key with a chosen probability. Second phase establishes the secure link between two nodes only when they carry secrete key common. Third phase assigns the path-key to selected pairs of sensor nodes in wireless communication range that do not share a key but are connected by two or more two or more links at the end of the shared key discovery phase.

4.3.1.1.4 q–composite Key Scheme [10]: In previous scheme, we require common single key from key rings of two communicating nodes in order to secure link in the key–setup phase. In q –composite key scheme, q>1 common keys are needed. In this way it increases the resilience of the network against node capture. This scheme uses Merkle puzzle in key set-up phase. After key set-up and discovery a new communication link key is generated as: K = hash ($K_1||K_2||\ldots||K_q$) and hashed in canonical order. This scheme has no resistance against node replication since node degree is not constrained and there is no limit on the number of times each key can be used.

4.3.1.1.5 Multipath Key Reinforcement [10]: This method conjunction with basic scheme strengthens the security of an established link key by establishing the link through multiple paths and improves the resilience against node capture. Key set-up is as per basic scheme. Then each link is secured using a single key from key pool. This single key may be the part of any other node and if that node is captured then the link may not be secured further. So to address this problem, multipath reinforcement scheme update the communication key to a random value after key set-up through multiple paths between the nodes. The more the path we can find between the nodes, the more security multipath key reinforcement provides for the link between any two nodes. A link is considered completely compromised if all its reinforcement paths are also compromised.

4.3.1.1.6 Random Pairwise Key scheme [10]: In initialization phase, a node can only store random set of np pairwise keys where n is total nodes can be used in sensor network and p is probability. Total n node unique identifiers are generated. Size of network may be less than n and other unused identifiers are used for future network expansion that means provides some range of scalability. In post-deployment key set-up phase, each node first broadcasts its node ID to its immediate neighbors. Scheme provides node-to-node authentication by using identifiers. Provides distributed node revocation with adding some overhead in key storage and provides perfect resilience against node capture as it does not reveals any information about links.

4.3.1.2 With Deployment Knowledge. In pre-distribution schemes the keys are stored into nodes before deployment into the field

4.3.1.2.1 ABAB Scheme [15]: This scheme uses approximate deployment prior knowledge to improve the performance of a random key pre-distribution scheme. Motivation of this scheme is to design simple, flexible key distribution scheme[14] that are easily applicable, extensible and sufficiently secure. This scheme uses two large key pools for overall network with some common keys in common. This scheme is totally based on “the basic scheme”.

4.3.1.2.2 ABCD Scheme [15]: ABAB scheme is easily applicable in sensor network but it has a resilience problem since same keys are used in different zones several times. ABCD scheme is more complex than ABAB but it is more efficient and resilient scheme as it uses 2r keys pools, where r is the number of rows of deployment. Direct key and path key establishment is as per the basic scheme.

4.3.2 Schemes for Heterogeneous Network

In homogeneous scheme, it is assumed that all sensor nodes are of same power and same capacity. But the works have suggested [16]that connectivity, lifetime, reliability and resilience can be improved substantially if few nodes are given greater power and transmission capacity.

Some Common Assumptions of heterogeneous sensor network environment are:

- There are two types of sensor nodes H (powerful and provided with temper-resistance) and L (ordinary).
- Each L nodes and H nodes have unique node ID.
- Routing in Heterogeneous sensor network consists of two phases:
  1. Intra cluster routing (each L sensor sends data to its cluster Head)
  2. Inter cluster routing (Each cluster head sends may aggregate data from multiple

L-sensors and then sends compressed data to sink via the H-sensor backbone.

Following are some heterogeneous key distribution schemes in heterogeneous wireless sensor network:

4.3.2.1 Routing Driven Key Management Scheme [17]: This scheme is referred as ECC based key management scheme. This scheme requires only small number of ECC computations in each L-sensor as compared to ECC public key cryptography. Server generates pair of ECC public and private keys, one pair for each L-sensor and H-sensor. Each H-sensor are pre-loaded with public keys of all the L-sensors, association between each L-sensor and its private key, and a special key K_h, which is used by a symmetric cryptography algorithm for verifying newly deployed sensors and for secure communications. Each L-sensor is pre-loaded with private key and public keys of H-sensors. In this scheme it is assumed that each L-sensor can determine its location. L-sensor sends key request message to H-sensor, which include its location and its ID via shortest distance path. After receiving the request message, H-sensor uses MST or SPT algorithm to determine the tree structure in the cluster. Then H-sensor generates shared keys for each L-sensor and its c-neighbors, Then H-sensor unicasts the message to respective L-sensor node with their private key. After receiving the message L-sensor decrypt the message and communicate securely with their neighbors. The scheme utilizes the fact that a sensor node communicates with a small portion of neighbors only and thus greatly reduces the communication and computation overheads of key set-up as compared to homogeneous schemes. It Stores small number of keys into the L-sensor.

4.3.2.2 Key Management Scheme Based on Random Key Distribution [3]: This scheme pre-load only one secrete key of key pool into L-sensor generate new key by applying one way function on key and its ID. H-sensors are pre-loaded with all keys of key pool along with a special master key for inter cluster communication. With Hello message L-sensor and H-sensor find their neighbors and then L-sensor sends the list of its neighbor to the H-sensor. After that H-sensor generates the data encryption key and integrity check key and forwards the MAC check along with nonce. After receiving the nonce L-sensor calculates the data encryption key and integrity check key. After setting the keys, Ha generates the shared pair-wise keys between a node and its neighbors. This scheme significantly reduces the storage requirement as compared to random key pre-distribution schemes.

4.3.2.3 A New Key Management Scheme [18]: During cluster formation this scheme scheme obtains the distance between the cluster head and other sensor nodes. This scheme uses the concepts of level, as each level has separate seed used for deriving the new keys that are only used in that level and neighboring level. The key pool consists of base key and derived keys. Derived key are hash of base keys with different seeds. In pre-distribution phase, scheme stores only base key and not derived key. It stores randomly k keys into each sensor and c base keys into each H-sensor where c>>k. Pair-wise key between sensor and base station is stored in L-sensor and will be used for authentication purpose. Each CH sends location to base station by GPS and obtains the maximum distance a point can have in his cluster. In this scheme, the number of base keys has effect on connectivity between nodes and number of seeds has effect on resiliency against node capture.

Table 2. Comparison of key distribution schemes

Scheme	**Pre-distribution**	**Deployment knowledge**	**Heterogeneity**	**Features**	**Drawback**
Trusted serevr [11]	No	No	No	1.Good Resilience to attack 2.Low memory required.	1.Require third party 2.Trust issue
Self enforcing [13]	No	No	No	1.Easy node addition. 2.Good Resilience to attack. 3.Most secure	1.High computational power 2.Large memory
One master key [11]	Yes	No	No	1.Easy node addition 2.Low Memory required	1.Bad Resilience to attack
N-1 pair-wise secrete key [11]	Yes	No	No	1.Better Resiliecne to attack	1. Node addition Difficult 2.Large memory required
Basic scheme [14]	Yes	No	No	1.Good Resilience to attack. 2.Easy Node addition.. 3.Simple method	1.Large Memory required
q-composite scheme [10]	Yes	No	No	1.More resilience to attack 2.Support Large network	1.Large memory required
Miltipath Key reinforcement [10]	Yes	No	No	1. Strongly secure links 2.Good resilience agianst node capture.	1.Add overhead key estabilishment traffic. 3.Large Memory required.
Random pair-wise scheme [10]	Yes	No	No	1.Provides node-to-node authentication. 2.Good resilince aginst node attack.	1.Large Memory required. 2.Scalable to some extend.
ABAB [15]	Yes	Yes	No	1.Very simple and flexible. 2.Less secure 3.Very much scalable.	1.Required prior deployment knowledge 2.Large Memory required
ABCD [15]	Yes	Yes	No	1.More secure than ABAB. 2.Highly scalable. 3.Requires less communicational cost.	1.Complicated than ABAB. 2.Required Prior deployment knowledge.
Routing driven [17]	Yes	No	Yes	1.Highly secure and scalable 2.Low memory storage.	1.Sensor node has to send its location through GPS.
Key mgnt. scheme based on random key distribution [3]	Yes	No	Yes	1.Better resillience to attack. 2.Low memory required. 3.Low computational cost. 4.Addition of node is easy.	1.Scalable to some extend. 2.H sensor exhuastion may occur with large network
A new key management scheme [18]	Yes	No	Yes	1.Reduces tradeoff between resilience and connectivity. 2.Require low memory.	1.Sensor node has to send its location through GPS.

Table 2 gives the comparison of the different key distribution schemes.

In homogeneous wireless sensor environment all sensor nodes have to store the large number of keys which may lead poor resilience to node capture attack. In heterogeneous wireless sensor environment the given schemes [3, 17, and 18] uses the

GPS unit to communicate to location to the cluster head. This adds additional overhead to the network. So such a hybrid key distribution scheme must be proposed which can be used for long lifespan and scalable network without additional overhead of GPS unit.

5 Conclusions

In wireless sensor network, encryption and authentication services are based on the operations involving keys. So energy efficient key distribution is an important issue. In this article we present a comprehensive survey of key distribution schemes in wireless sensor network. They have common objective of trying to distribute the keys to all sensor node with efficient use of the memory, computation power with consideration of the security aspect.

Overall, key distribution techniques can be classified on network structure as homogeneity, pre-distribution of keys and deployment knowledge basis.

Finally, we have given the comparison of all the key distribution schemes. Although, many of the techniques look promising, there are still many challenges that need to be solved in future key distribution scheme in wireless sensor network like large scalability and lifespan of the wireless sensor network.

References

1. Yong, W., et al.: A survey of security issues in wireless sensor networks. IEEE Communications Surveys & Tutorials 8, 2–23 (2006)
2. Al-Karaki, J.N., Kamal, A.E.: Routing techniques in wireless sensor networks: a survey. IEEE Wireless Communications 11, 6–28 (2004)
3. Kausar, F., et al.: Key Management and Secure Routing in Heterogeneous Sensor Networks. In: IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, WIMOB 2008, pp. 549–554 (2008)
4. Habib, A.: Sensor network security issues at network layer. In: 2nd International Conference on Advances in Space Technologies, ICAST 2008, pp. 58–63 (2008)
5. Karlof, C., Wagner, D.: Secure routing in wireless sensor networks: attacks and countermeasures. In: Proceedings of the First IEEE, International Workshop on Sensor Network Protocols and Applications, pp. 113–127 (2003)
6. Huijuan, D., et al.: Selective forwarding attack detection using watermark in WSNs. In: ISECS International Colloquium on Computing, Communication, Control, and Management, CCCM 2009, pp. 109–113 (2009)
7. Newsome, J., et al.: The Sybil attack in sensor networks: analysis & defenses. In: Third International Symposium on Information Processing in Sensor Networks, IPSN 2004, pp. 259–268 (2004)
8. Hu, Y.C., et al.: Packet leashes: a defense against wormhole attacks in wireless networks. In: Twenty-Second Annual Joint Conference of the IEEE Computer and Communications, INFOCOM 2003, vol. 3, pp. 1976–1986. IEEE Societies (2003)
9. Kaiping, X., et al.: Security improvement on an efficient key distribution mechanism for large-scale Wireless Sensor Network. In: 2nd International Conference on Anti-Counterfeiting, Security and Identification, ASID 2008, pp. 140–143 (2008)

10. Haowen, C., et al.: Random key predistribution schemes for sensor networks. In: Proceedings of 2003 Symposium on Security and Privacy, pp. 197–213 (2003)
11. Wenliang, D., et al.: A key management scheme for wireless sensor networks using deployment knowledge. In: Twenty-Third Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM 2004, 597 p. (2004)
12. Ibriq, J., Mahgoub, I.: A Hierarchical Key Establishment Scheme forWireless Sensor Networks. In: 21st International Conference on Advanced Information Networking and Applications, AINA 2007, pp. 210–219 (2007)
13. Pathan, A.S.K., Choong Seon, H.: Feasibility of PKC in resource-constrained wireless sensor networks. In: 11th International Conference on Computer and Information Technology, ICCIT 2008, pp. 13–20 (2008)
14. Eschenauer, L., Gligor, V.D.: A key-management scheme for distributed sensor network. In: 9th ACM Conference on Computer and Communications Security, Washington DC, pp. 41–47 (2002)
15. Tasci, S.E., et al.: Simple and Flexible Random Key Predistribution Schemes for Wireless Sensor Networks Using Deployment Knowledge. In: International Conference on Information Security and Assurance, ISA 2008, pp. 488–494 (2008)
16. Lu, K., et al.: A framework for distributed key management schemes in heterogeneous wireless sensor networks. In: 25th IEEE International Performance, Computing, and Communications Conference, IPCCC 2006, p. 7, 520 (2006)
17. Xiaojiang, D., et al.: A Routing-Driven Key Management Scheme for Heterogeneous Sensor Networks. In: IEEE International Conference on Communications, ICC 2007, pp. 3407–3412 (2007)
18. Banihashemian, S., Bafghi, A.G.: A new key management scheme in heterogeneous wireless sensor networks. In: 2010 The 12th International Conference on Advanced Communication Technology (ICACT), pp. 141–146 (2010)

An Integrated Intrusion Detection System for Credit Card Fraud Detection

M. Sasirekha, I. Sumaiya Thaseen, and J. Saira Banu

VIT University, Vellore -632014, TamilNadu, India

Abstract. Computer security is one of the key areas where lot of research is being done. Many intrusion detection techniques are proposed to ensure the network security, protect network resources and network infrastructures. Intrusion detection systems (IDS) attempt to detect attacks by gathering network data and analyze the information from various areas to identify the possible intrusions. This paper proposes an IDS combining three approaches such as anomaly, misuse and decision making model to produce better detection accuracy and a decreased false positive rate. The integrated IDS can be built to detect the attacks in credit card system using Hidden Markov approach in the anomaly detection module. The credit card holder's behaviours are taken as attributes and the anomalous transactions are found by the spending profile of the user. The transactions that are considered to be anomalous or abnormal are then sent to the misuse detection system. Here, the transactions are compared with predefined attack types and then sent to the decision making model to classify it as known/unknown type of attack. Finally, the decision-making module is used to integrate the detected results and report the types of attacks in credit card system. As abnormal transactions are analyzed carefully in each of the module, the fraud rate is reduced and system is immune to attacks.

Keywords: Intrusion detection, Anomaly detection, Misuse detection, Hidden Markov Model.

1 Introduction

The amount of online shopping is increasing day by day and millions of people are using the online services to fulfil their needs. As a result a large number of credit card transactions are being carried out in the net. These credit card transactions are vulnerable to malicious intruders attempting to negotiate on the integrity, confidentiality or any resource availability. The spending pattern of the card holder has to be analysed to determine if any inconsistency occurs in comparison with the usual pattern. Hence an Intrusion Detection System (IDS) is proposed to detect the attackers by analyzing the spending profile of the customer along with the type of purchase. Many fraud detection systems have been proposed using data mining and neural network approaches but an IDS combining such as anomaly detection, misuse detection and decision making model has not been developed for a credit card fraud system. As the system is of

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 55–60.
springerlink.com

hybrid nature, it attempts to increase the detection attack rate and also reduce the number of false positives which is of major concern in any IDS.

The rest of the paper is organized as follows. In section 2 we summarize the relevant work on intrusion and fraud detection systems. Section 3 discusses the detailed description of the proposed system. The experimental results and snapshots of anomaly detection module are discussed in section 4 and 5. Section 6 concludes the paper.

2 Related Work

Many anomaly IDS have been proposed in the literature. We briefly discuss some of the proposed solutions. Ghosh and Reilly [10] proposed a neural network for credit card fraud detection. Stolfo et al. [11] [12] developed a credit card fraud detection system (FDS) using meta learning techniques to study models of fraudulent credit card transactions. Performance metrics like True Positive—False Positive (TP-FP) spread and accuracy have been defined by them. The BOAT adaptive method was proposed by sherly et al [15]. Each individual transaction amount depends on the purchase of the corresponding type of item. Standard performance metrics, True Positive (TP) and False Positive (FP) are used to characterize the effectiveness of the system. Then the fraudulent transactions are identified. The difficulty with most of the above specified approaches is that they need labelled data for both real as well as fraudulent transactions to train the classifiers. In contrast, we present a Hidden Markov Model (HMM)-based credit card FDS, which does not need fraud signatures and yet it is able to identify frauds by considering the spending habit of the credit card holder.

Ourston et al. [13] have proposed the application of HMM in identifying multistage network attacks. Hoang et al. [14] present a innovative method to analyze series of system calls for anomaly detection using HMM. Another major advantage of the HMM-based approach is a severe decrease in the number of False Positives (FPs)—transactions detected as malicious by a FDS although they are actually genuine. Hence with the tremendous increase in attacks, there is a need to design an Intrusion Detection System that secures the credit card sector.

3 Proposed System

The proposed system uses the Hidden Markov Model to identify fraudulent transactions in the anomaly detection module. HMM is advantageous over other statistical approaches because it effectively reduces the false positive rate which is an important metric to measure the performance of Intrusion Detection System. The fraudulent transactions identified in the anomaly detection module are sent to Misuse detection module to identify the type of fraud. The role of anomaly module is to identify the fraud and the role of misuse module is to classify the fraud. Then the results are sent to decision making model.

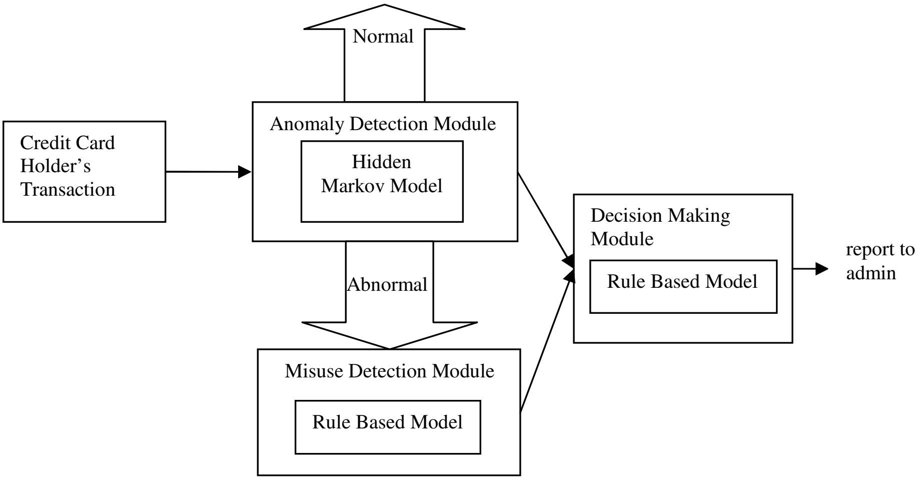


Fig. 1. Proposed Architecture of Integrated IDS

3.1 Anomaly Detection Module

The HMM model is mainly used to identify the false positive attacks. The false positive attack is the number of normal transactions that are identified as anomalous. The types of purchase are the hidden states. The new transaction is classified as anomalous or normal transaction based on the transaction history. Using the HMM [2], the user is grouped based on his spending profile. The false positive rate is the number of normal transactions identified as anomalous. The False Positive Rate (FPR) is identified using the following formula,

FPR=(Number of anomalous transactions/Number of normal transactions)* 100%

3.2 Misuse Detection Module

Misuse detection is an effective approach to handle attacks that are known by the system. When the particular type of attack is identified, the result is sent to Decision Making Module. The attacks are Cross-site, SQL, Path Traversal, etc can be identified using this module.

3.3 Decision Making Module

The decision making module provides the result as attack if both anomaly and misuse detection module identifies it as an attack. The Rule-based method is used for Decision making module. The rules are

- If anomaly detection model detects a fraud and misuse detection model does not detect the same fraud, then the detected fraud is not a fraud and it is an erroneous classification.
- If anomaly detection model detects a fraud and misuse detection model does detects the same fraud, then the detected fraud is a fraud and the fraud mode is classified.

- If anomaly detection model detects a fraud and misuse detection model finds it to be an unknown fraud, then the detected fraud is a new fraud.

4 Experimental Results

Initially the users are grouped based on his spending habit. The low range is between 0 and 5000, medium between 5000 and 12500 and high above 12500. The observation symbol for low cluster is denoted by 'l' and medium by 'm' and high by 'h'. The sequence length of 5-10 is used to identify the fraudulent transaction. The fraudulent transaction identification is as follows.

Low=(0-5000), Medium=(5000-12500), High=(above 12500).

The results below specify how fraudulent transactions are identified in each spending profile cluster. Calculations have been shown only for high spending profile. Low and medium profile can be calculated in the similar manner.

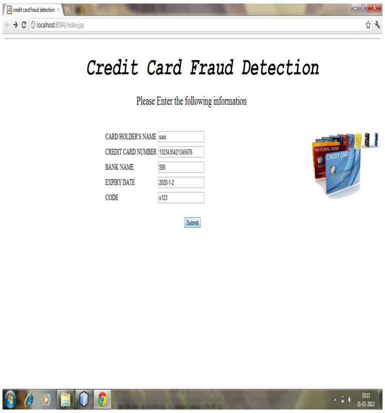
High Spender Profile(HS)

α_1: Transaction Sequence={15000,18000,6000,200,25000}. The state sequence is $\{s_3,s_1,s_2,s_3,s_1\}$and the observation sequence is {h,h,m,l,h}.

Table 1. Probability of observation sequence based on past history in high spending profile

	1	2	
α_1	1/3	0	1/6
α_2	0.16666666	0	0
α_3	0	0.1666667	0
α_4	0	0	0.08333333

5 Screen Shots



credit card fraud detection

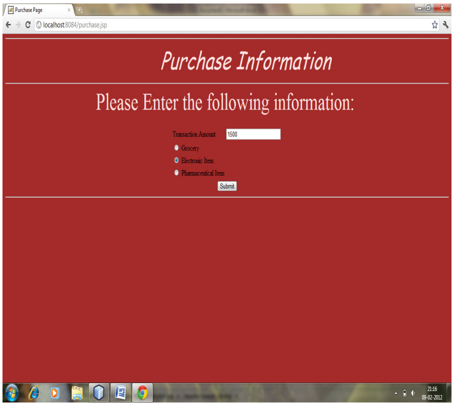
localhost:8084/index.jsp

Credit Card Fraud Detection

Please Enter the following information

CARD HOLDER'S NAME	sasi
CREDIT CARD NUMBER	1323435421345676
BANK NAME	SBI
EXPIRY DATE	2020-1-2
CODE	s123

Fig. 2. User Login



Purchase Page

localhost:8084/purchase.jsp

Purchase Information

Please Enter the following information:

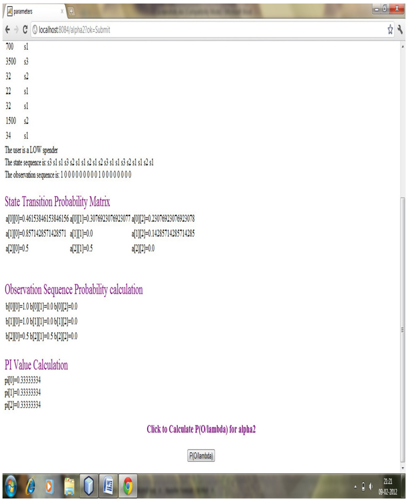
Transaction Amount 1500

○ Grocery

● Electronic Item

○ Pharmaceutical Item

Fig. 3. User's Purchase



```
700     s1
3500    s3
32      s2
22      s1
32      s1
1500    s2
34      s1
The user is a LOW spender
The state sequence is: s3 s1 s1 s3 s2 s1 s1 s2 s1 s2 s3 s1 s1 s3 s2 s1 s1 s2 s1
The observation sequence is: 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0
```

State Transition Probability Matrix

```
a[0][0]=0.46153846153846156 a[0][1]=0.3076923076923077 a[0][2]=0.23076923076923078
a[1][0]=0.8571428571428571  a[1][1]=0.0                a[1][2]=0.14285714285714285
a[2][0]=0.5                 a[2][1]=0.5                a[2][2]=0.0
```

Observation Sequence Probability calculation

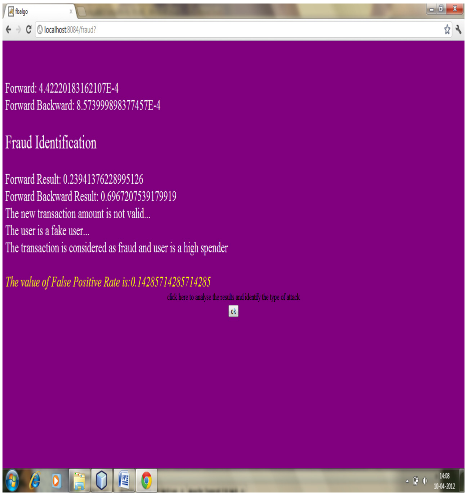
```
b[0][0]=1.0 b[0][1]=0.0 b[0][2]=0.0
b[1][0]=1.0 b[1][1]=0.0 b[1][2]=0.0
b[2][0]=0.5 b[2][1]=0.5 b[2][2]=0.0
```

PI Value Calculation

```
pi[0]=0.33333334
pi[1]=0.33333334
pi[2]=0.33333334
```

Click to Calculate P(O/lambda) for alpha2

Fig. 4. Estimating alpha2 using forward backward algorithm



```
Forward: 4.42220183162107E-4
Forward Backward: 8.573999898377457E-4
```

Fraud Identification

```
Forward Result: 0.23941376228995126
Forward Backward Result: 0.6967207539179919
The new transaction amount is not valid...
The user is a fake user...
The transaction is considered as fraud and user is a high spender
```

The value of False Positive Rate is:0.14285714285714285

click here to analyse the results and identify the type of attack

Fig. 5. Fraud Identification

6 Conclusion

This paper proposes an integrated intrusion detection system for credit card fraud detection by combining three approaches anomaly, misuse and decision making models. Anomaly detection module is implemented using Hidden Markov approach classifies the credit card transaction as normal or abnormal based on the threshold of the spending profile of the credit card user. Comparative studies reveal that the HMM technique results in higher accuracy over a wide variation in the input data and the proposed system can be scalable for handling large transaction data. The false positive rate (FPR) is calculated. Second, the Misuse detection module screens the abnormal transactions for type detection. The main aim of the attacker is to steal the details of the authorised user by using XSS and SQL Injection attack. Finally the results of the two detection modules are integrated by the decision making module to determine the fraud, type of fraud and return the same to the administrator for necessary action. The experimental results discussed are of anomaly detection module. Our future work will integrate the results of the anomaly module with the misuse module to produce effective detection accuracy.

References

1. Wang, S.-S., Yan, K.-Q., Wang, S.-C., Liu, C.-W.: An Integrated Intrusion Detection System for Cluster-based Wireless Sensor Networks (2011), doi:10.1016/j.eswa.2011.05.076
2. Srivastava, A., Kundu, A., Sural, S., Majumdar, A.K.: Credit Card Fraud Detection Using Hidden Markov Model. IEEE Transactions on Dependable and Secure Computing 5(1) (January-March 2008), doi:10.1109/TDSC.2007.70228

3. Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: A survey. ACM Comput. Surv. 41(3) (July 2009), doi:10.1145/1541880.1541882
4. Rabiner, L.R.: A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proc. IEEE 77(2), 257–286 (1989)
5. Filippov, V., Mukhanov, L., Shchukin, B.: Credit Card Fraud Detection System. Institute of Control Sciences Moscow, Russia
6. Sherly, K.K., Nedunchezhian, R.: Boat Adaptive Credit Card Fraud detection system. IEEE Transactions (2010)
7. Kruğel, C., Toth, T., Kirda, E.: Service specific anomaly detection for network intrusion detection. In: SAC 2002: Proceedings of the 2002 ACM Symposium on APPLIED Computing. ACM, New York (2002)
8. Mahoney, M.V.: Network traffic anomaly detection based on packet bytes. In: SAC 2003: Proceedings of the 2003 ACM Symposium on Applied Computing, pp. 346–350. ACM, New York (2003)
9. Wang, K., Stolfo, S.J.: Anomalous Payload-Based Network Intrusion Detection. In: Jonsson, E., Valdes, A., Almgren, M. (eds.) RAID 2004. LNCS, vol. 3224, pp. 203–222. Springer, Heidelberg (2004)
10. Ghosh, S., Reilly, D.L.: Credit Card Fraud Detection with a Neural-Network. In: Proc. 27th Hawaii Int'l Conf. System Sciences: Information Systems: Decision Support and Knowledge-Based Systems, vol. 3, pp. 621–630 (1994)
11. Stolfo, S.J., Fan, D.W., Lee, W., Prodromidis, A.L., Chan, P.K.: Credit Card Fraud Detection Using Meta-Learning: Issues and Initial Results. In: Proc. AAAI Workshop AI Methods in Fraud and Risk Management, pp. 83–90 (1997)
12. Stolfo, S.J., Fan, D.W., Lee, W., Prodromidis, A., Chan, P.K.: Cost-Based Modeling for Fraud and Intrusion Detection: Results from the JAM Project. In: Proc. DARPA Information Survivability Conf. and Exposition, vol. 2, pp. 130–144 (2000)
13. Ourston, D., Matzner, S., Stump, W., Hopkins, B.: Applications of Hidden Markov Models to Detecting Multi-Stage Network Attacks. In: Proc. 36th Ann. Hawaii Int'l Conf. System Sciences, vol. 9, pp. 334–344 (2003)
14. Hoang, X.D., Hu, J., Bertok, P.: A Multi-layer Model for Anomaly Intrusion Detection Using Program Sequences of System Calls. In: Proc. 11th IEEE Int'l Conf. Networks, pp. 531–536 (2003)
15. Sherly, K.K., Nedunchezhian, R.: BOAT adaptive credit card fraud detection system. In: 2010 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), December 28-29, pp. 1–7 (2010)

Analysis and Performance Evaluation of Application Specific Processors for Network-Based Intrusion Detection Systems

Majid Nezakatolhoseini[1], Sam Jabbehdari[2], and Mohammad Ali Pourmina[3]

[1] Computer and Mechatronic Department Science and Research Branch
Islamic Azad University Tehran, Iran
m.nezakat@srbiau.ac.ir
[2] Computer Department Tehran North Branch Islamic Azad University Tehran, Iran
sjabbehdari@gmail.com
[3] Electronic Department Science and Research Branch Islamic Azad University Tehran, Iran
pourmina@srbiau.ac.ir

Abstract. By growing and development of computer networks and generalizing the use of modern services on the information platform, the importance of communication and information security is considered more than the other times by network representations and users. Presented reports by response computer incident different groups show the wide growth of computer attacks in the recent years. In this case Network Intrusion Detection Systems (NIDS) as one of the Intrusion Detection System (IDS) types, are be transformed to the utilization systems for establishing the security levels and detecting the illegal activities in the network. This research includes an IDS which is written in C programming language that uses 15597 Snort rules and MIT Lincoln Lab network traffic. By running this security application on the V850, OR1K, MIPS32, ARM7TDMI and PowerPC32 microprocessors, their performance can be evaluated. For increasing the performance in this research, the GNU Compiler Collection (GCC) optimization levels are used and at the end, base on O2 optimization level a new combination of optimization flags is presented which the performance of ARM7TDMI microprocessor is increased.

1 Introduction

In 1980, James Anderson [1] was the first one that introduced the idea of using IDS, namely audit trail for misuse detection of information systems. He is known as a father of IDS. In 1987, Denning [2] presented the first model of IDS and afterwards several instances of IDS were created.

One of the main reasons for using the IDS even with firewall on the network is less security of firewalls against the attacks that occur by the different soft-wares to organization data and information. For example Nimda, Code red and Slammer worms can be reminisced. For connecting to the organization servers from Internet, at least one port should be opened on the firewall for accessing to servers (e.g. TCP 80 port for accessing via Web) that could be enough for entry of mentioned worms to Local Area

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 61–70.
springerlink.com

Network (LAN). Another reason is that firewall rules are simpler but more complex rules can be used in IDS.

The IDS in comparison with Intrusion Prevention System (IPS) has following capabilities: firstly, monitoring the network traffic without having an IP address secondly, working in passive form. However in this case, working with that is hard but at least we are sure that own IDS is not attacked.

In recent years the idea that integrated all network security in a box was raised but this issue conflicts to concepts such as defense in depth and security in layers because engender of each problem for that, means the failure in the whole network.

In this research, using the expandable and efficient microprocessors for implementation of NIDS is for two reasons: one for flexibility in system reconfiguration and the other is for performance. Note that the networks are vulnerable to new attack patterns, so updating the attack patterns in NIDS is inevitable. In the other hand achieving to high performance seems possible because of microprocessor hardware architectures.

In this article, IDS related works are considered in section 2. Section 3 evaluates the performance of mentioned microprocessors in execution of intrusion detection application in case of compiler doesn't perform optimization. In section 4 optimization topic in compiler is discussed and in continuance microprocessor evaluations are considered in case of compiler performs the optimization and then a novel offered combination is presented for optimization of ARM7TDMI microprocessor. In section 5 the results of the study are presented.

2 Related Works

In this section related works with IDS are considered in two groups of software and hardware.

2.1 Software Works

Since many NIDS software systems have been introduced in the form of open source or commercial but none of them have found the popularity and universality of Snort [3].

Snort is open source software and a network packet sniffer with a packet log recorder and IDS that attempts to detect the complex attacks to the network. This IDS has a huge database of attack patterns (rules or attack signatures) which is available to users. Snort compares character patterns in the network traffic with its own set of defined rules by pattern matching algorithms. Efficient algorithms is used at the heart of its detection engine for improving the pattern searching such as Boyer-Moore, Aho-Corasick and combination methods such as AC-BM [4].

Software intrusion detection on a conventional is executed on the General Purpose Processors (GPP) and therefore being slow of this method is its most important disadvantages.

2.2 Hardware Works

Hardware, especially FPGA-based methods can be provided much more efficiently through streaming, highly parallel architectures. Some of these are as follows.

In [5], a GB pattern matching tool that supports fully TCP/IP network has been described. This system divides TCP/IP stream to sub-streams and distributes the load to several pattern matching units which use Deterministic Finite state Automaton (DFA) pattern matching. When the number of rules increases, the number of states required to implement methods based on DFA is significantly increased. This can lead to reduce the performance of these systems.

Usually, DFA and Nondeterministic Finite state Automaton (NFA) are used for analysis of regular expressions. Traditional or uncompressed DFA problems which are mentioned above lead to other DFA including compressed DFA, D^{2}FA [6], TDP-DFA [7], CD^{2}FA [8], PDFA [9] and CDFA [10] with the aim of maintaining a minimum throughput of uncompressed or traditional DFA and reduction of memory space as well, are presented one after another.

The Bloom filters [11] as an efficient estimate memory have been used in the field of research related to pattern matching in intrusion detection. Bloom filters use a random technique for testing membership queries in the set of strings. Predefined set of signatures that have been grouped according to their length are stored in set of parallel Bloom filters in the form of hardware. Each of these Bloom filters includes the special length signatures. These Bloom filters are used for monitoring the network traffic and work on the strings with the same length of network data. Limitation in maintenance of long length strings such as some of viruses is among of challenges of this method.

In [12], by implementation of KMP algorithm on multi context FPGA, from own reconfiguration is exploited. The advantage of using the NFA is reduction of design area but its problem is the lack of scalability that system work frequency is decreased by increasing of rules (attack patterns). NFA versus DFA was able to reduce memory consumption, but will sacrifice throughput. This problem leads to other NFAs such as G-NFA [13] are introduced. G-NFA is provided a hardware architecture based on bit-design for Glushkov NFA which detects given regular expressions. As regards the main feature of NFAs is memory reduction and the main characteristics of DFA are performance and speed so hybrid methods are proposed for displaying of regular expression as well which for example can be mentioned to [14]. In this method, hybrid memory architecture is suggested to improve the ability of traditional memory architecture for considering the complex regular expressions. In [15] another type of finite automaton namely XFA or extended finite automaton has been introduced to take advantage of DFA and NFA.

Ternary Content Addressable Memory (TCAM) is a type of memory which performs parallel searching with high speed. Each cell in TCAM can take one of the three states 0, 1 or '?' (Don't care). It is assumed in [16] that if TCAM has several match states, the lowest match state index with input string is return as output. Ref. [16] for matching patterns with TCAM provides the algorithm in three states: simple patterns, long patterns and complex patterns, however different length of attack patterns is one of the challenges of this method. Ref. [17] including the study that uses the structures based on TCAM memory for matching all signature components (and not just a pattern or specific string) effectively.

3 Performance Evaluation

This section considers performance evaluation of V850, OR1K, MIPP32 from MIPS series, ARM7TDMI from ARM series and PowerPC32 from PowerPC microprocessors for execution of written network intrusion detection application. Fig. 1 illustrates the components of this IDS with its work flow.

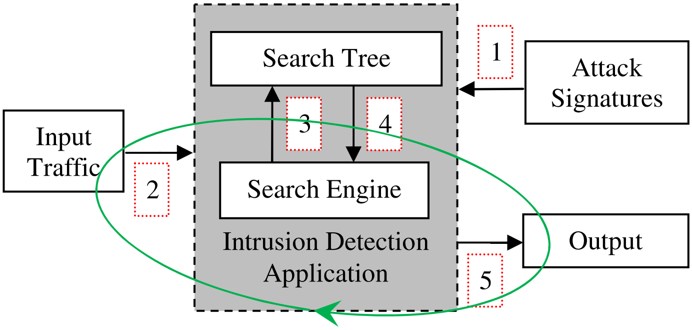


Fig. 1. IDS components and execution phases

Snort attack signatures are used for implementation of Fig. 1 IDS. Special format is used for writing the attack signatures in Snort. These signatures are divided to rule header and rule options sections logically [18]. Rule header includes rule operation, protocol, source and destination IP addresses and their netmask and source and destination port information. Rule option section includes alert messages and information that should be considered about some parts of packet. 15597 Snort rules are used in this research.

Aho-Corasick string matching algorithm is a string search algorithm (Important class of string algorithm which attempts to find the location of one or several strings that are named pattern in the longer string or text) which was invented by Alfred V. Aho and Margaret J. Corasick [19] in 1975. The search complexity of $T[1 \ldots m]$ with the Aho-Corasick automaton is $O(m + z)$ that z is the number of occurred patterns in T. Because of linear search that increases the search speed, Aho-Corasick automaton is used in this research. In [20], all of the documents for data structure, files and functions of Aho-Corasick are available in summary.

The Cyber Systems and Technology Group of MIT Lincoln Laboratory, under Defense Advanced Research Projects Agency (DARPA) and Air Force Research Laboratory sponsorship, has collected and distributed the first standard corpora for evaluation of computer network intrusion detection systems [21]. The 1998 DARPA evaluation was designed to find the strength and weaknesses of existing approaches and lead to large performance improvements and valid assessments of intrusion detection systems. This research uses five hundred thousand packets from simulation output traffic of the third week on Thursday, Lincoln Laboratory in 1999.

Intrusion detection application has two parts, creating a search tree and checking input traffic. In creating the search tree, according to attack rules in Snort that are divided to four sections consist of TCP, IP, ICMP and UDP based on their protocols, four search tree are made with the same names. Next, by calling the Create_Aho_Tree, reading the attack rules which already were downloaded from Snort.org site are started from *.rules files. In search engine, darpa_traffic function is called for inspecting the incoming traffic. This function reads the packets and then acquires the payload of them and determines their protocols, including ICMP, IP, TCP and UDP. By calling ahocorasick_KeywordTree_search _helper function, the packet payload is searched in corresponding tree. In this projects, only 6 (content, nocase, offset, depth, distance and within) of 37 rule options are examined.

If there is a definite attack pattern in packet payload, a report is recorded for the occurred attack in an output file. This report includes source and destination IP addresses, source and destination ports, the packet payload and alert message of found attack signature in packet.

Open Virtual Platform (OVP) [22] is a fast simulation with open source and free resource model and has Application Program Interfaces (API). The focus of OVP is to accelerate the adoption of the new way to develop embedded software, especially for System-on-Chip (SoC) and Multiprocessor System-on-Chip (MPSoC) platforms. OVP uses libraries of processor and behavioral models, and APIs for building the own processors, peripherals and platforms. OVP is flexible and is free for noncommercial usages. This simulation is a product of 2008 and used in this research.

In this research, version 2/23/2011 of OVP simulator program is used on a laptop with Windows XP SP2, 1.60 GHz CPU and 512 MB RAM. The simulation has used the basic microprocessors without cache and pipeline. All microprocessors have the same nominal speed, and are equal to 100 MHz.

Execution of intrusion detection application on V850 is encountered message “Heap and stack collision” because of memory shortage so this microprocessor is ignored. Fig. 2 illustrates the number of assembly instructions than the number of incoming packets for each microprocessor.

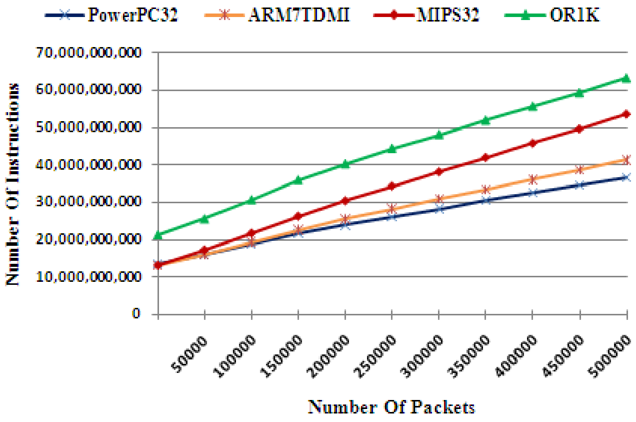


Fig. 2. Execution graph of intrusion detection application on four microprocessors

When there aren't any incoming packets, the graph shows the number of assembly instructions that are needed for making attack signature trees. However in making search tree section, ARM7TDMI, MIPS32 and PowerPC32 have almost the same performance but gradually with arrival of packets to the system, PowerPC32 microprocessor will be better. Because all of microprocessors have the same speed (100 MHz) to execute the instructions so run-time of intrusion detection application for five hundred thousand packets is explained in Table 1. Run-time of intrusion detection application in making search tree is not important because this tree is made just for one time, so Table 1. just shows run-time of intrusion detection application without the time that is needed for making search tree.

Table 1. Run-time of intrusion detection application for five hundred thousand packets

Microprocessors	Run-Time
PowerPC32	231.75 s
ARM7TDMI	280.95 s
MIPS32	404.66 s
OR1K	420.91 s

4 Optimization

A compiler is likely to perform many or all of the following operations: lexical analysis, preprocessing, parsing, semantic analysis (Syntax-directed translation), code generation, and code optimization.

In this research, code optimization as one of the compiler operations is used for increasing the performance.

Compilers bridge source programs in high-level languages with the underlying hardware. A compiler requires 1) determining the correctness of the syntax of programs, 2) generating correct and efficient object code, 3) run-time organization, and 4) formatting output according to assembler and/or linker conventions. A compiler consists of three main parts: the frontend, the middle-end, and the backend [23].

The front end checks whether the program is correctly written in terms of the programming language syntax and semantics. The middle end is where optimization takes place. The back end is responsible for translating the Intermediate Representation (IR) from the middle-end into assembly code.

This front-end/middle/back-end approach makes it possible to combine front ends for different languages with back ends for different CPUs. Practical examples of this approach are the GCC, LLVM, and the Amsterdam Compiler Kit, which have multiple front-ends, shared analysis and multiple back-ends.

4.1 GCC

The GCC is a compiler system produced by the GNU Project supporting various programming languages.

GCC has been ported to a wide variety of processor architectures, and is widely deployed as a tool in commercial, proprietary and closed source software development environments. GCC is also available for most embedded platforms, for example Symbian (called *gcce*), AMCC and Freescale Power Architecture-based chips [24].

GCC 1.0 which only handled the C programming language was released in 1987, and the compiler was extended to compile C++ in December of that year. Front ends were later developed for Fortran, Pascal, Objective-C, Java, and Ada, among others. The current stable version of GCC is 4.6.1, which was released on June 27, 2011.

4.2 Optimization with GCC Predefined Optimization Levels

Optimizations in GCC are done by the flags that use in gcc command line. –f<optimization name> is used for activating a flag and –fno–<optimization name> is used for deactivating a flag in command line. The GCC also has its own predefined levels of optimization [25] which begin with –O and include: –O or –O1, –O2, –O3, –O0 and –Os.

The performance of microprocessors are checked again for intrusion detection application but this time the optimization levels –O1, –O2 and –O3 are used. These three optimization levels reduce the run-time of applications. The performance is evaluated relative to the –O0 level which is the level without optimization. –O0 level results were shown in Table 1.

Table 2. shows the performance percentage of microprocessors with optimization levels relative to –O0 level.

Table 2. Performance increase percent of microprocessors by using predefined optimization levels for five hundred thousand packets

Microprocessors	O1 to O0	O2 to O0	O3 to O0
PowerPC32	20.32 %	15.47 %	15.34 %
ARM7TDMI	11.91 %	13.57 %	13.41 %
MIPS32	8.23 %	8.89 %	8.78 %
OR1K	28.36 %	28.10 %	27.87 %

4.3 Optimization with Offered Optimization Level

In order to increase performance, by focusing on ARM7TDMI microprocessor, it’s attempted to increase the run-time of intrusion detection application on this microprocessor. For this purpose, compiler improvement is used like the previous part.

As mentioned in Table 2. O2 level has the best functionality in the second section of intrusion detection application (Searching attack signature in packet payloads) in ARM7TDMI. For improving the performance, O2 level is intended as a baseline and by adding some other flags to this level, a new combination is presented that works better than O2 level in ARM7TDMI.

As regards the intrusion detection application has too loops with many iteration and the long jumps, first the flags in gcc that optimize the mentioned issues, are collected. By frequent running the intrusion detection application in binary combination (O2 with another flag), triad combination (O2 with two other flags) and so on, the following combination is obtained:

```
-O2 -freduce-all-givs -fmove-all-movables -mcpu=arm7 -fnew-ra
-fno-expensive-optimizations -fno-force-mem
-fno-guess-branch-probability -fno-if-conversion2 -fno-crossjumping
```

Table 3. shows the performance improvement percent (O2 to O0 and Offered to O0) for execution of intrusion detection application in ARM7TDMI. According to Table 3, proposed combination works 4.53 percent better than O2 level.

Table 3. Performance increase percent of ARM7TDMI microprocessor in O2 and offered level

Optimization Level	Improvement percentage to O0
O2	13.57 %
Offered	18.10 %

5 Conclusion

Using microprocessor for performing intrusion detection led to the problems such as attack signature updating are resolved which is in ASICs, because of the flexibility of microprocessors. This flexibility is related to the software which is run by microprocessor.

This study shows that just by using the predefined optimization levels, the performance of mentioned microprocessors can be increased between 8.23% to 28.36 which is fairly substantial. This is important because optimization is performed with lower cost and easier than other solutions such as hardware design changing.

The use of compiler optimization levels will not always improve the performance. Using the C functions, specially the functions that involve to string such as strcmp, strlen, strlwr and strstr in the written IDS wasn't increased more than 0.5% by predefined optimization levels. but with eliminating mentioned functions (If it is possible) or rewriting them with loops and conditions or using equivalent functions but with better performance led to performance range is putted in 8.23% to 28.36%. For example in the second part of intrusion detection application namely inspecting the packet payloads, Boyer Moore algorithm was used instead of strstr that almost works 50% better than it.

Unlike existent documents that know optimizing in O3 level is more than O1 and O2, this research shows that optimization of these levels are not deterministic and depends on application.

References

[1] Anderson, J.P.: Computer security threat monitoring and surveillance. Technical report. James P. Anderson Company, Fort Washington, Pennsylvania (April 1980)

[2] Denning, D.: An intrusion-detection model. IEEE Transactions on Software Engineering 13(2), 222–232 (1987)

[3] Sourcefire. Snort: The Open Source Network Intrusion Detection System (2009), http://www.snort.org

[4] Jason Coit, C., Staniford, S., McAlerney, J.: Towards Faster String Matching for Intrusion Detection or Exceeding the Speed of Snort. In: DARPA Information Survivability Conference and Exposition (DISCEX II 2001), vol. 1, p. 367 (2001)

[5] Moscola, J., Lockwood, J., Loui, R.P., Pachos, M.: Implementation of a Content-Scanning Module for an Internet Firewall. In: Proceedings of FCCM 2003 (April 2003)

[6] Kumar, S., et al.: Algorithms to Accelerate Multiple Regular Expressions Matching for Deep Packet Inspection. In: ACM SIGCOMM 2006, Pisa, Italy, September 12-15 (2006)

[7] Lu, H., Zheng, K., Liu, B., Zhang, X., Liu, Y.: A Memory-Efficient Parallel String Matching Architecture for High Speed Intrusion Detection. IEEE Journal on Selected Areas in Communications 24(10) (October 2006)

[8] Kumar, S., Turner, J., Williams, J.: Advanced algorithms for fast and scalable deep packet inspection. In: Proc. of ACM/IEEE Symposium on Architecture for Networking and Sommunications Systems (ANCS 2006), pp. 81–92. ACM Press, New York (2006)

[9] Jiang, J., Wang, X., He, K., Liu, B.: Parallel Architecture for High Throughput DFA-Based Deep Packet Inspection. In: Proc. of IEEE Int. Conf. on Communications (ICC), pp. 23–27 (May 2010)

[10] Song, T., Wang, D.: Another CDFA Based Multi-Pattern Matching Algorithm and Architecture for Packet Inspection. In: Proc. of 20th Int. Conf. on Computer Communications and Networks, ICCCN (2011)

[11] Dharmapurikar, S., Krishnamurthy, P., Sproull, T., Lockwood, J.: Implementation of a Deep Packet Inspection Ciruit using Parallel Bloom Filters in Reconfigurable Hardware. In: Proceedings of HOTi 2003 (2003)

[12] Sidhu, R., Mei, A., Prasanna, V.K.: String Matching on Multicontext FPGAs using Self-Reconfiguration. In: Proceedings of FPGA 2003 (February 1999)

[13] Lee, T.H.: Hardware architecture for high-performance regular expression matching. IEEE Trans. on Computers (July 2009)

[14] Lin, C.-H.: Hybrid memory architecture for regular expression matching. In: 52nd IEEE International Midwest Symposium on Circuits and Systems, MWSCAS, pp. 1159–1162 (2009)

[15] Smith, R., et al.: XFA: Faster Signature Matching with Extended Automata. In: 2008 IEEE Symposium on Security and Privacy (2008)

[16] Yu, F., Katz, R.H., Lakshman, T.V.: Gigabit Rate Packet Pattern-Matching Using TCAM. In: ICNP 2004 (2004)

[17] Taherkhani, M.A., Abbaspour, M.: An Efficient Hardware Architecture for Deep Packet Inspection in Hybrid Intrusion Detection Systems. In: Proc. 4th International Conference on Communications and Networking in China, August 26-28 (2009)

[18] Sourcefire, Inc. SNORT® Users Manual 2.9.0, The Snort Project (September 27, 2010)

[19] http://en.wikipedia.org/wiki/Aho-Corasick_algorithm

[20] Doxygen, FFPT Reference Manual 1.3 (July 2004), http://ffpf.sourceforge.net

[21] http://www.ll.mit.edu/mission/communications/ist/corpora/ideval/data/index.html
[22] OVP Simulation, http://www.ovpworld.org/aboutovp.php
[23] Compiler from Wikipedia, http://en.wikipedia.org/wiki/Compiler
[24] GNU Compiler Collection from Wikipedia, http://en.wikipedia.org/wiki/GNU_Compiler_Collection
[25] Optimize Options - Using the GNU Compiler Collection (GCC), http://gcc.gnu.org/onlinedocs/gcc-4.1.1/gcc/Optimize-options.html

ECDLP Based Proxy Multi-signature Scheme

Ramanuj Chouksey[1], R. Sivashankari[1], and Piyush Singhai[2]

[1] Name, VIT University, Vellore
ramanuj@vit.ac.in, sivashankari.r@vit.ac.in
[2] Name, Knowlarity Communications Private Limited
singhai.piyush@gmail.com

Abstract. A Proxy signature scheme enables a proxy signer to sign a message on behalf of the original signer. In this paper, we propose efficient and secure Proxy multi-signature scheme based on elliptic curve cryptosystem. Our scheme satisfy all the proxy requirements and require only elliptic curve multiplication and elliptic curve addition which needs less computation overhead compared to modular exponentiation also our scheme is withstand against original signer forgery and public key substitution attack.

1 Introduction

Signature is an authentication mechanism that enables the creator of a message to attach a code that acts as a signature. A digital signature allows an entity, called the designator or original signer, to generate a signature on any message using his private key such that the receiver can verify the validity of the signature and authentication of the signer using signer's certified public key. But in the case of absence of original signer, digital signature schemes are not applicable. Proxy signature schemes were proposed to address the problem in traditional signature schemes.

Proxy signature allows the original signer to delegate his signing power to a person called proxy signer who can replace the original signer, in case of say, temporal absence, lack of time or computational power, etc. Then the verifier can check the validity of signature, identity of the proxy signer and the original signer's agreement using original signers, proxy signer's certified public keys. Proxy signatures have been used in numerous practical applications, like e-commerce, electronic agreement, mobile agents, mobile communications, distributed computing, electronic voting, etc.

The concept of proxy signature was first proposed in 1996 by Mambo et. al [8]. Based on the delegations types, they classified proxy signature into*full delegation, partial delegation* and, *delegation by warrant* schemes. In a full delegation, As name implies, the complete delegation (the private key of original signer) is transferred to proxy signer. So a signature by proxy signer is indistinguishable from that created by an original signer. In a Partial delegation, A new secret key (proxy signature key) is computed by the the original signer using his private key. Using this secret key, the proxy signer can generate a proxy signature on any message. For security requirements, it is computationally infeasible for the proxy signer to derive the original signer's private key from the proxy signature key. However, in such schemes the range of messages a proxy signer can sign is not limited. This weakness eliminated by warrant schemes so using delegation by

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 71–79.
springerlink.com

warrant specifies the types of messages, delegation period and identity of signer (proxy or original), etc. In paper [1] they mention two kinds of proxy signature schemes depending on whether the original signer can generate the same proxy signature as the proxy signers do. The one is proxy-unprotected and other is proxy-protected in which anyone else, including the original signer, cannot generate the same proxy signatures. This difference is important in practical applications and this thing also avoid potential disputes between the original signer and proxy signer.

After the Mambo's scheme [8] that is one to one i.e. one original signer and one proxy signer there are so many proxy signature scheme have been proposed [9]. To meet the requirements of various rapidly growing applications, different types of proxy signature schemes have been evolved. Those are threshold proxy signatures, nominative proxy signatures, one-time proxy signatures, multi-proxy signatures, proxy multi-signatures, proxy blind signatures, etc. Unlike one to one scheme the proxy multi-signature scheme proposed by Yi et al's [4] allows two or more original signers to delegate his signing power to single proxy signer to sign the messages for all original signers. Yi et al's [4] proposed two types of proxy multi-signature scheme one is Mambo like proxy multi-signature scheme and another is Kim like proxy multi-signature scheme. Sun's [5] showed that both scheme are insecure. The Mambo-like proxy multi-signature scheme in [4] suffers from the public key substitution attack easily. The Kim-like proxy multi-signature scheme in [4] suffers from a kind of direct forgery. After that Sun [5] proposed two proxy multi-signature schemes one is Proxy protected proxy multi-signature scheme (Mambo like) and another is Proxy unprotected proxy multi-signature scheme (Kim like). Between these two schemes, one scheme provides the protection for proxy signers while another scheme does not. In these schemes, the secure channel is not necessary. However, Sun's [5] and Yi et al's [4] schemes have the common disadvantage that is size of the proxy signature depend on the number of original signers and both schemes involve exponential operation to verify proxy signature. Accordingly, an improvement is proposed to change the exponential operations into elliptic curve multiplicative ones. The elliptic curve cryptosystem can achieve a level of security equal to that of RSA or DSA but has a lower computational overhead and a smaller key size than both of these. Therefore, it is used in Sun's schemes [5] to improve their efficiency.

In light of the high computational overhead of Suns schemes [5] and Yi et. al's [4] scheme a new Efficient Multi signature scheme has been proposed by Tzer-shyong chen and Gwo-shiuan et. al's [10]. After that Tzer-shyong chen and Kuo-Hsuan et. al's [11] proposed A tracable proxy multi signature scheme. These scheme are based on ECC that can perform more efficiently then those based on DLP. These schemes are based on Elliptic curve discrete logarithm problem(ECDLP). These schemes makes size of the proxy signature independent of the number of original signers, so the computation overhead required for the verification is reduced. For improving Sun's [5] and Yi et al's [4] schemes so many DLP based schemes also proposed like Chien-Lung Hsu et. al's [2] and Guilin Wang et. al's [7]. But these schemes involve exponential operation to verify proxy signature. But Hsu et. al's [2] scheme and Tzer-shyong chen et. al's [11] schemes are insecure against malicious original signer. For the Hsu et. al's scheme in [2], which is suffer from cheat attack that is shown by Feng Cao and Zhenfu Cao [3] that means a malicious original signer can cheat the Certificate Authority into extracting a proxy signing key of a proxy signer. Furthermore, this attack can be used by proxy signer to cheat CA into extracting proxy signing key without the knowledge of the original signer. Yumin Yuan [12] also give improvement of this scheme. In addition to this Tzer-shyong chen and Gwo-shiuan et. al's. [10] and Tzer-shyong chen and Kuo-Hsuan et. al's [11] are also vulnerable to one original signer and all original signer proxy signing forgery attack respectively that is shown by Je Hong Park and Bo Gyeong Kang and Sangwoo Park in [6]. Original signer forgery attack means malicious original signer can generate valid proxy signature which looks like that it is generated by proxy signer. For generating valid proxy signature original signer forges proxy signing key and uses it to make a signature forgery.

In this paper we propose an efficient and secure proxy multi-signature scheme and analyze the security of the scheme. We show that our scheme are secure against the original signersforgery and public key substitution attack.

The rest of the paper is organized as follows. Sect. 2 we show the proxy requirements and security assumptions. Sect. 3 introduce the Tzer-shyong chen and Kuo-Hsuan et. al's proxy multi signature scheme and security analysis and possible attack in the scheme. In section Sect. 4 we show our proxy multi signature propose scheme and analyze its security and efficiency. Finally, Sect. 5 discusses some application.

2 Preliminaries

In 1996, Mambo, Usuda and Okamoto [8] first addressed the basic properties that a proxy signature scheme for partial delegation should satisfy, and defined them as follows:

- Verifiability
- Identifiably
- Unforgeability
- Undeniability
- Prevention of misuse

2.1 Security Assumption

The proxy multi signature scheme in this paper is based on some security assumption.

- **Elliptic curve Discrete logarithm Problem (ECDLP):** Consider the equation $Q = kP$ where Q, P, $E_p(a;b)$ and $k < p$. It is relatively easy to calculate Q given k and P, but it is relatively hard to determine k given Q and P. This is called the discrete logarithm problem for elliptic curves.
- **EC Diffie-Hellman Key Exchange:**
  - A's Private key and public key n_A and $P_A = n_A \times G$, This is point in $E_q(a;b)$.
  - B similarly selects a private key n_B and computes a public key P_B.
  - A generates the secret key $K = n_A \times P_B$ and B generates the secret key $K = n_B \times P_A$.
  - Both having same secret key, $n_A \times P_B = n_A \times (n_B \times G) = n_B \times (n_A \times G) = n_B \times P_A$

3 Tzer-Shyong Chen and Kuo-Hsuan et al.'s Proxy Multi-signature Scheme

This scheme is improved version of Tzer-shiuan et. al scheme [10]. In Tzer-shiuan et. al scheme each original signer sends information in individual manner but in this scheme [11] each original signer calculate some group commitment value that is common for all and then generate information using it and then send to proxy signer. This scheme has four phases.

1. proxy public key generation phase: All original and proxy signer generate their public and private key in this phase.
2. proxy signing key generation phase: For delegating signing power to proxy signer each original signer A_i performs following steps
   - A_i securely selects a random number $k_i \in \{1,2,\ldots,t-1\} \backslash d_i$ and computes R_i = k_iB = (x_{R_i}, y_{R_i}).
   - Broadcast R_i to the other original signer.
   - upon receiving R_j computes R = $\sum_{i=1}^{n} R_i$ = (x_R, y_R).
   - Then computes $s_i = d_i h(m_w, x_{Q_i}, x_{Q_P}, x_R)$ - k_i mod t.
   - Sends sub delegation parameter (m_w, s_i) to proxy signer.
3. Sub delegation parameter verification and secret key generation:
   - using Sub delegation parameter (m_w, s_i) proxy signer first calculate R'_i=($x_{R'_i}$,$y_{R'_i}$) as follows

$$R'_i = Q_i \times h(m_w, x_{Q_i}, x_{Q_P}, x_R) - s_i \times B.$$

and checks

$$x_{R'_i} = x_{R_i} \bmod t.$$

   - if all parameter is valid then proxy signer compute proxy signing key as follows

$$d = d_P + \sum_{i=1}^{n} s_i \mathrm{mod} t.$$

4. Proxy signature generation and verification: Proxy multi-signature is attached to the message m in the form of $(m, m_w, R, Sig_d(m))$, where $Sig_d(m)$ means the signature generated by designated scheme using the proxy signature key d. For verifying signature, verifier computes proxy public key Q corresponding to the proxy signing key d as

$$Q = Q_P + \sum_{i=1}^{n} h(m_w, x_{Q_i}, x_{Q_P}, x_R) Q_i - R.$$

   with this proxy public key the verifier confirms the validity of signature by validating the verification equation.

Now we discuss the security of this scheme. This scheme is suffer from one attack that is original signer forgery attack, that is described by Je Hong park, Bo Gyeong Kang et. al [6] Original signers forgery attack in which conspiracy of all original signers to generate valid proxy multi-signature without the agreement of proxy signer. They show how attack is possible as follows.

The original signer A_i select random number k_i and then compute

$$R_i = k_i B \text{ for } 1 \leq i \leq n.$$

Furthermore A_i adds Q_P to R_1 and then computes

$$R = \sum_{i=1}^{n} R_i = Q_P + \left(\sum_{i=1}^{n} k_i\right) B$$

The forged proxy signing key that is generated by all original signers is as follows

$$d = \sum_{i=1}^{n} d_i h(m_w, x_{Q_i}, x_{Q_P}, x_R) - k_i$$

from the verification equation, proxy public key Q computed by verifier as follows

$$\begin{aligned} Q &= Q_P + \sum_{i=1}^{n} Q_i h(m_w, x_{Q_i}, x_{Q_P}, x_R) - R \\ &= Q_P + \left(\sum_{i=1}^{n} d_i h(m_w, x_{Q_i}, x_{Q_P}, x_R)\right) B - \left(\sum_{i=1}^{n} k_i\right) B - Q_P = dB \end{aligned}$$

This means verifier will convinced that any proxy multi signature signed by using the forged signing key d are generated by agreement of all original signer and P. so this scheme is not proxy protected.

4 Our Proposed Proxy Multi-signature Scheme

A new ECDLP based proxy multi signature scheme is presented in this paper. The proposed scheme is independent to the number of original signer and we are also using the merits of ECC so the overhead of computation and communication cost due to modular exponential operation also reduced. Our scheme also secure against original signer forgery attack and public key substitution attack without using any encryption and decryption. This scheme involves three parties: original signers A_i $1 \leq i \leq n$, proxy signer p and verifier v and scheme is divided into four phases those are as follows:

1. System initialization phase: The following parameters over the elliptic curve domain must be known

   F_p: A field size p (a large prime number (512 bits)).

   E: An elliptic curve of the form $(y^2 = x^3 + ax + b \bmod p)$ over F_p, where $a, b \in F_p$ such that $4a^3 + 27b^2 \neq 0 (\bmod\ p)$. $E(F_p)$ represents a set of points $(x, y) \in F_p \times F_p$ which satisfy E and with an additional point called point at infinity O. The cardinality of E should be divisible by a large Prime number because of the issue of security raised by Pohlig and Hellman.

   B: $(B \neq O)$ A finite point on E with an order t (a large prime number).

Every participant has a private/public key pair $(d_Z, Q_Z = (x_Z, y_Z) = d \times B)$, where $d_i \in [1, t-1]$ such that $x_Z \neq 0$. Subscript Z indicates the identification of participant Z. $Q_{ZX} = Q_Z \times d_X = Q_X \times d_Z$ is the Diffie-Hellman shared secret key between the persons Z and X.

2. Key generation phase: In this phase original signer delegates his signing power to proxy signer by generating a key (proxy signing key) using his private key d_o and message warrant m_w. The value (proxy signing key) is equivalent to the signing of warrant by original signer using his private key. The Steps to generate the proxy key are shown below:

Part 1: Private/Public key generation phase: All original signers and the designated proxy signer are authorized to select their own individual secret keys. All original signer and proxy signer randomly selects a number $d_i \in [1, t-1]$ this number is his private key and then using this they calculate public key that is $Q_i = d_i \times B = (x_{Q_i}, y_{Q_i})$. If $x_{Q_i} \neq 0$, d_i is the secret key and Q_i is the public one.

Part 2: Proxy signing key generation phase

Step 1: Each original signer selects random number (secret key) $t_{o_i} \in \{1,2,3,\ldots,t-1\} \backslash d_i$, and then computes k_i = $t_{0_i} \times B$ = (x_{k_i}, y_{k_i}) and also computes

$$v_i = d_i h(m_w, x_{Q_i}) + t_{0_i} \times x_{Q_i} \bmod t.$$

Now each original signer broadcast (v_i, k_i) to other original signers. Each original use v_i for authenticate himself to the other original signers so that anyone except valid original signer cannot send these parameter and k_i use for generating group commitment value.

Step 2: Each original signer after receiving (v_i, k_i) first verify the parameter and checks all parameters are correct i.e.

$$R'_i = h(m_w, x_{Q_i}) \times Q_i + k_i \times x_{Q_i} \bmod t.$$

if $v_i \times B$ = (x_{v_i}, y_{v_i}) and x_{v_i}=$x_{R'_i}$ mod t then original signer accepts (v_i, k_i) as a valid parameter.

Step 3: Then all original signers calculate group commitment value K = $\sum_{i=1}^{n} k_i$ = (x_K, y_K).

Step 4: Now each original signer calculate sub delegation parameter using his secret key, private key and group commitment value that is

$$x_o = \sum_{i=1}^{n} x_{Q_i}$$

$$\sigma_i = d_i x_{Q_i} h(m_w, x_{Q_i}, x_K, x_p, x_o) + t_{O_i} x_K \bmod t.$$

Step 5: Now each original signer sends sub delegation parameter to proxy signer with the help of Diffie-Hellman key exchange so that their is mutual authentication between original signer and proxy signer. First original signer makes the parameter that has to be send as follows:

Step 6: Original signer choses a random number $\beta \in$ [1, t - 1] and calculates λ_1 = $\beta \times Q_{O_i}$, λ_2 = $\beta \times Q_{O_iP} = (x_2, y_2)$, and λ_3 = $\sigma_i \times Q_{O_iP}$ = (x_c, y_c)

such that $x_2 \neq 0$, otherwise he has to repeat this step with another random number.

Step 7: Sends the proxy share as (λ_1, λ_3, ($x_2 \times \sigma_i$)mod t, $x_2 \times k_i$, K, m_w) to p.

Step 8: Upon receiving the value (λ_1, λ_3, ($x_2 \times \sigma_i$)mod t, $x_2 \times k_i$, K, m_w) from the original signer, p gets the partial proxy share σ_i back as below: Calculates $\lambda_2 = \lambda_1 \times d_p = (x_2, y_2)$, using this λ_2 he will calculate $\sigma_i = \sigma_i \times x_2 \times x_2^{-1}$ mod t and $k_i = x_2 \times k_i \times x_2^{-1}$ and then verifies the validity of σ_i by checking the equation $\lambda_3 =^? \sigma_i \times Q_{O_iP}$ and $R'_i = h(m_w, x_{Q_i}, x_K, x_p, x_o) \times Q_i \times x_{Q_i} + k_i \times x_K$mod t. if $\sigma_i \times B = (x_{\sigma_i}, y_{\sigma_i})$ and x_{σ_i}=$x_{R'_i}$ mod t then proxy signer accepts sub delegation parameter. If the equality gets hold, both validity of the share and authentication of original signer are proved.

Step 9: Proxy multi signature secret key generation: After validating all sub delegation parameter proxy signer computes proxy signing secret key on behalf of all original signer as follows:
first proxy signer calculate

$$\sigma_0 = \sum_{i=1}^{n} \sigma_i$$

and then proxy signer generate random number l and calculate $L = l \times B = (x_L, y_L)$. finally calculate proxy signing secret key $d_{p'}$

$$d_{p'} = \sigma_0 \times x_{Q_p} + h(m_w, x_K, x_p, x_o, x_L) \times l + d_p \times x_L \text{mod}\, t.$$

3. Proxy multi signature generation phase: After generation of proxy signing key proxy signer sign the message m on a behalf of original signer using secret key $d_{p'}$ and resultant signature is $sign_{d_{p'}}(m)$ and later proxy signer p choses a random number $\beta' \in [1, t - 1]$ and calculates $\lambda_4 = \beta' \times Q_p$, $\lambda_5 = \beta' \times Q_{vp} = (x_3, y_3)$, such that $x_3 \neq 0$, otherwise he has to repeat this step with another random number. And proxy signer send the proxy multi signature (λ_4, $x_3 \times L$, $x_3 \times K$, m_w, m, $sign_{d_{p'}}(m)$) to v.
4. Proxy multi signature verification phase: Upon receiving proxy multi signature from proxy signer verifier first calculates $\lambda_5 = \lambda_4 \times d_v = (x_3, y_3)$, using this x_3 he will calculate $L = x_3 \times L \times x_3^{-1}$ mod t and $K = x_3 \times K \times x_3^{-1}$. Now verifier computes proxy public key using these value and public key of original signer and proxy signer.

$$Q_{p'} = \sum_{i=1}^{n} Q_i h(m_w, x_{Q_i}, x_K, x_p, x_o) x_{Q_p} + K \times x_K + \\ L \times h(m_w, x_K, x_p, x_o, x_L) + Q_p \times x_L.$$

using this proxy multisignature public key verifier will validates $sign_{d_{p'}}(m)$ and verify the correctness of the verification equation.

5 Security Analysis

Security analysis and some discussion are given below. First of all we show the correctness of verification equation that means derivation of proxy multi-signature public key

$$Q_{p'} = d_{p'} \times B = \sigma_0 \times x_{Q_p} \times B + h(m_w, x_K, x_p, x_o, x_L) \times l \times B + d_p \times B \times x_L \bmod t$$

$$Q_{p'} = d_{p'} \times B = \sum_{i=1}^{n} \sigma_i \times x_{Q_p} \times B + h(m_w, x_K, x_p, x_o, x_L)L + Q_p \times x_L \bmod t.$$

After putting σ_i value we get Q_p that is

$$Q_{p'} = \sum_{i=1}^{n} Q_i h(m_w, x_{Q_i}, x_K, x_p, x_o) x_{Q_p} + K \times x_K + L \times h(m_w, x_K, x_p, x_o, x_L) + Q_p \times x_L.$$

There are some security concern as follows:

1. ECDLP: The proposed scheme is based on ECDLP, therefore attacker has to face the difficulty of solving the ECDLP so that he will be unable to derive the secret key from public key so forging the signature is difficult for attacker.
2. Parameter passing using Diffie-Hellman: The original signer and proxy signer sends the parameter to proxy signer and verifier respectively in Diffie-Hellman fashion. In which they calculate $\lambda_1, \lambda_2, \lambda_3, \ldots$ etc, these values are calculated using the public key and the private key of the receiver and sender so that it is guaranteed that the parameter is coming directly from the particular sender and also the sender is assured that only the receiver can see the values. Similarly the Receiver has surety that the parameter is coming from the actual sender. Hence there is mutual authentication between the sender and the receiver. Forging the parameter is as difficult as solving ECDLP. With this method we can also stop "original signer forgery"-sometimes original signer generates proxy signature and sends to the verifier, in that case it is not verified whether the parameter is coming from the original signer. We can now trace who is the sender and who is the receiver and the original signer can not bypass the proxy signer.
3. Public key substitution attack: The proxy signature verification equation at the verifier side is combine with the public keys of the original signer, the proxy signer and one way hash function. Forging a public key in the verification equation the attacker has to face the problem of ECDLP and one way hash function that is more difficult. If we look at verification eq. that is

$$Q_{p'} = K \times x_K + L \times h(m_w, x_K, x_p, x_o, x_L) + Q_p \times x_L + Q_1 h(m_w, x_{Q_1}, x_K, x_p, x_o) x_{Q_p} + Q_2 h(m_w, x_{Q_2}, x_K, x_p, x_o) x_{Q_p} + \ldots + Q_n h(m_w, x_{Q_n}, x_K, x_p, x_o) x_{Q_p}$$

In this equation x_o is summation of all x-coordinate of all original signers public key. In one case original signer $Q1$ may forge his public key and randomly selects a pair $Q'_1 = (x_{Q'_1}, y_{Q'_1})$ as his public key, now for satisfying the verification equation original signer can change only value of K by changing his share in K but changing the value K that means in each term in the equation there will be a change and changing the public key also affects x_o in one way hash function hence the difficulty of so doing is harder then the ECDLP.

6 Conclusion

We have proposed a proxy multi signature scheme based on elliptic curve cryptosystem (ECC). The proposed scheme is secure then Tzer-Shyong Chen [11], Kuo-Hsuan et. al [2] and Sun et. al scheme [5] and computation cost is independent of the number of original signers. Our scheme uses Diffie-Hellman for sending the sub delegation parameter and we are not using any encryption and decryption method for sending parameter. So we also reduce the cost of encryption and decryption. Besides this Our scheme is able to withstand the public key substitution attack and original signer forgery attack.

References

1. Nonmember, B.L., Member, K.K.: Strong proxy signatures, December 13 (1999)
2. Hsu, C.-L., Wu, T.-S., He, W.-H.: New proxy multi-signature scheme. Applied Mathematics and Computation 162(3), 1201–1206 (2005)
3. Cao, F., Cao, Z.: Cryptanalysis on a proxy multi-signature scheme
4. Yi, G.X.L., Bai, G.: Proxy multi-signature scheme: a new type of proxy signature scheme. Electronics Letters 36, 134–138 (2000)
5. Sun, H.: On proxy multi-signature schemes. In: Proceedings of the International Computer Symposium, pp. 65–72 (2000)
6. Park, J.H., Kang, B.G., Park, S.: Cryptanalysis of Some Group-Oriented Proxy Signature Schemes. In: Song, J.-S., Kwon, T., Yung, M. (eds.) WISA 2005. LNCS, vol. 3786, pp. 10–24. Springer, Heidelberg (2006)
7. Guilin Wang, J., Bao, F., Deng, R.H.: Proxy signature scheme with multiple original signers for wireless e-commerce applications. Infocomm Security Department, Institute for Infocomm Research (I2R). 21 Heng Mui Keng Terrace, Singapore 119-613. IEEE (2004)
8. Mambo, M., Usuda, K., Okamoto, E.: Proxy signatures for delegating signing operation. In: Neuman, C. (ed.) Proceedings of the 3rd ACM Conference on Computer and Communications Security, pp. 48–57. ACM Press, New Delhi (March 1996)
9. Kim, S., Park, S., Won, D.: Proxy signatures. In: Proc. 1st International Information and Communications Security Conference, pp. 223–232 (1997)
10. Chen, T.-S., Chung, Y.-F., Huang, G.-S.: Efficient proxy multi-signature schemes based on the elliptic curve cryptosystem. Computers & Security 22(6), 527–534 (2003)
11. Chen, T.-S., Chung, Y.-F., Huang, K.-H.: A traceable proxy multi-signature scheme based on the elliptic curve cryptosystem. Applied Mathematics and Computation 159(1), 137–145 (2004)
12. Yuan, Y.: Improvement of Hsu-Wu-He's Proxy Multi-signature Schemes. In: Jonker, W., Petković, M. (eds.) SDM 2005. LNCS, vol. 3674, pp. 234–240. Springer, Heidelberg (2005)

SVIP-Enhanced Security Mechanism for SIP Based VoIP Systems and Its Issues

D. Chandramohan[1], D. Veeraiah[2], M. Shanmugam[1], N. Balaji[1],
G. Sambasivam[1], and Shailesh Khapre[1]

[1] School of Engineering, Department of Computer Science and Engineering,
Pondicherry University, Pondicherry
[2] Department of Computer Science, Vignan University
{pdchandramohan,d.veeraiah,maddy.shan,nbalajime1983,
gsambu,shaileshkhaprerkl}@gmail.com

Abstract. As the SIP based VoIP system is entirely based on IP network, the vulnerabilities in it affects the VoIP system. This may result in degrading of quality of service in three aspects such as confidentiality, integrity and availability. This paper diagnose the security issues such as registration hijacking, session teardown, message tampering, IP spoofed flooding, disguised proxy and Spam over Internet Telephony (SPIT) and as well as the enhancements to improve the security mechanism to overcome the issues and to provide the quality of service to the legitimate users.

Keywords: Voice over IP (VoIP), Session Initiation Protocol (SIP), User Agent (UA), Real Time Protocol (RTP), Spam over Internet Telephony (SPIT), Transport Layer Security (TLS), Secure Real Time Protocol (SRTP) Introduction (Heading 1).

1 Introduction

VoIP has reached rapid development in the IT sectors for communication which is associated with Session Initiation Protocol as a signaling mechanism for connection establishment. As the VoIP [4] relies on the IP network, the vulnerabilities of the IP network affect the VoIP system. The effects of vulnerabilities affect the three major services to the legitimate users such as confidentiality, integrity and availability. The interception of attackers may hack the user's data and the personal information which may affect the confidentiality to the legitimate users. Through Real Time Protocol (RTP) packets illegal users tap the phone conversations of the legitimate users which make the absence of integrity to authorized and authenticated users. The service provided by the VoIP system is denied indirectly by the VoIP system to the legitimate users because of message tampering and flooding of SIP requests to the VoIP system by the attacker which results in affecting the quality of service and availability of service to the authorized and authenticated users.

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 81–86.
springerlink.com

2 Functions of VOIP

VoIP stands for Voice over Internet Protocol which transmits voice as an IP packet through internet. VoIP digitalize the voice into data packets, sending them and reconverting them into voice at destination. The digitalized signal is more noise tolerant than the analog signal. For conversion of digital into analog and vice versa the VoIP [2] system uses the convertor as well as the compression algorithm like Pulse Code Modulation (PCM) and Adaptive Differential Pulse Code Modulation (ADPCM) at both ends of the communication media. VoIP data packets are transmitted through RTP associated by UDP/IP packets. VoIP doesn't use TCP/IP because it is too heavy for real time applications, so UDP/IP is used. UDP has no control over the order in which packets arrive at the destination and the time duration to deliver the packets. Both of these are important for voice quality and conversion quality. RTP solves the problem enabling the receiver to put the packets back into the correct order and not for waiting long time for delivery of packets

3 Essential Components of SIP

IETF developed SIP [4] for IP based multimedia communications control protocol in the following aspects of function user location, session setup and session management and performance management. There are several types of SIP components such as SIP User Agent (UA) and SIP Registration Server, SIP Proxy Server and SIP Redirect Server. SIP Proxy Server receives the call requests through hop-by-hop technique. The SIP Redirect Server performs the routing which routes the called IP address to the Caller IP address. SIP Registration Server receives the information about the particular user, who wants to make use of VoIP communication and it provide the location based services.

4 Security Threats in SIP

SIP signaling mechanism is subjected to six types of attack, registration hijacking, disguised proxy, and malformed message, IP Spoofed Flooding, Session Teardown and Spam over Internet Telephony (SPIT).

4.1 Registration Hijacking

In order to make voice communication media between two end users, users must register themselves with basic information with user name and password. The requests sent by the user is simply a SIP request message[2], where the body of the message contains the user information, the username as well as the password and the FROM field which contains the IP address of the user and the destination is mentioned in the TO field. Twist (0) = O (1) = Z/2. An attacker can intercept and performs the Man-in-Middle attack captures the packets modify the FROM field to their own IP address and gain the username and password of the legitimate user. Through which the attacker can modify, listen and crack the data of the particular user's signals and media packets. Model of Hijack threads involved indicated and described by following mathematical steps,

```
Twist (0) = O (1) = Z/2
Twist (1) = U (1) = SO (2)
Twist (2) = Sp (1) = SU (2)
Twist (3) = Sp (1) x Sp (1)
Twist (4) = Sp (2)
Twist (5) = SU (4)
{And && • ∩∩, Union → UU}
q□ (Hq • Im H∋p • qp□ (•┬q))
(q1, q2) • Im H∋p • q1pp□ (•┬q2))
⟦HP⟧^1 ≅ Twist•S^2
```

4.2 Disguised Proxy

Generally UDP packets are used for signaling between user agent and the proxy server where the security level is at low, this paves the way for the attacker to impersonate as a proxy through whom an attacker will be able to access all SIP messages and complete control of VoIP system call. Twist (1) = U (1) = SO (2). The illegal user disguises himself as a proxy by spoofed Domain Name Service (DNS) and camouflage Address Resolution Protocol (ARP) which is similar to the registration hijacking. If an attacker spoofed the DNS system of a particular domain, the attacker can make calls to any of the domain without any authentication and can also monitor and record the calls by intercepting the communication media.

4.3 Spam over Internet Telephony (SPIT)

Spam over Internet Telephony (SPIT) [1] is also called as VoIP Spam (VAM) is a bulk of message broadcasted over VoIP to phones connected to the internet. Marketers already using the voice mail for commercial messages, IP telephony makes a more effective channel because the sender can send messages in bulk instead of dialing each number separately. Twist (2) = Sp (1) = SU (2). Unscrupulous markets can use spam bots to harvest VoIP address or may hack into a computer used to route VoIP calls.

4.4 IP Spoofed Flooding

This is one of the denial of service where the attacker flood the SIP requests with different IP address spoofed from any mail server to the Twist (3) = Sp (1) x Sp (1) SIP server making the server overload and to terminate all calls which are currently in progress.

4.5 Malformed Message

The attacker sends a malformed or otherwise malicious SIP INVITE request to a telephony server which causes a crash of that server. Twist (4) = Sp (2). This is specifically a problem for operators that run their servers on the public Internet. Because SIP allows the usage of UDP packets, it is easy for an attacker to spoof any source address in the Internet and send the malformed message from untraceable locations. By flooding these requests periodically, attackers can completely interrupt the telephony service.

4.6 Session Tear Down

Session tear down occurs when an attacker observes the signaling for a call, and then sends spoofed SIP “BYE” messages to the UAs. Twist (5) = SU (4). Most SIP UAs do not require strong authentication, which allows an attacker to send a properly crafted BYE messages to the two UAs, tearing down the call. If the UA does not check the value of the call, the attacker simply sends the “BYE” message if the user is active by spoofing the IP address of the UA which causes the calls to be tear down.

5 Enhancing the Security Mechanism in SIP

5.1 TLS over SIP

Session Initiation Protocol (SIP) is not inherently secure. It is essentially a communications-specific version of the HTTP protocol that makes up the basis for web data. Just as HTTP uses Secure Sockets Layer (SSL) [4] and security certificates to encrypt communications and ensure secure data transmission on the Web, SIP needs some additional layer of protection to ensure that VoIP and other audio/video communications that rely on SIP are secure. The majority of VoIP communications are secured using Message Digest 5 (MD5) authentication. MD5 has some known weaknesses and recently vulnerable to spoofing which could allow an attacker to fake an MD5 certificate. The much more secure alternative is Secure Multipurpose Internet Mail Extensions (S/MIME) which does not have the weaknesses of MD5 and can encrypt data directly within the SIP packets. Basically, just as HTTP rides on SSL, SIP rides on Transport Layer Security (TLS). Encrypting SIP transmissions with TLS helps to protect communications from man-in-the-middle attacks, eavesdropping, or unauthorized access. Secure SIP (SIPS), or SIP over TLS [4], enables the session to be encrypted on a hop-by-hop basis between the source and destination, providing better security than basic MD5 authentication, but without the complexity and overhead imposed by S/MIME. The SIPS URI ensures that SIP over TLS is used to encrypt and protect communications between hops and provide a secure connection from end-to-end.

5.2 Internet Protocol Security (IPSec)

IPSec [6] may be used to secure data transmissions between SIP gateways and proxy servers within a network, but IPSec is not suitable for protecting VoIP and unified communications data from end to end. IPSec establishes a secure connection between the source and destination devices, meaning that SIP proxies and hops along the way are unable to decrypt or modify the information in the SIP packets. TLS is a less complex and easier to manage solution that accomplishes the protection of the SIP session while still allowing the interim hops to work with the SIP data.

5.3 Secure Real Time Protocol (SRTP)

SRTP [2] is a security profile for RTP that adds confidentiality, message authentication, and replay protection to that protocol. It is an action item in the IETF Audio-Video Transport Working Group. Similar to RTP, SRTP is primarily used in VoIP communications. SRTP uses authentication and encryption algorithms to reduce the risk of a denial of service attack. It has the ability to achieve a high throughput in numerous communication platforms, including both wireless and hard-wired devices. SRTP mainly aims at ensuring the authenticity of a communication partner. For this purpose, communication partners usually exchange individual keys for each connection, with which message transmission is encrypted. Consequently, intruding parties cannot send messages, the originator of which is apparently a different person. Securing the confidentiality of RTP payload.RTP data is encrypted before its transmission for this purpose. This makes it difficult for intruding parties to read along sent messages. SRTP [6] Guarantees the integrity of RTP payload and header. This disables an intruding party to alter a message unnoticed. The authentication algorithms used in SRTP are MD5 and Hash Message Authentication Code (HMAC).

5.4 S/MIME

The Session Initiation Protocol specification currently details optional support for the use of secure MIME [4]. In general, the encryption of SIP message end-to-end is problematic because there are certain SIP entities, which need to view and modify the SIP headers and bodies. However, there will still be two standards ways how to encrypt SIP messages at SIP layer. Firstly, S/MIME [6] can be used to encrypt the SIP bodies. This is the most common and traditional use of MIME. Secondly, S/MIME can also be used to encrypt confidential SIP headers together with SIP bodies using a special SIP S/MIME tunneling mechanism. In SIP S/MIME tunneling, a SIP message can be protected and wrapped in S/MIME.

Table 1. Table of Requirements

Requirements	Solutions
Communicating to correct callee	TLS, S/MIME, IPSec
Correct invoicing	TLS, S/MIME, IPSec
Signal Protection	TLS, S/MIME, IPSec
Authentication of both end-user	TLS, S/MIME, IPSec
Secured Communication Media	TLS, SRTP, IPSec

```
Q′ (Hq • IM H∋P • QP′ (•┬q))

(Q1, Q2) • Im , H ∋ P • Q1 PP′
(•┬Q2))

⟦HP⟧^1 ≅ Tw•S^2
```

Table 1 discusses about the issues and so the requirements for SIP based VoIP and as well as the solutions to tackle over the requirements needed for SIP based VoIP System to give the best quality of service to the authorized and authenticated users.

6 Conclusion

We have discussed about the functionality of VoIP, security issues in session initiation protocol and the enhancement of security mechanism in SIP based VoIP through which confidentiality, integrity and availability can be provided to the legitimate users and we have also discussed about the requirements and the protocols to overcome the requirements to provide a secure way of communication for SIP based VoIP.

References

1. Shan, L., Jiang, N.: Research on Security Mechanism of SIP basedVoIP System. In: 2009 IEEE Int. Conf. on Hybrid Intelligent Systems, pp. 408–410 (2009)
2. Albers, J., Hahn, B., McGann, S., et al.: An Analysis of Security Threats and Tools in SIP-Based VoIP Systems [EB/OL] (September 2005), http://www.colorado.edu/policylab/Papers/Univ_Colorado
3. Rosenberg, J., Schulzrinne, H., Camarillo, G., et al.: SIP: session initiation protocol [EB/OL] (June 2002), http://www.ietf.org/rfc/rfc3261.txt; Lucky, R.W.: Automatic equalization for digital communication. Bell Syst. Tech. J. 44(4), 547–588 (1965)
4. Zourzouvillys, T., Rescorla, E.: An Introduction to standards- Based VoIP. IEEE Internet Computing, 69–73 (2010)
5. TLS, SRTP, S/MIME, http://www.wikipedia.org (referred on January 2, 2011)
6. Salsano, S., Veltri, L., Papalilo, D.: SIP security issues: the SIP authentication procedure and its processing load. IEEE Network 16(6), 38–44 (2002)
7. Nanda Kishore, M.S., Jayakumar, S.K.V., Satya Reddy, G., Dhavachelvan, P., Chandramohan, D., Soumya Reddy, N.P.: Web service suitability assessment for cloud computing. In: Wyld, D.C., Wozniak, M., Chaki, N., Meghanathan, N., Nagamalai, D. (eds.) NeCoM 2011, WeST 2011, WiMoN 2011. CCIS, vol. 197, pp. 622–632. Springer, Heidelberg (2011)
8. Chandramohan, D., Jayakumar, S.K.V., Khapre, S., Nanda Kishore, M.S.: DWSE-Simulator For Distributed Web Service Environment. In: IEEE-International Conference on Recent Trends in Information Technology, ICRTIT-2011, pp. 1203–1208 (2011)
9. Chandramohan, D., Jayakumar, S.K.V., Khapre, S.: DTDWS-Design of TestBed for Distributed Web Service Environment. International Journal of Engineering Science and Technology (IJEST) 3(3), 2399–2404 (2011)
10. Chandramohan, D., Khapre, S., Ashokkumar, S.: A Study of Finding Similarities in Web Service Using Metrics. Selected for Publication in International Journal of Scientific & Engineering Research (IJSER) 2(6) (June 2011) ISSN 2229-5518
11. Khapre, S., Chandramohan, D.: Personalized Web Service Selection. International Journal of Web & Semantic Technology (IJWesT) 2(2), 78–93 (2011)

Host-Based Bot Detection Using Destination White-Lists for User's Profile

B. Soniya and M. Wilscy

Dept. of Computer Science, Kerala University, Karyavattom, Trivandrum, India
soniya.balram@gmail.com,
wilsyphilipose@hotmail.com

Abstract. Bots have become a popular vehicle for Internet crime. Bot detection is still a challenging task since bot developers come up with techniques for evading detection. Most bot detection techniques are network based and rely on correlation of behavior among similar hosts. Besides, network based systems deal with voluminous traffic and result in non-negligible false alarms. We propose a host-based detection technique leveraging the recurring patterns in the traffic generated by processes in a single user's profile. From outgoing traffic in an un-infected host, destination white-lists for a user profile are generated. These white-lists along with bot behavior are used for detection. We were able to detect two real life bots using our method.

1 Introduction

Computing has moved from supercomputers, workstations and PCs to laptops, netbooks, tablets and iphones and most of these devices have an always-on connectivity to the Internet. Such devices are being increasingly used for activities like banking, bill-payment, shopping and many other online transactions. But it is seen that users of these devices fail to protect and patch themselves, making them vulnerable to malwares. Criminals on the Internet can easily leverage such vulnerable machines to recruit them into botnets thereby making them victims of various attacks. So detection systems for such computing devices become an urgent need.

A botnet is a network of compromised machines under the influence of malware (bot) code. The botnet is commandeered by a "botmaster" and utilized as "resource" or "platform" for attacks such as distributed denial-of-service (DDoS) attacks, and fraudulent activities such as spam, phishing, identity theft, and information exfiltration. In order for a botmaster to command a botnet, there needs to be a command and control (C&C) channel through which bots receive commands and coordinate attacks and fraudulent activities. The C&C channel is the means by which individual bots form a botnet [1].

Botnet detection is an active area of research and detection systems can be broadly classified into network-based and host-based systems. Most of the botnet detection systems seen [1,2,5] are network based. Network based botnet detection systems analyze various features of network traffic like destination address, packet payload, sequence of bot events in traffic etc to pinpoint bots. The traffic is collected from network edges and tends to be voluminous resulting in higher processing and false

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 87–93.
springerlink.com

positives. Host-based systems monitor parameters on a single host. They are more suitable for personalized computing devices. Besides, the amount of data analysed by such a system is less and hence a more fine grained analysis is possible.

A survey of literature on bot analysis [6,7,8,9] reveal that more and more bots use HTTP as the command and control mechanisms adding more stealth to their operation. Data sent by bots over HTTP or P2P protocol are encrypted making detection more difficult. They are capable of downloading extra modules and executing them on the fly. It is also seen that bots do not generate processes of their own but have a tendency to inject themselves into other executing processes [8,9]. Considering this behavior, it is felt that the traffic generated by each of the processes on a machine could be studied to find the normal and contaminated behaviors. This paper proposes a host-based bot detection system leveraging the recurring patterns in the traffic generated by processes in a single user's profile. From outgoing traffic in an un-infected host, destination white-lists for a user profile are generated. These white-lists along with bot behavior are used for detection. We were able to detect two real life bots using our method. This paper is organized as follows: Section 2 considers existing host-based bot detection techniques and some limitations, Section 3 discusses the premise of the proposed system and the bot detection method is elaborated in Section 4. This is followed by details of the experimental set up in Section 5 and a discussion of results in Section 6. Section 7 provides the concluding remarks.

2 Existing Host-Based Detection Techniques

Thus far most researchers have proposed detection strategies for bots at the network level. Network-based systems attempt to detect bot infections by correlating similar behaviors among several different hosts on the monitored network. So they require that multiple hosts in the same network become infected for the bot to be detectable. It should be noted that botnets are constantly evolving and changing, e.g. from a centralized to distributed C&C structure, thereby increasing the complexity of network level only investigations. Bot C&C behavior is manifested equally well at the infected host. A finer grained analysis can be done at the host -level and can definitely complement the results of network-based systems.

Some of the host-based bot detection systems proposed in recent times are [3,4,10,11,12,13,14]. The research presented in [10] proposed a system which monitors outbound packets from a host and compares with destination-based white-lists. The white-lists are generated by observing an un-infected PC. Although this is a straightforward technique, the detection can be done only during the non-operating time of the PC. The research by J. A. Morales et al [12] tries to identify bot processes using a decision tree classifier which is input symptoms such as DNS activities of the host, digital signatures of file image and absence of GUI. The work in [13] analyses all outbound traffic to identify malicious ones but is intrusive to the user and is restricted to port 80 traffic. Jonghoon Kwon et al [14] detect bots based on flooding attacks on hosts.

3 Proposed System

The proposed method aims to identify bots by looking at outbound traffic on a single host. It is observed that users in general tend to use computers for specific applications – word processing, computations, gaming, entertainment, browsing the web, social networking and so on. The applications used and hence the processes created on a host machine by a user are restricted to a specific but slightly varying set. It logically follows that the traffic generated by the created processes also follows a pattern specific to the user. As mentioned in Section 1, malicious bots inject themselves into benign processes to hide themselves from detection. Although bots exhibit this and other stealthy behaviors, for any bot to serve its purpose, it needs to communicate with its bot master for updates and commands. At the same time, communication is required with other machines for further infection and propagation as well as launching attacks. It was felt that observing the packets generated by bot-infected hosts would provide a window to its detection.

We propose to monitor the outbound traffic from each of the processes on an uninfected host and generate a white-list of domain names. A change in traffic generated due to malware ingestion will indicate an infected bot process.

The detection method proposed is based on the following points:

1. In a successfully configured botnet system, the bots communicate periodically with the bot master for updating their status information as well as getting updated configuration information.
2. Bots generate repeated DNS queries to resolve the domain names of their bot masters.
3. On unsuccessful name resolution using DNS, NetBIOS name resolution is also repeatedly tried.
4. Some bots also repeatedly try to generate scan traffic in order to contact the bot masters.
5. On unsuccessful name resolution, some bots try to contact bot masters through hard coded IP addresses.
6. Bot code might be injected into one or more benign processes on an infected host.
7. The destinations being contacted by the bots are not the ones commonly accessed by a normal host.

Based on the above observations a bot detection method is being proposed.

4 Bot Detection Method

1. The following parameters of traffic generated in the host system is captured:
   a. Process Name & Process ID of the process generating traffic.
   b. Source Name , IP Address and Port
   c. Destination Name, IP Address and Port.
   d. Protocol
   e. Time of packet generation
2. The traffic is continuously captured and aggregated for specific time slots.

3. For a time slot t, the outgoing flows from each process are identified.
4. Per-Process and Per-Profile destination domain name lists are generated.
5. The lists are compared with the previously generated white-list from un-infected host.
6. If a previously unseen destination domain name is found, it is designated as suspicious. Outgoing traffic to the suspicious domain is analysed as follows.
   a. To detect successfully connected bots

      Look for periodic communication to the suspicious destination. This indicates C&C behavior.
   b. To detect bots which are unable to locate their C&C server, look for
      i. Repeated DNS queries to the suspicious destination domains.
      ii. Repeated NetBIOS queries for the suspicious destination domains.
      iii. Repeated failed TCP/UDP connections to unknown destinations indicating the bot trying to contact the C&C server using hard coded IP addresses

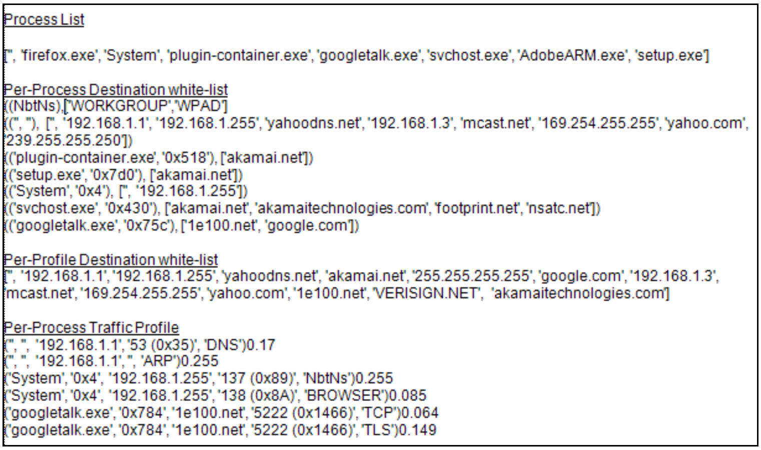
5 Experimental Setup and Data Collection

DETER testbed [16] is used to setup a botnet for observing the bot behavior. A Zeus [8] botnet is set up with one botmaster and 9 bots on Windows XP SP2 machines. The traffic generated by bots is observed. It is seen that the bots periodically communicated with the botmaster to update their status as well as to get updates and configuration information. The domain names of the botmaster are configured into the bot clients. In cases where the bots are not able to contact the bot master, they periodically tried to generate scan traffic, DNS queries, NetBIOS queries. It is also noticed that no new process was created in the bot client, but the bot injected itself into bot processes services.exe and explorer.exe.

A prototype detection system is implemented on a Windows XP host machine. Microsoft Network Monitor 3.4 is used to capture traffic generated by the processes on the host. Testing is done for traffic generated by Zeus and BlackEnergy [15] bots.

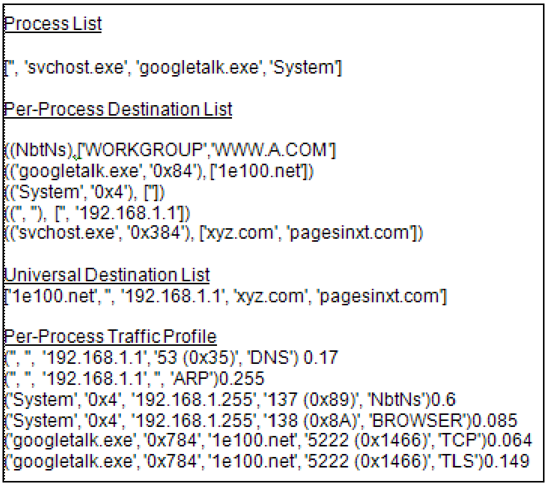
6 Results and Discussion

Figures 1 and 2 show the profiles for an un-infected and infected system respectively. The profile for the un-infected system is obtained by analysing data over several time slots. The profile for the infected system in Figure 2 is for a single time slot. The profile data for an un-infected system includes Process List, Per-Process Destination white-list and Per-Profile Destination white-lists. The Figure 2 showing a profile for an infected host shows the suspicious domains *xyz.com* and *pagesinxt.com* generated by the process svchost.exe. These domains are further investigated as in 6a and 6b of our Bot Detection Method on the data generated by svchost.exe in the specified time slot. This two step method of analysis greatly reduces the amount of data to be analysed for detection.



```
Process List

['', 'firefox.exe', 'System', 'plugin-container.exe', 'googletalk.exe', 'svchost.exe', 'AdobeARM.exe', 'setup.exe']

Per-Process Destination white-list
((NbtNs),['WORKGROUP','WPAD']
(('', ''), ['', '192.168.1.1', '192.168.1.255', 'yahoodns.net', '192.168.1.3', 'mcast.net', '169.254.255.255', 'yahoo.com',
'239.255.255.250'])
(('plugin-container.exe', '0x518'), ['akamai.net'])
(('setup.exe', '0x7d0'), ['akamai.net'])
(('System', '0x4'), ['', '192.168.1.255'])
(('svchost.exe', '0x430'), ['akamai.net', 'akamaitechnologies.com', 'footprint.net', 'nsatc.net'])
(('googletalk.exe', '0x75c'), ['1e100.net', 'google.com'])

Per-Profile Destination white-list
['', '192.168.1.1', '192.168.1.255', 'yahoodns.net', 'akamai.net', '255.255.255.255', 'google.com', '192.168.1.3',
'mcast.net', '169.254.255.255', 'yahoo.com', '1e100.net', 'VERISIGN.NET',  'akamaitechnologies.com']

Per-Process Traffic Profile
('', '',  '192.168.1.1', '53 (0x35)', 'DNS')0.17
('', '',  '192.168.1.1', '', 'ARP')0.255
('System', '0x4', '192.168.1.255', '137 (0x89)', 'NbtNs')0.255
('System', '0x4', '192.168.1.255', '138 (0x8A)', 'BROWSER')0.085
('googletalk.exe', '0x784', '1e100.net', '5222 (0x1466)', 'TCP')0.064
('googletalk.exe', '0x784', '1e100.net', '5222 (0x1466)', 'TLS')0.149
```

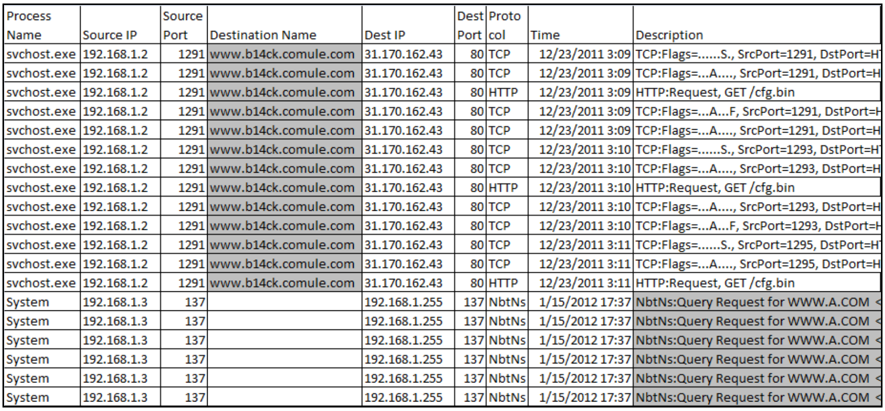
Fig. 1. Normal Profile Sample



```
Process List

['', 'svchost.exe', 'googletalk.exe', 'System']

Per-Process Destination List

((NbtNs),['WORKGROUP','WWW.A.COM']
(('googletalk.exe', '0x84'), ['1e100.net'])
(('System', '0x4'), [''])
(('', ''), ['', '192.168.1.1'])
(('svchost.exe', '0x384'), ['xyz.com', 'pagesinxt.com'])

Universal Destination List
['1e100.net', '', '192.168.1.1', 'xyz.com', 'pagesinxt.com']

Per-Process Traffic Profile
('', '',  '192.168.1.1', '53 (0x35)', 'DNS') 0.17
('', '',  '192.168.1.1', '', 'ARP')0.255
('System', '0x4', '192.168.1.255', '137 (0x89)', 'NbtNs')0.6
('System', '0x4', '192.168.1.255', '138 (0x8A)', 'BROWSER')0.085
('googletalk.exe', '0x784', '1e100.net', '5222 (0x1466)', 'TCP')0.064
('googletalk.exe', '0x784', '1e100.net', '5222 (0x1466)', 'TLS')0.149
```

Fig. 2. Malicious Profile Sample

Figure 3 shows a sample of data generated by a Zeus bot. It shows the bot periodically asking for configuration information from the botmaster, www.b14ck.comule.com.The figure also illustrates another bot trying to contact the botmaster www.a.com using the Netbios protocol.



Process Name	Source IP	Source Port	Destination Name	Dest IP	Dest Port	Protocol	Time	Description
svchost.exe	192.168.1.2	1291	www.b14ck.comule.com	31.170.162.43	80	TCP	12/23/2011 3:09	TCP:Flags=......S., SrcPort=1291, DstPort=HT
svchost.exe	192.168.1.2	1291	www.b14ck.comule.com	31.170.162.43	80	TCP	12/23/2011 3:09	TCP:Flags=...A...., SrcPort=1291, DstPort=H
svchost.exe	192.168.1.2	1291	www.b14ck.comule.com	31.170.162.43	80	HTTP	12/23/2011 3:09	HTTP:Request, GET /cfg.bin
svchost.exe	192.168.1.2	1291	www.b14ck.comule.com	31.170.162.43	80	TCP	12/23/2011 3:09	TCP:Flags=...A...F, SrcPort=1291, DstPort=H
svchost.exe	192.168.1.2	1291	www.b14ck.comule.com	31.170.162.43	80	TCP	12/23/2011 3:09	TCP:Flags=...A...., SrcPort=1291, DstPort=H
svchost.exe	192.168.1.2	1291	www.b14ck.comule.com	31.170.162.43	80	TCP	12/23/2011 3:10	TCP:Flags=......S., SrcPort=1293, DstPort=HT
svchost.exe	192.168.1.2	1291	www.b14ck.comule.com	31.170.162.43	80	TCP	12/23/2011 3:10	TCP:Flags=...A...., SrcPort=1293, DstPort=H
svchost.exe	192.168.1.2	1291	www.b14ck.comule.com	31.170.162.43	80	HTTP	12/23/2011 3:10	HTTP:Request, GET /cfg.bin
svchost.exe	192.168.1.2	1291	www.b14ck.comule.com	31.170.162.43	80	TCP	12/23/2011 3:10	TCP:Flags=...A...., SrcPort=1293, DstPort=H
svchost.exe	192.168.1.2	1291	www.b14ck.comule.com	31.170.162.43	80	TCP	12/23/2011 3:10	TCP:Flags=...A...F, SrcPort=1293, DstPort=H
svchost.exe	192.168.1.2	1291	www.b14ck.comule.com	31.170.162.43	80	TCP	12/23/2011 3:11	TCP:Flags=......S., SrcPort=1295, DstPort=HT
svchost.exe	192.168.1.2	1291	www.b14ck.comule.com	31.170.162.43	80	TCP	12/23/2011 3:11	TCP:Flags=...A...., SrcPort=1295, DstPort=H
svchost.exe	192.168.1.2	1291	www.b14ck.comule.com	31.170.162.43	80	HTTP	12/23/2011 3:11	HTTP:Request, GET /cfg.bin
System	192.168.1.3	137		192.168.1.255	137	NbtNs	1/15/2012 17:37	NbtNs:Query Request for WWW.A.COM <
System	192.168.1.3	137		192.168.1.255	137	NbtNs	1/15/2012 17:37	NbtNs:Query Request for WWW.A.COM <
System	192.168.1.3	137		192.168.1.255	137	NbtNs	1/15/2012 17:37	NbtNs:Query Request for WWW.A.COM <
System	192.168.1.3	137		192.168.1.255	137	NbtNs	1/15/2012 17:37	NbtNs:Query Request for WWW.A.COM <
System	192.168.1.3	137		192.168.1.255	137	NbtNs	1/15/2012 17:37	NbtNs:Query Request for WWW.A.COM <
System	192.168.1.3	137		192.168.1.255	137	NbtNs	1/15/2012 17:37	NbtNs:Query Request for WWW.A.COM <

Fig. 3. Malicious Data Sample

7 Conclusion

Botnets are serious threats to all network connected computing devices. Most bots use HTTP and P2P for their command and control. It is difficult to distinguish such C&C channels from normal traffic making detection challenging. In this paper, we present a host-based bot detection technique based on the assumption that the destination domains generated by processes on a host will belong to a specific set for a particular user. Two real world HTTP-based bots, Zeus and Blackenergy, were used to generate the dataset and we were able to detect the presence of bots on the host. A two step process of identifying unusual and suspicious destinations first and then going for a detailed analysis of flows to these destinations reduces the overhead in the detection process.

Attack behaviors of bots like flooding attacks or spam generation were not considered in this work. Browser traffic also shows a pattern for a particular user. This needs to be separately analysed and studied.

References

1. Gu, G., Perdisci, R., Zhang, J., Lee, W., et al.: BotMiner: Clustering analysis of network traffic for protocol-and structure-independent botnet detection. In: Proceedings of the 17th Conference on Security Symposium (SS 2008). USENIX Association, Berkeley (2008)
2. Zang, X., Tangpong, A., Kesidis, G., Miller, D.J.: CSE Dept Technical Report on Botnet Detection through Fine Flow Classification Report No. CSE11-001 (2011)
3. Law, F.Y.W., Chow, K.P., Lai, P.K.Y., Tse, H.K.S.: A Host-Based Approach to BotNet Investigation? In: Goel, S. (ed.) ICDF2C 2009. LNICST, vol. 31, pp. 161–170. Springer, Heidelberg (2010)
4. Fedynyshyn, G., Chuah, M.C., Tan, G.: Detection and Classification of Different Botnet C&amp;C Channels. In: Calero, J.M.A., Yang, L.T., Mármol, F.G., García Villalba, L.J., Li, A.X., Wang, Y. (eds.) ATC 2011. LNCS, vol. 6906, pp. 228–242. Springer, Heidelberg (2011)

5. Strayer, W., Lapsley, D., Walsh, B., Livadas, C.: Botnet Detection Based on Network Behavior. In: Botnet Detection. Advances in Information Security, vol. 36, pp. 1–24. Springer, Heidelberg (2008)
6. Borgaonkar, R.: An Analysis of the Asprox Botnet. In: 4th International Conference on Emerging Security Information Systems and Technologies (2010)
7. Stone-Gross, B., et al.: Your Botnet is My Botnet: Analysis of a Botnet Takeover. In: CCS 2009 Proceedings of the 16th ACM Conference on Computer and Communications Security. ACM, New York (2009)
8. Binsalleeh, H., Ormerod, T., Boukhtouta, A., Sinha, P., Youssef, A., Debbabi, M., Wang, L.: On the Analysis of the Zeus Botnet Crimeware Toolkit. In: Eighth Annual International Conference on Privacy, Security and Trust
9. Sinha, P., Boukhtouta, A., Belarde, V.H., Debbabi, M.: Insights from the Analysis of the Mariposa Botnet. In: Fifth International Conference on Risks and Security of Internet Systems (2010)
10. Takemori, K., Nishigaki, M., Takami, T., Miyake, Y.: Detection of Bot Infected PCs using Destination-based IP and Domain Whitelists during a non-operating term. In: IEEE Global Telecommunications Conference, IEEE GLOBECOM (2008)
11. Liu, L., Chen, S., Yan, G., Zhang, Z.: BotTracer: Execution-Based Bot-Like Malware Detection. In: Wu, T.-C., Lei, C.-L., Rijmen, V., Lee, D.-T. (eds.) ISC 2008. LNCS, vol. 5222, pp. 97–113. Springer, Heidelberg (2008)
12. Morales, J.A., Kartaltepe, E., Xu, S., Sandhu, R.: Symptoms-Based Detection of Bot Processes. In: Kotenko, I., Skormin, V. (eds.) MMM-ACNS 2010. LNCS, vol. 6258, pp. 229–241. Springer, Heidelberg (2010)
13. Xiong, H., Malhotra, P., Stefan, D., Wu, C., Yao, D.: User-Assisted Host-Based Detection of Outbound Malware Traffic. In: Qing, S., Mitchell, C.J., Wang, G. (eds.) ICICS 2009. LNCS, vol. 5927, pp. 293–307. Springer, Heidelberg (2009)
14. Kwon, J., Lee, J., Lee, H.: Hidden Bot Detection by Tracing Non-human Generated Traffic at the Zombie Host. In: Bao, F., Weng, J. (eds.) ISPEC 2011. LNCS, vol. 6672, pp. 343–361. Springer, Heidelberg (2011)
15. Nazario, J.: Blackenergy DDoS bot analysis. Arbor Networks, Tech. Rep. (2007)
16. DETERlab, http://www.isi.deterlab.net/

Effective Implementation and Evaluation of AES in Matlab

Amish Kumar and Namita Tiwari

Department of Computer Science Engineering,
Maulana Azad National Institute of Technology, Bhopal, India
amish_aks@yahoo.co.in,
namitatiwari@manit.ac.in

Abstract. Efficient implementation of block cipher is critical towards achieving high efficiency with good understandability. Numerous number of block cipher including Advance Encryption Standard have been implemented using different platform. However the understanding of the AES algorithm step by step is very tipical. This paper presents the efficient implementation of AES algorithm with the increase of understandability with the use of MATLAB platform. Mainly use of MATLAB in Algorithm development, Data analysis, exploration, visualization, modeling, simulation, prototyping, application development including GUI building and computation. Taking advantages of understandability in MATLAB. Implementation result of this approach shows the interest and confirms the contribution of MATLAB for robust and optimal implementation.

Keywords: AES, MATLAB, S-Box.

1 Introduction

Cryptography plays main role in information security. Many cryptographic algorithms have been proposed, such as the Data Encryption Standard (DES), the Elliptic Curve Cryptography (ECC), the Advanced Encryption Standard (AES) and other algorithms. Many researchers and hackers are always trying to break these algorithms using brute force and side channel attacks. Some attacks were successful as it was the case for the Data Encryption Standard (DES) in 1993, where the published cryptanalysis attack [2] could break the DES. Nowadays the Advanced Encryption Standard (AES) is considered as one of the strongest published cryptographic algorithmswhere it was adopted by the National Institute for Standards and Technology (NIST) after the failing of the Data Encryption Standard (DES).

This paper discusses a Matlab implementation of the Advanced Encryption Standard (AES) [6]. AES is based on the block cipher Rijndael [4][5] and became the designated successor of the Data Encryption Standard (DES) [7] which has been implemented in a tremendous number of cryptographic modules worldwide since 1977.

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 95–101.
springerlink.com

Matlab [3] is a matrix-oriented programming language, perfectly suited for the matrix-based data structure of AES.

Even though this implementation is fully operational, (i. e. it can be utilized to encrypt arbitrarily chosen plaintext into cipher text and vice versa), the main optimization parameter of this implementation has not been execution speed but understandability.

In this paper first section contains the introduction of AES algorithm and Matlab, section 2 contain the internal structure of AES with algorithm, section 3 describes results and observation of the paper and section 4 contain the conclusion part and last section is reference.

2 Internal Structure of AES

AES is symmetric key block cipher. It uses a fixed 128-bit block cipher and three key lengths supported by AES as this was an NIST design requirement. The number of internal rounds of the cipher is a function of the key length according to Table 1.

Table 1. Key length and number of rounds of AES

key lengths	# rounds = *nr*
128 bit	10
192 bit	12
256 bit	14

There are three different types of layers to perform AES operation and the function of the different layers is:

1. Key Addition layer: A 128-bit round key, or sub key, which has been derived from the main key in the key schedule, is XORed to the state.
2. Byte Substitution layer (S-Box): Each (fig 1) element of the state is nonlinearly transformed using lookup tables with special mathematical properties. This introduces confusion to the data.

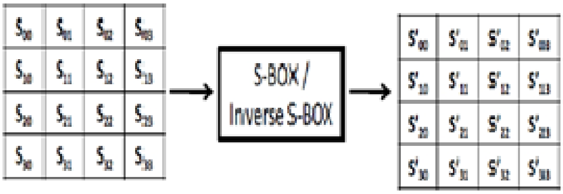


Fig. 1. S-box and inverse S-box

3. Diffusion layer: It provides diffusion over all state bits. It consists of two sub layers, both of which perform linear operations:
   (a) The Shift Rows layer provides the mechanism for shifting the rows (Fig 2) of the above layer output.

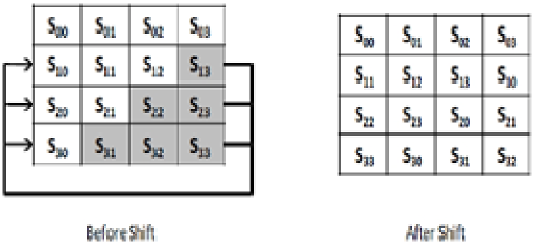

S_{00} S_{01} S_{02} S_{03}
S_{10} S_{11} S_{12} S_{13}
S_{20} S_{21} S_{22} S_{23}
S_{30} S_{31} S_{32} S_{33}
Before Shift
S_{00} S_{01} S_{02} S_{03}
S_{11} S_{12} S_{13} S_{10}
S_{22} S_{23} S_{20} S_{21}
S_{33} S_{30} S_{31} S_{32}
After Shift

Fig. 2. Shift row

   (b) The Mix Column layer is a matrix operation where each 4-byte column is considered as a vector and multiplied by a fixed 4×4 matrix. The matrix contains constant entries (Fig 3).

$$\begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \\ b_7 \end{pmatrix} \equiv \begin{pmatrix} 1&0&0&0&1&1&1&1 \\ 1&1&0&0&0&1&1&1 \\ 1&1&1&0&0&0&1&1 \\ 1&1&1&1&0&0&0&1 \\ 1&1&1&1&1&0&0&0 \\ 0&1&1&1&1&1&0&0 \\ 0&0&1&1&1&1&1&0 \\ 0&0&0&1&1&1&1&1 \end{pmatrix} \begin{pmatrix} b'_0 \\ b'_1 \\ b'_2 \\ b'_3 \\ b'_4 \\ b'_5 \\ b'_6 \\ b'_7 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \bmod 2.$$

Fig. 3. Mix column

The complete process of the AES is depicted in the (fig 4). Last round of the AES does not contain mix column step which makes it more strong. Decryption process is reverse of encryption processes which perform inverse byte substitution operation, inverse shift row and inverse mix column.

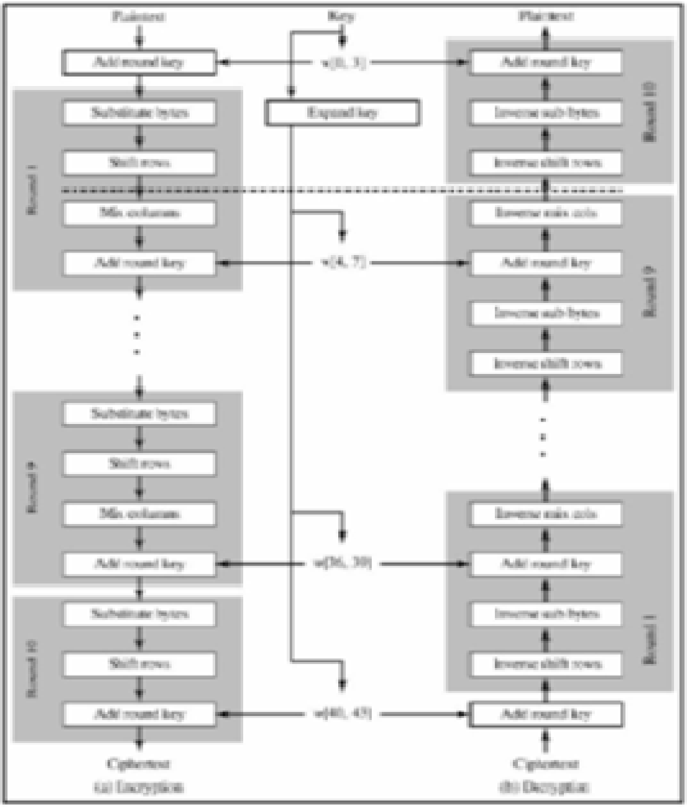


Fig. 4. AES encryption and decryption

```
   Algorithm: AES-Cipher
   input: ByteA[4× nb], WordK[nb ×(nr+ 1)];
   output: ByteC[4× nb],
   Byte state[4, nb]; state := A;
   AddRoundKey(state, K[0, nb– 1];
   forround := 1 to nr– 1
   do
   SubBytes(state);
   ShiftRows(state);
   MixColumns(state);

   AddRoundKey(state,K[round×nb,nb(round+1))
   End loop
SubBytes(state);
ShiftRows(state);
AddRoundKey(state, K[nr × nb, nb(nr + 1) – 1]);
```

```
C := state;
returnC;
end
```

In this algorithm A shows the input plain text which is the size of 128 bit, K shows key which is either 128 bit, 192bit or 256 bit long, C shows the cipher text which is of the length 128 bit long. The number of rounds depends upon the size of the key which varies from 10 to 14rounds.

3 Observation and Results

This section contains result of AES algorithm implemented in MATLAB.

```
Key: {'00' '01' '02' '03' '04' '05' '06' '07' '08' '09''0a' '0b' '0c' '0d' '0e' '0f'}
Plaintext: {'00' '11' '22' '33' '44' '55' '66' '77' '88''99' 'aa' 'bb' 'cc' 'dd' 'ee' 'ff'}
Cipher text: {69 C4 E0 D8 6A 7B 04 30 D8 CDB7 80 70 B4 C5 5A}

Round 1:
s_box={63 CA B7 04 09 53 D0 51 CD 60 E0E7 BA 70 E1 8C}
shift_row={63 53 E0 8C 09 60 E1 04 CD 70 B751 BA CA D0 E7}
mix_col={5F 72 64 15 57 F5 BC 92 F7 BE 3B29 1D B9 F9 1A}
add_round_key={89 D8 10 E8 85 5A CE 68 2D18 43 D8 CB 12 8F E4}

Round 2:
s_box:{A7 61 CA 9B 97 BE 8B 45 D8 AD 1A61 1F C9 73 69}
shift_row={A7 BE 1A 69 97 AD 73 9B D8 C9CA 45 1F 61 8B 61}
mix_col={FF 87 96 84 31 D8 6A 51 64 51 51FA 77 3A D0 09}
add_round_key={49 15 59 8F 55 E5 D7 A0 DACA 94 FA 1F 0A 63 F7}

Round 3:
s_box:{3B 59 CB 73 FC D9 0E E0 57 74 22 2DC0 67 FB 68}
shift_row={3B D9 22 68 FC 74 FB 73 57 67CB E0 C0 59 0E 2D}
mix_col={4C 9C 1E 66 F7 71 F0 76 2C 3F 868E 53 4D F2 56}
add_round-key={FA 63 6A 28 25 B3 39 C9 4066 8A 31 57 24 4D 17}

Round 4:
s_box:{2D FB 02 34 3F 6D 12 DD 09 33 7E C75B 36 E3 F0}
shift_row={2D 6D 7E F0 3F 33 E3 34 09 36 02DD 5B FB 12 C7}
mix_col={63 85 B7 9F FC 53 8D F9 97 BE 478E 75 47 D6 91}
add_round_key={24 72 40 23 69 66 B3 FA 6ED2 75 32 88 42 5B 6C}

Round 5:
s_box={36 40 09 26 F9 33 6D 2D 9F B5 9D 23C4 2C 39 50}
shift_row={36 33 9D 50 F9 B5 39 26 9F 2C 092D C4 40 6D 23}
mix_col={F4 BC D4 54 32 E5 54 D0 75 F1 D6C5 1D D0 3B 3C }
add_round_key={C8 16 77 BC 9B 7A C9 3B 2502 79 92 B0 26 19 96}
```

Round 6:
s_box={E8 47 F5 65 14 DA DD E2 3F 77 B64F E7 F7 D4 90}
shift_row={E8 DA B6 90 14 77 D4 65 3F F7 F5E2 E7 47 DD 4F}
mix_col={98 16 EE 74 00 F8 7F 55 6B 2C 049C 8E 5A D0 36}
add_round_key={C8 16 77 BC 9B 7A C9 3B 2502 79 92 B0 26 19 96}

Round 7:
s_box={B4 15 F8 01 68 58 55 2E 4B B6 12 4C5F 99 8A 4C}
shift_row={B4 58 12 4C 68 B6 8A 01 4B 99 F82E 5F 15 55 4C}
mix_col={C5 7E 1C 15 9A 9B D2 86 F0 5F 4BE0 98 C6 34 39}
add_round_key={C6 2F E1 09 F7 5E ED C3CC 79 39 5D 84 F9 CF 5D}

Round 8:
s_box={3E 17 50 76 B6 1C 04 67 8D FC 22 95F6 A8 BF C0}
shift_row={3E 1C 22 C0 B6 FC BF 76 8D A850 67 F6 17 04 95}
mix_col={BA A0 3D E7 A1 F9 B5 6E D5 512C BA 5F 41 4D 23}
add_round_key={D1 87 6C 0F 79 C4 30 0A B455 94 AD D6 6F F4 1F}

Round 9:
s_box={54 11 F4 B5 6B D9 70 0E 96 A0 90 2FA1 BB 9A A1}
shift_row={54 D9 90 A1 6B A0 9A B5 96 BBF4 0E A1 11 70 2F}
mix_col={E9 F7 4E EC 02 30 20 F6 1B F2 CCF2 35 3C 21 C7}
add_round_key={FD E3 BA D2 05 E5 D0 D735 47 96 4E F1 FE 37 F1}

Round 10:
s_box={7A 9F 10 27 89 D5 F5 0B 2B EF FD9F 3D CA 4E A7}
shift_row={7A D5 FD A7 89 EF 4E 27 2B CA10 0B 3D 9F F5 9F}
add_round_key={69 C4 E0 D8 6A 7B 04 30 D8CD B7 80 70 B4 C5 5A}

Observation: From the above result we can see the cipher text is very strong for very simple plain text.

4 Conclusions

This paper presents the MATLAB implementation of AES algorithm which increases the understandability of the program. AES is the very strong cipher and impossible to break without knowing the key so the importance of AES algorithm is high security. The complex process of AES algorithm can be comfortably implemented in MATLAB.

References

1. Kumar, A., Tiwari, N.: Performance Evaluation of AES methods: Review. In: ICCC (2012); paper id1365
2. Eli, B., Shamir, A.: Differential Cryptanalysis of the Data Encryption Standard. Springer (1993)

3. The Mathworks: Matlab, The Language of Technical Computing (2001), http://www.mathworks.com/products/matlab
4. Rijmen, V.: The block cipher Rijndael (2001), http://www.esat.kuleuven.ac.be/~rijmen/rijndael/
5. Daemen, J., Rijmen, V.: AES proposal: Rijndael (1999), http://www.esat.kuleuven.ac.be/~rijmen/rijndael/rijndaeldocV2.zip
6. National Institute of Standards and Technology: Specification for the Advanced Encryption Standard, AES (2001), http://csrc.nist.gov/publications/fips/fips197/fips-197.pdf
7. National Institute of Standards and Technology: Data Encryption Standard, DES (2001), http://csrc.nist.gov/publications/fips/fips46-3/fips46-3.pdf
8. Tomoiaga, R., Stratulat, M.: AES Performance Analysis on Several Programming Environments, Operating Systems or Computational Platforms. In: 2010 Fifth International Conference on Systems and Networks Communications (2010)
9. Parikh, C., Parimal Patel, M.S.: Ph.D. Performance Evaluation of AES Algorithm on Various Development Platforms

Low Overhead Handoff Based Secure Checkpointing for Mobile Hosts

Priyanka Dey and Suparna Biswas

Department of Computer Science & Engineering
West Bengal University of Technology
priyankadey24@yahoo.co.in,
mailtosuparna@gmail.com

Abstract. An efficient fault tolerant algorithm based on movement-based secure checkpointing and logging for mobile computing system is proposed here. The recovery scheme proposed here combines independent checkpointing and message logging. Here we consider mobility rate of the user in checkpointing so that mobile host can manage recovery information such as checkpoints and logs properly so that a mobile host takes less recovery time after failure. Mobile hosts save checkpoints when number of hand-off exceeds a predefined hand-off threshold value. Current approaches save logs in base station. But this approach maximizes recovery time if message passing frequency is large. If a mobile host saves log in its own memory, recovery cost will be less because log retrieval time will be small after failure. But there is a probability of memory crash of a mobile host. In that case logs can not be retrieved if it is saved only in mobile node. Hence in this algorithm mobile hosts also save log in own memory and base station. In case of crash recovery, log will be retrieved from base station and in case of transient failure recovery logs will be retrieved from mobile host. In this algorithm recovery probability is optimized and total recovery time is reduced in comparison to existing works. Logs are very small in size. Hence saving logs in mobile hosts does not cause much memory overhead. This algorithm describes a secure checkpointing technique as a method for providing fault tolerance while preventing information leakage through the checkpoint data.

Keywords: Fault-Tolerance, Mobile Computing, Checkpointing, Logging, hand-off, recovery time, crash failure, transient failure etc.

1 Introduction

Fault tolerant mobile computing systems are increasingly being used in such application as e-commerce, banking, different mobile monitoring devices in hospital and mission critical application, where privacy and integrity of data as important as uninterrupted operation of services provided.

The checkpointing and logging technique is one such distributed service to provide fault tolerance for the system. Checkpoint which is a consistent snapshot of the system contains process's state, register, segment and actual data of process [1]. Many checkpointing-recovery schemes have been proposed for the distributed systems.

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 103–112.
springerlink.com

However, these schemes cannot be directly used in the mobile environment because of mobile computing system has many constraints [2][3] e.g. mobility, low bandwidth, less stable storage and frequent disconnection.

In light of the above constraint, this paper presents a movement-based checkpointing strategy combined with logging for recovery of individual hosts in mobile computing environments. In this approach Mobile host takes a checkpoint when handoff exceeds threshold of mobility. Our proposed approach mobile host saves log in its own memory to reduce the recovery time after failure. This type of log saving schemes is applicable to such kind of mobile device which can support large amount of storage of message and in which instant recovery is required after failure. This log saving approach is applicable in hospital while several mobile monitoring devices are attached to patient to monitor their temperature, pulse and so on. If failure is occur these devices should be recovered instantly [4]. But due to probability of memory crash, mobile hosts save log in both own memory and base station. This case is suitable in a mission critical application providing communications and shared situational awareness to an active military unit where quick and perfect recovery is required.

The network is one of the common places where security threats exist [5][6]. In this algorithm each mobile host encrypts checkpoint and saves in base station. This prevents information leakage from checkpoint while being transferred through wireless channel.

2 Related Works

An overview of different types of checkpointing and logging techniques and roll back recovery techniques based on the different types of checkpoint based and log based can be found in [7] [8].Prakash and Singhal describe in [9] a checkpointing algorithm for Mobile Computing System. This checkpoint collection algorithm is synchronous and non-blocking. A minimum number of nodes are forced to take checkpoints. Each MH maintains a dependence vector .MHs.T.Park et.al has presented an efficient movement based recovery scheme in [10]. Main feature of this algorithm is that a host carrying its information to the nearby MSS can recover instantly in case of a failure. An MH moving inside a range, recovery information remains in host MSS otherwise it moves recovery information to nearby MSS.In [11] Sapna E. George et .al describes a movement based checkpointing and logging scheme based on mobility of mobile hosts. A checkpoint is saved in mobile support station when hand-off count exceeds a predefined optimum threshold. Recovery cost is minimized in this scheme by computing optimal movement threshold.

2.1 Our Observations

In [3] checkpoint is taken on periodic basis. But this checkpoint may not be suitable for mobile environment due to following reason-(a) if the frequency of check pointing is high, the additional overhead is large. (b)If the frequency is low, the recovery cost may be very large. In [11] log is saved in BS. So recovery time increases due to log retrieval cost.

2.2 Problem Definition

Based on the above analysis we propose an efficient fault tolerant algorithm based on movement-based secure checkpointing and logging which identifies problems and tries to provide proper solutions.

(i) **Transient failure of mobile hosts:** Occurrence of transient fault can not affect memory content of mobile host. No need to save log in base station. Log can be saved and retrieved from memory of mobile host itself.

(ii) **Crash recovery:** To reduce recovery overhead we keep log in mobile host's own memory. If mobile host's memory is crashed, log saved inside memory can't recover. So we consider another case in log saving techniques where log will be saved in both base station and mobile host's memory. If memory crash occur log can be retrieved from base station.

(iii) **Security attack to checkpoint in wireless channel:** To provide fault tolerance secure check pointing is an important issue. Checkpoint is transferred to base station when it is saved and transferred from base station when mobile host is failed over insecure wireless channel. The checkpoint can be attacked, compromised and corrupted in-transit by any intruder or malicious node. The mobile can not recover using modified checkpoint data. Hence rollback recovery of a failed mobile host from last saved state will not be possible. So, security of checkpoint data in wireless network is required to ensure fault tolerance of the processes running on the mobile hosts. Here we encrypt checkpoint so that checkpoint data does not changed.

(iv) **Random movement and handoff of mobile hosts:** Mobility of mobile host is an important concern in case of mobile environment because depending on mobility, logs and checkpoints of a mobile host are saved in different base stations. So, recovery cost depends on checkpoint and log retrieval cost. The optimum movement threshold value ensures that checkpoints can be saved nearer to recovery base station, logs are not scattered too much so that overhead of unnecessary checkpoints and logs can be avoided.

(v) **Overhead:** Overhead is optimized by saving log in mobile host's own memory so that log retrieval cost will be zero during recovery from transient failure.

3 System Model

The mobile computing system considered here consists of n number mobile host (MH) and m number of base station (BS). MHs are connected to BSs through wireless network and BSs are connected with each other though wired network. An MH can communicate with other MHs and BSs only through the BS to which it is directly connected. In this system the channel between an MH is connected to a BS also ensures FIFO communication in both the directions. Proposed algorithm is non-blocking while taking checkpoint.

3.1 Assumption

- During checkpoint interval messages sent, received are saved into log file.
- Every MH takes and transfers encrypted checkpoint.

4 Data Structure and Notation

$\text{BS}_{\text{current}}$= currently connected base station id. BS_{prev}=previous visited base station id.h_{c}= handoff counter which keeps the record of number of handoff occurs. $\text{BS}_{\text{log=}}$base station id where logs are saved$_{..}$$\text{BS}_{\text{chkp=}}$base station id where checkpoint is saved$_{..}$hc = number of hopcount.lc = count the number of log. chkpt_intv=checkpoint interval. r = Ratio of bandwidth of wireless network to wired network. $\text{T}_{\text{chkp\_take\_time}}$ = time required to take checkpoint. $\text{T}_{\text{chkpencrypt\_time}}$ = time required to encrypt checkpoint. T_{rec}= time required to recovery after failures. $\text{T}_{\text{save\_chkp\_bs}}$ = time required to transfer checkpoint.T1=time required to send the checkpoint request to BSchkp. T2=time required to transfer checkpoint from BS_{chkp} to $\text{BS}_{\text{current.}}$T3$_{=}$ time required to save checkpoint. T4=time required to send the request of BS to send log file. T5 = time required to transfer log from BS_{log} to $\text{BS}_{\text{current .T6=time required to transfer log to MH from}}$ $\text{BS}_{\text{current.}}$ T7 = time required to replay log. T8= time required to decrypt checkpoint.

5 Proposed Work

Every time MH will connect with any BS after successful authentication. There is a handoff counter which will increment every time when MH will move one BS to another. When MH's handoff counter will exceed predefine threshold value MH will take checkpoint. During checkpointing, MH will first save the states and encrypt it and then save it in the stable storage .Then the handoff counter will initialized to zero.MH can communicate with other MH through message. The receiver MH will save the received message in log file. When failure is occurred MH will connect to any BS (not necessary the BS where failure is occurred). MH then gives the BS id to the $\text{BS}_{\text{current}}$ to get the checkpoint. The $\text{BS}_{\text{current}}$ will then send request to the BS_{chkp} where checkpoint is saved to get checkpoint. BS_{chkp} then response the request and send the checkpoint. The current BS then sends the encrypted checkpoint to the MH. MH then decrypts the checkpoint and rollback to its last taking checkpoint. In case of logging we consider two cases based on log saving techniques of different mobile computing devices. Two cases are illustrated in the following-

Case 1: MH will save log in its own memory. It will not copy the log in BS.

Case 2: MH will save log in its own memory and also copy the log in BS_{log}. In case1 there is a probability of MH memory crash. So if memory crash occurs MH can retrieve log from BS_{log} otherwise from memory.

5.1 Working Example

The above proposed work is illustrated through the following example cellular network. For example we consider 3 number of MH in this figure is. Here $C_{1,1}$ denotes

MH_1 takes its checkpoint at interval 1. When MH1 moves from one BS to the other BS its handoff counter i.e. h_c is incremented by 1.

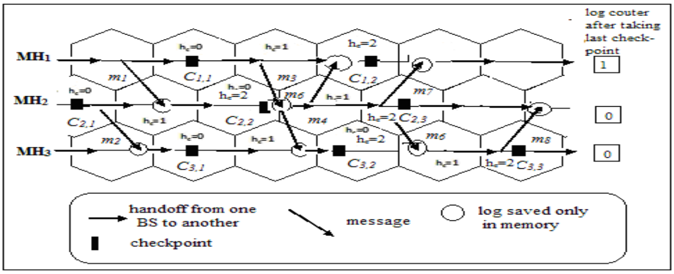


Fig. 1. Working example of our proposed algorithm

MH_1 will take checkpoint when handoff counter i.e. h_c=2.In this example we consider threshold value 2.When $MH_{1,}$ $MH_{2,}$ MH_3 will receive message they will first save in memory and then save to BS. Every MH also maintain log counter which count number of message received till last checkpoint. For MH_1 after taking $C_{1.2}$ it receives m_7 so log counter will be 1.After taking checkpoint log counter will be initialized to zero. Suppose MH_3 fails before taking checkpoint $C_{3,3}$ then MH_3 will rollback to $C_{3,2}$ and replay one log i.e. m_6 as its log counter will be 1 .Thus MH_3 will recover from failure.

5.2 Algorithm

1. MH initially connects to BS after successful authentication.
2. MH sends message to the another MHs. Receiver MH saves the messages in log. Then for case1 MH can save log in memory and for case2 in $BS_{log.}$
3. MH moves one BS to another BS and increment handoff counter.
4. If value of handoff counter is greater than movement threshold Checkpoint will be taken.
5. MH recovers after failure as explained in proposed work section:

6 Correctness and Proof

Theorem 1: The Proposed algorithm ensures consistent global checkpointing

Lemma 1: No orphan or lost message is generated by the technique

Proof: To prove this we consider two cases on the basis of Fig.1.

Case 1: MH_1 fails after taking checkpoint $C_{1,2}$ and receiving m_7 from MH_2.
In this case m_7 is considered as lost message because the state of MH_2 reflects sending it but MH_1 does not reflects receive it. So m_7 can be considered as loss message because it can't replay after failure. According to our algorithm m_7 is not considered as lost message because when MH_1 receives m_7 it saves in its own memory. After failure MH_1 rollbacks to $C_{1,2}$ and replays m_7.

Case 2: MH_2 fails before taking checkpoint $C_{2,3}$ after sending m_6 to MH_1
In this case m_6 considers as orphan message because the state of MH_3 saves m_6 is received from MH_2 but as MH_2 fails , its state has no record about the event that MH_2 sends m_6 to MH_3. So m_6 can be considered as orphan message. In our algorithm m_6 is not considered as orphan message because when MH_1 receives m_6 it saves in its own memory and. can replay after failure. So theorem1 is proved.

7 Performance Analysis

We simulate this algorithm for mobile environment using C language to get different parameters for the performance analysis of the proposed algorithm.. We simulated encryption and decryption of checkpoints using a very simple cryptography technique to verify the working of the proposed secure checkpointing algorithm. In Practical purpose strength of security technique is a big issue. Public key cryptography using elliptic curve cryptography (ECC) is already well established for PDAs. Implementation of ECC algorithm is out of scope of this work. Only encryption and decryption times of checkpoint are taken same as ECAES encrypt and ECAES decrypt time mentioned in [12]. In our system the following parameter values are kept constant. The ratio of bandwidth of wireless to wired network is .1 [11]. Size of each character of each message is considered as 1 bit. Failure rate and log arrival rate are considered to be of exponential distribution. These values do not assume a specific application or environment and are chosen for simulation and performance analysis only.

In this performance analysis we compare the result of log taking scheme used in [11] with our log taking scheme. For this comparison we consider three cases. We consider this to compare with our own case to calculate the overhead. These cases are:

Case 1: MH will save the log to the BS where it will save checkpoint and delete the message from own memory. MH will retrieve log from BS_{log} for recovery.

Case 2: MH will save log in its own memory. It will not copy the log in BS.

Case 3: MH will save log in its own memory and also copy the log in BS_{log}.In case memory is crashed it can retrieve log from BS_{log} otherwise from MH's memory.

- **Secure Checkpoint Cost**

This cost will be same for case1, case2 and case3 because checkpoint taking techniques for three cases are same.

$$T_{chkp\_take\_time} + T_{chkpencrypt\_time} + T_{save\_chkp\_bs}$$
$$= .003+.234+1.759 = 1.996s$$

If we consider encrypted checkpoint overhead will increase by 1.759s.

- **Recovery Time**

For case 1: $T_{rec=}(T1_{+}hc*(T2))*r+T3_{+}T4_{+}lc_{*}hc*(T5)*r+T6_{+(\,lc\,*}\;T7_{)}$

For case 2: $T_{rec\,=}(T1_{\,+}hc*(T2))*r+\;T3_{\,+}T8_{+}\;T1_{\,+(}\,lc_{\,*}\,T7_{)}$

For case 3: $T_{rec\,=}(T1_{\,+}hc*(\,T2)\,)*r+T3_{\,+}T8_{+}p*_{(}lc_{*}T7_{)})+((1-p)*(T4_{+}lc_{*}hc*(\,T5)*r$
$+\;\;T6_{+(}\,lc_{\,*}\,T7_{))}$

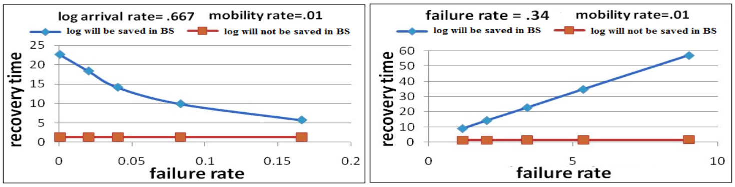


Fig. 2. (a) Failure vs. recovery time log. (b) log arrival rate vs. recovery time.

In fig.2. (a) we take log arrival rate .667 as constant. By varying failure rate we get recovery time we see that if failure rate increases recovery time will decreases because if failure rate increases less log will be taken. In this figure we compare the result of case1 and case2. In Fig.2.(b).We get recovery time by varying log arrival rate and keep failure rate .34 as constant. We see that if log arrival rate increases recovery time will increase because if log arrival rate is increased less number of logs will be taken. So recovery time will be less. In this figure we compare the result of case1 and case2. In both figure we see that recovery time overhead increases in case1 than case2 as in case2 log arrival cost does not consider.

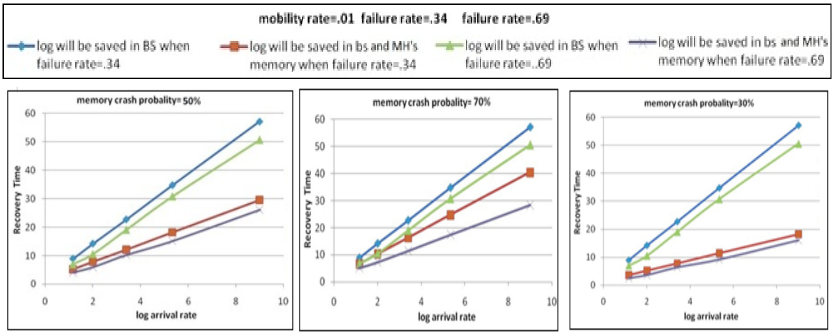


Fig. 3. Log arrival rate vs. recovery time (a) when memory crash probability=50% (b) when memory crash probability=70% (c) when memory crash probability=30%

In the above all the figure we compare the recovery time of case1 with case3 by varying log arrival rate and keeping constant the failure rate .34 and .69.In Fig.3 (a) according to our result the recovery time overhead for case1 is increases 13.006s than case3 for failure rate=.34 and 6.8257s than case3 for failure rate=.69. In Fig.3 (b) 7.832s for failure rate=.34 and 5.2794s for failure rate=.69.In Fig.3 (c) 18.27s for failure rate=.34 and 11.774s for failure rate=.69.

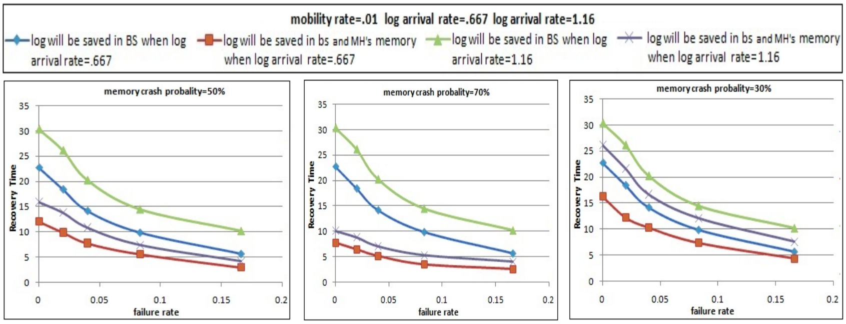


Fig. 4. Failure rate vs.recovery time (a) when memory crash probability=50% (b) when memory crash probability=70% (c) when memory crash probability=30%

In all the above figure we compare the recovery time of case1 with case3 time by varying failure rate and keeping constant the log arrival rate.667 and 1.16.In Fig.4(a) we see that the recovery time overhead for case1 increases 7.6625s than case3 for log arrival rate=.667 and 9.8685s for log arrival rate=1.16.In the Fig.4(b) 4.for log arrival rate=.667 and 5.6672s for log arrival rate=1.16.In the Fig.4(c) 8.77s for log arrival rate=.667 and 13.221s for log arrival rate=1.16.

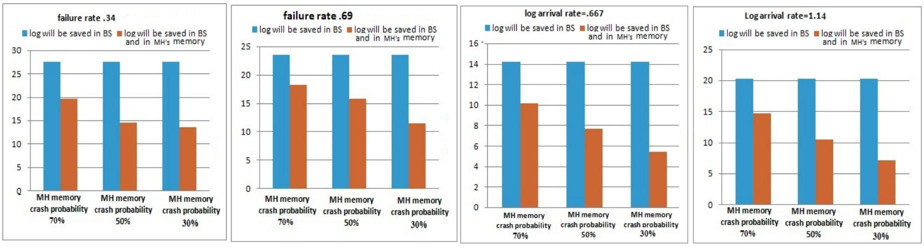


Fig. 5. Recovery time overhead comparison between case 1 and case 3 for memory crash probability 30%, 50%, & 70% (a) for failure rate .34 and .69. (b) log arrival rate .667 and 1.16.

In the above Fig.5.(a) and 5.(b) as memory crash probability increases we can see that recovery time overhead decreases because probability of retrieve log from BS increases with memory crash probability. So we can see that case3 also gives less recovery time than case1.We encrypts checkpoint in our proposed algorithm. In the below we analyze for encryption how much time is increased.

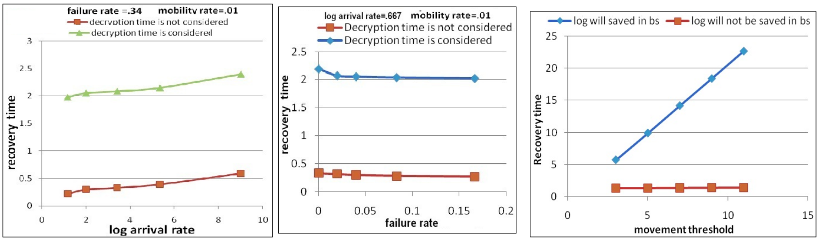


Fig. 6. (a) Log arrival rate vs recovery time and failure rate vs recovery time. (b) Recovery time vs. movement threshold.

In the above Fig.6.(a) we calculate recovery time by considering decryption time of checkpoint.and not considering decryption.We can see that recovery time overhead increases by 1.759s.In Fig.6.(b) we compare recovery time of case1 with case2 by varying movement threshold. We can see if movement threshold increases recovery time also increases because if movement threshold increases checkpoint interval between two checkpoint increases so checkpoint transfer cost will be high and more number will be taken between this interval.Recovery time for case1 is more than case2 because in case2 log is saved in own memory.

8 Conclusions

Mobile computing has been developing very rapidly in recent years. Some of the checkpointing and recovery techniques proposed for mobile computing systems did not take checkpoints regard to the mobility rate of the user and unnecessarily incur additional overhead in maintaining recovery data. In our proposed approach in case of transient failure logs are retrieved from mobile host's own memory to reduce the recovery time after failure. But due to probability of memory crash, mobile hosts save log both in their own memory and base station. In case of crash failure logs are retrieved from base station. Saving two copies of log may seem to cause memory overhead but log searching and transfer cost from base station gets reduced. This algorithm proposes a secure checkpointing system as a method for providing checkpointing capability while simultaneously preventing information leakage of application data saved in checkpoint.

References

1. Gman, H., Ahn, S.J., Han, S.C., Park, T., Yeom, H.Y., Cho, Y.: Kckpt: Checkpoint and Recovery Facility on UnixWare Kernel. Computers and Applications, 303–308 (2000)
2. Forman, G.H., Zahorjan, J.: Challenges of Mobile Computing. Journal Computer 27(4), 31–40 (1994)
3. Schiller, J.: Mobile Communications, 2nd edn. Addison Wesley
4. Adis, W.: Mobile Computing for Hospitals: Transition Problems. Communications IIMA 5(2), 67–76 (2005)
5. Nam, H., Kim, J., Hong, S.J., Lee, S.: Secure checkpointing. Journal of Systems Architecture 48, 237–254 (2003)

6. Agrawal, D.P., Deng, H., Poosarla, R., Sanyal, S.: Secure Mobile Computing. In: Das, S.R., Das, S.K. (eds.) IWDC 2003. LNCS, vol. 2918, pp. 265–278. Springer, Heidelberg (2003)
7. Mootaz Elnozahy, E.N., Alvisi, L., Wang, Y.-M., Johnson, D.B.: A Survey of Rollback Recovery Protocol in Message Passing System. ACM Comput. Surv. 34(3), 375–408 (2002)
8. Kumar, P., Garg, R.: Checkpointing Based Fault Tolerance in Mobile Distributed Systems. International Journal of Research and Reviews in Computer Science (IJRRCS) 1(2), 83–93 (2010)
9. Prakash, R., Singhal, M.: Low Cost Checkpointing and Failure Recovery in Mobile Computing Systems. IEEE Transacrions on Parallel and Distrinuted Systems 7(10), 1–38 (1996)
10. Park, T., Woo, N., Yeom, H.Y.: An Efficient recovery scheme for fault-tolerant mobile computing systems. Future Generation Computer System 19(1), 37–53 (2003)
11. George, S.E., Chen, I.-R., Jin, Y.: Movement-Based Checkpointing and Logging for Recovery in Mobile Computing Systems. In: MobiDE 2006 Proceedings of the 5th ACM International Workshop on Data Engineering for Wireless and Mobile Access, pp. 51–58 (2006)
12. Lopez, J., Dahab, R.: An overview of Elliptic Curve Cryptography, Technical Report IC-00-10. State University of Campinas, pp. 1–34 (2000)

An Obfuscated Implementation of RC4

Roger Zahno and Amr M. Youssef

Concordia Institute for Information Systems Engineering, Concordia University,
Montreal, Quebec, Canada
r_zahno@ciise.concordia.ca, youssef@ciise.concordia.ca

Abstract. Because of its simplicity, ease of implementation, and speed, RC4 is one of the most widely used software oriented stream ciphers. It is used in several popular protocols such as SSL and it has also been integrated into many applications and software such as Microsoft Windows, Lotus Notes, Oracle Secure SQL and Skype.

In this paper, we present an obfuscated implementation for RC4. In addition to investigating different practical obfuscation techniques that are suitable for the cipher structure, we also perform a comparison between the performance of these different techniques. Our implementation provides a high degree of robustness against attacks from execution environments where the adversary has access to the software implementation such as in digital right management applications.

1 Introduction

Because of its simplicity, ease of implementation and robustness, RC4 [1] has become one of the most commonly used stream ciphers. In its software form, implementations of RC4 appear in many protocols such as SSL, TLS, WEP and WPA. Furthermore, it has been integrated into many applications and software including Windows, Lotus Notes, Oracle Secure SQL, Apple AOCE, and Skype. Although the core of this two decade old cipher is just a few lines of code, the study of its strengths and weaknesses as well as its different software and hardware implementation options is still of a great interest to the security and research communities.

Cryptographic techniques are traditionally implemented to protect data and keys against attacks where the adversary may observe various inputs to and outputs from the system, but has no access to the internal details of the execution. On the other hand, several recently developed applications require a higher degree of robustness against attacks from the execution environment where the adversary has closer access to the software implementation of key instantiated primitives. Digital Rights Management (DRM) is an example of such applications where one of the main design objective is to control access to digital media content. This can be achieved through white-box implementations where encryption keys are hidden, using obfuscation techniques, within the implementation of the cipher. White-box implementations for block ciphers, such as AES and DES, are widely available [2] [3]. However, to the authors' knowledge, there is no published white-box implementation for any dedicated stream ciphers including RC4. Directly applying the techniques developed for white-box implementation of block ciphers to stream ciphers does not seem to work, mainly, because normal operation of

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 113–129.
springerlink.com

stream ciphers requires us to always maintain the inner state of the cipher. Recovering the inner state of the stream cipher usually sacrifices the security of the cipher even if the attacker is not able to recover the key. In fact, it seems to be an open research problem to determine if a practical white box implementation (in a cryptographic sense) would ever exist for stream ciphers such as RC4.

On the other hand, obfuscation can be used to transform a program from an easily readable format to one that is harder to read, trace, understand and modify. This offers an additional layer of security as it protects the program by increasing the required human and computational power that is needed to reverse engineer, alter, or compromise the obfuscated program. Obfuscation techniques can be categorized into automated and manual methods. The former systematically, using specific tools, modify the source code (e.g. [4]) or the binary executable file (e.g. [5],[6]). The latter techniques rely on the programmer to follow obfuscation techniques during coding. These techniques are governed by the programming language in use. For example, languages like C and C++, which are commonly used in the implementation of cryptographic primitives due their high performance, are flexible in syntax and allow the use of pointers.

Obfuscated programs can be reverse engineered using techniques such as static and dynamic analysis. Static analysis techniques analyze the program file by performing control flow and data flow analysis ([7],[8],[9]) without running the program. Dynamic analysis, on the other hand, takes place at runtime and addresses the followed execution path.

In this paper, we investigate several obfuscation techniques that are suitable for applications to array-based stream ciphers such as RC4. We also perform a comparison between the performance of these different techniques when applied to RC4. Although our proposed implementation does not provide the same level of theoretical security provided by white-box implementations for block ciphers, it still provides a high degree of robustness against attacks from execution environments where the adversary has access to the software implementation such as in digital right management applications.

The rest of this paper is organized as follows. A brief review of the RC4 key scheduling algorithm and pseudorandom generation algorithms is provided in the next section where we also briefly describe the Skype attempt to obfuscate the RC4 software implementation. Our proposed obfuscated implementation is described in section 3. Section 4 shows a performance comparison between different implementation options. Finally, our conclusion is provided in section 5.

2 The RC4 Cipher

In this section, we briefly review the Key Scheduling Algorithm (KSA), and the Pseudorandom Generation Algorithm (PRGA) of RC4. We also describe the Skype attempt to provide an obfuscated software implementation for RC4.

2.1 Standard RC4 Implementation

While traditional feedback shift register based stream ciphers are efficient in hardware, they are less so in software since they require several operations at the bit level. The design of RC4 avoids the use of bitwise operations as it requires only byte manipulations which makes it very efficient in software. In particular, RC4 uses 256 bytes of memory for the state array, $S[0]$ through $S[255]$, k bytes of memory for the key, $key[0]$ through $key[k-1]$, and two index pointers: a sequential index i, and quasi random index j.

Algorithm 1 shows the KSA of RC4, where the permutation S is initialized with a key of variable length, typically between 40 to 256 bits. Once this is completed, the key stream is generated using the PRGA shown in Algorithm 2. The generated key stream is combined with the plaintext, usually, through an XOR operation.

Algorithm 1. RC4 Key Scheduling Algorithm (KSA) [1]

```
for $i = 0 \rightarrow 255$ do
    S[i] := i
end for

j := 0
for $i = 0 \rightarrow 255$ do
    j := (j + S[i] + key[i mod keylength]) mod 256
    swap values of S[i] and S[j]
end for
```

Algorithm 2. RC4 Pseudo-Random Generation Algorithm (PRGA) [1]

```
i := 0
j := 0
while GeneratingOutput do
    i := (i + 1) mod 256
    j := (j + S[i]) mod 256
    swap values of S[i] and S[j]
    K := S[(S[i] + S[j]) mod 256]
    output K
end while
```

Analyzing the KSA and PRGA algorithms of RC4 yields the following observations:

1. As with most stream ciphers, an adversary does not have to find the key in order to break the cipher. In other words, recovering the initialized inner-state S allows the adversary to efficiently generate the keystream output of the cipher and decrypt the target ciphertext even without knowing the *key* array.
2. An adversary who is able to observe the values of the index pointer j in the PRGA can efficiently recover the whole inner state S.

2.2 Skype's RC4 Implementation

The only reference to obfuscated RC4 implementation in the open literature appeared in a Blackhat publication that describes the leaked implementation used in Skype [10] [11]. By analyzing this leaked implementation, we observe the following:

- Regarding the cipher itself, the cryptographic key used is 80 bytes in length whereas standard implementations use a key of 40 to 256 bits in length (i.e. 5 to 32 bytes).
- Regarding key management, the Skype's implementation selects a key from a pool of 2^{32} keys.
- Regarding the use of the cipher, RC4 itself is used as an obfuscation technique to hide the network layer.

The implementation utilizes a macro called *RC4_round* that is used in both the KSA and the PRGA. Therefore, we first describe this macro. This macro is shown in Algorithm 3. When called, the macro is passed the following parameters:

i**:** the sequential index

j**:** the quasi random index

$RC4$**:** an array corresponding to S in the standard implementation

t**:** a variable for swapping the i^{th} and j^{th} element in S

k**:** the cryptographic key used in this iteration

Lines 1, 3 and 4 describe the swapping operation, line 2 evaluates the new quasi random index j, and line 5 evaluates the key for this iteration. The main difference in the use of this macro between KSA and PRGA is in the key value passed and the action based on the return value. In the PRGA, the value for k is always zero, whereas in the KSA, the key used in this iteration is passed. Furthermore, in the KSA, the returned value of this macro is discarded, whereas in the PRGA the returned value is used as part of the key stream.

Algorithm 3. Round Macro: RC4_round(i,j,t,k,RC4) [10]

```
1: t :=RC4[i],
2: j := (j + t + k) mod 256,
3: RC4[i] := RC4[j],
4: RC4[j] := t,
5: RC4[(RC4[i] + t) mod 256] {Output to be returned}
```

The KSA is shown in Algorithm 4. In this implementation, the inner-state of the cipher is stored in a data structure called rc4. This structure contains an array (representing the array S from Algorithms 1 and 2) which holds the random ordered values, a sequential index i, and quasi random index j. The "for" loop in lines 1 through 6 initializes the array rc4.s sequentially from 0 to 255. The array rc4.s is allocated exactly 256 sequential bytes in memory. The implementation takes advantage of the array structure in memory by initializing 4 bytes in each iteration rather than a single byte. Therefore, the index j in the loop is incremented in each iteration by 4 bytes (0x04040404), and is assigned to the array at the i^{th} position. Consequently, the index i has to be incremented by 4 in each iteration. The reordering step (lines 9 through 11) is done by the macro *RC4_round*.

Algorithm 4. Skype Implementation for the RC4 KSA [10]

```
1:  j := 0x03020100
2:  for $i = 0 \to 255$ do
3:      i := i + 4
4:      j := j + 0x04040404
5:      rc4.s[i] := j
6:  end for
7:
8:  j := 0
9:  for $i = 0 \to 255$ do
10:     RC4_round(i, j, t, byte(key,i%80), rc4.s)
11: end for
```

The PRGA is shown in Algorithm 5 where the ”for” loop iterates over the data stream (buffer, of size bytes) performing the encryption/decryption operation. This is done by XORing the data (in buffer) and the key stream generated by the macro *RC4_round*. Furthermore, the indices i and j in the data structure rc4 are updated.

Algorithm 5. Skype Implementation for the RC4 PRGA [10]

```
for $(;bytes;bytes--)$ do
    i := (i + 1) mod 256
    buffer++ {positioning the pointer to the new value to be en-/decrypted}
    *buffer = *buffer XOR RC4_round(i, j, t, 0, s)
    rc4.i := i
    rc4.j := j
end for
```

Clearly, the Skype implementation described above does not provide enough level of protection for the inner state of the cipher. It should be noted, however, that Skype uses the RC4 cipher itself as an obfuscation technique for the network layer but the effective data stream (voice, chat, video) is encrypted with AES [11].

3 Proposed Implementation

In this section we present our obfuscated implementation of RC4. Throughout our work, we assume that the cipher is implemented as a standalone module, i.e., the implemented code contains only the functionality of the cipher. Another approach would be to mix the implementation of the cipher with parts of a larger application. Such a needle in the haystack approach can add more security as the containing application offers more obfuscation space. However, since an implementation of this approach is highly dependent on the containing application, it is therefore not considered in this work.

In our proposed obfuscated implementation, we first eliminate the use of the array S by using independent set of variables. To improve the efficiency of this approach, we use function pointers. Following that, we utilize multithreading to provide security against dynamic analysis attacks. Finally, We present other generic techniques used to further obfuscate the proposed implementation. Throughout the remaining of this paper, for illustrative purposes, we use a toy implementation of RC4 (with an array of size $N = 4$). However, performance measures have been made based on a RC4 with standard parameters (i.e. with an array of size $N = 256$).

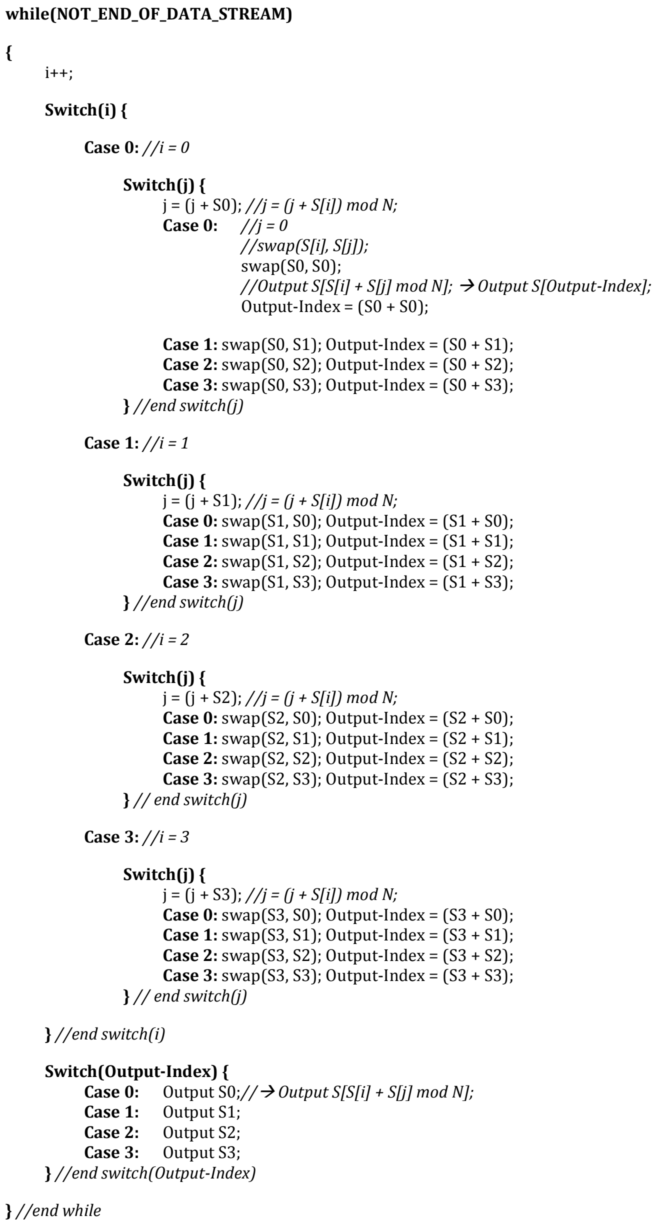
3.1 Eliminating the S Array Data Structure

As the KSA and PRGA algorithms show, standard RC4 implementation requires only a few data objects, namely two index pointers i and j, and an array S.

As a first step towards obfuscation, we substitute the array data structure S by N independent variables, where N is the number of elements in the array S. Unlike the elements of an array which are stored in consecutive memory locations, these independent variables can be scattered throughout the program memory. On the other hand, working with such independent variables eliminates our ability to dynamically address them using a loop structure since we no longer have an array index that can be related to loop counters. To address these variables, we use the loop unrolling technique, also known as loop unwinding. This is illustrated for the toy implementation in Figure 1. As depicted in the figure, the implementation is based on two nested switch/case structures, where the outer structure operates over the index i and the inner structure operates over the index j. It is worth noting that the inner switch/case structures are almost identical for various outer switch/case structures. However, since the i^{th} element in array S has been substituted with an independent variable, this variable has to be correctly referenced in each inner switch/case structure. This nested structure cannot dynamically evaluate expressions such as $S[S[i]+S[j]]$. For such expressions, a dedicated switch/case structure, $Switch(Output-Index)$, is used. This structure is passed an intermediate value, $Output-Index = S[i]+S[j]$, and evaluates the expression above. In an 8 bit implementation with an array of size $N = 256$, the number of possible combinations is given by $N \times N = 2^8 \times 2^8 = 2^{16}$. This implies that a total of 2^{16} case statements are needed to represent all the possible combinations. Thus, despite its conceptual simplicity, the use of nested switch/case structures results in a prohibitively large program (e.g. for $N = 256$, the program size exceeds 12 MB). In the next subsection we show how this obfuscation approach can be enhanced to yield a more practical program size.

3.2 The Use of Function Pointers

Function pointers are pointers that hold addresses of functions, and can be used to execute them. Depending on the address assigned to the pointer, a single function pointer



```
while(NOT_END_OF_DATA_STREAM)
{
    i++;

    Switch(i) {

        Case 0: //i = 0

            Switch(j) {
                j = (j + S0); //j = (j + S[i]) mod N;
                Case 0:    //j = 0
                           //swap(S[i], S[j]);
                           swap(S0, S0);
                           //Output S[S[i] + S[j] mod N]; → Output S[Output-Index];
                           Output-Index = (S0 + S0);

                Case 1: swap(S0, S1); Output-Index = (S0 + S1);
                Case 2: swap(S0, S2); Output-Index = (S0 + S2);
                Case 3: swap(S0, S3); Output-Index = (S0 + S3);
            } //end switch(j)

        Case 1: //i = 1

            Switch(j) {
                j = (j + S1); //j = (j + S[i]) mod N;
                Case 0: swap(S1, S0); Output-Index = (S1 + S0);
                Case 1: swap(S1, S1); Output-Index = (S1 + S1);
                Case 2: swap(S1, S2); Output-Index = (S1 + S2);
                Case 3: swap(S1, S3); Output-Index = (S1 + S3);
            } //end switch(j)

        Case 2: //i = 2

            Switch(j) {
                j = (j + S2); //j = (j + S[i]) mod N;
                Case 0: swap(S2, S0); Output-Index = (S2 + S0);
                Case 1: swap(S2, S1); Output-Index = (S2 + S1);
                Case 2: swap(S2, S2); Output-Index = (S2 + S2);
                Case 3: swap(S2, S3); Output-Index = (S2 + S3);
            } // end switch(j)

        Case 3: //i = 3

            Switch(j) {
                j = (j + S3); //j = (j + S[i]) mod N;
                Case 0: swap(S3, S0); Output-Index = (S3 + S0);
                Case 1: swap(S3, S1); Output-Index = (S3 + S1);
                Case 2: swap(S3, S2); Output-Index = (S3 + S2);
                Case 3: swap(S3, S3); Output-Index = (S3 + S3);
            } // end switch(j)

    } //end switch(i)

    Switch(Output-Index) {
        Case 0:    Output S0;//→ Output S[S[i] + S[j] mod N];
        Case 1:    Output S1;
        Case 2:    Output S2;
        Case 3:    Output S3;
    } //end switch(Output-Index)

} //end while
```

Fig. 1. The implementation of the PRGA when replacing the array data structure by independent variables

can be used to call multiple functions. Normally, a designated array is used to hold the addresses of the functions and when a function is to be called, its address is assigned to the pointer and the function is executed. A visualization of function pointers is shown

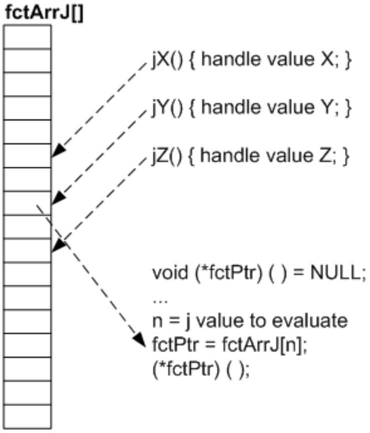


Fig. 2. Array of Function Pointers

in Figure 2 where the array $fctArrJ[]$ is the designated array that holds the addresses of functions $jX(), jY(), jZ()$. As the code fragment shows, the address of the desired function in loaded into the function pointer $fctPtr$, and then executed.

T. Ogiso *et al.* [12] analyze the use of function pointers in software obfuscation. Specifically, they prove that when using arrays of function pointers, determining the address a pointer points to is NP-hard.

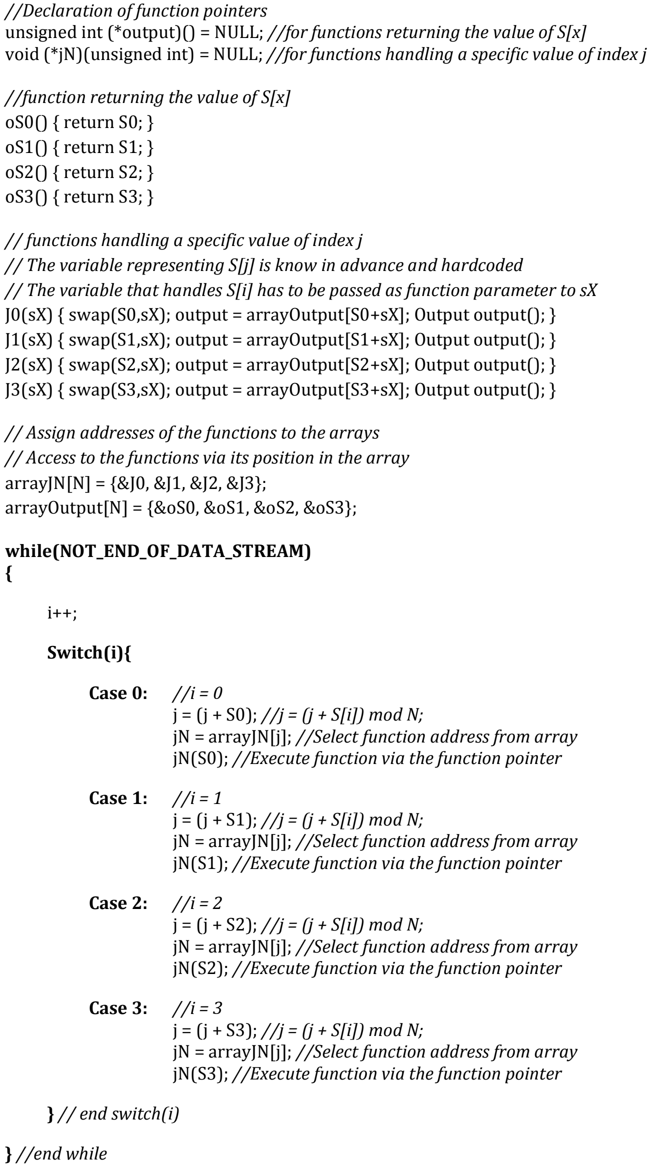
Arrays of function pointers can be used to replace the inner switch/case structures in our obfuscation technique presented in the previous subsection. In addition to the security advantage resulting from the difficulty of determining the address a function pointer points to, implementing function pointers requires much less space than switch/case structures described in the previous section. This enables us to maintain the complexity introduced by the nested switch/case structure while reducing the program size.

The use of function pointers as a replacement of the inner switch/case statement requires the following:

1. For each index j, there exists a function in which the variable that substitutes $S[j]$ is hard coded. Furthermore, the variable that substitutes $S[i]$ is passed as a parameter to this function. With these variables, the functionality of RC4 can be easily realized.
2. There exists a designated array that holds the addresses of the functions described above.

The array of function pointers can be directly used to replace the inner switch/case structures operating on index j. The inner switch/case statements are replaced by functions in which the variable representing $S[j]$ is hard coded. To evaluate the inner structure, we first compute the new j value. This is used to retrieve the corresponding function address from the designated array. Finally, the retrieved function is called using a function pointer and the variable that substitutes $S[i]$ is passed as a parameter.

Figure 3 shows the use of function pointers as a replacement of the inner switch/case structures of our obfuscated toy implementation of RC4. As shown in the figure, the



```
//Declaration of function pointers
unsigned int (*output)() = NULL; //for functions returning the value of S[x]
void (*jN)(unsigned int) = NULL; //for functions handling a specific value of index j

//function returning the value of S[x]
oS0() { return S0; }
oS1() { return S1; }
oS2() { return S2; }
oS3() { return S3; }

// functions handling a specific value of index j
// The variable representing S[j] is know in advance and hardcoded
// The variable that handles S[i] has to be passed as function parameter to sX
J0(sX) { swap(S0,sX); output = arrayOutput[S0+sX]; Output output(); }
J1(sX) { swap(S1,sX); output = arrayOutput[S1+sX]; Output output(); }
J2(sX) { swap(S2,sX); output = arrayOutput[S2+sX]; Output output(); }
J3(sX) { swap(S3,sX); output = arrayOutput[S3+sX]; Output output(); }

// Assign addresses of the functions to the arrays
// Access to the functions via its position in the array
arrayJN[N] = {&J0, &J1, &J2, &J3};
arrayOutput[N] = {&oS0, &oS1, &oS2, &oS3};

while(NOT_END_OF_DATA_STREAM)
{

    i++;

    Switch(i){

        Case 0:    //i = 0
                   j = (j + S0); //j = (j + S[i]) mod N;
                   jN = arrayJN[j]; //Select function address from array
                   jN(S0); //Execute function via the function pointer

        Case 1:    //i = 1
                   j = (j + S1); //j = (j + S[i]) mod N;
                   jN = arrayJN[j]; //Select function address from array
                   jN(S1); //Execute function via the function pointer

        Case 2:    //i = 2
                   j = (j + S2); //j = (j + S[i]) mod N;
                   jN = arrayJN[j]; //Select function address from array
                   jN(S2); //Execute function via the function pointer

        Case 3:    //i = 3
                   j = (j + S3); //j = (j + S[i]) mod N;
                   jN = arrayJN[j]; //Select function address from array
                   jN(S3); //Execute function via the function pointer

    } // end switch(i)

} //end while
```

Fig. 3. Implementation of the PRGA using switch/case for *i* and array of function pointer for *j*

loop over index *i* remains unrolled using the switch/case structure proposed initially. The setup for the array of function pointers requires:

1. The declaration of two function pointers ($*output$ and $*jN$)
2. The definition of output functions $oS0(), oS1(), os2()$ and $oS3()$ that return the value $S[x]$

3. The definition of functions $j0(sX), j1(sX), j2(sX)$ and $j3(sX)$ that replace the inner switch/case structure
4. The definition of arrays used to hold the addresses of the functions stated in steps 2 and 3 above

Next, we illustrate the combination of switch/case structures with array function pointers. The outer structure used to unroll the loop over the index i remains unchanged. However, the inner switch/case structure is replaced by using the function pointers concept. To do this, we first compute the new j value. This is used to retrieve the corresponding function address from the designated array. Finally, the retrieved function is called using a function pointer where the variable that represents $S[i]$ is passed as a parameter. With these variables, the functionality of RC4 can be realized. When implemented, this approach while improving the obfuscation level, significantly reduces the obfuscated program size . In comparison to the initial attempt (in section 3.1), the program size is reduced from 12 MB to about 450 KB.

The techniques used so far have mainly increased the resilience of the implementation against static analysis. We, next, introduce further obfuscation with the objective of increasing its resilience against dynamic analysis.

3.3 Multithreading

Traditionally, multithreading allows various parts of a program to run simultaneously. Each such part is called a thread and although these are functionally independent, they share some resources such as processing power and memory, with other threads. Shared resources, such as data, code, and heap segments allow communication and functional synchronization between threads. Furthermore, constructs such as *critical sections*, and *semaphores* enable the realization of atomic units which, in turn, prevent corruption of shared resources. Because of this parallelism and the randomness in order of execution, analyzing multithreaded programs is much harder than their single threaded counterparts [13]. In this section, we capitalize on this and utilize the randomness in the order of execution introduced by multithreading to further obfuscate our implementation. To do so, we require the following:

1. There exists a multithreaded environment where each thread implements the RC4 functionality (key stream value) for a specific subset of index values i. That is, each thread contains an implementation of a subset of switch/case statement for the corresponding values of i. Furthermore, for each implemented switch/case statement, let the thread implement the function pointer concept for all values of j, as described in 3.2.
2. The sets of index values i are assigned to the threads such that each value of i is assigned randomly to at least two threads. This introduces randomness in the threads that have the capability of implementing the RC4 functionality for a given value of i. Since at least two threads have this capability, the execution path cannot be determined with certainty which introduces an additional layer of obfuscation.

If one uses only 2 threads, requirement 2 above would result in identical functionality for both threads, which simplifies conducting static analysis on the cipher. Thus, in our

```
Thread1() {

    while(NOT_END_OF_DATA_STREAM)

    {
        EnterCriticalSection; //enter the coherent, not interruptible code sequence

        Switch(i) {

            //Case 0: i = 0 is not handled in this thread

            Case 1:    //i = 1
                       j = (j + S1); //j = (j + S[i]) mod N;
                       jN = arrayJN[j]; //Select function address from array
                       jN(S1); //Execute function via the function pointer
                       i++;
                       //Leave the coherent, not interruptible code sequence
                       LeaveCriticalSection;

            Case 2:    //i = 2
                       j = (j + S2); //j = (j + S[i]) mod N;
                       jN = arrayJN[j]; //Select function address from array
                       jN(S2); //Execute function via the function pointer
                       i++;
                       //Leave the coherent, not interruptible code sequence
                       LeaveCriticalSection;

            Case 3:    //i = 3
                       j = (j + S3); //j = (j + S[i]) mod N;
                       jN = arrayJN[j]; //Select function address from array
                       jN(S3); //Execute function via the function pointer
                       i++;
                       //leave the coherent, not interruptible code sequence
                       LeaveCriticalSection;

            default:   //Handles the non-available 'cases'
                       //Leave the critical section
                       LeaveCriticalSection;

        } // end switch(i)
    } //end while
} //end Thread1

Thread2() {
    //same structure like Thread1(), but
    //Case 0: i = 0 is handled
    //Case 2: i = 2 is handled
    //Case 3: i = 3 is handled
    //and
    //Case 1: i = 1 is not handled
}

Thread3() {
    //same structure like Thread1(), but
    //Case 0: i = 0 is handled
    //Case 1: i = 1 is handled
    //Case 3: i = 3 is handled
    //and
    //Case 2: i = 2 is not handled
}
```

Fig. 4. Implementing the PRGA using multithreading

implementation, the minimum number of threads used is set to 3. This ensures that the implementations of various threads differ. The specific number of threads to be used is left as a design parameter.

To compute the key stream value for the current value of i, the running thread enters a critical section and retrieves i. If this thread does not implement the switch/case statement for this value of i, the critical section exits without affecting the cipher inner state and without producing any new keystream words. On the other hand, if the thread implements the switch/case statement for this value of i, the key stream value is evaluated and returned.

Figure 4 illustrates a toy implementation, with $N = 4$, of the multithreading implementation described above. In this example, the sequential index i can have the values $\{0,1,2,3\}$, and 3 threads are used. The first thread implements the switch/case statements for $i \in \{1,2,3\}$; the second thread implements the statements for $i \in \{0,2,3\}$, and the third thread implements the statements for $i \in \{0,1,3\}$. For standard RC4 parameters, this implementation increases the program size to 650 KB but offers an additional obfuscation layer and enhances the implementation's resilience to dynamic analysis attacks.

3.4 Handling the Key Scheduling Process

As shown in section 2.1, RC4 runs two main algorithms, the PRGA, and the KSA. While the implementation of the KSA can be obfuscated using the same techniques discussed above, one weakness of this approach is that the cryptographic key has to be passed in the clear to the KSA algorithm. In this section, we discuss a possible extension of the above implementation in order to mitigate this vulnerability.

In the white-box implementations of AES [2] and DES [3], the cryptographic key is integrated into the lookup tables of the implemented algorithms. Furthermore, the lookup tables are pre-created outside the users' environment. Applying this off-line generation technique to the inner states of RC4 can be used to eliminate the need for a KSA algorithm and consequently mitigate the vulnerability described above. This shifts our objective from protecting the key and key scheduling algorithm to protecting the process of securely assigning the off-line generated values to the inner-state.

Assume a setup where the user receives some encrypted data stream, and pre-created inner-state for the cipher from some service provider. In this case, the inner-state can be transferred from the provider to the customer in the form of an array. Instead of generating the inner-state by the KSA, the inner-state is initialized by directly assigning the values from the array to the corresponding variables. In other words, the array from the provider contains 256 values corresponding to S and the two index pointers i and j in a random order. Those 258 values are directly assigned to the variables representing the inner-state on the customer side. It should be noted that as long as the order of the values in the array is not known, an adversary cannot gain any useful information about the inner state of the cipher. Furthermore, assuming that the service provider knows the memory structure of the user's cipher, the service provider can produce a formatted memory dump that can be loaded directly into the user's cipher.

To this point, our obfuscation approaches structurally altered the implementation of the cipher. Additional obfuscation techniques, that are deployable on a smaller scale can also be utilized. These techniques do not significantly change the implemented structure and are easily applied. In the next subsection, we briefly explore the application of such techniques to our RC4 implementation.

3.5 Generic Obfuscation Techniques

In this section, we introduce a set of standard techniques that can be used to further obfuscate our implementation of RC4. These techniques are independent of the structure of the implemented program and do not significantly change the structure of the resulting obfuscated program.

3.5.1 Order and Dimension Change of Arrays

When using arrays of function pointers, assigning the pointers to the array in a sequential order introduces a $1:1$ mapping between the j value and the index of the array. This mapping can be further obfuscated using a random allocation table or by modifying the array structure. In [14], Zhu, *et al.* address this problem by changing the index order and the arrays' dimension. Transforming an array $A[N]$ into an array $B[M]$, where$M > N$ and M is relatively prime to N, can be done by applying the mapping $B[i] = A[i \times N \bmod M]$. We have used this methodology to obfuscate the array of function pointers in our implementation.

3.5.2 Variable Aliases

When two or more variables address the same memory location, they are called aliases. Introducing aliases to a program reduces the effectiveness of static analysis techniques as they increase the data flow complexity [7] since the attacker has to identify and track all aliases that manipulate a specific memory location. The larger the program, the harder it is for the attacker to identify and keep track of all the aliases. The pointers used in our implementation are an extensive form of aliasing, and therefore, introduce an additional level of obfuscation.

3.5.3 Scattering the Code for the Swap Operations

The RC4 cipher makes extensive use of swapping in both the KSA and the PRGA algorithms. In a standard implementation, monitoring the swapping function easily reveals the position of j and consequently, its value, which compromises the security of the implementation. To address this problem, in our implementation, we scatter the steps of the swap functionality throughout the program. In addition, we use a pool of temporarily swap variables rather than a single variable.

3.5.4 Opaque Constructs

Opaque predicates are expressions that evaluate either to true or false upon a given condition, but their outcome is known/controlled in advance. These constructs introduce confusion and are widely used in obfuscation [7], [15]. Opaque predicates can be classified based on their possible outcome into two types. In the first type, the outcome is always either 'true' or always 'false'. An example of such predicate would be $j > 255$, which is always evaluated 'false' in an 8 bit environment. In the second type, the output could be either 'true' or 'false', but is controlled by adjusting the statement that it evaluates. To increase the complexity of control flow analysis, we implemented a similar approach where either the real function or some other dummy is executed.

3.5.5 Evaluation of the Index Pointer j

Normally, the index pointer j, at step i, is calculated as

$$j_i = j_{i-1} + S[i] \tag{1}$$

where j_i denotes the value of the index pointer j at iteration i and $S[i]$ denotes the value of the inner state array during the i^{th} iteration of the cipher. Monitoring the value of j_i, while knowing the value of i, allows reproducing the inner state. In our implementation, we obfuscate the computation of j_i by introducing three intermediate variables (a,b,c) that are initialized as follows:

$$\begin{aligned} b &= \text{random}, \\ a &= b - j_{i-1}, \\ c &= 2a - b - S[i+1]. \end{aligned}$$

Then we calculate the value of j_i as $j_i = a - c$.

4 Performance Evaluation

In this section we compare the execution costs (i.e., program size and execution time) for various combinations of obfuscation techniques proposed in the previous sections.

The reported timing, as summarized in Table 1, are measured on an HP PC with a quad core Intel 2.67 GHz processor, and 8 GB of RAM. The prototype was implemented in C using Microsoft Visual C++ that was running on Windows 7 Enterprise platform. As expected, obfuscated implementations impose a penalty on the resulting program size and execution speed.

From Table 1, the slowdown factor varies highly between the obfuscation configurations. For configuration 1, the slowdown factor is about 16, whereas the slowdown factor when implementing all the described obfuscation techniques is about 481. The slowdown factors for the white-box implementations of AES and DES found in the published literature are 55 for AES and 10 for DES [16],[17].

In the following subsections, we highlight the main causes of this performance impairment.

Table 1. Program size and throughout for different obfuscation options

RC4 Implementation options	Program Size (KB)	Throughput (KB/sec)	a	b	c	d
No obfuscation	8	288,700	-	-	-	-
Configuration 1	514	17,850	x	-	-	-
Configuration 2	518	15,450	x	x	-	-
Configuration 3	537	11,500	x	x	x	-
Fully obfuscated	969	600	x	x	x	x

a: Loop unrolling over index pointer *i*, Array of function pointers, Variable aliases, Opaque constructs

b: Order and dimension change of arrays

c: Evaluation of index pointer *j*

d: Multithreading

4.1 Multithreading

Although multithreading is typically used to enhance the performance of a program, in our implementation it is used only as an obfuscation technique. Multithreading, as implemented, significantly enhances the programs resilience against dynamic analysis attacks but it also slows down the resulting program because of the following reasons:

1. The threads are essentially executed in sequence as the key stream value is evaluated only within a critical section. Therefore, when this section is in use by a given thread, other threads remain idle.
2. The use of a critical section introduces an additional overhead. Furthermore, as the design of RC4 dictates, only a single index can be evaluated at a time, which offers no room for parallelism.
3. Threads that do not implement the switch/case statement for the given value of the index *i* introduce an additional delay. These threads lock the critical section and consequently prevent other threads from operation and, yet, produce no keystream words.

4.2 Excessive Use of Context Switches

The proposed implementation extensively uses branch statements to switch between real and dummy functionality. This context switching introduces a delay since each switch/case structure is evaluated before the real functionality is executed. In addition, each function, whether dummy or real, introduces further delay as its parameters have to be pushed or pulled from the stack. Furthermore, due to the use of arrays of function pointers, in each iteration, two additional functions have to be handled.

4.3 Additional Calculations Overhead

Many mathematical calculations are required to obfuscate the value of *j* and the array indices. This includes the additional calculations used to obfuscate the index pointer *j*.

Furthermore, when obfuscating the order of arrays by changing their dimensions, additional computations are used to select the correct index for the address of each function.

5 Conclusion

In this work, different practical obfuscation techniques for RC4 implementation were investigated and compared. Our implementation provides a high degree of resiliency against attacks from the execution environment where the adversary has access to the software implementation such as digital right management applications. Furthermore, while the focus of this paper was RC4, these techniques can be applied to other array-based stream ciphers such as HC-128, HC-256 and GGHN.

Providing a white box implementation with more theoretical foundations for different cryptographic primitives, including RC4, is an interesting research problem. Exploring how to use multithreading to speed up the RC4 implementation is another challenging research problem.

References

1. Menezes, A., van Oorschot, P.C., Vanstone, S.A.: Handbook of Applied Cryptography. CRC Press (1996)
2. Chow, S., Eisen, P.A., Johnson, H., van Oorschot, P.C.: White-Box Cryptography and an AES Implementation. In: Nyberg, K., Heys, H.M. (eds.) SAC 2002. LNCS, vol. 2595, pp. 250–270. Springer, Heidelberg (2003)
3. Chow, S., Eisen, P., Johnson, H., van Oorschot, P.C.: A White-Box DES Implementation for DRM Applications. In: Feigenbaum, J. (ed.) DRM 2002. LNCS, vol. 2696, pp. 1–15. Springer, Heidelberg (2003)
4. STUNNIX. C++ Obfuscator - Obfuscate C and C++ Code, http://www.stunnix.com/prod/cxxo/overview.shtml (accessed September 2011)
5. UPX: the Ultimate Packer for EXecutables, http://upx.sourceforge.net/ (accessed September 2011)
6. SecuriTeam. SecuriTeam - Shiva, ELF Encryption Tool, http://www.securiteam.com/tools/5XP041FA0U.html (accessed September 2011)
7. Collberg, C.S., Nagra, J.: Surreptitious Software: Obfuscation, Watermarking and Tamperproofing for Software Protection. Addison-Wesley (2010)
8. Bergeron, J., Debbabi, M., Desharnais, J., Erhioui, M., Lavoie, Y., Tawbi, N.: Static detection of malicious code in executable programs. Int. J. of Req. Eng. (2001)
9. Wang, C., Hill, J., Knight, J., Davidson, J.: Software Tamper Resistance: Obstructing Static Analysis of Programs. Technical Report CS-2000-12. Univ. of Virginia (2000)
10. Reddit: the Front Page of the Internet. Skype's Obfuscated RC4 Algorithm Was Leaked, so Its Discoverers Open Code for Review: Technology, http://www.reddit.com/r/technology/comments/cn4gn/skypes_obfuscated_rc4_algorithm_was_leaked_so_its/ (accessed September 2011)
11. Biondi, P., Desclau, F.: Silver Needle in the Skype, http://www.secdev.org/conf/skype_BHEU06.pdf (accessed September 2011)

12. Ogiso, T., Sakabe, Y., Soshi, M., Miyaji, A.: Software obfuscation on a theoretical basis and its implementation. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences E86-A, 176–186 (2003)
13. Collberg, C.S., Thomborson, C., Low, D.: A taxonomy of obfuscating transformations. Technical Report 148, Department of Computer Science. University of Auckland (1997)
14. Zhu, W., Thomborson, C.D., Wang, F.-Y.: Obfuscate arrays by homomorphic functions. In: GrC, pp. 770–773 (2006)
15. Collberg, C.S., Thomborson, C.D., Low, D.: Manufacturing Cheap, Resilient and Stealthy Opaque Constructs. In: POPL, pp. 184–196 (1998)
16. Park, J.-Y., Yi, O., Choi, J.-S.: Methods for practical whitebox cryptography. In: 2010 International Conference on Information and Communication Technology Convergence (ICTC), pp. 474–479 (November 2010)
17. Link, H.E., Neumann, W.D.: Clarifying obfuscation: Improving the security of white-box encoding, cryptology eprint archive. In: Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC 2005), vol. I (2005)

A Statistical Pattern Mining Approach for Identifying Wireless Network Intruders

Nur Al Hasan Haldar[1], Muhammad Abulaish[2,⋆], and Syed Asim Pasha[3]

[1] Comviva Technologies Pvt. Ltd., Gurgaon, Haryana, India
nurjamia@gmail.com
[2] Center of Excellence in Information Assurance, King Saud University, Saudi Arabia
(On leave from Jamia Millia Islamia, New Delhi, India)
abulaish@ieee.org
[3] Ericsson India, Gurgaon, Haryana
asim.pasha2@gmail.com

Abstract. In this paper, we present a statistical pattern mining approach to model the usage patterns of authenticated users to identify wireless network intruders. Considering users activities in terms of ICMP packets sent, DNS query requests and ARP requests, in this paper a statistical approach is presented to consolidate authenticated users activities over a period of time and to derive a separate feature vector for each activity. The proposed approach also derives a local threshold for each category of network data analyzed. The learned features and local threshold for each category of data is used during detection phase of the system to identify intruders in the network. The novelty of the proposed method lies in the elimination of redundant and irrelevant features using PCA that often reduce detection performance both in terms of efficiency and accuracy. This also leads our proposed system to be light-weight and deployable in real-time environment.

1 Introduction

Due to easy installation, portability and mobility features, Wireless Local Area Networks (WLANs) are frequently being used by users from different walk of life. Although, WLANs solve some problems that exist in traditional wired LANs, there still exists certain vulnerabilities due to flaws in some IEEE 802.11 standard protocol which makes wireless network a highly desirable target in terms of security breach. Vulnerability exploits based remote attacks are one of the most destructive security issues faced by the security community as most of the high profile security threat of automatic dissemination attacks or worms are based on remote exploitation of vulnerabilities in compromised systems [10]. The open nature of wireless medium makes it easy for attacker to listen and analyze network traffic. Various useful security techniques like, AES, WEP, WPA or WPA2 can protect data frames, but an attacker can still spoof control or management

⋆ Corresponding author.

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 131–140.
springerlink.com

frames to damage security. These attributes makes wireless network potentially vulnerable to several different types of attacks. Although, a number of security measures have been already proposed by security researchers and they are operational to protect wireless networks, one cannot make the assumption that wireless users are trusted. Malicious individuals can easily sit outside or inside an organization's premises, and freely connect to a wireless network to do malicious activities.

Intrusion detection is generally used to secure any system in a network by comparing the set of baselines of the systems with their present behavior [8]. The main two known models of intrusion detection system (IDS) are (i) signature-based intrusion detection, and (ii) anomaly-based intrusion detection. The first one uses signatures of the well-known attacks to detect intrusion. A signature-based detection technique facilitates intrusion detection by investigating various routing protocols when attack signatures are completely known. This type of detection monitors the wireless networks for finding a match between the network traffic and a well-known attack pattern. But, this approach suffers with a number of limitations including the inability to cope up with exponential increase in new malicious exploits and consequently the huge size of signature database, inability to detect new and unfamiliar intrusions even though they are very similar to the known attacks. Further, the signature-based techniques can not be used to detect zero-day exploits. In contrast to signature-based IDS, anomaly-based IDS creates profiles that are based on the normal behavior of the underlying network [2]. Anomaly-based IDS first establishes a model of normal system behavior and anomaly events are then distinguished based on this model. So in this approach, the detection is performed by learning the normal behavior of a network and comparing it with the behavior of monitored network. This type of detection has the ability to detect new unknown attacks (zero-day attacks) without any prior knowledge about them. However, the false alarm rate in such systems is usually higher than that in the signature-based systems [11].

Considering this scenario which necessitates to monitor users activities of a network to identify possible intruders, in this paper we have proposed a statistical pattern mining approach to design a wireless intrusion detection system. A poster version of this work appears in [14]. The proposed approach gathers authenticated users activity data in the network and applies statistical pattern mining to derive feature vectors to characterize them in the system. The characterization is later used during detection phase of the system to identify malicious users or intruders whose characteristics substantially deviates from the activity patterns of the normal users. Some of the features monitored by the proposed IDS are ICMP packets sent, DNS query requests, and ARP requests. In order to enhance the performance of the proposed system both in terms of efficiency and detection accuracy, Principal Component Analysis (PCA) is applied on captured traffic data to filter out irrelevant components and map them into a lower-dimensional space. The method described in [1] is found to be very sensitive to small changes in a noisy environment at a reasonable label, which causes many non-negligible numbers of false alarms. In order to reduce such false alarms and to increase the effectiveness, we enhance the PCA-based detection to identify anomalous nodes. Also the threshold estimation approach presented in this paper helps in reducing the false alarm rate.

Starting with a brief review of the existing intrusion detection techniques in Section 2, section 3 introduces the process of dimension reduction and pattern extraction using PCA. Section 4 presents the design of the proposed system, followed by the experimental results in section 5. Finally, section 6 concludes the paper with future directions of work.

2 Related Work

Due to increasing security issues related to wireless networks, a number of research efforts have been directed towards identification and prevention of network attacks. As a result, there exists many intrusion detection systems that can detect an attack at the IP layer or above [3]. Hu and Perrig [3] proposed a signature-based intrusion detection method to detect wormhole attacks. In [13], Mukkamala et al. proposed a method for audit trail and intrusion detection using SVM and neural network approach. Some other approaches including statistical methods, wavelet analysis, data mining techniques, and Kolmogorov-Smirnov test have been also proposed by various researchers. Besides its usefulness to detect the anomalous traffic, the Kolmogorov-Smirnov test [4] fails to identify the cause of the anomaly. Similarly, the anomaly detection methods using wavelet analysis suffers with heavy computational overheads and consequently they can not be deployed in real-time environment. In [12], Portnoy et al. proposed an algorithm to detect both known and new intrusion types using a clustering approach which does not need labeled training data, but it requires a predefined parameter of clustering width which is not always easy to determine.

In can be seen that the above-described methods are mostly heavy-weighted and most of them cannot be deployed in real-time as those require complete dataset for processing. On the other hand, PCA-based methods are relatively light-weight and processing speed is also high. The proposed work by Feather et al. in [6] is very similar to PCA-based method. They used some signature matching algorithm to detect anomalies. Another work is reported by Dickinson in [5], where anomalous traffic is compared against certain threshold. In the combined work reported by Wang et al. [7], the intrusion detection techniques are developed using profile-based neighbor monitoring approach. They used Markov Blanket algorithm for the purpose of feature selection to decrease the number of features dramatically with very similar detection rate. But the profile-based neighbor monitoring intrusion detection approach requires a lot of network features to monitor. For improved detection rate, our method identify effective features by ignoring uncorrelated variables using PCA.

3 Dimensionality Reduction and Pattern Extraction Using PCA

Generally, consideration of a large set of features results in too many degrees of freedom, which leads to poor statistical coverage and thus poor generalization. In addition, each feature introduces a computational overhead both in terms of efficiency and storage. Principal Component Analysis (PCA) is a statistical method for dimension reduction that maps high-dimensional data points onto a lower-dimensional set of axes that best explain the variance observed in the dataset [9]. The purpose behind using PCA is to replace the original (numerical) variables with new numerical variables called "principal components" that have the following properties. (i) The principal components can be ranked by decreasing order of importance, and the first few most important principal components account for most of the information in the data. (ii) There is almost as much information in the principal component variables as there are in the original variables. In this way, PCA can be perceived as a technique for dimensionality reduction to eliminate irrelevant data from the original dataset. The working principle of PCA can be summarized as follows.

- Organize data as an $m \times n$ matrix $A_{m \times n}$, where m is the number of measurements and n is the number of trials (dimensions).
- Subtract mean from each of the data dimensions, where mean is the average across each dimension. This produces a data set whose mean is zero.
- Generate covariance matrix $Cov_{n \times n}$ for $A_{m \times n}$ using equation 1 to calculate the covariance between a pair of dimensions dim_i and dim_j.

$$cov(dim_i, dim_j) = \frac{\sum_{k=1}^{m}(dim_{i_k} - \overline{dim_i})(dim_{j_k} - \overline{dim_j})}{m-1} \tag{1}$$

- Calculate the eigenvectors and eigenvalues of the covariance matrix $Cov_{n \times n}$. Since $Cov_{n \times n}$ is a square matrix, we can calculate the eigenvectors and eigenvalues, describing important information about the patterns in data.
- Select eigenvalues and form feature vectors.

Here the role of PCA in data compression and dimensionality reduction comes in picture. Generally, different eigenvalues are quite different values and the eigenvector with the highest eigenvalues is the principal component of the dataset. For feature selection, we order the eigenvectors by their eigenvalues (highest to lowest), which gives us the components in order of their significance and ignore the components of lesser significance. Finally, we form the feature vector by arranging the selected eigenvectors in the form of a matrix.

4 Proposed Statistical Pattern Mining Based IDS

In this section, we present the procedural detail of the proposed statistical pattern mining based intrusion detection system, which is shown in figure 1. Our system processes the following three types of data in the network to learn the normal usage pattern of authorized users using PCA and generates a profile as a feature vector for each of them. The generated profiles are then used as a baseline to identify intruders in the network.

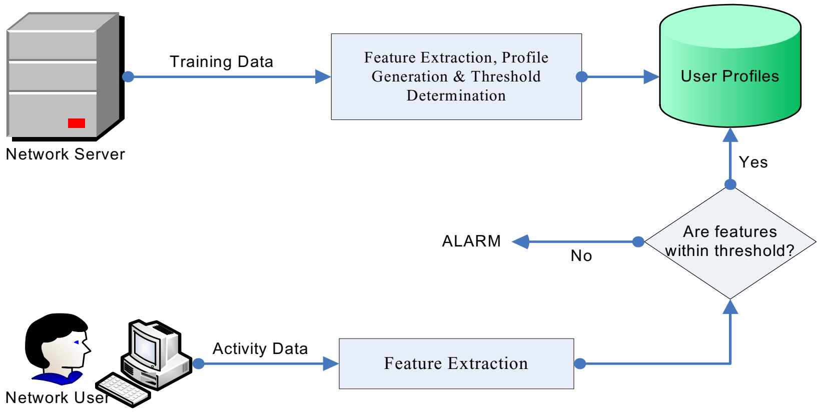

Network Server
Training Data
Feature Extraction, Profile Generation & Threshold Determination
User Profiles
Yes
Are features within threshold?
No
ALARM
Network User
Activity Data
Feature Extraction

Fig. 1. Architecture of the proposed intrusion detection system

ICMP packets sent – Internet Control Message Protocol (ICMP) is a network layer protocol, which is generally responsible to relay query message and also to send some error message if either a requested service is not available or a host or a router could not be reached. Attacks like *Smurf* and *Papasmurf* happen when intruder sends a large amount of ICMP packets (ping) to a subnet broadcast address. ICMP can be used as a first step in an attack because it can determine the alive hosts before attacking. Therefore, ICMP packet is considered as a parameter and the number of IMCP requests sent and received is monitored by our system.

DNS query requests – Domain Name System (DNS) query is used to translate domain names into the actual IP address of destination machine. DNS spoofing attack changes the entry of IP address of a domain name to some other IP address. Therefore, the DNS requests of a user are considered to model his/her domain of interests.

ARP requests – Address Resolution Protocol (ARP) is responsible to map an IP address to its associated data link layer address (MAC address). The ARP requests can be spoofed by the intruders to divert the packets to a wrong destination. Also, attacker can block traffic resulting in Denial of Service (DoS) attack. So, ARP request is also an important feature to characterize intruders in the system.

Like standard classification methods such as Naive Bayes, Decision Tree, Support Vector Machine (SVM) and Neural Network, our intrusion detection system is implemented as a two-phase process given below.

Phase-1 (Training phase): This is basically profile generation phase. In this phase, the activity data of authorized users are collected and their profiles are generated using PCA algorithm. Threshold value to determine the maximum profile deviation of a normal user from the profiles of the authenticated users is also determined during this phase.

Phase-2 (Detection phase): This is also called profile detection phase. In this phase, the learned profiles are used as a baseline to identify possible intruders in the system. If the Euclidean distance of a user profile with the learned profiles of the authentic users is greater than the pre-determined threshold then an alarm is generated and the user is classified as a possible intruder and consequently the packets are dropped. Otherwise, the user is classified as a normal user and its profile is added in the profile set of the authenticated users as shown in figure 1. In this way, the profile set and thereby the intrusion detection accuracy of the system increases with time. Further detail about the profile generation and threshold determination phase is presented in the following sub-section.

4.1 Profile Generation and Threshold Determination

In this phase, we train the application to learn the usage patterns of the known users after analyzing their activity data. For this, we capture the activity data related to different type of features for each authenticated user u_i in the system and organize them into an $n \times m$ matrix A, where n is the number of slots and m represents the number of days for which data are captured. Thereafter, we apply PCA on matrix A to select $p < m$ eigenvectors corresponding to high eigenvalues, which forms an $n \times p$ matrix. We say this matrix a weight matrix and represent it using W. In order to get an aggregated pattern for the user under consideration over the considered time-period, we apply scalar product between each column vectors of W and matrix A using equation 2. This scalar product gives a matrix P of order $m \times p$ in which the columns correspond to the selected eigenvectors and the rows corresponds to the days considered for profile generation. Finally, each column of matrix P is averaged and used to generate a profile of the user u_i as a p-dimensional vector $\psi(u_i) =< w_1, w_2, \ldots, w_p >$, where the value of w_k $(k = 1, 2, \ldots, p)$ is calculated using equation 3.

$$P = A^T.W \tag{2}$$

$$w_k = \frac{\sum_{j=1}^{m} col_{k_j}(P)}{m} \tag{3}$$

For each class of user activity data, this process is repeated to generate the profile of all authenticated users. All generated profiles are considered to form a single cluster and its centroid is calculated as an average of the corresponding elements in the profile vectors. Dissimilarity between a pair of two profile vectors

$\psi(u_i) =< w_1, w_2, \ldots, w_p >$ and $\psi(u_j) =< w'_1, w'_2, \ldots, w'_p >$ is calculated as an Euclidean distance between them using equation 4. And, the maximum of these distances over all user profiles is decided as threshold value as given in equation 5. This threshold value is used in profile detection phase to identify possible intrusions in the network. For a given user, his/ her profile is generated using the above-discussed method and its distance from the generated centroid is calculated. If the distance value is within the threshold then the profile and thereby the data packet is considered as benign, otherwise it is considered as malicious and an alarm is raised.

$$\delta(\psi(u_i), \psi(u_j)) = \sqrt{(w_1 - w'_1)^2 + (w_2 - w'_2)^2 + \ldots + (w_p - w'_p)^2} \tag{4}$$

$$\theta = \left\lceil \max_{1 \leq i < n,\ i < j \leq N} \{\partial(\psi(u_i), \psi(u_j))\} \right\rceil \tag{5}$$

5 Experimental Setup and Results

In this section, we discuss our experimental setup and results. All experiments are performed on a PC with Intel Core Duo 1.66 GHz processor, and 2 GB RAM. For simulation, we have used SIGCOMM[1] 2008 dataset, which contains three types of anonymized traces – 802.11a, Ethernet and Syslog. After downloading dataset, we discarded other irrelevant data like wired traces, Syslog traces, etc. and took only wireless data for our experiment.

Since the wireless dataset contains "pcap" files, we have used Wireshark[2] for analysis and filtering purposes. For all days, the overall start packet time and last packet time are noted at 10:32:32 hrs and 19:24:12 hrs respectively, resulting in total 31900 seconds data per day. After filtering the dataset, we merged all individual file data day-wise with respect to available packets' time. The all days data were sorted after merging in ascending order with respect to start packet time. Then we calculated the number of packets in a fixed interval (100 seconds) for each day. Some sample instances along with data file name and no. of different types of packets are given in table 1.

$$W = 1.0e + 003 * \begin{pmatrix} -0.2458 & 0.1224 & 0.0033 \\ -0.2458 & 0.1227 & 0.0042 \\ -0.2458 & 0.1227 & 0.0042 \\ \ldots & \ldots & \ldots \\ -0.2449 & 0.1167 & -0.0064 \\ -0.2457 & 0.1198 & 0.0051 \\ -0.2458 & 0.1227 & 0.0042 \end{pmatrix} \tag{6}$$

We have selected a particular machine (MAC address) from the wireless dataset and gathered the total number of packets (ICMP, DNS and ARP) transmitted by

[1] http://uk.crawdad.org/meta.php?name=umd/sigcomm2008

[2] http://www.wireshark.org/download.html

Table 1. Sample data files and the no. of filtered ICMP, DNS and ARP packets

Data File Name	**ICMP packets**	**DNS packets**	**ARP packets**
sigcomm08_wl_10_2008-08-19_11-22_23_2008-08-19_18-07_56_b88796010035_36.pcap	565	8250	709
sigcomm08_wl_13_2008-08-19_13-10_42_2008-08-19_19-24_12_b88796010035_36.pcap	379	4530	409
sigcomm08_wl_10_2008-08-20_10-50_35_2008-08-20_17-46_56_b88796f1e2c7_52.pcap	4616	18655	1044
sigcomm08_wl_16_2008-08-20_11-29_26_2008-08-20_17-52_37_b887965a5527_36.pcap	35188	16724	1092
sigcomm08_wl_9_2008-08-20_10-46_31_2008-08-20_17-44_57_b887965a5527_36.pcap	22610	10367	733
sigcomm08_wl_8_2008-08-20_10-58_15_2008-08-20_17-50_06_b88796a13ec7_149.pcap	1536	1982	495

Table 2. Weight components for day-wise data

	w_1	w_2	w_3
Day 1	1562891	-916210	-202630
Day 2	2173000	-6086500	545010
Day 3	39677000	485680	104580
Day 4	-49121	-236750	-17780
Day 5	2735800	-1673300	-1838700

that machine. We took five days data (31900 seconds data each day) for profile generation from the above-mentioned dataset. After observing the captured data pattern, we manipulated them to fit the total number of packets in time interval of 100 seconds. The time interval may vary according to the packet rates of the captured dataset for better result. Thus, we got total 319 slots and calculated the number of packets in each slot of 100 second interval. In this way, the length of the column vector, representing each day data, is 319 resulting in 319×5 matrix for the five days data. After applying PCA on its covariance matrix, the first three eigenvectors corresponding to top-three eigenvalues forming 319×3 weight matrix is shown in equation 6.

Thereafter, the weight components corresponding to the highest three eigenvalues for each day is calculated using equation 2. The resultant matrix data after this operation is shown in table 2. Finally, each column of table 2 is averaged to generate the user profile as a 3-dimensional vector. This process is repeated for all the authenticated users and for all three different types of data - ICMP, DNS, and ARP. The centroid vector for each data type is calculated by taking the average of the corresponding components of the user profile vectors. Similarly, the threshold value for each type of data packet is calculated using equations 4 and 5.

Once the centroids and threshold values for all datasets are fixed they are used to identify possible intrusion attempts in the system. For a user under consideration, if any of the corresponding thresholds is crossed then this indicates that a probable intrusion attempt is made by the user, whereas on the other

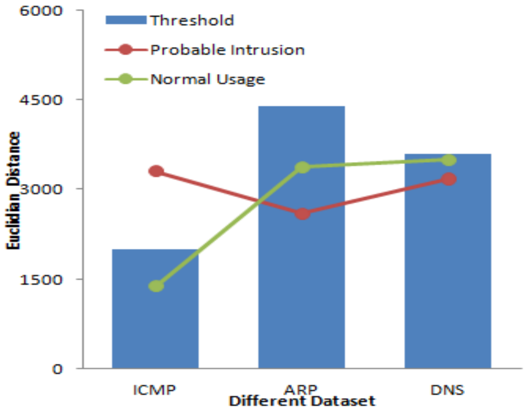


Fig. 2. Visualization of threshold values for ICMP, ARP and DNS datasets, and two user profiles - one (green curve) benign and other (red curve) malicious

hand if the user profile regards all threshold values, it is considered as benign and added in the profile set to derive finer tuned user profile. Figure 2 shows the threshold values for ICMP, DNS and ARP datasets, and two user profiles – one (green color) identified as benign user, whereas the other one (red color) as a malicious user.

6 Conclusion and Future Work

In this paper, a light-weight statistical pattern mining based wireless intrusion detection system is proposed to model authenticated users activities to identify intrusions in wireless networks. Various network layer data are collected and statistical pattern mining approach is applied to identify usage patterns of authenticated users for their characterization. For efficiency purpose, the proposed system applies PCA algorithm on traffic data to map them into lower-dimensional space. Presently, we are working towards the evaluation of the proposed system on different real datasets. The proposed method can be extended to protect wireless routers from malicious attacks.

References

1. Lakhina, A., Crovella, M., Diot, C.: Diagnosing network-wide traffic anomalies. In: Proc. of the ACM SIGCOMM 2004, NY, USA, pp. 219–230 (2004)
2. Sekar, R., Gupta, A., Frullo, J., Shanbhag, T., Tiwari, A., Yang, H., Zhou, S.: Specification based anomaly detection: a new approach for detecting network intrusions. In: Proc. of the 9th ACM CCS, NY, USA, pp. 265–274 (2004)
3. Hu, Y.C., Perrig, A., Johnson, D.B.: Wormhole attacks in wireless networks. Journal on Selected Areas in Communications 24(2), 370–380 (2006)

4. Caberera, J.D., Ravichandran, B., Mehra, R.K.: Statistical traffic modeling for network intrusion detection. In: Proc. of the 8th Int'l Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems, pp. 466–473 (2000)
5. Dickinson, P., Bunke, H., Dadej, A., Kraetzl, M.: Median graphs and anomalous change detection in communication networks. In: Proc. of the Information, Decision and Control, Australia, pp. 59–64 (2002)
6. Feather, F., Siewiorek, D., Maxion, R.: Fault detection in an ethernet network using anomaly signature matching. In: Proc. of the ACM SIGCOMM 1993, NY, USA, pp. 279–288 (1993)
7. Wang, X., Lin, T.L., Wong, J.: Feature Selection in intrusion detection system over mobile ad-hoc network. Technical Report. Iowa State University, USA (2005)
8. Mishra, A., Nadkarni, K., Patcha, A.: Intrusion detection in wireless ad-hoc networks. IEEE Wireless Communications 11(1), 48–60 (2004)
9. Smith, L.I.: A tutorial on Principal Components Analysis (2002)
10. Wang, H.J., Guo, C., Simon, D., Zugenmaier, A.: Shield: vulnerability-driven network filters for preventing known vulnerability exploits. SIGCOMM Comput. Commun. Rev. (2004)
11. Garcia-Teodoro, P., Diaz-Verdejo, J., Macia-Femandez, G., Vezquez, E.: Anomaly-based network intrusion detection: techniques, systems and challenges. Computers and Security, 18–28 (2009)
12. Portnoy, L., Eskin, E., Stolfo, S.: Intrusion detection with unlabeled data using clustering. In: Proc. of ACM CSS Workshop on Data Mining Applied to Security, pp. 5–8 (2001)
13. Mukkamala, S., Janoski, G., Sung, A.: Intrusion detection using neural networks and support vector machines. In: Proc. of the Int'l. Joint Conf. on Neural Networks, pp. 1702–1707 (2002)
14. Haldar, N. Al-H., Abulaish, M., Pasha, S.A.: An activity pattern based wireless intrusion detection system. In: Proc. of the 9th Int'l. Conf. on Information Technology–New Generations, Las Vegas, USA (2012)

SLA for a Pervasive Healthcare Environment

J. Valarmathi, K. Lakshmi, R.S. Menaga, K.V. Abirami, and V. Rhymend Uthariaraj

Computer Center,
Madras Institute of Technology, Anna University
Chennai, India
{valar.shakthi,kasinathan.lakshmi,
rsmenu20,abivijay21}@gmail.com

Abstract. The use of Pervasive computing in Healthcare field is an evolving paradigm and is under constant development in the recent years. Pervasive computing allows information processing and interactions to be carried out by smart devices autonomously. The advent of wearable computers or wearable devices has enabled continuous monitoring of patients and has changed the face of healthcare systems in many parts of the world. Since these devices carry sensitive data and interactions are initiated and carried out without much human intervention, the need for trust and security is crucial. This paper proposes a Service Level Agreement (SLA) based approach to trust, that relies on credential matching to form a notion of trust. The number of interactions and computational overhead is reduced without compromising safety. The aim is to provide secure and reliable healthcare services without any arbitration. The security and power issues in a pervasive environment are also addressed.

Keywords: SLA, trust, Pervasive Computing, Credentials.

1 Introduction

Pervasive computing is an open and decentralized environment. The communication between devices in a pervasive environment take place with, if at all any, minimum human supervision. The devices, which could be heterogeneous in nature, interact without prior knowledge about each other. More often than not, these interactions involve large amount of data or sensitive data. Hence, proper mechanisms need to be in place to protect the integrity of the interactions and safety of the data transfer. The devices need to be trained to make a distinction between the trustworthy devices and the selfish and malicious ones.

Wearable computing devices have taken healthcare scenarios to an altogether new level. Most common examples of wearable devices comprise of Personal Digital Assistants(PDAs),EEG or ECG sensors and blood pressure meters. With the help of such devices, critical bio metrics can be monitored and signals can be set of in critical situations. Owing to the mobility factor that is dominant in any pervasive environment, conventional cryptographic and trust models do not gain significance in this type of network. Another important feature that characterizes the pervasive environment is the limited processing capability in terms of battery power. Therefore, the resources utilized by the algorithms to compute the trustworthiness of a node, in comparison to

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 141–149.
springerlink.com

the resources utilized for the purpose for which the network is set up, should be negligible.

When a node enters a new community or when the history of interaction between two devices are barely existent, most trust models resort to an arbitrary initial trust assignment or reputation based trust calculation. But reputation based trust increase the number of interactions significantly. Hence a credential matching approach is used to identify the best possible device that could be trusted. Negotiations are conducted to resolve any conflicts in choosing a device based on optimal pricing, time of delivery and availability. A fast and effective negotiation system is provisioned by means of a Service Level Agreement that records the finalized negotiation deal and the penalty that any party is likely to attract on grounds of a violation.

The rest of the paper is organized as follows. Section 2 consists of related works in the areas of trust and pervasive computing , Section 3 explains about the proposed work. In section 4, the performance evaluation as compared with the deterministic model is shown. Finally section 5 concludes the work and shows future work.

2 Related Work

Trust has been dealt with in [1] as multicast scheme. Security is identified as essential as many patients have privacy issues and concern when it comes to transmitting their data over a network. Trust is considered as function of time, longer the residence time in an environment, greater the trust. In [2], two major case studies are performed in the light of assessing the importance of Pervasive healthcare scenario. Proof of medical benefit, user participation and financial clarification are identified as major factors that influence adoption of pervasive computing in healthcare. [3] uses a rapid trust sniffer module to enhance security. It involves moving trust partially from application layer to Operating System layer. The trust sniffer validates an application by comparing sample measurement with reference measurements. In [4], the wearIT@work project is discussed which explains work of the open wearable computing group. [5] points out that trust based on Bayes theorem,Pempster Shafter Theory of evidence and subjective algebra methods to trust computation are very complex and suggests considering trust in a multi disciplinary approach, such as economics and IT etc. In [6],the need for a unified trust relationship model is highlighted and a context specific approach to multi hop interactions is proposed. [7] assures availability of resources, trustworthiness by creating an explicit trust binding between the components that may participate in a service composition. A CTB is a set of rules which define the collection of allowable components for a particular service. A service-invoking node can distribute a CTB to the service-providing node which is then expected to enforce the CTB policy as to which components are permissible for use in delivery of the service. Similarly a content-owning node can distribute a CTB to a service which processes the content, which is then expected to enforce the CTB policies during access to that content. [8] lists recommendation mechanism is introduced on assumption that trust is inherently transitive as issues and shortcomings in previous models. Further, [8] uses an approach that provides a pervasive trust model between autonomous entities without central security managers. A recommendation mechanism of rewards and punishments is based on recommendation-credibility. [9] suggests a novel trust propagation model which can get higher quality of trust by combining both the trust value and trust strength which is nothing but the credibility of the recommended trust value. Its main idea is to increase the trustworthiness of the target node by not considering just the trust value from direct experiences or by recommendation but also considering the credibility of the trust value.

2.1 SLA in Other Domains

In [10], roles of two parties and focus on how service providers generate offers upon receiving the requests from service requesters are studied. From the provider's perspective, the provider has to decide the right values to offer based on its current resource availability while aiming to satisfy the requester requirements. [10] has proposed an approach to addressing offer generation, including the architecture, the information modeling and the generation algorithm. [11] considers problem of mediated group decision making where a number of agents provide a preference function over a set of alternatives. A mediation step is applied to aggregate the individual preferences in order to obtain a group-preference function. And the most supported alternative is selected. In a cloud environment, [12] focuses on cost-aware scheduling of queries. iCBS takes the query costs derived from the service level agreements (SLAs) between the service provider and its customers into account to make cost aware scheduling decisions. iCBS is an incremental variation of an existing scheduling algorithm, CBS. In [13], PROTUNE, a rule-based Trust Negotiation system is described and the advantages that arise from an advanced rule-based approach in terms of deployment efforts, user friendliness, communication efficiency, and interoperability are illustrated. [14] mainly aims in employing SLA under a web-service domain and it proposes a design of a service oriented SLA based monitoring system that helps not only in administering the web services but also in evaluating the quality of those services.

3 Proposed Work

3.1 Need for SLA Based Management

Every expectation that the customer has are written in the service level agreement or SLA. The SLA in general describes the core services provided to customers and details their responsibilities in making sure that every concept is met in accordance to SLA. Customer feedback is one of the reasons why products and services are constantly innovating. All the services that the customer needs are included in the SLA.

3.2 Pervasive Community Model

In this trust management system for a Pervasive Environment, arbitrary trust assignment and recommendation based interactions for trust gathering are not used. Arbitrary trust assignment could botch the integrity of an honest device or reward untrustworthy devices. Also, since devices in a pervasive environment are not conventional computing devices, but are largely mobile and hence possess limited processing capability in comparison to their conventional counterparts, in terms of both power and memory, a recommendation based trust gathering mechanism would only increase the number of interactions between devices.

The proposed trust management system is based on a community model that comprise of nodes. The communities are geographically designed and maintained and a node in the community, which is the central agent for the community, takes the responsibility to maintain all registration details of all the nodes that reside in a community. Every node that wants to register in a community provides details of itself along with the accredited credentials based on which it will be assessed mainly. This information is stored locally by the community head, that acts as a trusted agent to mediate requests and services between nodes. When a node is in need of a service, it sends out a message requesting the service. The message consists of

- Service(S)
- Cost(C) , willing to be paid
- On or before time(obt)
- Required credential list , in an order denoting priority
- Least value for a match (X) , for ascertaining a credential match.

As suggested by [15] , a global data store is used to store the trust values of all the nodes across various communities. Recommendation overhead is reduced to a remarkable extent. Instead of pooling recommendations from n different recommenders about a node, the global trust value of the node is obtained by accessing it from the global data store, which effectively provides the same result with just one interaction. The records of the global data store are stored as tuples of the format<T,n,t>, where T is the aggregated trust from n interactions and t is the latest time of update of the tuple. Also, all the nodes, locally maintain relevant trust values that are indicators of how other nodes have behaved with any particular node.

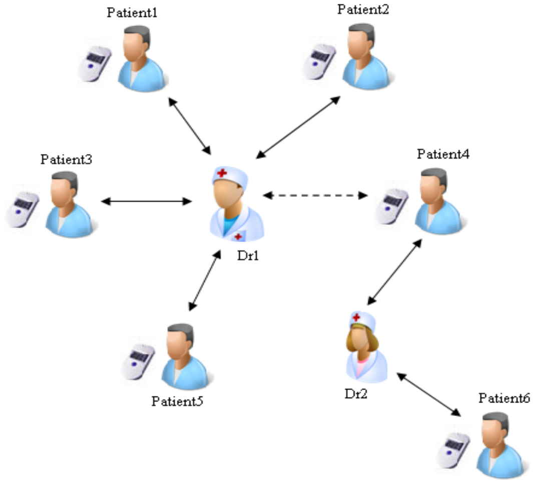


Fig. 1. Pervasive community

3.3 Request Processing

Node A, when it wants to avail a particular service from the community, it sends a message in the format specified previously to the community head. On receiving the message, the community head now retrieves the required credential list from the message. From this list, it derives a priority or weight for every corresponding policy j, for every node i as $w_{ij.}$

ALGORITHM : Request Processing.

Input : Service request message

Output : Set N that contains nodes after successful credential matching

```
For every node i,
   assign priority $w_{ij}$ to all the credentials;
   1<=j<=n
P : = Nodes providing service S
For every node i in P,
    $p_{ij}$ ←credential value for node i under credential
    j;1<=j<=n
$X_i = \sum w_{ij} p_{ij} / \sum w_{ij}$
If($X_i$ >=X)
    Then N=N U {i}
If N = ∅
    Forward request to neighbouring communities.
End
```

For all nodes i in N,

$$\Psi_i = X_i + \left(\alpha * \frac{N_S}{N_S+N_F}\right) + \beta * T_i \tag{1}$$

If, $\Psi_i < \text{Treshold}_A$, remove i from N.

N_S, is the number of successful interactions,

N_F, is the number of failed interactions.

$$\alpha + \beta = 1 \tag{2}$$

Equation (1) takes into account the credential matching factor computed in the algorithm, previous history of node A with node i, and the global trust of node i, T(i). Both N_S and N_F are obtained from history of interaction that is stored locally by node A. α and β are weighing factors, that are node-specific. α = 0, if there is no previous interaction between node A and node i.

3.4 Negotiation

Instead of going by the node by the maximum value of Ψi ,Node A negotiates with the nodes in N to find out the best offer that is optimized in time and cost.

```
while (!isEmpty(N))
    j← retrievenode(N);
   if(time(j)<=obt && cost(j)<=C)
         Z=Z U {j}
A= AU {Zi}  where time(Zi)  is minimum.
If |A|>=1
    Find j in A, such that cost(A) is min.
```

3.5 Monitor SLA Violation

After a successful negotiation stage, an SLA is created and maintained with the resource and job information. While monitoring, any overuse in the guaranteed parameters than the agreed values or reduce in the quality parameters agreed upon triggers a SLA violation. The requester node, can then decide the reward of punishment factor that would influence the future trust of the provider node.

3.6 Trust Evaluation and Updation

After every interaction, the trust value a node has on another, has to be updated based on the how well the provider responded. A direct trust computation is performed after two nodes have finished interacting with each other. Let us consider two devices A and B. $T_A(B)$ is used to represent the trust that node A has on node B. After every interaction between A and B, their trust is aggregated and stored locally by A, and also contributes to global trust value of B. This value is calculated based on the outcome of the interaction, the satisfaction of node A with node B's service. Basically node A performs an evaluation of the service provided by node B, based on the compliance of the agreement made beforehand, and gives a rating value r. time, which is the time taken by the node to respond is also included in the evaluation process. The more the time, less satisfied node A would be.

$$T'_A(B) = \omega * \frac{1-T_A\,(B)^{r/time}}{1-T_A(B)} \tag{3}$$

Where $T'_A(B)$ is the updated trust value and ω is the quantization factor that is used to keep trust in the range 0 to 1.

To update the global trust value,

STEP I: Get tuple corresponding to node B <T(B),n,t>
STEP II: Calculate $T'_A(B)$

$$T'(B) = \frac{n*T(B)+T'_A(B)}{n+1} \tag{4}$$

STEP III: Get current time,t
STEP IV: Update tuple in global set as < T'(B),n+1,t>

The Algorithm makes use of weighted arithmetic mean to update the trust value globally by maintaining trust as an average of all the users in the environment. The global trust of a node or device builds up only gradually and there is no way of one particular entry, malicious or not, to malign the existing values.

4 Performance Evaluation

For the evaluation of the proposed model, a series of performance evaluations is performed in a pervasive healthcare environment. The healthcare environment, consists of a number of healthcare providers and a number of requesters who wish to avail these services.

The following performance parameters are used in the evaluation process.

- Number of nodes , denote the number of active nodes in the environment
- Throughput, considered as a network parameter, is the number of packets processed per unit time or the number of requests processed per unit time.
- Number of Interactions , is used to judge the traffic density of the network. It is directly proportional to the number of nodes in the network.

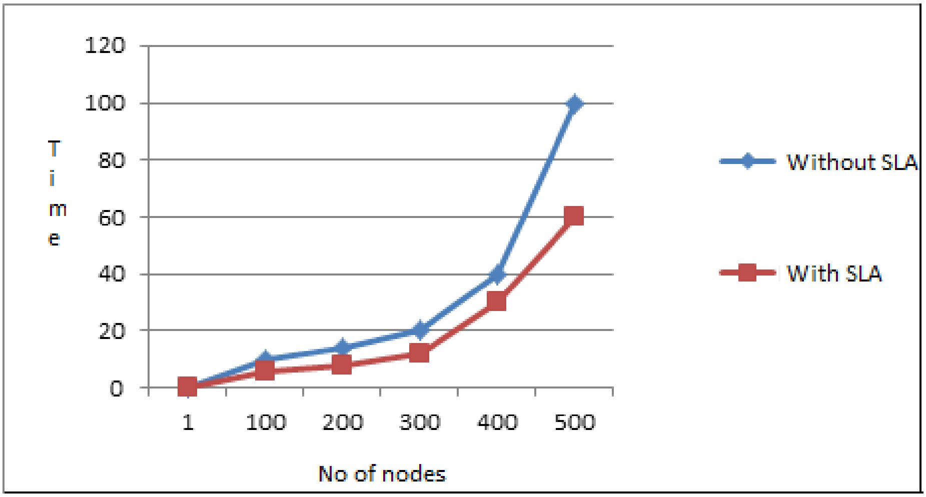


Fig. 2. Time Vs. No. of nodes

In this section, we have compared the time taken for trust value computation for a deterministic model without SLA to the time taken for trust computation with SLAs. In the conventional model, since recommendations are used, it leads to more network traffic density and more time is taken when number of recommending nodes increases. With SLA, the trust computation depends on a device based trust as well as network characteristics. Thus, in the proposed model, number of interactions is kept very low and the time taken for trust computation is also considerably reduced.

5 Conclusion

In this paper, the concept of using Service Level Agreements in a Pervasive environment for a healthcare scenario is proposed. Using Service Level Agreements, both the parties have an understanding of what is expected out of them and what price they have to pay in case of a violation by either party. The customer knows exactly what to expect in terms of Quality of Service. Experimentally, it is observed that trustworthiness of node is maintained steadily with the number of interactions.

References

1. Boukerche, A., Ren, Y.: A Secure Mobile Healthcare System using Trust-Based Multicast Scheme. IEEE Journal of Selected Areas in Communications 27, 387–399 (2009)
2. Orwat, C., Rashid, A., Holtmann, C., Wolk, M., Scheermesser, M., Kosow, H., Graefe, A.: Adopting Pervasive Computing for Routine Use in Healthcare. IEEE Pervasive Computing 9, 64–71 (2010)
3. Surie, A., Perrig, A., Farber, D.J.: Rapid Trust Establishment for Pervasive Personal Computing. IEEE Pervasive Computing 6, 24–30 (2007)
4. Lawo, M., Herzog, O., Boronowsky, M., Knackfuss, P.: The Open Wearable Computing Group. IEEE Pervasive Computing 10, 78–81 (2011)
5. Trcek, D.: Trust Management in the Pervasive Computing Era. IEEE Security & Privacy 9, 52–55 (2011)
6. Ahamed, S.I., Haque, M.M., Hoque, E., Rahman, F., Talukder, N.: Design, analysis and deployment of omnipresent Formal Trust Model with trust bootstrapping for pervasive environments. Journal of Systems and Software 83 (February 2010)
7. Buford, J., Kumar, R., Perkins, G.: Composition Trust Bindings in Pervasive Computing Service Composition. In: Fourth Annual IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops 2006, pp. 260–266 (March 2006)
8. Zeng, S., Xin, Y., Yang, Y.-X., Hu, Z.-M.: TM-RMRP: A Trust Model Based on a Recommendation Mechanism of Rewards and Punishments. In: 2010 Second International Workshopon Education Technology and Computer Science (ETCS), pp. 31–34 (March 2010)
9. Liu, Y.L., Guo, S.Z., Tang, Y., Ding, Y.Z.: Trust Propagation Model in Pervasive Computing Environment. Pervasive, 120–123 (September 2010)
10. Ismail, A., Yan, J., Shen, J.: An offer generation approach to SLA negotiation support in service oriented computing. In: Service Oriented Computing and Applications, vol. 4, pp. 277–289. Springer, London (2010)
11. Pelta, D.A., Yager, R.R.: Decision Strategies in Mediated Multiagent Negotiations: An Optimization Approach. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans 40, 635–640 (2010)

12. Chi, Y., Moon, H.J., Hacigumus, H.: iCBS: Incremental Costbased Scheduling under Piecewise Linear SLA. Proceedings of the VLDB Endowment 4 (2011)
13. Bonatti, P.A., De Coi, J.L., Olmedilla, D., Sauro, L.: A Rule-Based Trust Negotiation System. IEEE Transactions on Knowledge and Data Engineering 22, 1507–1520 (2010)
14. Li, L., Song, M., Zhang, X.: An SLA Based Web Service Quality Monitor System. Pervasive, 661–664 (2009)
15. Valarmathi, J., Uthariaraj, D.V.R., Arjun Kumar, G., Subramanian, P., Karthick, R.: A Novel Trust Management Scheme In Pervasive Computing. In: 2010 The 2nd IEEE International Conference on Information Management and Engineering (ICIME), pp. 141–145 (April 2010)

Composing Signatures for Misuse Intrusion Detection System Using Genetic Algorithm in an Offline Environment

Mayank Kumar Goyal and Alok Aggarwal

Dept. of CSE/IT, JIIT University, Noida, UP, India
mayankrkgit@gmail.com, alok.aggarwal@jiit.ac.in

Abstract. In recent years Internet has experienced a rapid expansion and also facing increased no. of security threats. However many technological innovations have been proposed for information assurance but still protection of computer systems has been difficult. With the rapid growth of Internet technology, a high level of security is needed for keeping the data resources and equipments secure. In this context intrusion detection (ID) has become an important area of research since building a system with no vulnerabilities has not been technically feasible.

In this paper, a Genetic Algorithm based approach is presented for network misuse intrusion detection system. The proposed genetic algorithm uses a set of classification rules which are generated from a predefined intrusion behavior. From the results it could be concluded that by applying proposed rule based network intrusion detection algorithm, more no. of intrusions can be detected.

Keywords: Genetic algorithm, misuse intrusion detection, information assurance, data set.

1 Introduction

Computers store, process and retrieve the data. Data is an invaluable asset for every organization, company, enterprise or even for an individual. Availability, integrity and confidentiality are the most important requirements for data handling. Earlier computers were isolated and usually not connected with other computers and does not have a modem. During those days the most common attack to the data stored in computers was the physical use of computer system. Thus in those days security of the room where computer system is placed was enough to secure the data. But with the growth of Internet during the last one decade gave many issues to security of data. Now-a-days computer break-in and misuse has become a common feature.

Intrusion is an activity performed by a person by breaking into an information system or performing an illegal action. Such person is termed as an intruder [1]. Intruders can be classified as external and internal. External intruders are the person

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 151–157.
springerlink.com

who do not have authorized access to the system and who tries to access the system illegally by using different saturation methods. Internal intruders are the persons who have authorized access to the system but carry out the illegal/unauthorized activities. Different methods have been used by intruders whether internal or external for intrusion like password cracking, software bug exploitation, mis-configuration of the system, sniffing unsecured traffic, utilizing the specific protocols design flow etc. No. of such attacks have been increased exponentially during the last one decade.

The two generally accepted categories of intrusion detection are misuse detection and anomaly detection. Former refers to techniques that characterize known methods for penetration into the system which are characterized as a 'pattern' or a 'signature' that the intrusion detection systems look for. These signature or pattern can be a static string or a set sequence of actions. Later refers to techniques that characterize normal or acceptable behaviors of the system like CPU utilization, job execution time, system calls etc. Behaviors that deviate from the expected normal behavior are considered intrusions [2].

Genetic algorithm has been used by many researchers in different types of intrusion detection systems. Genetic algorithms apply biological evolution theory to computer systems [3,4,5]. Genetic algorithm is a method of data analysis and can be termed as analogous to Darwinian evolution [6]. These are research techniques used in computer science and are implemented as a computer simulation and the approximate to combinatorial optimization problems [3].

In this paper, a genetic algorithm based approach is presented for network misuse intrusion detection system. The proposed genetic algorithm uses a set of classification rules which are generated from a predefined intrusion behavior. By applying proposed rule based network intrusion detection algorithm, more no. of intrusions can be detected. This paper is organized as follows. Section 2 describes genetic algorithms, section 3 gives genetic algorithm based intrusion detection. In section 4, different operators of Genetic algorithm used in the proposed algorithm are presented. Section 5 describes the proposed algorithm and experimental assumptions. Finally, results and discussions are given in section 6.

2 Genetic Algorithms

Genetic algorithms are computerized search and optimization methods that work very similar to the principles of natural evolution. Based on Darwin's survival-of-the-fittest principles, GA's intelligent search procedure finds the best and fittest design solutions which are otherwise difficult to find using other techniques. Metaphor from biology is used in genetic algorithms and genetics are used to iteratively evolve a population of initial individuals to a population of high quality individuals [14]. Here each individuals is composed of a fixed number of genes and represents a solution of the problem to be solved. The implementation of Genetic algorithms starts with a population of randomly selected chromosomes. The chromosomes which represent a better solution to target problem are given more opportunities to reproduce in comparison to those provide poorer solutions. The fitness of a solution is typically defined with respect to the current population. These chromosomes are representations of the problem to be solved. According to the attributes of the problem different positions of each chromosome are encoded as numbers. These positions are referred to as genes and are changed randomly within a range during evolution. The set of chromosomes during a stage of evolution are called a population. An evaluation function is used to calculate the fitness of each chromosome.

During the process of evaluation, two basic operators, crossover and mutation, are used to simulate the natural reproduction and mutation of genes. The selection of chromosomes for survival and combination is biased towards the fittest chromosomes. Genetic algorithm begins with a randomly generated population, evolves through selection, crossover, and mutation. Finally, the best chromosome is picked up as the final result. The working principle of a Genetic Algorithm is illustrated in figure1.

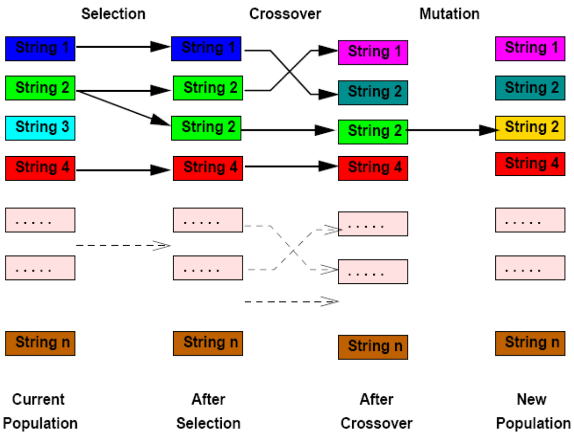


Fig. 1. Working principal of Genetic Algorithm

3 Genetic Algorithm Based Intrusion Detection

The rules stored in the rule base are in the following form:

If (condition) then (do)

The condition usually refers to a match between current network connection and the rules in intrusion detection system, such as source and destination IP addresses and port numbers, protocol, no of bytes of data sent by sender and responder indicating the probability of an intrusion. Several network features have higher possibilities that can be put in network intrusion detection identification. In our approach, six out of those features are taken to compose a classification rule. Table 1 shows the features. The first column in table 1 represents the feature name and second column represents no. of genes.

Table 1. Features and No. of Genes

Feature Name	No. of Genes
Source IP	4
Destination IP	4
Destination port	1
Protocol	1
Sender data amount	1
Responder data amount	1

For example, a rule can be defined as:

if (source-ip=9.9.12.19 and destination-ip=172.16.115.50 and destination-port=79 and protocol=http and sender byte=15000 and responder byte=15000) then (intrusion=attack A).

Each rule is encoded as a chromosome where each network features is encoded using one or more genes of different types.

The final goal of applying genetic algorithm is to generate rules that match only the anomalous connections.

4 Genetic Algorithm Operators

Encoding chromosomes and genetic operators are as follows:

Encoding Chromosomes: Encoding chromosomes in genetic algorithm means represents the set of events to the problem in one string of values.

Binary Encoding: Represent chromosome gene in binary numbers (0's and 1's).

Chromosome M	101010101010101011111
Chromosome N	010101111110000111100

Permutation Encoding: Represent chromosome gene in integer numbers.

Chromosome M	1 3 5 2 4 6 8 9 7
Chromosome N	9 4 1 3 2 7 8 5 6

Fitness Function: Fitness function measures the performance of all chromosomes in the population. To determine the fitness of a rule, the support and confidence framework is used. If a rule is represented as if X then Y, then the fitness of the rule is determined using the function of java genetic algorithm package in java.

Support = |X and Y| / Z

Confidence = |X and Y| / |X|

Here, Z is the total number of network connections in the audit data, |X| stands for the number of network connections matching the condition X, and |X and Y| is the number of network connections that matches the rule if X then Y.

Selection: The application of the fitness criterion to choose which individuals from a population will go on to reproduce.

Crossover: The parent's chromosomes are recombined by one of the crossover methods. It produces one or more new chromosome(s). A crossover operator is used to recombine two strings to get a better string. In crossover operation, recombination process creates different individuals in the successive generations by combining material from two individuals of the previous generation.. It takes two chromosomes and cut their strings at some randomly chosen position and swaps the tail positions.

Mutation: New genetic material is introduced into the new population through mutation process. This will increase the diversity in the population. It is an operator that introduces diversity in the population whenever the population tends to become homogeneous.

```
01101011110110
   |
01111011110110
```

5 Algorithm

Rule set generation using genetic algorithm.

Input : Population size, number of generations
Output : A set of classification rules

1. Initialize the blacklisted classification rules (population)
2. Initialize the population

```
3. N= total no of records in rule set
4. For each chromosome in the population
   4.1 calculate the fitness
5. Select the chromosome into new population
6. For each chromosome in the population
   6.1 Apply a crossover rate of 10 to the chromosome
   6.2 Apply a mutation rate of 1/50 to the chromosome
7. If number of generations is not reached, go to line 4
```

The training process begins by randomly generating an initial population of rules (Step 2). Step 3 calculates the total number of records. Step 4 calculates the fitness of each rule. Step 5 selects the best chromosome. Step 6 applies the crossover and mutation operators to each rule in the new population. Finally, step 7 checks and decides whether to terminate the process or to enter the next generation to continue.

6 Results and Discussions

Table 2 represents the new generated classification rules for signature based intrusion detection system using genetic algorithm which provide if implemented in existing intrusion detection system provide more efficient intrusion detection system. Obtained results are shown in Table 2.

Table 2. Newly generated classification rules for misuse intrusion detection system using genetic algorithm

Chromsome	Evolution time	Fitness value	Source IP	Destination IP	Port no	Protocol	Originator byte	Responder byte
1	1697	2526	125.15.34.137	119.127.190.239	52077	DNS	70141	697519
2	1505	26	125.119.15.58	119.141.38.12	313	SMTP	146869	431444
3	1388	26	125.81.1.146	119.81.43.93	34346	DHCP	59736	42788
4	1380	26	125.51.243.187	119.76.82.40	43980	LDAP	237241	454894
5	1399	26	125.79.240.135	119.129.132.234	51715	FTP	195534	538333

7 Conclusion

A method of genetic algorithm of rule base generation for misuse intrusion detection system is presented in this paper. Experiments have been carried out using a predefined dataset. The major advantage of using a genetic algorithm comes from the fact that in the real world the types of intrusions are dynamic. The proposed system can develop new rules to the systems so as the new intrusions become known. Therefore, it is adaptive and cost effective. Genetic algorithms are potential solutions for optimized rules sets and the determination of potential and actual network intrusions. If only mutation is used, the algorithm is very slow. Crossover makes the algorithm significantly faster.

References

[1] Satya Keerthi, N.V.L.C., Prasanna, P.L., Priscilla, B.M.: Ïntrusion Detection system Using Genetic Algorithm. Int. Journal of P2P Network Trends and Technology 1(2), 1–7 (2011)

[2] Jiang, M., Munawar, M., Reidemeister, T., Ward, P.: Efficient Fault Detection and Diagnosis in Complex Software Systems with Information–Theoretic Monitoring. IEEE Trans. on Dependable and Secure Computing (99) (2011)

[3] Owais, S.S.J., Krömer, P., Snášel, V.: Implementing GP on Optimizing Boolean and Extended Boolean Queries in IRs with Respect to Users Profiles. In: Proc. IEEE ICCES 2006, Egypt, pp. 412–417 (2006)

[4] Owais, S.: Optimization of Boolean Queries in Information Retrieval Systems Using GAs-Genetic Programming and Fuzzy Logic. In: CSIT 2006, Jordan, vol. 2, pp. 303–314 (2006)

[5] Owais, S., Krömer, P., Snašel, V.: Query Optimization by Genetic Algorithms. In: DATESO, pp. 125–137 (2005) ISBN: 80-01-03204-3

[6] Koza, J.: Genetic Programming: On the Programming of Computers by Means of Natural Selection. The MIT Press (1992)

[7] Zhao, J.L., Zhao, J.F., Li, J.J.: Intrusion Detection Based on Clustering Genetic Algorithm. In: Proc. Int. Conf. on Machine Learning and Cybernetics, vol. 6, pp. 3911–3914 (2005)

[8] Diaz-Gomez, P.A., Hougen, D.F.: Three Approaches to Intrusion Detection Analysis and Enhancements. In: Proc. VI National Computer and Information Security Conference ACIS, Colombia (2006)

[9] Li, W.: Using Genetic Algorithm for Network Intrusion Detection. In: Proc. of the United States Department of Energy Cyber Security Group (2004)

[10] Gong, R.H., Zulkernine, M., Abolmaesumi, P.: A software Implementation of a Genetic Algorithm Based Approach to Network Intrusion Detection. In: Proc. Int. Workshop on Self-Assembling Wireless Networks, pp. 246–253 (2005)

[11] Chen, Y., Abraham, A., Yang, B.: Hybrid Flexible Neural-Tree-Based Intrusion Detection Systems. International Journal of Intelligent Systems 22, 337–352 (2007)

[12] Abraham, Grosan, C.: Evolving Intrusion Detection Systems. Studies in Computational Intelligence (SCI) 13, 57–79 (2006)

[13] Sinclair, L.P., Matzner, S.: An Application of Machine Learning to Network Intrusion Detection. In: Proc. 15th Annual Conf. on Computer Security Applications (ACSAC), pp. 371–377 (1999)

[14] Pohlheim, H.: Genetic and Evolutionary Algorithms: Principles, Meth-ods and Algorithms, http://www.geatbx.com/docu/index.html

SRSnF: A Strategy for Secured Routing in Spray and Focus Routing Protocol for DTN

Sujoy Saha[1], Rohit Verma[1], Satadal Sengupta[1], Vineet Mishra[2], and Subrata Nandi[1]

[1] Dept. of Computer Science & Engineering, National Institute of Technology, Durgapur
[2] Dept. of Computer Applications, National Institute of Technology, Durgapur,
India–713209
{sujoy.ju,satadal.sengupta.nit,rohitverma.kgp,vineet0309,
subrata.nandi}@gmail.com

Abstract. This paper deals with the aspect of security in Delay Tolerant Networks (DTN). DTNs are characterized with decentralized control. Network performance and trustworthiness of transmitted information in DTNs depend upon the level of co-operation among participating nodes. As a result, DTNs are vulnerable towards untoward activities arising out of node selfishness as well as malicious intentions. In this paper, we limit our focus to the *Black Hole Denial-of-Service* attack. We develop a table-based strategy to record network history and use this information to detect discrepancies in the behavior of nodes, followed by elimination of those detected as malicious. We explain our detection mechanism considering *Spray & Focus* routing protocol as the representative routing scheme. The detection mechanism has been described in detail with examples pertaining to various case scenarios. Furthermore, we study the effect of variation of various parameters on detection efficiency and message transmission through simulation results.

1 Introduction

Wireless ad hoc networks are able to perform message transmission without fixed network infrastructure. However, in practice, ad hoc routing protocols do not work efficiently due to high node mobility, low node density, and short radio ranges. To deal with such loopholes, the concept of Delay Tolerant Network [1] (DTN) was introduced. These types of networks use store-and-forward approach to deal with challenging networking scenarios, e.g., battlefield and deep-space communication.

In DTNs, the packet is stored in the buffer of a node if it does not find the next-hop node en route to the destination. The metric-based routing protocols in DTN, e.g. Spray & Focus [2], PRoPHET [3], MaxProp [4], firstly store the packet in memory; then transmit it to particular nodes based on some delivery metric, such as some utility value (determined using some pre-defined utility function).

However, in DTN, it is a challenging task to verify the integrity and trustworthiness of transmitted information, mainly due to network constraints like low power availability, low node density, etc. Therefore, DTNs are vulnerable to adversary attacks and internal compromises, taking advantage of which, insider attacks may be launched. Insider attacks can cause significant disturbances.

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 159–169.
springerlink.com

In this paper, we propose a scheme which uses *Spray & Focus* [2] routing protocol. Spray & Focus exhibits better performance that other utility-based mechanisms in most practical applications, which is the reason for it being our choice. The strategy is to check each node by its peer node for maliciousness. Detection is based on entries in tables maintained at each node which get verified and updated during each encounter between 2 nodes. Each node is capable of such integrity verification. Like in *Spray & Focus*, utility values have been calculated using transitivity too. Detailed description of strategy followed by comprehensive evaluation through simulations using ONE [5] has been carried out.

The paper has been divided into 5 sections. Section-2 gives a brief description of related work in maliciousness detection. Section-3 describes in detail the proposed strategy and incorporates case scenarios and algorithms. In section-4, we evaluate performance of our strategy by presenting relevant simulation results. Finally, in section-5, we conclude and point out areas for further research.

2 Related Work

A large amount of work has been done in the field of security in ad hoc wireless networks. A large number of threats that have been identified in case of ad hoc networks are applicable in some form or another in DTNs as well; however, due to non-availability of an end-to-end path in case of DTNs, those either fail completely, or perform miserably in a DTN scenario. Therefore, security approaches which are tailored to suit DTN characteristics are required. Strategies that have been proposed till date can be broadly classified into 3 categories: *reactive approach* which tries to mitigate the effects of malicious activities by identifying their source and taking prohibitory action, *proactive* approach which provides incentives to co-operative nodes thus encouraging participating nodes to forward messages whenever possible, and *user-interest oriented approach* where rational behavior of users based on personal interests or social ties are considered as design constraints.

The schemes that follow a *reactive approach* are as follows. Chuah et. al. proposed a ferry based detection scheme (FBIDM) [6]. In this scheme, detection is carried out by trusted examiners, i.e., ferry nodes. This strategy was improved upon by Ren et. al. by considering the property of transitivity (shown to be an important attribute for calculating delivery probability) in their paper called MUTON [7]. MUTON, like FBIDM, also used ferries to detect malicious nodes; MUTON was shown to perform much better. Meanwhile, Li et. al. proposed a strategy with encounter tickets [8] to thwart blackhole attacks. The scheme required forwarding nodes to generate signed encounter tickets to build an encounter history, which was later interpreted appropriately to make forwarding decisions. Ren et. al. came up with a packet exchange recording scheme [9] later, which required each node to maintain 2 tables for storing records of packet exchange; each node was taken through a sanity check at every encounter, failing which the node was blacklisted.

In the *proactive approach*, Zhu et. al. proposed SMART [10], a secure, multi-layered coin based approach which provided incentives to forwarding nodes if a packet relayed by them reached the destination successfully. However, forwarding nodes were deprived of credit when successful delivery failed due to faulty action of another node or expiration of time-to-live. This problem was tackled by Lu et. al. in

Pi [11], which, in addition to providing credit for successful delivery, ensured a boost in reputation for each forwarding action by a node.

Rational nature of users is given priority in the *user-interest oriented approach*. Li et. al. proposed socially selfish routing [12] which ensured that nodes made forwarding decisions based on strength of their social ties with other nodes. Ning et. al. proposed an incentive aware DTN [13] where user interest, e.g. news, sports, or entertainment, was taken into consideration to ensure user satisfaction.

In [14], we present a concise description of the work done so far by us on this field and results achieved, as well as indications for future study.

3 SRSnF: Secured Routing for *Spray & Focus* Protocol

We propose a table based scheme for ensuring secured routing using *Spray & Focus* routing protocol. Our scheme follows the *reactive approach* described in the previous section. In our scheme, each node maintains a table with entries corresponding to each encounter in the network; this is called *Network Records Table* (NRT). Additionally, each node stores a *Malice Identification List* (MIL) with node IDs of detected nodes in the network. Whenever 2 nodes (say A & B) encounter each other, they perform the following actions, in the specified order:

1. Node A checks its *Malice Identification List* for node ID of B. Node B performs the same *MIL* check for A. If any of the two tests positive, no further interaction takes place. If both test negative, they move on to the next step.
2. Node A and node B exchange their respective *Network Records Tables*.
3. Node A performs a *Maliciousness Test* on node B by matching its own NRT entry by entry with that of node B. B performs the same test on A.
4. If one of the nodes (say, node A) finds the other node (node B) to have engaged in malicious behavior, then A appends the node ID of B to its MIL. No further interaction takes place under this circumstance.
5. If each node passes the *Maliciousness Test* performed by the other node, then they update their *Network Record Tables* by inserting entries corresponding to latest information about the network. So, entries in NRT of A which are not present in NRT of B, are inserted into NRT of B and vice-versa.
6. Also, the nodes exchange and update their MIL in the same way that they do in case of the NRTs in order to know about all detected nodes.
7. Finally, they will exchange messages according to *Spray & Focus* protocol; then generate and exchange corresponding entries to be stored in their NRTs.

For our scheme, we shall consider the *Black Hole Denial-of-Service* (DoS) attack, i.e., malicious nodes will drop all packets they receive. Each such node will participate in routing of the message but are not going to forward the packet any further. Our strategy consists of 3 phases: 1st phase states pre-requisites for the network; next 2 phases describe detection in the *spray* and *focus* phases of *Spray & Focus*.

A. Authentication Phase: Before routing starts in the network, it is essential for each participating node to be aware about all other participating nodes; otherwise, external intrusion would become rampant. There is an Authentication Authority (AA) to ensure this. It assigns a unique node ID and a pair of public and private keys to each participating node. Also, AA ensures that each node is preloaded with the list of node IDs of all participating nodes, and corresponding public keys. We assume that either this list doesn't change throughout the working of the DTN, or AA is able to modify suitably settings of all nodes in case of changes.

Whenever a node generates an entry corresponding to an encounter with another node, it is signed and encrypted using its pair of public & private keys. Such an entry can be decrypted by another authenticated node using its set of keys. Securing information using such techniques is a different field altogether and we do not discuss such techniques in this paper; rather we state that by using such a technique, we can ensure that modification of the contents of any entry is rendered impossible, thus securing the network from information forging attacks.

Furthermore, let us assume that AA generates a unique sink node ID, which is never assigned to any participating node, but is included in the list loaded into all nodes. Whenever a node is forced to purge a message from its memory due to buffer constraints, it generates an entry with this sink node ID as the receiver node. During Maliciousness Test, the checking node counts all such entries and includes this figure in its evaluation to ensure that there is no false report.

B. Spray Maliciousness Detection Phase: In spray phase, maliciousness can be shown by accepting packets from neighboring nodes, then dropping them instantly. Let such behavior be known as Spray Maliciousness; nodes exhibiting it be known as spray malicious nodes. Let us consider an example with these events:

1. E_0: Node A generates 3 copies of a message with ID 7002#.
2. E_1: Node A encounters node B and transfers 1 token for the message to it.
3. E_2: Node A encounters node C and transfers 1 token for the message to it.

During E_1, A & B update their NRTs as in Tab.1. During E_2, A & C update their NRTs as in Tab.2. Note that only the sections of NRTs relevant to the Spray Phase have been shown in Tabs.1,2,3. Let us consider following possible events after E_2:

CASE-1: ***Node B encounters A or C before encountering other node***- If node B is malicious, it will drop message after E_1. Now it does not contain any copy of the message, although its NRT has an entry showing reception from A. Whenever B comes in contact with A or C and poses as a relay node, its NRT will be checked. It will be found that B has an entry from A but no message to show for it; hence B is malicious, and is blacklisted. If B is regular, no anomaly would be found.

CASE-2: ***Node B encounters node D before encountering A or C-*** If node B is regular, and D is a better relay node than B itself, it will pass its message copy on to D. Then B and D will update their NRTs as shown in Table 3. So, when A or C encounters it at a later stage, it will know from B's NRT that no malicious activity has been exhibited. If B is malicious, node D will be able to ascertain that B had received a message copy from A but no longer has it; hence D will blacklist it.

Table 1. Spray Phase section of Network Record Table with node A and node B after E_1

Sender	Receiver	Message ID	Time to live	Copies with sender	Copies with receiver
A	B	7002#	5	2	1

Table 2. Spray Phase section of Network Record Table with node A and node C after E_2

Sender	Receiver	Message ID	Time to live	Copies with sender	Copies with receiver
A	B	7002#	5	2	1
A	C	7002#	5	1	1

Table 3. Spray Phase section of NRT with node B & D if they exchange copy

Sender	Receiver	Message ID	Time to live	Copies with sender	Copies with receiver
A	B	7002#	5	2	1
B	D	7002#	5	0	1

Algorithm 1. Spray Phase Detection Algorithm

```
//Key: Checker = checking node, peer = checked node
BEGIN
If checker has encountered peer previously, then
              If time_to_live has not expired, then
                 If(no.of copies received - no.of copies delivered -
                 no.of entries to sink node) <> 0
                    If lack of sufficient contact time is not the rea-
                    son for message drop/s, then
                      Peer is malicious, append peer to MIL;
                    EndIf
                 EndIf
              EndIf
Else
              Follow Spray phase mechanism;
EndIf
END
```

C. Focus Maliciousness Detection Phase: In Spray & Focus, a utility value $\tau_i(j)$ is defined for each pair of nodes, which indicates the prob. of node i to deliver the message to node j. When a node has only 1 copy of the message left, it passes it on to a node with better utility for destination. We assign the utility values for each pair of different nodes initially as infinity. $\tau_i(j)$ increases by 1 at every clock tick.

Let us consider 2 nodes A and B. At the beginning, $\tau_A(B) = \infty$ and $\tau_B(A) = \infty$. When they encounter $\tau_A(B) = \tau_B(A) = 0$. As soon as the connection is lost, this value starts increasing by 1 at every clock tick. It is also updated by transitivity:

$$\text{If } (\tau_B(C)) < (\tau_A(C) - t_m(d_{AB})), \text{ then } \ \tau_A(C) = \tau_B(C) + t_m(d_{AB}) \qquad (1)$$

where $t_m(d_{AB})$ is time to cover distance AB under given mobility model m. This can be evaluated through calculation of velocity of node movement using traces.

Maliciousness can be shown by nodes by declaring fake lesser utility values (less being better in our example) for the destination. Let such behavior be known as *Focus Maliciousness*; nodes behaving in such fashion be known as *focus malicious nodes*. If successful, such nodes will receive a large fraction of packets from regular nodes, and drop them. In our strategy, a record for every change in utility value of every node is stored. An encountering node can determine the actual utility value of the node from these records. If the declared utility value is less than the calculated one, then it is tagged as malicious by the encountering node.

Let us take an example and calculate utility values of nodes for every other node. Let there be 4 nodes - A, B, C, D. Let *starting time* denote time when first contact is disconnected. $t_m(d_{ij})$ is assumed as 8 throughout. Let the events be:

1. Connection between A and B is lost at the *starting time*.
2. After 15 time units connection between A and C is lost.
3. After 10 time units connection between C and D is lost.
4. After 20 time units connection between A and D is lost.
5. After 5 time units connection between B and D is lost.
6. After 7 times units connection between B and C is lost.

Final status of NRTs relevant to *Focus* will be as in Tables 4, 5, 6 respectively.

Table 4. Focus Phase section of Network Record Table with node A

Sender	Receiver	Time Stamp	Low Utility Value	High Utility Value
A	B	00	--	--
A	C	15	$\tau_A(B)$ / 15	$\tau_C(B)$ / ∞
C	D	25	$\tau_C(A)$ / 10, $\tau_C(B)$ / ∞	$\tau_D(A)$ / ∞, $\tau_D(B)$ / ∞
A	D	45	$\tau_D(C)$ / 20, $\tau_A(B)$ / 45	$\tau_A(C)$ / 30, $\tau_D(B)$ / 61

Table 5. Focus Phase section of Network Record Table with nodes B and C

Sender	Receiver	Time Stamp	Low Utility Value	High Utility Value
A	B	00	--	--
A	C	15	$\tau_A(B)$ / 15	$\tau_C(B)$ / ∞
C	D	25	$\tau_C(A)$ / 10, $\tau_C(B)$ / ∞	$\tau_D(A)$ / ∞, $\tau_D(B)$ / ∞
A	D	45	$\tau_D(C)$ / 20, $\tau_A(B)$ / 45	$\tau_A(C)$ / 30, $\tau_D(B)$ / 61
B	D	50	$\tau_D(A)$ / 5, $\tau_D(C)$ / 20	$\tau_B(A)$ / 50, $\tau_B(C)$ / ∞
B	C	57	$\tau_B(D)$ / 7, $\tau_B(A)$ / 20	$\tau_C(D)$ / 32, $\tau_C(A)$ / 65

Table 6. Focus Phase section of Network Record Table with node D

Sender	Receiver	Time Stamp	Low Utility Value	High Utility Value
A	B	00	--	--
A	C	15	$\tau_A(B)$ / 15	$\tau_C(B)$ / ∞
C	D	25	$\tau_C(A)$ / 10, $\tau_C(B)$ / ∞	$\tau_D(A)$ / ∞, $\tau_D(B)$ / ∞
A	D	45	$\tau_D(C)$ / 20, $\tau_A(B)$ / 45	$\tau_A(C)$ / 30, $\tau_D(B)$ / 61
B	D	50	$\tau_D(A)$ / 5, $\tau_D(C)$ / 20	$\tau_B(A)$ / 50, $\tau_B(C)$ / ∞

If the node is malicious, it will try to come in path of the message delivery route by falsely declaring a lesser utility value. Now, let us assume that A and B encounter each other after 7 more time units, and B declares its utility value for D (where D is destination) as 5. Now, A and B follow the *focus detection algorithm.*

CASE-1: ***Detection of node B if it is malicious-*** From the point where contact of node B and node D is searched node A will check downwards in the table and check whether $\tau_B(D)$ has changed by transitivity or not. As we can see that it has not been changed, therefore at present its value should be 14 (7 when B and C meet after 7 sec, another 7 when we are assuming that A and B meet). Now the declared value is 5, calculated value is 14, hence B is malicious and A blacklists it.

CASE-2: ***Detection of C if it is malicious-*** D & C meet each other after 10secs of B & C meeting (at time 67). Let declared value of $\tau_C(A)$ be 26. D will search last encounter of A & C, then it moves downward and checks for any change in $\tau_C(A)$. It's evident that when B & C meet at 57, $\tau_C(A)$ changes from 65 to 28 and thus at 67, its value should be 38, which is greater than declared value; C is detected.

CASE-3: A case may arise where the regular node has a less updated NRT than a peer malicious node, which is not sufficient for detection. Even in that case, since nodes update NRTs at every encounter, it won't long before an encounter with a sufficiently updated regular node takes place, and the malicious node is detected.

Algorithm 2. Focus Phase Detection Algorithm

```
//Key: checker = checking node, peer = checked node
BEGIN
If checker & peer NRT entries match till last entry, then
                If peer has met destination before, then
                    If $\tau_{peer}$(destination) has changed by transitivity, then
                      Calculate actual value of $\tau_{peer}$(destination) using up-
                      date through transitivity values;
                    Else
                      Calculate actual value of $\tau_{peer}$(destination) using
                      time difference from last encounter;
                    EndIf
                EndIf
                If actual value of $\tau_{peer}$(destination) & declared value of
                $\tau_{peer}$(destination) do not match, then
                      Peer is malicious, append peer to MIL;
                Else
                      Follow Focus phase mechanism;
                EndIf
EndIf
END
```

4 Evaluation through Simulation

A. Simulation Setup & Methodology: We carried out relevant simulations using *ONE simulator*[5]; it was customized to impart malicious behaviour to randomly chosen nodes. Modifications were made to *Spray & Focus* code. Setup in Tab.7.

Table 7. Simulation Setup

Parameter	Value	Parameter	Value	Parameter	Value
Simulation Area	4500x3000	Packet Size	500KB – 1MB	Message TTL	480min
No. of nodes	100	Buffer Size	512 MB	No. of simulations	15/case
Mobility Model	Shortest Path Map Based	Interfaces (Range,Speed)	Bluetooth (10m,2Mbps), High-speed (1km,40Mbps)	Msg Generation Interval	25–35 secs
Simulation Time	12 hours	Node Speed	0.5m/s – 1.5m/s	Copies/msg	3

The following scenarios were considered: (1) Percentage of malicious nodes was varied among 10%, 20%, 30% & 40% of the total no. of nodes; (2) Total no. of nodes in the network was varied among 100, 90, 80, 70 & 60. In (1), we study the effect of varying percentage of malicious nodes on detection time & delivery prob. In (2), we perform comparative study between metrics obtained using (n,p) nodes & (n-p,0) nodes where, (x,y) = (total no. of nodes, no. of malicious nodes). In each simulation, an appropriate no. of nodes have been chosen randomly and made to behave maliciously: 50% as *spray malicious*, rest as *focus malicious*.

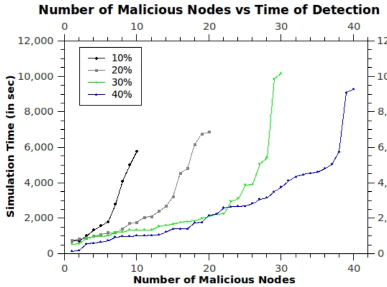


Fig. 1. Time v/s Average no. of detections

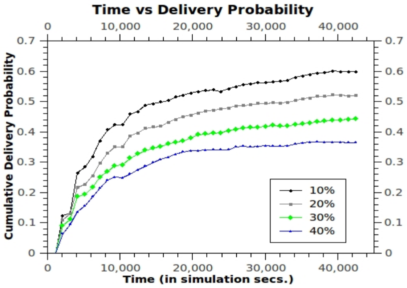


Fig. 2. Cumulative del. prob. v/s Time

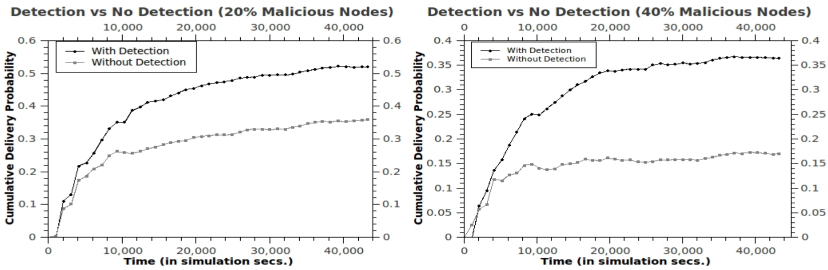


Fig. 3(a,b). Improvement in del. prob. in 20% & 40% cases using our proposed mechanism

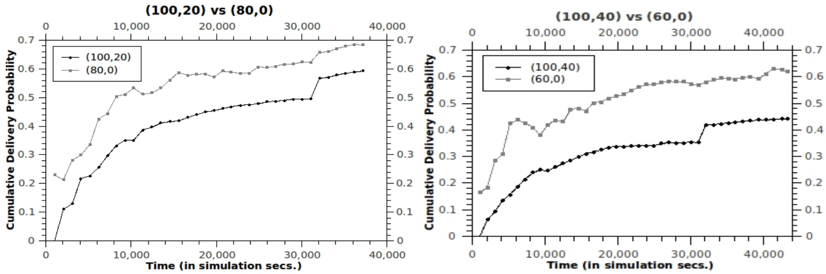


Fig. 4(a,b). Comparison between maliciousness inclusive & exclusive scenarios

B. Metrics Observed: The metrics used to evaluate performance and efficiency of the proposed detection strategy are: (1) *average number of detections* - it gives the avg. no. of detections made successfully at certain time intervals, thus showing how quickly our strategy is able to detect and eliminate malicious nodes; (2) *total delivery probability* - it gives ratio between total no. of messages delivered during entire simulation time & total no. of messages created during entire simulation time; (3)

cumulative delivery probability - it gives ratio between no. of messages delivered & no. of messages created till present simulation instance thus enabling us to study the pattern in which delivery ratio increases or decreases upon detection and elimination of more and more malicious nodes from the network.

C. Simulation Results and Analysis: The simulation results are as follows:

1. Efficiency of the proposed Detection Mechanism: Fig.1 presents the graph obtained by plotting *average number of detections* against time in seconds. We have plotted values obtained against all 4 cases of scenario (1). From the graph, we can conclude that the detection algorithm works effectively as it detects all malicious nodes in the network within satisfactory time. As expected, more the no. of malicious nodes in the network, more is the time required to detect and eliminate all of them. It is also noteworthy that detections occur very quickly when the simulation time is at lesser values, whereas as higher values of time,detection slows down. This can explained in the following way initially, node movement is less and the number of messages in the network is also quite meager; therefore it is easier to detect anomalies within NRTs due to lesser number of entries. On the contrary, after an appreciable period of time, as node mobility and message transmissions increase, it becomes more difficult to acquire sufficient knowledge about the entire network to detect anomalies.

2. Impact on delivery probability: Fig.2 depicts a comparative picture of delivery probabilities obtained in all 4 scenarios. Fig.3 consists of 2 different graphs, each representing the efficiency of our proposed strategy in improving delivery prob. The impact is most profound when malicious activity is rampant. If we look at any individual curve, we shall notice that it consists of 3 phases: phase-1 is when del. prob. is significantly low due to setback from malicious activities; phase-2 is when the network gradually recovers from the effect and shows improvement; phase-3 is when the nework is in a steady state and shows increase in del. prob. at a decreasing rate until it almost becomes parallel to horizontal axis.

3. Comparison between Maliciousness Inclusive & Exclusive Scenarios: Fig.4 deals with yet another absorbing aspect of performance of our proposed strategy. It depicts comparative curves between *(n,0)* & *(n-p,0)* scenarios. Such a study compares between situations when a network is simulated with *(n-p)* nodes all regular, against when a network is simulated with *n* nodes which effectively comes down to *(n-p)* after the *p* malicious nodes have been detected & eliminated. The curves show that our scheme achieves fairly close delivery probs. compared to normal scenario; difference increasing as density of malicious nodes increases.

4. Limitations: Having discussed the advantages of our detection scheme in terms of performance, let's take a look at the limitations and drawbacks involved. As can be observed from Fig.5, the overhead ratio incurred in case of detection using our detection mechanism is significantly higher compared to that incurred when the network runs with malicious activities but without detection. This is due to the extra computation time required during each encounter in the network.

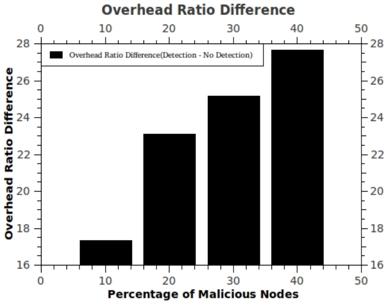


Fig. 5. Increase in overhead involved

5 Conclusion and Future Work

In this paper, we have taken a look at the proposed schemes which aim at combating various malicious attacks in a DTN environment. We have proposed a strategy for secured routing in the *Spray & Focus* routing scheme, which uses a table-based approach for detection of malicious nodes in the network. The intricacies of the mechanism have been described in detail. Simulations show us that the proposed scheme is effective in detecting malicious nodes and ensuring that those take no further part in the routing process. However, simulations also indicate significant increase in average latency and overhead involved. In order to curb such ill-effects, we are investigating a more balanced approach that allows only certain trusted checkers to perform *Maliciousness Tests*.

Other than the nature of attack considered and dealt with in this paper, there are a variety of attacks possible in a delay tolerant environment. Although a large amount of work has been carried out with regards to security in wireless ad-hoc networks, much remains to be done to combat similar and dissimilar security threats to disruption tolerant networks. Development of strategies to deal with a multitude of other attacks can be fodder for future research.

References

[1] Fall, K.: A Delay Tolerant Network Architecture for Challenged Internets. In: Proc. ACM SIGCOMM, pp. 27–34 (2003)

[2] Spyropoulos, T., et al.: Spray & Focus: Efficient Mobility-Assisted Routing For Heterogeneous & Correlated Mobility. In: Proc. Fifth IEEE PERCOM Workshops 2007 (2007)

[3] Lindgreny: Probabilistic Routing in Intermittently Connected Networks. In: Proc. ACM SIGMOBILE Mobile Computing & Communications Review, vol. 7 (July 2003)

[4] Burgess, J., et al.: Maxprop: Routing for vehicle-based disruption-tolerant networking. In: Proc. INFOCOM (April 2006)

[5] Kernen, et al.: The ONE Simulator for DTN Protocol Evaluation. In: Proc. of the 2nd Int‘l Conf. on Simulation Tools & Techniques, Simutools 2009, Belgium (2009)

[6] Chuah, M., et al.: A Ferry based Intrusion Detection Scheme for Sparsely Connected Ad Hoc Networks. In: Proc. MOBIQUITOUS (August 2007)

[7] Ren, Y., et al.: MUTON: Detecting Malicious Nodes in Disruption-Tolerant Networks. In: Proc. IEEE WCNC (2010)

[8] Li, F., et al.: Thwarting Blackhole Attacks in Distruption-Tolerant Networks using Encounter Tickets. In: Proc. INFOCOM (2009)

[9] Ren, Y., et al.: Detecting Blackhole Attacks in Disruption-Tolerant Networks through Packet Exchange Recording. In: Proc. IEEE WoWMoM (2010)

[10] Zhu, H., et al.: SMART: A Secure Multilayer Credit-Based Incentive Scheme for Delay-Tolerant Networks. IEEE Trans. on Vehicular Tech. 58 (October 2009)

[11] Lu, R., et al.: Pi: A Practical Incentive Protocol for Delay Tolerant Networks. IEEE Trans. on Wireless Communications 9(4) (April 2010)

[12] Li, Q., et al.: Routing in Socially Selfish Delay Tolerant Networks. In: Proc. of INFOCOM, pp. 857–865 (2010)

[13] Ning, T., et al.: Incentive-Aware Data Dissemination in Delay-Tolerant Mobile Networks. In: Proc. of SECON (2011)

[14] Saha, S., et al.: Secured Routing in DTNs: Threats & Counter-measures. In: Ph.D. Forum, ICDCN (2011)

Multi Tree View of Complex Attack – Stuxnet

Shivani Mishra[1], Krishna Kant[2], and R.S. Yadav[3]

[1] Research Scholar, CSED, Motilal Nehru National Institute of Technology,
Allahabad, Uttar Pradesh, India
[2] Professor, Computer Engineering and Applications Department, GLA
University, Mathura, India
[3] Professor, CSED, Motilal Nehru National Institute of Technology,
Allahabad, Uttar Pradesh, India
shivaniallld@gmail.com, krishna.kant@gla.ac.in, rsy@mnnit.ac.in

Abstract. Stuxnet attack on critical infrastructures is considered as paradigm shift in malware attack approach. The complexity and sophistication involved in this attack make it unique. Attacking approach of the malware, on control infrastructures, is a motivation for academic research. This paper describes the application of the Attack Tree methodology to analyze Stuxnet attack on SCADA system. Root node of the Attack Tree represents the major goal of an attacker and branches represent sub goals. The authors have identified six major goals to penetrate SCADA system, and then have built Attack Trees which demonstrate step by step activity to achieve these goals and sub goals. For each such sub goal, we have found several common categories of attacks which make Stuxnet attack successful and are used to analyze those components of control infrastructure which are susceptible to attacks.

Keywords: Malware, Stuxnet, SCADA , Control infrastructures, Attack Trees, Attack Goal, Attack Sub Goal.

1 Introduction

Nowadays, sustained cyber attacks against critical infrastructure have been escalated to an alarming situation. They are causing havocs to digital installations with a very sophisticated manner and enormous frequency. A complex computer worm, discovered in June 2010, effectively disabled Iran's nuclear program for more than a year. It happened even when their nuclear facilities were highly secured, located underground physically and electromagnetically isolated from insecure networks known as Air Gapped from the Internet (AGI) [1].

The recent findings, as documented in reports published by Symantec [2], ICS-CERT [3], and Eset [4] indicate that worm was propagated to ins and outs of the facility from universal serial bus, USB, using thumb drive technology through AGI. The AGI was used as via media to allow worm to penetrate the SCADA system. The highly complex computer worm called Stuxnet was designed to spread until it found specific control system as its target. Investigating experts have opined that the worm, so used is the first weaponized virus [5] [6].

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 171–188.
springerlink.com

In the present research paper, our focus is to analyze the notorious Stuxnet worm (WIN 32/Stuxnet) attack on SCADA system using the Attack Tree approach. Bruce Schenier is the one and the first researcher who introduced Attack Tree modeling to analyze attacks [7]. Early approaches to this problem through the Attack Tree modeling heavily relied upon

1. Categorization of attack sequences and modeling the path traced by the attacker to exploit the system [8].
2. Assuming all elementary attacks took place simultaneously, some researchers adopted parallel modeling, which consist of attack models parallel to the actual incidents so assumed [9].

In the context we assume, ab nitio, that the attacker had different attack alternatives, if some initiations fail to succeed, at least one would definitely succeed. In our work we propose detection oriented security modeling approach using Attack Trees for Stuxnet attack on SCADA system. In this track of security modeling evaluation process we first determine,

1. Goals and Sub Goals of an Attacker to penetrate the system
2. Vulnerabilities in the system
3. Common Category of attacks which enable Stuxnet to execute
4. Resources exploited by these attacks

The organization of the paper is as follows. Section 1 is the Introduction to Stuxnet attack scenario in critical SCADA systems. Section 2 gives motivational background for the present work. A brief review of the relevant work is made in Section 3. Section 4 presents the proposed work – Attack Tree structure for Stuxnet malware attack. Conclusions are drawn in Section 5.

2 Background

2.1 Introduction of SCADA Systems

The systems that control critical infrastructures are Industrial Control Systems (ICS). There are several types of control systems in it including Supervisory Control and Data Acquisition (SCADA) systems, Distributed Control Systems (DCS), and other smaller control systems such as Programmable Logic Controllers (PLC). Industries associated with critical infrastructure include power generation and distribution, oil and gasoline refining and distribution, water and waste systems, manufacturing, telecommunications, and banking infrastructures.

SCADA systems, an architectural block diagram of which is shown in Fig 1, consist of:

- One or more field data interface devices, usually Remote Terminal Unit (RTU), or PLCs, which interface to field sensing devices and local control switchboxes and valve actuators.
- A communications system used to transfer data between field data interface device and control units and the computers in the SCADA central host. The system can be radio, telephone, cable, satellite, etc., or any combination of these.

- A central host computer server or servers (Also known as SCADA Center, master station, or Master Terminal Unit (MTU)
- A collection of standard and/or custom software (Known as Human Machine Interface (HMI) software or Man Machine Interface (MMI) software) systems used to provide the SCADA central host and operator terminal application, support the communications system, and monitor and control remotely located field data interface devices

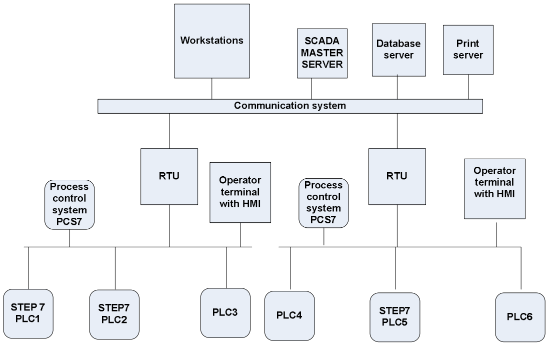


Fig. 1. Typical SCADA architecture

Software products typically used within a SCADA system are for the following purpose:

- **Central host computer operating system**: Software used to control the central host computer hardware. The software can be based on Windows, UNIX or other popular operating systems.
- **Operator terminal operating system**: Software used to control the central host computer hardware. The software is usually the same as the central host computer operating system. This software, along with that for the central host computer, contributes to the networking of the central host and the operator terminals.
- **Central host computer application**: Software that handles the transmission and reception of data to and from the RTUs and the central host. The software also provides the graphical user interface which offers site mimic screens, alarm pages, trend pages, and control functions.
- **Operator terminal application**: Application that enables users to access information available on the central host computer application. It is usually a subset of the software used on the central host computers.
- **Communications protocol drivers**: Software that is usually based within the central host and the RTUs, and is required to control the translation and interpretation of the data between ends of the communications links in the system. The protocol drivers prepare the data for use either at the field devices or the central host end of the system.

- **Communications network management software**: Software required to control the communications network and to allow them to be monitored for performance and failures.
- **RTU automation software**: Software that allows engineering staff to configure and maintain the application housed within the RTUs (or PLCs). Most often this includes the local automation application and data processing tasks that are performed within the RTU [10].

2.2 Stuxnet Attack on SCADA

Stuxnet is a computer worm, discovered in June 2010. It targets Siemens industrial software and equipment running on Microsoft Windows. It is the first discovered malware that spies on and subverts industrial systems, and the first to include a PLC root kit. The worm initially spreads indiscriminately, but includes a highly specialized malware payload that is designed to target only Siemens SCADA systems that are configured to control and monitor specific industrial processes. Stuxnet infects PLCs by subverting the Step-7 software application (SIMATIC Step7 is an integral component of the centralized Totally Integrated Automation Portal engineering framework) that is used to reprogram these devices [11].

Stuxnet worm is so intelligent that it makes itself neutral if Siemens software is not found on infected computers, Stuxnet contains code for a man-in-the-middle (MITM) attack through hooking that jukes industrial process control signals so as the infected system does not alarm due to this abnormal behavior. Such complexity is very unusual for malware. The attacking strategy of this malware can be divided in two parts, they are as follows-

1. Gain access on Windows operating system,
2. Disrupt Siemens PCS 7, WinCC and STEP7 industrial software applications that run on Windows.

Windows infection

Stuxnet attack occurs on Windows systems using four zero-day attacks. A zero-day attack is exploitation of computer system software vulnerabilities that are not known to others. Initially it spreads using infected removable drives such as USB flash drives, and then uses other exploits and techniques such as peer-to-peer RPC to infect and update other computers inside private networks that are not directly connected to the Internet.

SIMATIC WinCC is a supervisory control and data acquisition system from Siemens. It can be used in combination with Siemens PCS 7 and Teleperm, a process control system. WinCC is written for Microsoft Windows operating system. WinCC uses Microsoft SQL Server for logging and comes with a VBScript and ANSI C application programming interface. WinCC and PCS 7 are the first SCADA systems which were targeted by malware to subvert control system [12].

3 Related Work

The complex infrastructure of SCADA systems provides great capabilities for operation, control and availability of resources, simultaneously it also increases security risk due to cyber related vulnerabilities. Disruption, information leakage, malfunctioning of system process can cause havoc in the system which in turn causes negative financial impact for the system. Because of critical nature of these systems, they are now targets for adversaries.

Digital security for critical infrastructure requires extensive research methodology for securing control system rather than focusing only on securing IT system associated with SCADA systems. In [13], Manimanran et al. (2010) has proposed SCADA security framework with the following four major components. They are Real time monitoring, Anomaly detection, Impact Analysis and Mitigation strategies.

The Attack tree modeling is used for impact analysis. System-, scenario-, and leaf-level vulnerabilities are discovered by identifying system's adversary objectives. The leaf vulnerability involves port auditing or password strength evaluation which is a measure of intrusion in a system. The approach used in this paper is extensive and can be applied for evaluation of different parameters of security like confidentiality, integrity, availability and authentication so that possibility of attack related to parameters such as eavesdropping, modification, malware attacks, like, virus, worm, Trojan horse, bypassing controls can be measured and countermeasures are predetermined.Malware attacks like Stuxnet is considered as paradigm shift in critical infrastructure threats. Unlike most malwares, Stuxnet has great capabilities to attack in depth for controlling physical machinery. At the same time the sophistication is found in Stuxnet programming which is considered as "a new class and dimension of malware" [14] [15]. The worm's code was written in multiple programming languages. Research has also speculated that Stuxnet was professionally developed and would have required access to SCADA hardware for testing. Stuxnet employs multi vector approach by exploiting four zero-day vulnerabilities, out of which two were privilege escalation, and remaining were printer spooler flaw & USB flash drive [16]. Stuxnet is considered as epochal because of its ability to infiltrate networks, find control system of supervisory and data acquisition industry and reprogram the hardware control system. A nuclear facility was attacked by Stuxnet [17].

The sophistication and capability to penetrate the system invisibly make the Stuxnet worm attack as complex attack. A complex attack can be described as

Multi Agent: For Stuxnet Attack, threat agent can be one or many who want to sabotage the facility of SCADA system, at the same time threat agent can take one or more actions for the successful completion of an attack.

Multi Phase: Stuxnet attack has several attack stage. Specifically these stages are exploits at different junctures.

Multi Pace: Stuxnet choice of target is very specific. For successful attack it involves multiple *intrusion* activities in a SCADA system to reach the intended target.

An Attack Tree approach to analyze Stuxnet worm activity inside SCADA systems is the main contribution of our work. In the next section a brief description of Attack Tree and Stuxnet attack approach visualization through Attack Tree is mentioned

4 Attack Tree Approach for Problem Solving

Attack Trees were introduced by B. Schneier, as a way of formally analyzing the security of systems. Since this methodology is efficient to model the behavior of an attacker while attacking on a system, security researchers are gaining interest to use this technique as a tool for evaluation of different aspects of security[18] [19].

Basically an Attack Tree approach is systematic categorization of different ways to attack on a system. In this structure a root node represents main goal of an attacker and intermediate nodes are representative of sub goals to achieve main goal. Branches of an Attack Tree represent different paths to achieve a goal and are termed as sub goals. Leaf nodes represent attack as a compromised state of system as an event and moving upward in a tree gives the ways to reach to the sub goals and finally to the main goal. Attack sub goals and Attack main goals are also connected with AND/OR Nodes. Tree traversal is from left to right and is shown by Ordered –AND nodes and Ordered –OR nodes (O-AND/OR). It is necessary to mention here that all nodes of a Tree may or may not have precedence order (O). An attack goal is successfully achieved when all of its AND children or at least one of its OR children are accomplished. This is same for all sub goals down to leaves of the tree [20].

We have followed various reports published by Symantec, and Eset, on Stuxnet attack, for analyzing Stuxnet attack strategy through Attack Tree. This attacking strategy is divided in six major goals and is represented in Fig 2.

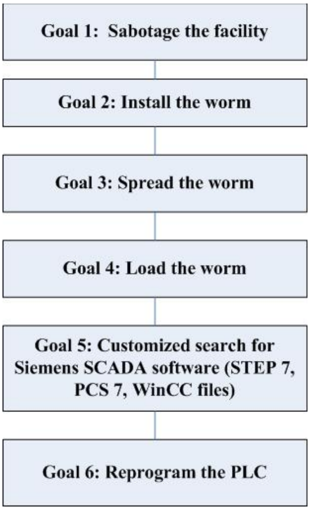

Goal 1: Sabotage the facility
Goal 2: Install the worm
Goal 3: Spread the worm
Goal 4: Load the worm
Goal 5: Customized search for Siemens SCADA software (STEP 7, PCS 7, WinCC files)
Goal 6: Reprogram the PLC

Fig. 2. Attack strategy of Stuxnet

Attack Goal 1: Sabotage the Facility

The main goal of an attacker was to sabotage the SCADA system. For this an attacker can target SCADA center system access, communication system, or disrupt field data interface. A central host computer server and HMI software's on host computers constitute SCADA central systems. SCADA central host computer servers are Master server, database servers, print servers and network server etc. A Human Machine interface (HMI) i.e. application software installed on central host computer presents process data to human operator. To disrupt HMI an attacker can gain access on central host computer, Operator terminal O.S, central host computer applications, communication protocol drivers, Operator terminal applications, communication network management software & RTU Automation systems. To disrupt communication systems an attacker can disable transmission media or exploit any vulnerability in communication media. Lastly an attacker can destruct field data interface by modifying the code written for PLC's and corrupt Remote telemetry unit's interfaces. In case of Stuxnet the main target of an attacker is to reprogram Industrial control system by modifying code on Programmable logic controllers (PLC's), specifically for Siemens SIMATIC/WINCC PCS7 software.

The Attack Tree, shown in Fig 3 and Fig 4, outlines the methods of gaining access to the SCADA system. The intention is to determine all the possible goals of an attacker to deploy Stuxnet in SCADA system.

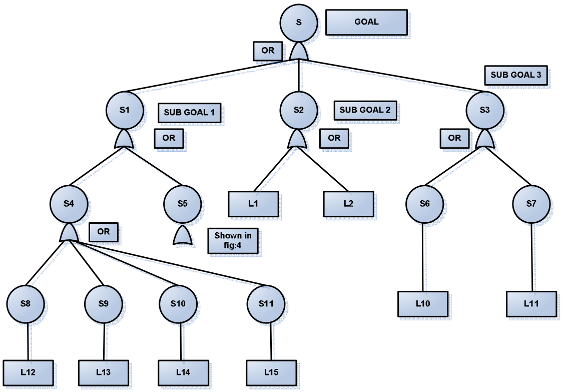


Fig. 3. Methods to sabotage the SCADA system

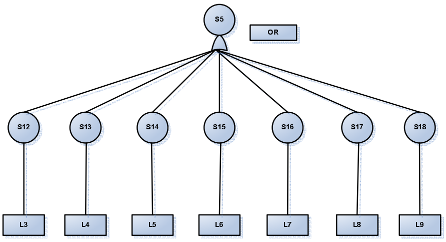


Fig. 4. Methods to Disrupt HMI

Attack Goal 1: Sabotage the SCADA System (S)

Sub Goal 1: Gain SCADA center system access (S1) OR

- 1.1. Gain access to central host computer servers (S4) OR
  - 1.1.1. Compromise SCADA Master server (S8) OR
  - 1.1.2. Compromise SCADA Master Database server (S9)OR
  - 1.1.3. Compromise SCADA Master print server (S10) OR
  - 1.1.4. Compromise SCADA Master network server (S11)
- 1.2. Disrupt HMI (S/w products used within SCADA) (S5) OR
  - 1.2.1. Gain control on central host computer O.S (S12, L3) OR
  - 1.2.2. Gain control on Operator terminal O.S. (S13, L4) OR
  - 1.2.3. Gain control on central host computer applications. (S14, L5) OR
  - 1.2.4. Disrupt communication protocol drivers (S15, L6) OR
  - 1.2.5. Gain control on Operator terminal applications (S16, L7) OR
  - 1.2.6. Disrupt communication network management software (S17,L8) OR
  - 1.2.7. Subvert RTU Automation (S18,L9)

Sub Goal 2: Disrupt communication system (S2) OR

- 2.1. Disable Transmission media (L1)
- 2.2. Exploit Vulnerabilities in SCADA communication protocols (L2)

Sub Goal 3: Disrupt field data interface (S3)

- 3.1 Disrupt PLC (S6) OR
  - 3.1.1Corrupt software at operator terminal (SIMATICPCS7) (L10)
- 3.2 Disrupt RTU (S7)
- 3.3 Subvert RTU Automation software (Modular controllers SIMATIC S7) (L11)

Analysis

On analyzing Attack Tree we have derived three sub goals a) Gain access to SCADA center system, b) Disrupt communication system and PLC, c) Disrupt field data interface, aimed by attackers for deploying Stuxnet malware in SCADA systems.

We have also found several attack methods used for achieving Goal1 of stuxnet. Critical resources which were targeted by these attacks are also listed. A One-to-Many relationship is characterized between attacks and resources, because for an attack, one or more resources can be exploited.

Attacks={MITM, Denial of service against Network, Privilege escalation, Unauthorized access, Flooding attack for PLC's, Root kit Attack, Buffer overflow Attack, Export hooking to gather PLC information, Resource Exhaustion}

Resources={Hardware servers, Print servers, Operating system, Database software, Automation and control Software, Communication channels, Communication protocols, Remote Terminal units, PLC software, Process control system}

Descriptions are provided for the following generic attack types.

1. MITM Attack

Man in the middle attack is a form of active eavesdropping in which the attacker makes independent connections with the victims and relays messages between them. The attacker must be able to intercept all messages going between the two victims and inject new ones. A man-in-the-middle attack succeeds only when the attacker impersonates each endpoint to the satisfaction of the other; it is an attack on mutual authentication.

The two compromised certificates from Jimicron and Realtek enabled this malware to execute user mode and kernel mode root kit under Windows, it is by-passing authentication mechanism. Data communication between a PLC and an HMI was targeted through MITM attack. Stuxnet caused modified data to be reported at HMI, which lead to the undesirable events for operator. At this juncture control systems started behaving abnormally, which resulted in further sabotage.

2. **Denial of Service Against Network**

   A Denial of service (DOS) attack is an attempt by attacker to prevent intended users from using the service by consuming network bandwidth, disrupting connections between machines.

   In case of Stuxnet attack an attacker has used printer and network connections to spread the information. The vulnerability exploited is MS10-061, The Vulnerability in printer spooler service (CVE-2010-2729) could allow remote code execution, if a system shares a printer over the network. Attackers have found a way to exploit MS08-67 vulnerability, an old RPC vulnerability in Windows server service for peer-to-peer servers to initialize DOS attacks. This form of attack basically enabled to get Stuxnet version, injection of specific module and send the malware through peer to peer network. In this way Stuxnet were identified, communicated and updated to each other.

3. **Resource Exhaustion**

   Resource exhaustion is a kind of denial of service attack, in which particular resource is consumed so that it is not available at the time it is required or system is not responding for legitimate requests. Vulnerabilities found in system architecture, protocol services, are subject to resource exhaustion attack.

   In case of Stuxnet attack on control infrastructure, different servers, communication bandwidth, vulnerabilities, found in MODBUS protocol and some application programs like Siemens S7 programs were targeted for resource exhaustion.

4. **Sniffing**

   Sniffing is a kind of passive eavesdropping. The Stuxnet code contains Trojan to sniff data to send and receive information from remote systems. The vulnerability MS10-046 in Windows shell allowed Remote code execution. Another vulnerability MS08-67 in server service could allow remote code execution based on sniffed information.

5. **Privilege Escalation**

   The Privilege Escalation via keyboard layout file (MS10-073) and via task scheduler, were seen in Stuxnet attack for gaining access to resources. These vulnerabilities were exploited to get permission from normal user level to system level permissions. MS08-67 vulnerability in server service allows remote code execution and MS10-073 vulnerability in Windows kernel mode drivers could allow elevation of privileges. Both vulnerabilities are responsible for privilege escalation by Stuxnet.

6. **Unauthorized Access**

   Unauthorized access to a *SCADA* network through an unprotected USB port is the method used to launch the Stuxnet worm. The MS10-046 Windows shortcut vulnerability allowed unauthorized access to spread via removable drives even if auto run was disabled.

7. **Flooding Attack for PLCs**
   Flooding attacks are also DOS attacks in which some specific ports are flooded with packets. An easy way to perform DOS attack on PLCs could be flooding PLC with data or command. Once the Stuxnet worm located in a targeted SCADA system uploads its own program into PLCs to control the automation process and finally flood the PLC with command, causing damage to resources.
8. **Root Kit Attack**
   The Stuxnet attack is a kind of attack which includes PLC root kit. This malware has both user mode and kernel mode root kit capability under Windows. Thus modification in code remains undetected for long period of time.
9. **Buffer Overflow Attack**
   Buffer overflow attacks take advantage of known vulnerabilities within operating system and applications. A buffer overflow occurs when an application receives unexpected data. Stuxnet used malicious code injection in Siemens system and caused buffer overflow attack.
10. **Export Hooking to Gather PLC Information**
   API hooking is a technique used to intercept and alter the command or functions behavior of operating system or other application software. Basically a code has written (termed as hook) to intercept function calls. Stuxnet hooked Ntdll.dll file to monitor for requests to load specially crafted file names [21].

An evaluation table has been established which demonstrates the severity of each attack, i.e. its likelihood of occurrence while achieving a sub goal. The assessment is qualitatively, and on ordinal scale let us assume severity is rated as “High”, “Medium”, and “Low” and relative numeric scale is 9 , 7 , 5 respectively.“High” level of an attack manifests that an adversary uses this category to its optimum precision and concerned application or resources need to be addressed significantly. Medium level attacks need to be addressed with less urgency depending upon how much effort and cost is required to provide countermeasure to stop the attack and low level of attack means at earlier stage it was high, but at present stage it causes to create a path for another attack.

Table 1. Attack severity for Goal 1

Attack	**Severity for sub goal1**		**Severity for sub goal2**		**Severity for sub goal3**	
	Level	Scale	Level	Scale	Level	Scal e
MITM Attack	High	9	Medium	7	Low	5
DOS	Medium	7	High	9	High	9
Sniffing	Medium	7	Low	5	Medium	7
Privilege escalation	High	9	High	9	Medium	7
Unauthorized access	High	9	low	5	Medium	7
Flooding attack(PLC)	High	9	Low	5	High	9
Rootkit Attack	High	9	Low	5	High	9
Bufferoverflow Attack	High	9	Low	5	Low	5
Export hooking	Medium	7	Low	5	High	9
Resource Exhaustion	Medium	7	High	9	High	9

Folowing is a graph plot to quickly analyze the above facts.

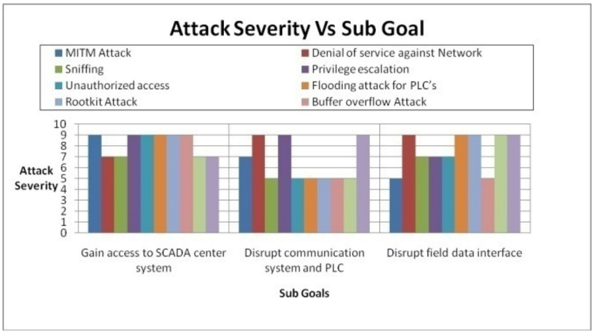


Fig. 5. Sub Goals and Attacks for achieving Goal 1

Attack Goal 2: Install the worm

Stuxnet has a complex architecture. The heart of Stuxnet consists of a large ".dll" file that contains many different exports and resources. Initialization of main installation of malware in the system starts with export 16. Export 16 first checks the value NTVM trace in registry. If the value is 19790509 it confirms existence of threat. This strategy was used to confirm whether Stuxnet is existing or not. After ensuring this, Stuxnet disables firewall settings to start installation procedure by dropping required files in Windows directory. Following Attack Tree demonstrates the procedure. All these activities are possible only when an attacker gains system level permission. Privilege escalation, unauthorized access, a form of MITM attack and root kit attack is seen at this juncture.

Attack Goal 2: Install Stuxnet O-OR
Sub goal 1: Begin main Installation (Invoke Export function 16#) O-OR
1.1. Check for NTVDMTRACE AND
1.2. Check for value 19790509
Sub goal 2: Write into Windows directory
2.1. Modify Windows firewall settings O-OR
2.2. Disable the Windows firewall

Analysis: For installing worm, the sub goal a) Begin main Installation (Invoke Export function 16#) and b) Write into Windows directory, have to be achieved. Possible attacks and exploited resources are as follows.
Attacks = {MITM, Privilege escalation, Unauthorized access, Root kit Attack}
Resources = {Hardware servers, Operating system, Database software, Automation and control Software, Communication and control Software}
Here we are showing all categories of attack because severity of some attacks are low but favors another category of attack for its successful completion.

Table 2. Attack severity for Goal 2

Attack	Severity for sub goal1		Severity for sub goal2	
	Level	Scale	Level	Scale
MITM Attack	High	9	Medium	7
DOS	Low	5	Low	5
Sniffing	Medium	7	Medium	7
Privilege escalation	High	9	High	9
Unauthorized access	High	9	Medium	7
Flooding attack (PLC)	Low	5	Low	5
Rootkit Attack	Medium	7	High	9
Buffer Overflow	Low	5	Low	5
Export hooking	Low	5	low	5
Resource Exhaustion	Low	5	Low	5

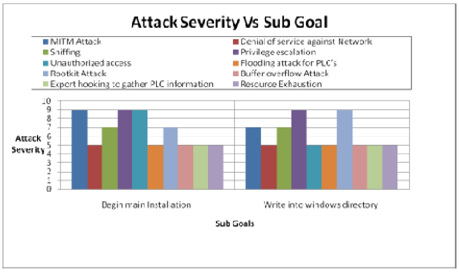


Fig. 6. Sub Goals and Attacks for achieving Goal 2

Attack Goal 3: Spread the worm

Stuxnet propagates by infecting removable drives and also by copying itself over the network using a variety of means. In addition, Stuxnet has the ability to copy itself into Step 7 projects using a technique that causes Stuxnet to auto-execute on opening of the project.

Attack Goal 3: Spread the worm
Sub goal 1: USB Drive infection OR
1.1. Spread via Exploiting .LNK vulnerability (CVE-2010-2568) AND
1.2. Stuxnet verify it is running under services.exe AND
1.2.1. Create a file ~ WTR412.tmp O-AND
1.2.2. Create a file ~ WTR4141.tmp
1.2.3. Create shortcut to .lnk
1.2.4. Create a copy of shortcut to .lnk
1.2.5. Create a copy of copy of shortcut to .lnk
1.2.6. Create a copy of copy of copy of shortcut to .lnk
1.2.7. Check for versions of Windows
1.2.8. Create a hidden window and wait for USB to be inserted
1.3. Use Export 19
1.3.1. Copying Routine
Sub goal 2: Spread Via peer to peer communication
Sub goal 3: Spread via Network share

Analysis: Spreading of the worm was seen in three ways, the sub goal a) USB Drive infection b) Spread via peer to Peer communication and c) Spread via Network share. Possible attacks and resources are as follows.

Attacks = {MITM, Denial of service against Network, Privilege escalation, Unauthorized access, Root kit Attack, Export hooking to gather PLC information} **Resources** = {Hardware servers, Print servers, Operating system, Database software, Automation and control Software, Communication channels, Communication protocols, Remote Terminal units, PLC software, Process control system}

Table 3. Attack severity for Goal 3

Attack	Severity for sub goal1		Severity for sub goal2		Severity for sub goal3	
	Level	Scale	Level	Scale	Level	Scale
MITM Attack	High	9	High	9	High	9
DOS	Medium	7	High	9	High	9
Sniffing	Low	5	Low	5	Medium	7
Privilege escalation	High	9	Medium	7	Medium	7
Unauthorized access	High	9	Medium	7	Medium	7
Flooding Attack(PLC)	Low	5	Low	5	Low	5
Root kit Attack	Medium	7	Low	5	Low	5
Buffer overflow	Low	5	Low	5	Low	5
Export hooking	High	9	Medium	7	High	9
Resource Exhaustion	Low	5	Low	5	Low	5

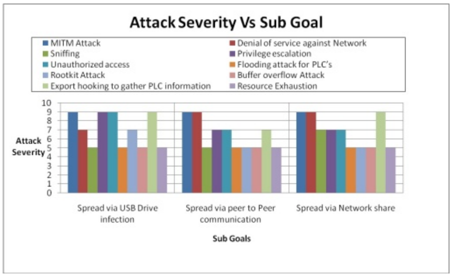


Fig. 7. Sub Goals and Attacks for achieving Goal 3

Attack Goal 4: Load the worm

At this juncture the main objective of attacker was execution of Stuxnet malware every time whenever infected system boots up. This task was performed by Mrxcls.sys driver. The goal of driver was to inject and execute copies of Stuxnet in to specific processes. The code snippet written for this driver has user mode and kernel mode privileges.

Attack Goal 4: Load the worm

Sub goal 1: Load Stuxnet module into a process by Mrxcls.sys driver O-AND

1.1 Get process address

Sub goal 2: Inject Stuxnet module into the process in kernel mode and user mode

2.1. Allocate memory AND

2.2. Execute Stuxnet.dll AND

2.2.1. Call exports AND

2.2.2. Call resources

2.3. Create MZ and PE files.

Analysis: Sub goals discovered at this stage are a) Load Stuxnet module into a process by Mrxcls.sys driver, b) Inject Stuxnet module into the process in kernel mode and user mode, for loading Stuxnet malware in SCADA systems.
Attacks = {MITM, Privilege escalation, Unauthorized access, Flooding attack for PLC's, Root kit Attack, Export hooking to gather PLC information, Resource Exhaustion}
Resources = {Hardware servers, Print servers, Operating system, Database, Remote Terminal units, PLC software}

Table 4. Attack severity for Goal 4

Attack	Severity for sub goal1		Severity for sub goal2	
	Level	**Scale**	**Level**	**Scale**
MITM Attack	Medium	7	Medium	7
DOS	Low	5	Low	5
Sniffing	Medium	7	Medium	7
Privilege escalation	High	9	High	9
Unauthorized access	High	9	Medium	7
Flooding attack (PLC)	Low	5	low	5
Rootkit Attack	Medium	7	High	9
Buffer overflow	Low	5	Medium	7
Export hooking	High	9	High	9
Resource Exhaustion	Low	5	Medium	7

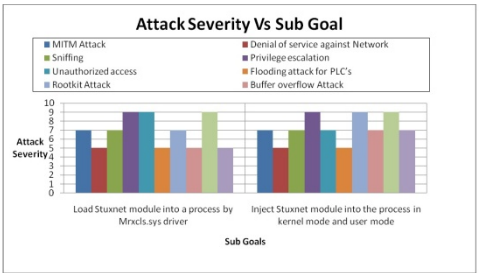


Fig. 8. Sub Goals and Attacks for achieving Goal 4

Attack Goal 5: Searching for step 7 project files

WinCC Simatic manager, used to manage a WinCC/Step7 project. Stuxnet monitor step7 projects (.S7P files). Hooking method is used to open specific APIs from project files inside the s7tgtopx.exe process. Following is Attack Tree representation of searching and infecting step7 project files.

Attack Goal 5: Search for step 7 project files
Sub goal 1: Search for .S7P file extensions OR
1.1. Create Following files O-AND
1.1.1. xutils\listen\xr000000.mdx O-AND
1.1.2. xutils\links\s7p00001.dbf
1.1.3. xutils\listen\s7000001.mdx

1.2. Scan Sub folders Under h0msave7 O-AND
1.2.1. Drop Resource 202
1.3. Modify step 7 project files
1.3.1. Modified step7 data files
Sub goal 2: Search for .mcp file extensions and infect project and WINCC database OR
2.1. Create Following File O-AND
2.1.1. GracS\cc_alg.sav O-AND
2.1.2. GracS\db_log.sav O-AND
2.1.3. GracS\cc_alg.sav xutils\listen\s7000001.mdx
2.2. Scan Sub folder GracS
2.2.1. Dropped a copy of resource 203
Sub goal 3: Search for .tmp file extensions
3.1 Validate file name OR
3.2 Examined contents of the file OR
3.3 Update for newer versions

Analysis: On analyzing Attack Tree we have derived three sub goals a) Search for .S7P file extensions, b) Search for .mcp file extensions and infect project and WINCC database, and c) Search for .tmp file extensions, aimed by attackers for searching the targeted step7 project files in SCADA systems.

Attacks = same as mentioned for Goal 4 and

Resources={Hardware servers, Operating system, Database software, Automation and control Software, Communication channels, Remote Terminal units, PLC software, Process control system}

Table 5. Attack severity for Goal 5

Attack	Severity for sub goal1		Severity for sub goal2		Severity for sub goal3	
	Level	Scale	Level	Scale	Level	Scale
MITM Attack	Medium	7	Medium	7	Medium	7
DOS	Low	5	Low	5	Low	5
Sniffing	Medium	7	Medium	7	Medium	7
Privilege escalation	High	9	High	9	Medium	7
Unauthorized access	Medium	7	Medium	7	Low	5
Flooding attack (PLC)	Medium	7	Medium	7	Low	5
Rootkit Attack	Medium	7	Medium	7	Low	5
Buffer overflow	Medium	7	Low	5	High	9
Export hooking	High	9	High	9	High	9
Resource Exhaustion	High	9	Medium	7	Low	5

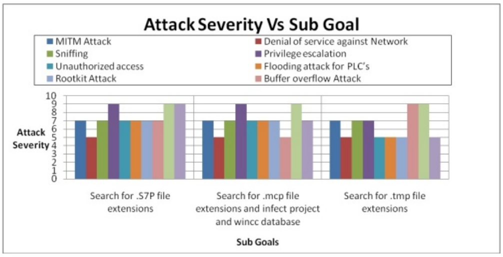


Fig. 9. Sub Goals and Attacks for achieving Goal 5

Attack Goal 6: Modify PLC code

The end goal of Stuxnet is to infect specific types of Simatic PLC devices. PLC devices are loaded with blocks of code and data written using a variety of languages, such as STL or SCL. The compiled code is an assembly called MC7. These blocks are then run by the PLC, in order to execute, control, and monitor an industrial process.

Resource 208 is dropped by export #17 and is a malicious replacement for Simatic's s7otbxdx.dll file. The original s7otbxdx.dll is responsible for handling PLC block exchange between the programming device (i.e., a computer running a Simatic manager on Windows) and the PLC. By replacing this .dll file with its own, Stuxnet is able to perform the following actions:

1. Monitor PLC blocks being written to and read from the PLC.
2. Infect a PLC by inserting its own blocks and replacing or infecting existing blocks.
3. Mask the fact that a PLC is infected.

Attack Goal 6: Modify PLC code

Sub goal 1: Monitor PLC Blocks O-AND
- 1.1. Target WinCC/Step7 and s70tbxdx.dll

Sub goal 2: Infect PLC Blocks O-AND
- 2.1. Rename s70tbxdx.dll to s70txsx.dll O-AND
- 2.2. Replace the original dll with its original malicious code. O-AND
  - 2.2.1. Intercept hooking process O-AND
  - 2.2.2. Start malicious thread O-AND
    - 2.2.2.1. Run an infection routine in every 15 seconds O-AND
    - 2.2.2.2. Infect CPU's O-AND
    - 2.2.2.3. Regularly query PLC for specific block O-AND
    - 2.2.2.4. Customize code block O-AND

Sub goal 3: Mask PLC infection (PLC Root kit)
- 0.1. Monitor and read requests O-AND
  - 0.1.1. Read requests for its own malicious block OR
  - 0.1.2. Read requests for infected blocks OB1, OB35, DP-RECV
- 0.2. Intercept the code O-AND
  - 0.2.1. Write requests that could overwrite Stuxnet's own code
- 0.3. Modify code
  - 0.3.1.Modify requests to ensure new version of code block are infected OAND
  - 0.3.2.Read block of code O-AND
  - 0.3.3.Delete block of code

Analysis: At this juncture sub goals are a) Monitor PLC Blocks, b) Infect PLC Blocks, c) Mask PLC infection, for modifying PLC's in SCADA systems.
Attacks = {MITM, Privilege escalation, Unauthorized access, Flooding attack for PLC's, Root kit Attack, Buffer overflow Attack, Export hooking to gather PLC information, Resource Exhaustion}
Resources = {Operating system, Database software, Automation and control Software, Remote Terminal units, PLC software, Process control system}

Table 6. Attack severity for Goal 6

Attack	Severity for sub goal1		Severity for sub goal2		Severity for sub goal3	
	Level	**Scale**	**Level**	**Scale**	**Level**	**Scale**
MITM Attack	Low	5	Medium	7	Low	5
DOS	Low	5	High	9	Low	5
Sniffing	High	9	High	9	Medium	7
Privilege escalation	Medium	7	Medium	7	Medium	7
Unauthorized access	Medium	7	High	9	Medium	7
Flooding attack (PLC)	Low	5	Medium	7	High	9
Rootkit Attack	Low	5	Medium	7	High	9
Buffer overflow	Low	5	Medium	7	Low	5
Export hooking	High	9	High	9	High	9
Resource Exhaustion	Low	5	High	9	Low	5

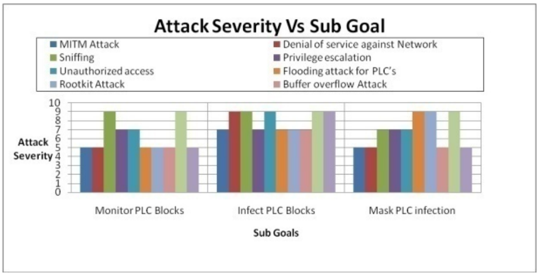


Fig. 10. Sub Goals and Attacks for achieving Goal 6

5 Conclusion

Stuxnet attack on SCADA system reemphasizes on extensive research in cyber security aspect of critical infrastructures. In this paper we have focused on one threat modeling technique to detect possible vulnerabilities attacks, and exploited resources in a system. An Attack Tree is a static threat modeling technique to capture the attacker's behavior as well as system behavior, this facilitates detailed analysis of an attack. We have designed novel Attack Trees for post Stuxnet attack scenario. For each Attack Tree, Goals and Sub goals are highlighted to clearly indicate the attacking approach. Common category of attacks has been discovered which were used for execution of Stuxnet attack. A one-to-many relationship between attacks and resources are discovered. A graphical representation of attack severity versus sub goals is helpful for quickly analyzing attacks in any component of SCADA system. This also shows how one attack opens the door for another attack and what resources were used for successful completion of attack. Using this information one can determine possible countermeasures depending upon cost and effort estimation.

References

1. Langner, R.: Stuxnet: Dissecting a Cyber warfare Weapon. IEEE Security & Privacy 9(3), 49–51 (2011)
2. http://www.wired.com/images../Symantec-Stuxnet-Update-Feb-2011.pdf
3. http://www.us-cert.gov/control_systems/pdf/ICSA-10-272-01.pdf
4. http://go.eset.com/us/resources/white-papers/Stuxnet_Under_the_Microscope.pdf
5. Stuxnet: The first weaponized software?, http://www.cs.columbia.edu/~smb/blog//2010-09-27.html
6. Cárdenas, A.A., Amin, S., Lin, Z.-S., Huang, Y.-L., Huang, C.-Y., Sastry, S.: Attacks against process control systems: risk assessment, detection and response. In: Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security (ASIACCS 2011), pp. 355–366. ACM, New York (2011)
7. Schneier, B.: Attack Trees. Dr. Dobb's Journal 24(12), 21–29 (1999)
8. Khand, P.A.: System level security modeling using Attack Trees. In: 2nd International Conference on Computer, Control and Communication, IC4 2009, pp. 1–6 (February 2009)
9. Thesis, Efficient Semantics of Parallel and Serial Models of Attack Trees, http://www.cyber.ee/publikatsioonid/20-magistri-ja../JurgensonPhD.pdf
10. Supervisory Control and Data Acquisition (SCADA) Systems, http://www.ncs.gov/library/tech_bulletins/2004/tib_04-1.pdf
11. Chen, T.M., Abu-Nimeh, S.: Lessons from Stuxnet. Computer 44(4), 91–93 (2011)
12. Paulson, L.D.: Worm Targets Industrial-Plant Operations. Computer 43(11), 15–18 (2010)
13. Ten, C.-W., Manimaran, G., Liu, C.-C.: Cybersecurity for Critical Infrastructures: Attack and Defense Modeling. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans 40(4), 853–865 (2010)
14. Greengard, S.: The new face of war. Commun. ACM 53(12), 20–22 (2010)
15. Chen, T.M.: Stuxnet, the real start of cyber warfare? [Editor's Note]. IEEE Network 24(6), 2–3 (2010)
16. Stuxnet: rumors increase, infections spread. Network Security (10), 1–2 (October 2010)
17. Jeong, O.-R., Kim, C., Kim, W., So, J.: Botnets: threats and responses. International Journal of Web Information Systems 7(1), 6–17 (2011)
18. Morais, A., Martins, E., Cavalli, A., Jimenez, W.: Security Protocol Testing Using Attack Trees. In: International Conference on Computational Science and Engineering, CSE 2009, August 29-31, vol. 2, pp. 690–697 (2009)
19. Saini, V., Duan, Q., Paruchuri, V.: Threat modeling using Attack Trees. J. Comput. Sci. Coll. 23(4), 124–131 (2008)
20. Camtepe, S.A., Yener, B.: Modeling and detection of complex attacks. In: Third International Conference on Security and Privacy in Communications Networks and the Workshops, SecureComm 2007, September 17-21, pp. 234–243 (2007), doi:10.1109/SECCOM.2007.4550338
21. Sungmo Jung,S. K., Song, J.-G.: Design on SCADA Test-bed and Security Device. International Journal of Multimedia and Ubiquitous Engineering 3(4) (October 2008)

Secure Peer-Link Establishment in Wireless Mesh Networks

Swathi Bhumireddy, Somanath Tripathy, and Rakesh Matam

Department of Computer Science and Engineering
Indian Institute of Technology Patna
Patna, Bihar-800013
India
{som,b.swathi,m.rakesh}@iitp.ac.in

Abstract. Wireless Mesh Network (WMN) has become popular, as it allows fast, easy and inexpensive network deployment. It is observed that the current peer link establishment mechanism presented in IEEE 802.11s draft standard is vulnerable to various kinds of relay and wormhole attacks. In this paper, we propose a certificate-based peer link establishment protocol that employs location information to prevent such attacks. The security analysis shows that the proposed mechanism is resistant against different kinds of wormhole, relay and Sybil attacks.

1 Introduction

Wireless mesh networks (WMNs) have emerged as a promising technology to provide low-cost high-bandwidth wireless access services in a variety of application scenarios. A typical WMN is comprised of a set of stationary mesh routers (MRs) that form the mesh backbone and a set of mesh clients that communicate via mesh routers. WMNs have several advantages such as low-setup cost, increased coverage and most importantly reliable and flexible [1]. In spite of the above benefits, they are also constrained by the open wireless medium, varying channel conditions and interference. In addition to the above specified constraints, failing to meet the security requirements further restricts the extensive deployment of WMN. Designing of effective security mechanisms is a challenging task in WMN due to the open wireless medium which is more susceptible to attacks. The other limitation is the multi-hop cooperative communication that makes services more vulnerable to attacks especially coming from within the network.

The authenticated mesh peer-link exchange (AMPE) protocol presented in IEEE 802.11s standard (draft) [2] forms an key component in the deployment of a WMN. Vulnerabilities in the peer-link establishment mechanism makes the network susceptible to various kinds of attacks and wormhole attack is one such attack that has severe impact on performance of WMN. In a wormhole attack, an adversary can capture packets from one location and replay them at another location with the help of an out-of-band channel, simple packet encapsulation or high-powered transmission to establish a wormhole link. The route via the wormhole link would be naturally preferred by legitimate nodes as it offers a path

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 189–198.
springerlink.com

with lower hop-count and latency than any other multi-hop routes. The existing AMPE protocol cannot prevent such a wormhole attacks, as the attacker can relay peer-link establishment messages without modifying the packet contents.

Typically, a wormhole attack is handled at the routing layer in wireless mobile ad hoc networks. Various secure routing protocols have been proposed to enhance the security of such network. Few of the existing secure routing protocols are SAODV [3], SEAODV [4], LHAP [5] and ARAN [6]. It has been already shown that these protocols are vulnerable to wormhole attacks, an attack launched by colluding malicious nodes. Even though various other mechanisms exist in literature to detect and prevent wormhole attacks, the most effective way to prevent them in WMN is to secure the peer-link establishment process in WMN.

Therefore, in this paper we initially show that the current peer-link establishment process is vulnerable to relay and wormhole attacks and later propose a peer-link establishment mechanism to secure the WMN against such attacks. The rest of the paper is organized as follows. In section 2, we describe the wormhole attack in detail and its adverse effects on the network. In section 3, we present the related work on the detection and prevention of wormhole attacks. In section 4, we describe present our peer-link establishment mechanism to prevent wormhole attacks. In section 5, we present a detailed security analysis and finally section 6 concludes the paper.

2 Related Work

Several works have been carried out to specially address a wormhole attack in ad hoc networks. Most of the proposed mechanisms try to detect or prevent wormhole attacks during route discovery or data transmission. Packet Leashes [8] is one such mechanism that defends against wormhole attacks in a network. This mechanism can be used with any of the existing routing protocols. Typically with each packet, a leash(an information) is added to a packet to restrict the packet from travelling more than the maximum allowed transmission distance. The two types of leash are geographical and temporal leashes. To accommodate a geographical leash a node must know its location and all nodes must have loosely synchronized clocks. The sender includes in the packet, its own location and the time it sent the packet. The receiver compares these values to its location and the time it receives the packet. If the clocks of both sender and receiver are synchronized within some predefined bounds, then the receiver can compute a distance between itself and the sender. From the distance the receiver can estimate the minimum number of hops between the sender and itself, thereby detecting the presence of wormhole link.

Temporal leashes require nodes to have tightly synchronized clocks such that the maximum difference between any two nodes clocks is δ and δ must be known by all network nodes. The sending node includes in the packet, the time at which it sent the packet and this value is compared by the receiving node to the time it receives the packet. The receiver can determine whether the packet travelled further based on the supposed transmission time and the speed of light. The sender could also include an expiration time in the packet so that the receiver does not accept the packet after this time.

Another alternate wormhole detection approach that uses the nodes location information is proposed by lazos et.al [9] where only a small fraction of the nodes need to be equipped with a GPS receiver. These special nodes are called guards and it is also assumed that the guards have a larger radio range (denoted by R) than the other nodes. The guards broadcast their positions in their one hop neighborhood. Two nodes consider each other neighbor only if they hear a threshold number of common guards. The nodes use the location information broadcast by the guards to detect wormholes based on the following two principles: (i) since any guard heard by a node must lie within a range of radius R around the node, a node cannot hear two guards that are 2R apart from each other; and (ii) since the messages sent by the guards are authenticated and protected against replay, a node cannot receive the same message twice from the same guard. It is shown that based on these principles, wormholes can be detected with probability close to one. However, the disadvantage of this approach is that the guards are distinguished nodes in the network that differ from the regular nodes.

EDWA [10] is an end-to-end detection of wormhole attack (EDWA) in wireless ad-hoc networks. It proposes a wormhole detection which is based on the smallest hop count estimation between source and destination. If the hop count of a received shortest route is much smaller than the estimated value an alert of wormhole attack is raised at the source node. Then the source node will start a wormhole TRACING procedure to identify the two end points of the wormhole. Finally, a legitimate route is selected for data communication. Distance between the source and destination is estimated using Euclidean Distance Estimation technique. The protocol is proposed specifically proposed for a source routing protocol and does not work with other routing protocols and also requires the length of the wormhole to be large to accurately detect and identify wormhole links.

The protocol proposed in [11] to detect wormholes attacks employs local neighborhood information. The network topology is assumed to be static and the links are assumed to be bi-directional. However, they assume that the wormhole must change the topology structure of the network and they compute edge-clustering coefficient. The assumption is that in a dense network every two neighbours must have a common neighbour. A wormhole node is detected by one of its neighbours if that neighbour cannot reach one of the wormhole neighbours without using that node. However, it is very possible to come up with many scenarios with wormholes that will not satisfy any of the necessary conditions with this approach to detect the wormhole. This will only successfully detect open wormholes or closed wormholes that only connect one single node with another single node. If the wormhole connects a group of nodes (2) with another group of nodes, which is the most common form of wormhole, then the protocol will not detect the wormhole. The protocol can be shown to report high false-positive ratio due to the kind of design methodology employed by the protocol.

Thaier et.al. propose DeWorm [12], a protocol that uses routing discrepancies between neighbours along a path from the source to destination along a path from the source to destination to detect a wormhole. It is based on the observation that to have a successful impact on the network the wormhole must attract significant amount of traffic in the network and the length of the wormhole is significantly large. Most of the existing work tries to address the wormhole attack at the network layer while the most effective way to prevent a wormhole attack in WMN is to secure the AMPE protocol.

3 Network Assumptions and Adversary Model

3.1 Network Model and Assumptions

We consider a typical WMN architecture, where a set of mesh routers (MR's) form the backbone of the WMN. We assume that a public-key infrastructure administered by a Certificate Authority (CA) exists in the network that allows nodes to obtain and authenticate using digital certificates. The MR's are equipped with GPS systems to facilitate them to determine their location information. Few of the MR's are designated with an additional functionality and act as gateway (ROOT) nodes by connecting to the Internet. Each node obtains a valid certificate from the *CA* in prior to joining the network. We also assume that a root node maintains a local directory of the certificates of all the nodes in the network and they are updated whenever a node issues or re-issues a certificate. Every root node also maintains a directory about the nodes that are under its sub-network.

3.2 Adversary Model

We consider an adversary model where an adversary is capable of relaying packets with the help of an out-of-band high speed transmission link. The adversary can also compromise an MR and collude with it to launch an FRI-Attack [7]. We also assume that an adversary is capable of generating multiple identities to establish peer-links with nodes in the different parts of the network.

4 Proposed Peer-Link Establishment Mechanism

The proposed peer-link establishment mechanism prevents various kinds of wormhole attacks during the authenticated mesh-peering exchange in WMN.

4.1 Existing Peer-Link Establishment Mechanism

The mesh peering management framework enables mesh STAs to establish, manage and tear down peering between mesh STAs. The AMPE protocol uses Mesh Peering Open frames, Mesh Peering Confirm frames, and Mesh Peering Close frames to establish, manage, and tear down a mesh peering. Mesh STAs shall not transmit frames other than the ones used for candidate peer mesh STA discovery, mesh peering management, and simultaneous authentication of equals (SAE) to a neighboring mesh STA until a mesh peering has been established with the mesh STA. This prevents a mesh STA from gaining unauthorized access to the network resources. When a mesh STA discovers one or more neighbor mesh STAs through scanning process either through beacon or probe response frames, it may try to become a member of the mesh BSS of which the discovered mesh STA is already a member, and establishes a mesh peering with that neighbor mesh STA with the help of authenticated mesh peering exchange protocol (AMPE). The AMPE protocol requires the existence of shared pair wise master key (PMK) security association established between two candidate peer mesh STAs. If the shared Mesh PMKSA is not identified, the mesh STA shall execute an authentication protocol to mutually authenticate with the candidate peer mesh STA.

As a mesh STA accepts and processes information elements from only STAs that are registered as peers, most of the external attacks are effectively prevented. To launch any kind of external attack, the attacker should manipulate information elements to register himself as a legitimate peer. This can be carried out by exploiting the vulnerabilities in the peer-link management protocol that allows an attacker to convince two nodes located far-away as peers by relaying peer-link messages between them. An attacker can also gain access to the network with the help of a single compromised node.

Table 1. Notations employed by the proposed peer-link establishment mechanism

Notation	**Meaning**
CA	Certification Authority
Pos_A	Position locations of MR_A
$Cert_A$	Initial certificate issued by the CA to MR_A
$CertPos_A$	Certificate with position information of MR_A
R_i	i_{th} Root Node in the network
ReqCertPos	An attribute for requesting a certificate with position information
RepCertPos	An attribute in the packet the CertPos of a node RemLink
RemLink	An attribute to tear down a peer-link

4.2 Proposed Peer-Link Establishment

For a candidate MR to establish a peer link, it should pocess a certificate with position information (CertPos). To obtain such a certificate, it initiates a peer-link establishment process by presenting its initial certificate with location information. A MR needs to revoke a CertPos whenever it changes its location. We present a detailed procedure carried out by a node to obtain a CertPos. Later, we present the revoking of the CertPos. Finally, we present how a node uses position enabled certificate (CertPos) to form wormhole free peer-links.

- **Initial Certificate Request with Position Information** A node A intended to join the mesh network presents a valid certificate (CERT) issued by CA. A node that receives a request to establish a peer-link checks for the position information. Lack of position information indicates a node trying to establish a peer-link for the first time. A peer node that receives CERT, requests for a new certificate with position information (CertPos) on behalf of the potential node intended to join the network on receiving a authentication frame with the contents ReqCertPos, Source address, Target Address, CERT and Pos from the potential node wishing to join the network. The request process is shown in Fig.1(a).

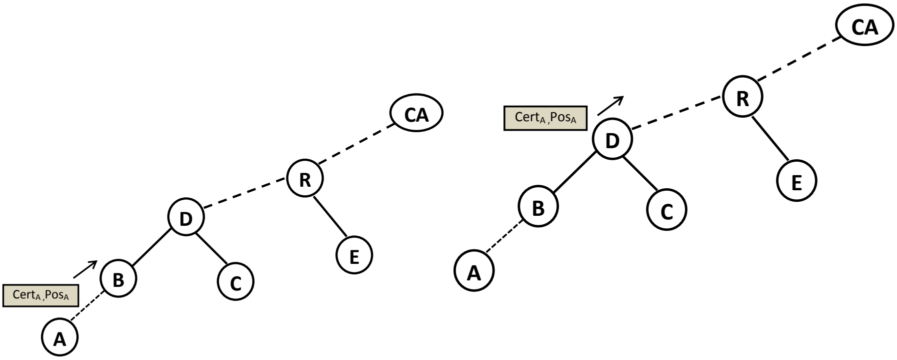


Fig. 1. (a) Step one (b) Step two

On receiving a authentication request, a peer node B checks the position information presented by A by verifying whether A is in its range of transmission. On validating the position information it propagated the authentication request to the nearest root node as only root nodes are capable of communicating with CA. The entire request process is shown in Fig.1(b). The peer node caches the information about A till it receives a CertPos. The root node propagates the authentication request to the CA as shown in Fig.2 and caches the information similar to node B.

On receiving a authentication request, the CA verifies the CERT and issues a certificate with position information in it (CertPos). The process is shown Fig.2. CA then multicast the CertPos to every root node in the network as shown in Fig.3.

All the root nodes that receive the packet from CA update their information about certificates of all the nodes in the network. The root nodes

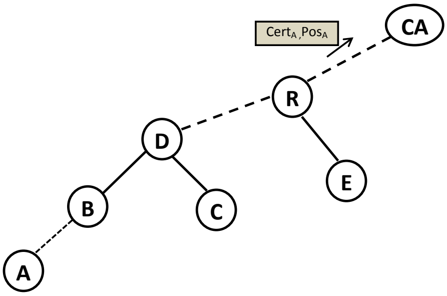


Fig. 2. (a) Step three

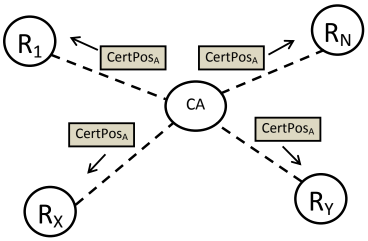


Fig. 3. Step four

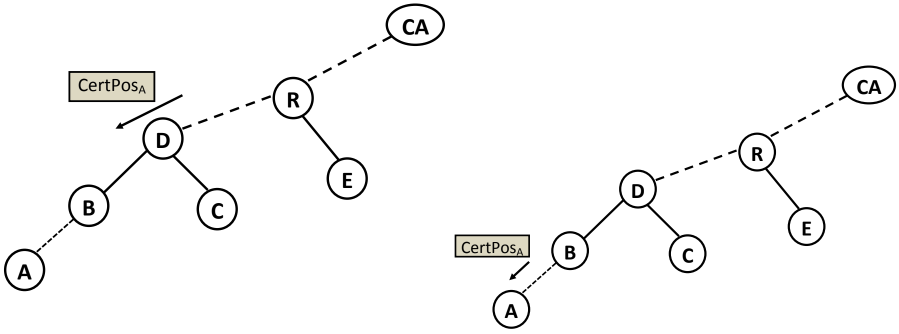


Fig. 4. (a) Step five (b) Step six

matches the CertPos with the cached information while sending the request. The only root node that has propagated the request packet forwards the authentication reply packet to the node from which it has received the request packet. As the requesting node most probably joins its sub-network, it stores the information about the node in an active directory containing the information of nodes in its sub-network.

In the final step, when the peer node receives the reply packet with the CertPos of node A, it checks its local cache to verify whether it has forwarded such a request. After verifying it propagates the reply to the node that has requested one. Finally, the new node A that has requested for a CertPos receives one form the CA through the intermediate nodes.

– **Re-issuing Certificate with modified Position Information**

Whenever a node part of the network changes its position, it has to re-issue the CertPos for establishing mesh peer links with the new peer nodes. The procedure is similar to that of a node requesting for CertPos. It only differs in the fourth and fifth steps. In the fourth step, when the root nodes receives a request with the RepCertPos attribute, it checks if there exists a node in its sub-network for which the CertPos has been already issued. If it does exist, then the root node creates a packet with the address of the node that requested CertPos and a RemLink attribute. The root node multicast's the packet to all the nodes in its sub-network.

In the fifth step, nodes in the network that receive a packet with RemLink attribute act on the information if they share peer-links with the node's present in the packet and is marked as stale. Once this process is completed, a new certificate with position information is issued to the requesting node.

– **Peer-Link Establishment** A node that possesses a CertPos wishes to establish a peer link with the nodes that it has discovered through beacon frames or probe response frames, it sends a mesh peer link establishment request packet with its CertPos appended to the authentication frame. The peer nodes that receive this packet, verifies whether the node is in its range of transmission by using the locations of the node that can it obtained from CertPos. If it does not fall in its transmission range, it simply discards the packet. On receiving a valid CertPos (meeting transmission range constraints), it sends a request to the nearest root node asking to validate the CertPos of the node with which it wishes to establish a mesh peer link. The root node that receives the validation request, verifies the CertPos and replies accordingly. In case of a valid certificate, it updates the sub-network directory and replies to the requesting node. On receiving the reply that the CertPos is genuine, the node forms a mesh peer link with the requested node.

5 Security Analysis

Security analysis of the proposed peer-link establishment mechanism depends on the ability to successfully thwart hidden wormhole, FRI-attack and sybil attack. It allows MRs to accurately verify the geographical position of nodes making authentication requests. The security of the proposed scheme is analysed for all the above mentioned attacks.

- **Hidden Wormhole Attack:** In a hidden wormhole attack, an adversary captures a peer-link establishment message and relays it to the other part of the network. A MR that receives such a peer-link message verifies the certificate and confirms the peer link establishment process thus forming a non-existent peer-link. The position certificate (CertPos) issued by the CA prevents a MR from establishing a peer-link with a node outside its transmission range. Even though an adversary can successfully relay (in-band or out-of-band) peer-link establishment messages, the CertPos prevents nodes from establishing such non-existent peer-links.
- **FRI-Attack:** In fraudulent routing information attack, an adversary has access to all the keying-material of a legitimate MR that is required to obtain a valid certificate from the CA. The proposed mechanism does not prevent an adversary from obtaining a certificate instead it prevents the adversary from colluding with the compromised MR_C. The CA that generates a revised CertPos for MR_C, broadcasts it to all the ROOT nodes to allow them to update the current node information. A ROOT node that has an entry for MR_C broadcasts a message in the sub-network to allow the MRs to invalidate the existing peer-links with such a node MR_C. Even in a case where an external adversary gains access to the network with the help of compromised MR, all the nodes that share peer-links with such a compromised MR are invalidated. Thus the proposed mechanism prevents the adversary from exploiting and disrupting the network with the help of compromised MRs effectively.
- **Sybil Attack:** A Sybil attack is the form of attack where a malicious node creates multiple identities in the network, each appearing as a legitimate node. It can disrupt network services like packet forwarding, routing, and collaborative security mechanisms. The proposed peer-link establishment mechanism inherently prevents a sybil attack as it allows a single active instance of a node identity in the WMN.

6 Conclusion

In this paper we proposed an improved peer-link establishment protocol that successfully prevents the peers from forming links with far-away nodes in the network. The proposed mechanism incurs additional overhead when compared to the existing peer-link management mechanism but it out-weighs the existing mechanism in terms of security. It also prevents stealthy attacks such as sybil attack, relay, FRI-attack and wormhole attack.

References

1. Akyildiz, I.F., Wang, X., Wang, W.: Wireless mesh networks: A survey. Computer Networks and ISDN Systems (2005)
2. IEEE P802.11s/D5.0 Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY)specifications, Amendment 10: Mesh Networking
3. Zapata, M., Asokan, N.: Securing ad-hoc routing protocols. In: Proceedings of ACM Workshop on Wireless Security, pp. 1–10 (September 2002)
4. Li, C., Wang, Z., Yang, C.: Secure Routing for Wireless Mesh Networks. International Journal of Network Security 13(2), 109–120 (2011)
5. Zhu, S., Xu, S., Setia, S., Jajodia, S.: LHAP: A Lightweight Hop-by-Hop Authentication Protocol for Ad-Hoc Networks. In: Proceedings of ICDCS International Workshop on Mobile and Wireless Network, Providence, Rodhe Island, pp. 749–755 (May 2003)
6. Sangiri, K., Dahil, B.: A secure routing protocol for ad hoc networks. In: Proceedings of 10th IEEE International Conference on Network Protocols, pp. 78–89 (2002)
7. Matam, R., Tripathy, S.: FRI Attack: Fraudulent Routing Information Attack on Wireless Mesh Networks. In: Proc. of IEEE Xplore, ACWR (2011)
8. Hu, Y., Perrig, A., Johnson, D.: Packet leashes: A defence against wormhole attacks in wireless networks. In: Proc. of the Twenty-second IEEE International Conference on Computer Communications (April 2003)
9. Poovendran, R., Lazos, L.: A graph theoretic framework for preventing the wormhole attack in wireless ad hoc networks. ACM Journal of Wireless Networks 13(1), 2759 (2005)
10. Wang, X., Wong, J.: An end-to-end detection of wormhole attack in wireless ad-hoc networks. In: Proc. of the Thirty-First Annual International Computer Software and Applications Conference (July 2007)
11. Wang, Y., Zhang, Z.: A Distributed Approach for Hidden Wormhole Detection with Neighborhood Information. In: IEEE Fifth International Conference on Networking, Architecture and Storage, NAS (2010)
12. Hayajneh, T., Krishnamurthy, P., Tipper, D.: Deworm: A simple protocol to detect wormhole attacks in wireless ad hoc networks. In: Proceedings of the IEEE Symposium on Network and System Security (2009)

Secret Image Embedded Authentication of Song Signal through Wavelet Transform (IAWT)

Uttam Kr. Mondal[1] and Jyotsna Kumar Mandal[2]

[1] Dept. of CSE & IT,
College of Engg. & Management, Kolaghat,
Midnapur (W.B), India
[2] Dept. of CSE,
University of Kalyani,
Nadia (W.B), India
uttam_ku_82@yahoo.co.in, jkm.cse@gmail.com

Abstract. In this paper, an algorithm has been proposed to provide security to digital songs through wavelet transform with the help of a secure image embedded with coefficients of it without changing audible quality. Sampling the hidden image with the help of amplitude coding for generating lower magnitude values is the first phase of proposed technique followed by fabrication of authenticating code by embedding into lower magnitude values with selected coefficients of song signal generated via wavelet transform [symmetrization mode]. The embedded hidden secure image as well as authenticating code is used to detect and identify the original song from similar available songs. A comparative study has been made with similar existing techniques and experimental results are also supported with mathematical formula based on Microsoft WAVE (".wav") stereo sound file.

Keywords: Average absolute difference (AD), maximum difference (MD), mean square error (MSE), normalized average absolute difference (NAD), normalized mean square error (NMSE), sampling image, wavelet transform.

1 Introduction

Improvement in digital signal processing, entertainment reaches to a new higher dimension with more enriched quality audio medium like song, voice, speech etc. It creates an opportunity for million of people to create good songs for commercial purpose. Creative organizations are facing competitive market for spreading business as quality products involved lot of investments. People are finding easier way to put less effort or investing money and producing products for existence in this contemporary market. Even some of them are applying technology to produce similar audio songs in lower cost with partially or fully modification of original songs as well as ignoring the copy right property of audio. This intension is a frequent phenomenon for digital audio/video industry with improvement of digital editing technology [4]. It is quite harder for a listeners to find the original from similar versions which is a big challenge for business men, computer professionals or concerned people to ensure the

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 199–208.
springerlink.com

security criteria of originality of songs [1, 2], protecting from releasing the duplicate versions. The purpose of this paper is to review this authentication problem and propose a technique to solve it.

In this paper, a framework for protecting originality of a particular song with the help of secret image embedded with coefficient values of wavelet transform without changing its audible quality has been presented. Sampling the hidden image with the help of amplitude coding for generating lower magnitude distribution is the first phase of proposed technique followed by fabrication of authenticating code by embedding lower magnitude values with selected coefficients of song signal generated via wavelet transform [symmetrization mode]. The embedded hidden secure image as well as authenticating code will use to detect and identify the original song from similar available songs. It is experimentally observed that added extra values will not affect the song quality but provide a level of security to protect the piracy of song signal.

Organization of the paper is as follows. Sampling image and converting into lower magnitude values (in the range of 0 to 0.0011) are presented in section 2.1. Embedding secret image is performed in section 2.2. The extraction is shown in section 2.3. Experimental results are given in section 3. Conclusions are drawn in section 4. References are given at end.

2 The Technique

The scheme fabricates the secure hidden image with help of wavelet transform followed by generating authenticating code. Sampling image and Embedding Secret Image (IAWT-ESI), two methodologies of the technique are given in section 2.1 and section 2.2 respectively.

2.1 Sampling Image

Sampling of image is to produce set of lower intensity values to add with the selected coefficient values of song signal in wavelet domain has been fabricated with authenticated code. Let, I_m is an two dimensional image with length p and width q. $I_m(x,y)$ is the intensity value of pixel position(x, y). If I_m is two dimensional gray image, it should contain only the color value B (Black). If I_m is a color image, it should contain the color information of R (Red), G (Green), B. Let, I_m is a gray image of size (p X q). Each pixel position should contain the color information of B. Let, V is the color value of pixel position (x, y). It, therefore, should hold the value in the range of 0 to 255 i.e. 8 bits information. The intensity value of each pixel of taken gray image can be represented into a set of small amplitude values as described in the following steps.

Step 1: Let, V is the intensity value of pixel position (x, y). Separate V into individual digits. Represent the generated individual digits by small magnitude values as follows

0 will be represented by 0, 1 by 0.0001, 2 by 0.0010, 3 by 0.00011, 4 by combination of two values (0.0010, 0.0010), similarly n ($3<n<=9$) is represent by following way

i. Divide n by 2 to produce two integers (result+0.5) and (result-0.5) respectively.

ii. Repeat above step i for each produced integer until generated results come under the range [0, 3] and then, represent the result as a set of small magnitude values as above representation.

Step 2: Repeat step 1 until all intensity values of pixels of taken image is converted into set of small magnitude values.

Let, Val is set of small magnitude values generated from converting intensity of image of size (l × k), where values of (l × k) would be calculated based on division of individual digits as well as (p × q). The process will be same if the image is color image where above process will be applied repetitively for color values R, G and B respectively.

2.2 Embedding Secret Image (IAWT-ESI)

The method of embedding secret image with original song is described with the help of wavelet transform is fabricated with embedding image with coefficient components of the amplitude signal without affecting its audible quality. The procedure of generating secret signal is depicted in the following algorithm.

Algorithm

Input: Original song signal and sampled values of image (Val).
Output: Modified song signal with embedding image.
Method: Separating amplitude and phase of original song and applying single-level discrete 2-D wavelet transform [symmetrization mode] are described in the following steps.

Step 1: Separating phase and amplitude signal from original song.
Step 2: Apply single-level discrete 2-D wavelet transform [symmetrization mode] over amplitude and phase signals to generate the approximation coefficients matrix approximation coefficients (CA) and details coefficients matrices horizontal detail coefficients (CH), vertical detail coefficients (CV), diagonal detail coefficients (CD) for each of them. The mathematical expressions of the wavelet transform are as follows.

A signal, f(t), can be represent using wavelet transform as given in equation 1.

$$f(t) = \sum_{m} \sum_{n} C_{m,n} \psi_{m,n}(t) \tag{1}$$

Where both m and n are integer and $\psi_{m,n}(t)$ represents successive scaled and dilated versions of a single wavelet function $\psi(t)$ [mother wavelet] [5]. The wavelet basis functions can be formulated with the help of mother wavelet as given in equation (2).

$$\psi_{m,n}(t) = 2^{-m/2} \psi(2^{-m} t - n) \tag{2}$$

The above factor $2^{-m/2}$ usually maintains a constant relationship, i.e., the wavelet basis is independent of scale m and represents the successive partitioning of signal's spectrum throughout wavelet scaling and dilations. In case of iterative wavelet transform, real-valued coefficients, i.e., $\alpha_{m,n}$ are used. The approximation and wavelet coefficients $c_{m,n}$ and $\alpha_{m,n}$ can be represented as follows:

$$\alpha_{m,n}(t) = \sum_{k} h_{2n-k}\alpha_{m-1,k} \tag{3}$$

$$c_{m,n}(t) = \sum_{k} g_{2n-k}c_{m-1,k} \tag{4}$$

Where h and g are low pass FIR and high pass FIR filters respectively [7]. For the 2-D DWT separate horizontal and vertical audio signal filtering and downsampling are needed. The 2-D DWT produces four sub-bands at each level of analysis. The first is the scaling of the input audio signal (CA) and the remaining three are the detail coefficients at the horizontal (CH), vertical (CV) and diagonal (CD) directions respectively.

Step 3: It is experimentally observed the values of diagonal detail coefficients (CD) for both amplitude and phase signals of original song are less difference. If we make the both values of CDs equal then, no affect on audible quality has been proved [8]. If the difference values of two CDs become higher for a particular audio, then more higher level discrete 2-D wavelet transform in symmetrization mode may be applied as required.

Step 4: Embedding image sampled values with diagonal detail coefficients (CD) of amplitude signal as given below.

i. Add ith sampled value ,Val, with kth position of CD as follows
CD[k,C] =CD[k,C] + Val [i] , where C indicates the channel number,
In case of mono type song C is not
needed

ii. Add next value of Val after n positions from kth position of the alternative channel (if previous inserted value at C=1, then C=2, or vice versa) of CD, where n represents gaps between two consecutive embedded positions. The value of n is after adding 0 is 1, after 0.0001 is 2, after 0.0010 is 3 and after 0.0011 is 4. 2 additional gaps will added when all parts of magnitude values of a particular intensity value is appended with CD for separating from next intensity value as well as another 2 gaps will be added after inserting all individual digits of each intensity value of image.

ii. Repeat above step i until all values of Val are not added with CD.

Step 5: Reconstruct the both amplitude and phase signal using the inverse DWT out of calculated wavelet and scaling coefficients and generate song signal with reconstructed amplitude and song signals.

Though the inserted magnitude values of Val (of image) are very small, therefore, appending extra values with original song (cover song) will not affect its overall audible quality but able to separate the original song from the similar songs.

2.3 Extraction

The decoding is performed using similar mathematical calculations. The algorithm of the same is given below.

Algorithm:
Input: Modified song signal with embedded image sampled values.
Output: Original song signal.
Method: The details of extraction of original song signal are given below.

Step 1: Separate amplitude and phase from song signal and find the diagonal detail coefficients (CD) as step 1 and 2 of embedding secret image [section 2.2].

Step 2: if 1 gap found between unequal values of two CDs then added digit is 0 , if n gaps found then digit is (n+1), where 0>=n>=3, if additional 2 gaps found, add the all previous digits to get the individual digit of intensity value. Again additional gaps 2 found then put previous individual digit side by side to get the final intensity value.

Step 3: Deduct the extra value from particular channel of CD[extra value should be added or subtracted based on negative or positive magnitude value of the channel] of amplitude signal comparing with CD of phase signal.

Step 4: Repeat step 2-3 until all extra values are deducted from CD's component of amplitude signal.

Step 5: Reconstruct the both amplitude and phase signal using the inverse DWT out of calculated wavelet and scaling coefficients and generate song signal with reconstructed amplitude and song signals.

3 Experimental Results

Encoding and decoding technique have been applied over 10 seconds recorded songs, the song is represented by complete procedure along with results in each intermediate step has been outlined in subsections 3.1.1 to 3.1.4. The results are discussed in two sections out of which 3.1 deals with result associated with IAWT and that of 3.2 gives a comparison with existing techniques.

3.1 Results

For experimental observation, a strip of 1 minute song ('The Catalyst', sang by Linkin Park) has been taken. Figure 1 shows amplitude-time graph of the original signal. Figure 2 is representing the concealed image. IAWT is applied on this signal and the output generated in the process is shown in figure 3. Figure 4 shows the difference of amplitude values before and after modification of original song. From figure 4 it is seen that the deviation of the modified song signal is very less, i.e., its audible quality will not be affected at all.

3.1.1 Original Recorded Song Signal (1 Minute)

The graphical representation of the original song, considering sampled values (441000) of x(n,2) [stereo type song] is given in the figure 1(a). Figure 1(b) and 1(c) are shown amplitude and phase signals respectively.

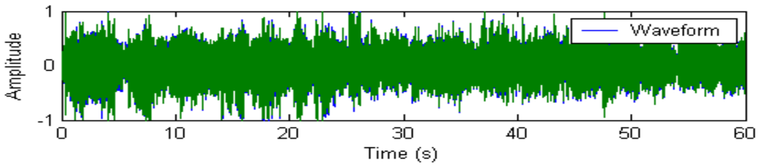
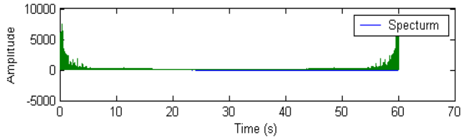


Fig. 1(a). Original song ('The Catalyst', sang by Linkin Park)



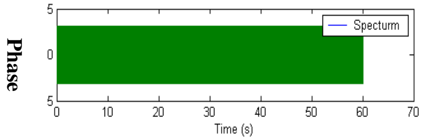


Fig. 1(b). Amplitude signal

Fig. 1(c). Phase signal

3.1.2 Hidden Image

The concealed image is shown in the figure 2.

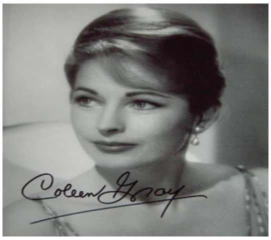


Fig. 2. Hidden Image ('Coleen Gray')

3.1.3 Modified Song after Embedding Secret Image (1 Minute)

The graphical representation of the modified authenticated song signal is shown in the figure 3.

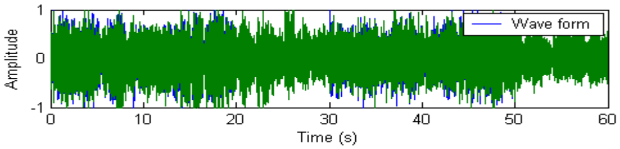


Fig. 3. Modified song with embedded image

3.1.4 The Difference of Magnitude Values Between Original and Modified

The graphical representation of the difference of magnitude values of original and modified songs is shown in the figure 4.

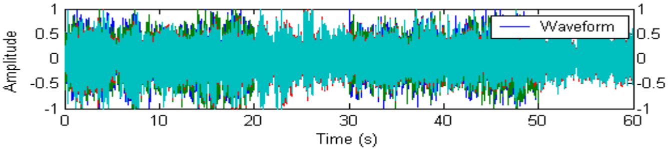


Fig. 4. The sampled values difference between signals figure 1 and 2

3.2 Comparison with Existing Systems

Various algorithms [5] are available for embedding information with audio signals. They usually do not care about the quality of audio but we are enforcing our authentication technique without changing the quality of song. A comparison study of properties of our proposed method with Data Hiding via Phase Manipulation of Audio Signals (DHPMA)[3] before and after embedding secret message/modifying parts of signal (16-bit stereo audio signals sampled at 44.1 kHz) is given in table 1, table2 and table3. Average absolute difference (AD) is used as the dissimilarity measurement between original song and modified song to justify the modified song. Whereas a lower value of AD signifies lesser error in the modified song. Normalized average absolute difference (NAD) is quantization error is to measure normalized distance to a range between 0 and 1. Mean square error (MSE) is the cumulative squared error between the embedded song and the original song. A lower value of MSE signifies lesser error in the embedded song. The SNR is used to measure how much a signal has been tainted by noise. It represents embedding errors between original song and modified song and calculated as the ratio of signal power (original song) to the noise power corrupting the signal. A ratio higher than 1:1 indicates more signal than noise. The PSNR is often used to assess the quality measurement between the original and a modified song. The higher the PSNR represents the better the quality of the modified song. Thus from our experimental results of benchmarking parameters (NAD, MSE, NMSE, SNR and PSNR) in proposed method obtain better performances without affecting the audio quality of song.

Table 1. Metric for different distortions

Sl No	**Statistical parameters for differential distortion**	**Value using IAWT**	**Value using DHPMA**
1	MD	0.1431	0.1854
2	AD	0.0320	0.0294
3	NAD	0.1421	0.1910
4	MSE	8.341e-004	0.0017
5	NMSE	0.0835	1.5271e-005

Table 2 gives the experimental results in terms of SNR (Signal to Noise Ratio) and PSNR (Peak signal to Noise Ratio). Table 3 represents comparative values of Normalized Cross-Correlation (NC) and Correlation Quality (QC) of proposed algorithm with DHPMA.

Table 2. SNR and PSNR

Sl No	Statistical parameters for differential distortion	Value using IAWT	Value using DHPMA
1	Signal to Noise Ratio (SNR)	18.4331	20.0130
2	Peak Signal to Noise Ratio (PSNR)	31.5487	27.5231

Table 3. Representing NC and QC

Sl No	Statistical parameters for correlation distortion	Value using IAWT	Value using DHPMA
1	Normalised Cross-Correlation (NC)	1	1
2	Correlation Quality (QC)	-0.0719	-0.0721

The Table 4 shows PSNR, SNR, BER (Bit Error Rate) and MOS (Mean opinion score) values for the proposed algorithm. Here all the BER values are 0. The figure 5 summarizes the results of this experimental test. It shows this algorithm's performance is stable for different types of audio signals.

Table 4. Showing SNR, PSNR BER, MOS

Audio (1s)	SNR	PSNR	BER	MOS
Song1	18.4331	31.5487	0	5
Song2	13.5429	31.2436	0	5
Song3	16.4528	32.3481	0	5
Song4	22. 9687	35.3215	0	5
Song5	14.2487	28.5905	0	5

This quality rating (Mean opinion score) is computed by using equation (5).

$$Quality = \frac{5}{1 + N * SNR} \tag{5}$$

Where N is a normalization constant and SNR is the measured signal to noise ratio. The ITU-R Rec. 500 quality rating is perfectly suited for this task, as it gives a quality rating on a scale of 1 to 5 [6]. Table 5 shows the rating scale, along with the quality level being represented.

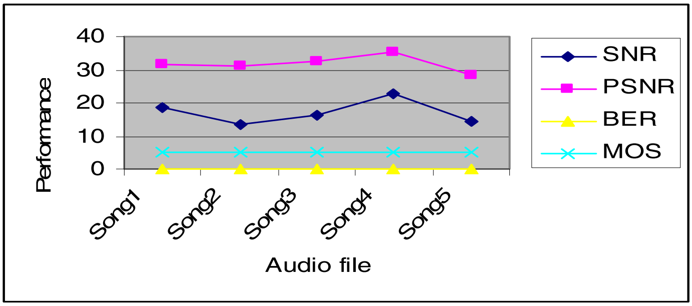


Fig. 5. Performance for different audio signals

Table 5. Quality rating scale

Rating	Impairment	Quality
5	Imperceptible	Excellent
4	Perceptible, not annoying	Good
3	Slightly annoying	Fair
2	Annoying	Poor
1	Very annoying	Bad

4 Conclusions and Future Work

In this paper, an algorithm for generating the hidden authenticating code with the help of wavelet transform for selected coefficient of song signal and embedding secret image in the specified region of that coefficient has been proposed which will not affect the song quality but it will ensure to detect the distortion of song signal characteristics. The modified song with authenticated secret image will not affect the song quality but ensure to detect the distortion of song signal characteristics.

This technique is developed based on the observation of characteristics of different songs but the mathematical model for representing the variation of those characteristics after modification may be formulated in future. It also can be extended to embed an audio into song signal instead of image or numeric values. The perfect estimation of percentage of threshold numbers of sample data of song that can be allow to change for a normal conditions will be done in future with all possibilities of errors.

Acknowledgments. The authors express gratitude to the Dept. of CSE & IT, College of Engg. & Management, Kolaghat and Dept of CSE and the PURSE scheme of DST, Govt. of India of University of Kalyani, under which the research has been carried out.

References

1. Mondal, U.K., Mandal, J.K.: A Practical Approach of Embedding Secret Key to Authenticate Tagore Songs(ESKATS). In: Wireless Information Networks & Business Information System Proceedings (WINBIS 2010), vol. 6(1), pp. 67–74. Rural Nepal Technical Academy (Pvt.) Ltd., Nepal (2010) ISSN 2091-0266
2. Mondal, U.K., Mandal, J.K.: A Novel Technique to Protect Piracy of Quality Songs through Amplitude Manipulation (PPAM). In: International Symposium on Electronic System Design (ISED 2010), pp. 246–250 (2010) ISBN 978-0-7695-4294-2
3. Xiaoxiao, D., Mark, F., Bocko, Z.I.: Data Hiding Via Phase Manipulation of Audio Signals. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2004), vol. 5, pp. 377–380 (2004) ISBN 0-7803-8484-9
4. Erten, G., Salam, F.: Voice Output Extraction by Signal Separation. In: ISCAS 1998, vol. 3, pp. 5–8 (1998) ISBN 07803-4455-3
5. Pohl, C., Van Genderen, J.L.: Multisensor image fusion in remote sensing concepts, methods and applications. International Journal of Remote Sensing 19(5), 823–854 (1998)
6. Arnold, M.: Audio watermarking: Features, applications and algorithms. In: IEEE International Conference on Multimedia and Expo., New York, NY, vol. 2, pp. 1013–1016 (2000)
7. Ioannidou, S., Karathanassi, V., Sarris, A.: The optimum wavelet-based fusion method for urban area mapping. In: WSEAS International Conference on Environment, Ecosystems and Development (2005)
8. Mondal, U.K., Mandal, J.K.: Generation and Fabrication of Authenticated Song Signal through Preserving Balance Ratio of Song's Components (BRSAS). In: International ssConference on Communication and Industrial Application (ICCIA 2011), Kolkata, India, December 26-28 (2011)

Specification Based IDS for Power Enhancement Related Vulnerabilities in AODV

Chaitali Biswas Dutta[1] and Utpal Biswas[2]

[1] Research Scholar, Dept of
CSE, University of
Kalyani, India
Asst. Prof., Dept of CA,
GIMT, Guwahati, India
mail.chaitali@yahoo.in
[2] University Of Kalyani
Dept of CSE, University of Kalyani,
Nadia, West Bengal, India
utpal01in@yahoo.com

Abstract. Wireless sensor network (WSN) is basically a wireless network, comprised of a large number of sensor nodes which are densely deployed, small in size, lightweight and portable. AODV is a well known, standardized routing protocol used in WSNs. AODV is subject to several attacks like black hole, worm hole, mad in the middle etc. Several Intrusion detection systems (IDS) have been proposed which successfully detect these attacks. Among these IDSs signature based and anomaly based are simple in nature but generate false alarms. To cater to this issue, recently specification based IDS is proposed for WSNs which have low false alarms yet detect most of the attacks. Lots of works have been reported on enhancement of AODV to improve throughput, PDR, NRO, End to End delay, power etc. Power Aware AODV (POW-AODV), enhances WNSs from the perspective of lifetime of nodes (in terms of power). In this paper we show that POW-AODV gets subject to more vulnerability, compared to AODV, in the effort to reduce power. Such attacks reduce life time of nodes instead of increasing them. Following that we propose a specification based IDS for this protocol to detect these vulnerabilities.

Keywords: Wireless Sensor Network, Ad-Hoc on Demand routing protocol, Fault-Tolerance.

1 Introduction

Now-a day's wireless technology has become very popular because of the convenience that comes with its use. Wireless sensor network (WSN) [1], [2] is basically a wireless network, is comprised of a large number of sensor nodes which are densely deployed, small in size, lightweight and portable. The WSNs are used in various important fields, like forest fire detection, flood detection, military purposes, tracking and monitoring doctors and patients inside a hospital, home application,

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 209–218.
springerlink.com

commercial application etc. Wireless network is highly dynamic. Topology changes, link breakage, node failure happen quite frequently. That is why routing is an important factor in case of wireless network. If nodes are within the range then routing is not required. Otherwise routing protocol is necessary because routing protocols specify that how routers communicate with each other. Routing protocols in wireless sensor networks are subdivided into proactive routing protocol and reactive routing protocol. Reactive protocols find the route only when there is data to be transmitted. As a result, it generates low control traffic and routing overhead. On the other hand, proactive protocols find paths in advance for all source and destination pairs. Also periodically exchange topology information to maintain them. AODV, DSR are the examples of reactive protocol and OLSR and DSDV are examples of proactive protocol.

Wireless network is not controlled in a centralized manner. It is really tuff to give protection the individual nodes. Inherent properties of ad hoc networks make them vulnerable. Malicious nodes can exploit these vulnerabilities to launch various kinds of attacks. So, Intrusion Detection Systems (IDS) [3] have become an essential component of computer security to detect attacks. IDSs are categorized in two major ways, the first one is based on location of deployment and the second one is based on attack detection methodology. Depending on the location of deployment, again IDSs are classified as Host-based Intrusion Detection System (HIDS) and Network-based Intrusion Detection System (NIDS). Depending on attack detection methodology, IDSs are divided into signature based, anomaly based and specification based. A signature based IDS will monitor data packets of the network and try to match them with the attributes of known malicious threats. Anomaly detection is a process which compares the data packets with some statistics. Signature based and anomaly based IDSs are very popular for a decades. But these two IDSs generate high number of false alarm. Also a few types of attacks are there which neither match with the pattern nor follow the statistics. To avoid this problem specification based IDSs are introduced.

In this paper we give a look on Ad-hoc On-Demand Vector (AODV) [4] routing protocol. Here we will introduce specification based IDS for this protocol. AODV is a very well known standardized reactive routing protocol. AODV's performance is measured in terms of parameters, like throughput, Packet Delivery Ratio (PDR), Normalized Routing Overhead (NRO), End to End delay, power etc. Lots of work has been done on enhancement of AODV to improve the quality of service of these parameters. While these enhancements have improved quality of service in WSNs, they have also lead to new vulnerabilities. Tan et al. have proposed POW-AODV in [32], which is able to provide more throughput than AODV. In this paper we will concentrate on POW-AODV and try to find out the vulnerabilities, also try to develop an event based IDSs for WSNs targeting attacks that arise due to enhancements of routing protocols.

2 IDS for Wireless Sensor Network

Intrusion stands for unauthorized access. The purpose of IDS is to define a boundary between authorized and unauthorized activity. Intrusion Detection Systems (IDSs) detect this type of access. A number of IDSs have been proposed to mitigate various kinds of attacks in WSNs. So, IDS have become an essential component of computer security to detect malicious attacks before they affect the wide network and/or system. There are two major ways for categorization of IDSs – i) based on location of deployment and ii) based on attack detection methodology. Depending on the location of deployment, IDS are classified as Host-based Intrusion Detection System (HIDS) and Network-based Intrusion Detection System (NIDS). A Host-based IDS monitors and analyzes internals of a computing system but a network-based IDS try to detect malicious activities by monitoring network traffic. Depending on attack detection methodology, IDSs are categorized as signature based [5], anomaly based [6] and event/specification based [7]. Signature based IDSs evaluate network traffic for well-known patterns or signatures to detect attacks. Unlike anomaly based IDSs, signature based IDSs have good detection rate only for known attacks. Its drawback is the inability to detect previously unseen attacks. Anomaly based IDS refer to the problem of identifying patterns in data that do not conform to expected or normal behavior. These non-conforming patterns are often referred to as anomalies, outliers and exceptions. This IDS has two phase–training phase and testing phase. These two detection systems are simple and very well known but both of them generate a high degree of false alarm. But specification based has been proposed as a promising alternative that combine the strengths of misuse and anomaly detection. Specification based IDSs use a set of rules to detect attacks. This IDS has the potential to detect previously unknown attacks. Anomaly based detection is capable of detecting novel attacks, but suffers from a high rate of false alarms. This is the main advantage of specification based system. It generates minimum false alarm in comparison of other two IDS. In this paper, we will apply specification-based techniques to monitor the AODV routing protocol.

3 The Ad-Hoc On-Demand Distance Vector Protocol

AODV is a well-known reactive and stateless routing protocol used in sensor network. The reactive routing protocols create and maintain routes only on demand. That is, routes between nodes are built when the source node desire. Reactive protocols usually use distance-vector routing algorithms. In our paper we will mainly concentrate on AODV protocol. It uses traditional routing tables and sequence numbers to determine the freshness of routes. An important feature of AODV is that a routing entry not recently used is expired. AODV uses mainly three control packets:

- Routing request message (RREQ) is broadcasted by a node requiring a route to another node,
- Routing reply message (RREP) is unicasted back to the source of RREQ,
- Route error message (RERR) is sent to notify other nodes of the loss of the link

AODV uses periodic HELLO messages to inform the neighbors that the link is still alive. AODV is vulnerable to different kinds of attacks.

When a source node requires a route to a destination it broadcasts a route request (RREQ) packet across the network. After receiving the RREQ a node may unicast a reply message (RREP). This node may be the destination node or may be an intermediate node. If it is a intermediate node then it only rebroadcast the message. If the nodes receive a RREQ which they have already processed, they do not forward it again and discard the RREQ. In AODV sequence number plays an important role. Sequence number Sequence number is increased under two conditions: i) when the source node initiates RREQ and ii) when the destination node replies with RREP.

Figure.1 illustrates the flow of the RREQ and RREP messages in a scenario wherein a node A wants to find a route to a node D. (Initially, nodes A, B, C and D do not have routes to each other). Node A broadcasts a RREQ message (a), which reaches to node B. Node B then re-broadcast the request (b). Node C receives the messages and broadcasts the message (c), which arrives at the destination node D. Last, D unicasts back the RREP message to A. We call these RREQ and RREP packets a request-reply flow.

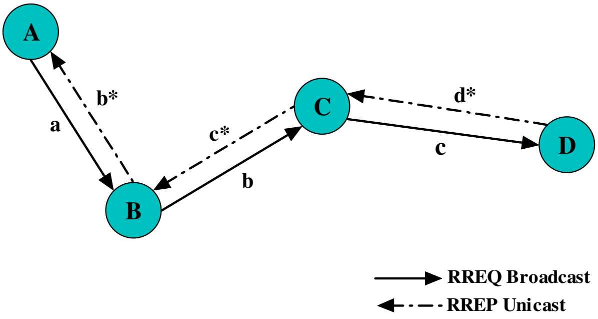


Fig. 1. Example of an AODV Scenario

Lots of work has been done on enhancement of AODV to improve quality of service. Performance of AODV have been measured in terms of parameters, like throughput [8], Packet Delivery Ratio (PDR) [9], Normalized Routing Overhead (NRO) [10], End to End delay [11], power [13]etc. In this paper we will consider the parameter power.

4 Enhancements of AODV for Power Aware Routing and Vulnerabilities

Tan et al. [13] have proposed Power Aware AODV (POW-AODV), which is able to provide more throughput than AODV and can make better use of the limited battery power available, is an extension or modified version of AODV.

POW-AODV approach considers a cost function based on the availability of the battery power. Here, the cost function of the overall route is the sum of the cost functions of the individual nodes along the route. The aim of PAW-AODV is to apply such an algorithm on AODV that can find a route with the least cost.

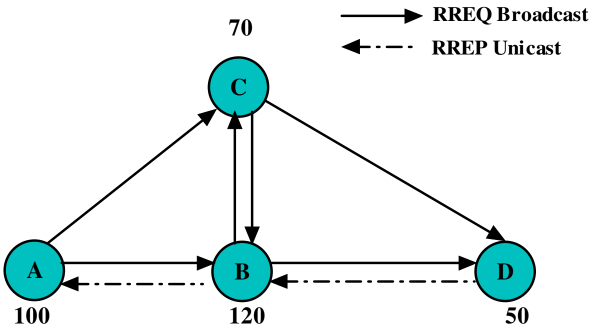


Fig. 2. Example to illustrate PAW-AODV

In Figure.2, A, B, C, D are some nodes of the WSN. Values of remaining battery life time corresponding to the nodes are also given in the figure (e.g., 100 Joules for A). RREQ and RREP sequences when A wants to communicate with D are shown in the same figure. In PAW-AODV more than one path from A to D will be found because intermediate nodes do not drop multiple RREQs with same source IP and Request ID.

i) RREQ ABD
ii) RREQ ABCD
The paths are generated are listed below:
i) ABD
ii) ABCD
iii) ACD
iv) ACBD

As discussed before, in POW-AODV cost function depends on the minimum remaining power of a node in a path, which is given in Table 1 for the example of Figure.2.

Table 1. Minimum remaining power of a node in a path for the example of Figure.2

Route	Minimum Remaining Power (Joules)
ABD	100
ABCD	70
ACD	70
ACBD	70

As path ABD has the maximum remaining power it will be considered and RREP will be unicasted through the path DBA.

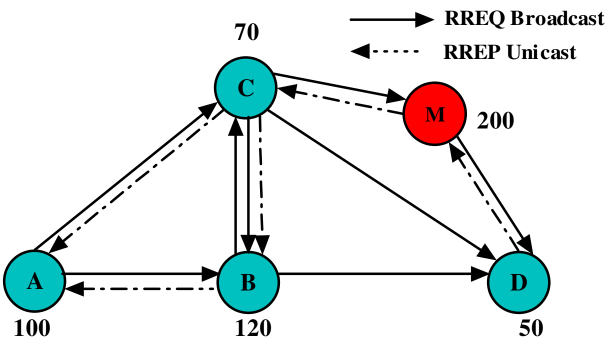


Fig. 3. Illustration of an attack against POW-AODV

POW-AODV enhances WSNs from the perspective of lifetime of nodes (in terms of power). However, POW-AODV can be easily exploited by attackers to do the reverse, i.e., use paths having nodes with less remaining battery power, thereby reducing lifetime of some nodes. Figure 3 illustrates such a case. In Figure.3, again we consider the same situation of Figure 2. But here one malicious node M exists between node C and D. In this case, when RREQ reaches D via A,C,M the cost function value should be 70 (due to low remaining power of C), however, M makes it 200 Joules.

This erroneously tells D that path "ACMD" can be taken considering remaining battery life time perspective. It may be easily noted that this path would lead to depletion of power of node C much faster than expected.

In this paper we will propose a scheme which will detect that type of malicious attacks. A number of Sensor Monitors observe the networks and maintain some search table from which monitor can say that this is a malicious attack or not. Then we try to resolve this problem.

5 Specification Based IDS for Vulnerabilities of POW-AODV

A sensor network is composed of a large number of sensor nodes. Sensor Monitors (SMs) observe and tracing RREQ and RREP messages in a request-reply flow and also try to detect unauthorized access. One SM covered a few number of sensor nodes. Each request-reply flow could have several branches and network monitor maintains a sensor table to trace the branches. When this monitor observed a new request (RREQ) packet, it searches the sensor table and tries to collect the details of the new packet. If SM is failed to match this current packet with the previous packet in sensor table, it contacts with its' neighbor SMs. If one of neighboring monitor give response, SM receives the details of its' previous packet as well as insert a new entry into the sensor table. Otherwise monitor consider it as an attack. Similarly, SM is also efficient to detect anomaly in case of reply (RREP) message. In case of AODV nodes receive a RREQ which they have already processed, they just drop the packet. Since, PAW-AODV wants to find a route with the least cost. So, it allows RREQ message to proceed even if this message has already processed. A SM then employs a finite state machine (FSM) to find out unmatched request or/and reply messages. Before discussing about the function of this state machine we have to enlist some basic assumption. May be these assumptions minimize the area of applications but in this paper we wish to work with these limitation. Basic assumptions are listed below:

- The MAC address of the nodes & the associated IP addresses are fixed & real, a table is maintained by the network monitors which contain the addresses.
- The network monitors pass messages securely & are authenticated to protect against spam.
- Every node must forward or respond messages to its neighbor using some protocol which must be performed within some time threshold.
- Every node must be under the preview of one network monitor at a particular instant of time which can be changed dynamically.
- There may be few nodes not responding the broadcast message, which will not affect the usual functionality.

5.1 Basic Block Diagram of a Sensor Monitor

Sensor network is comprised of a large number of sensors which is monitored by some sensor monitors. Sensor monitors observe route request (RREQ) and route reply (RREP) messages. Figure 4 illustrates the block diagram of sensor monitor.

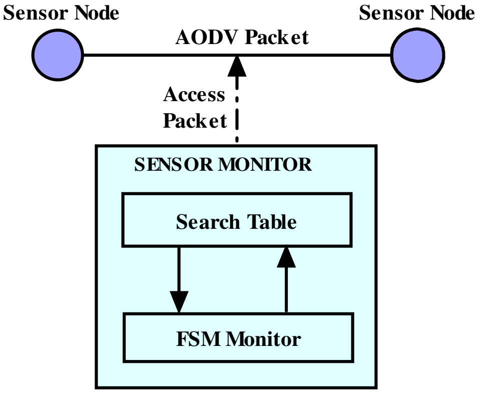


Fig. 4. Basic Block Diagram of Sensor Monitor

A sensor monitor maintains the record of the RREQ and RREP messages last received by each monitored node. Sensor monitors are stored the records as in a table format, which is called sensor table. When SM observes a new AODV packet it searches its sensor table and try to find out the previous packet of this packet. If sensor table is cannot provide the data to match the current packet with its previous packet then it will ask its neighboring monitors. If SM receives answer from one of its neighboring SM then no problem but if SM is unable to trace the previous packet then it detects it as an unauthorized access. SM also can mark out the problem like node failure or link failure and mention these as a RRER. Another part of SM is *FSM Monitor*. Search table decides the state with the help of this FSM monitor.

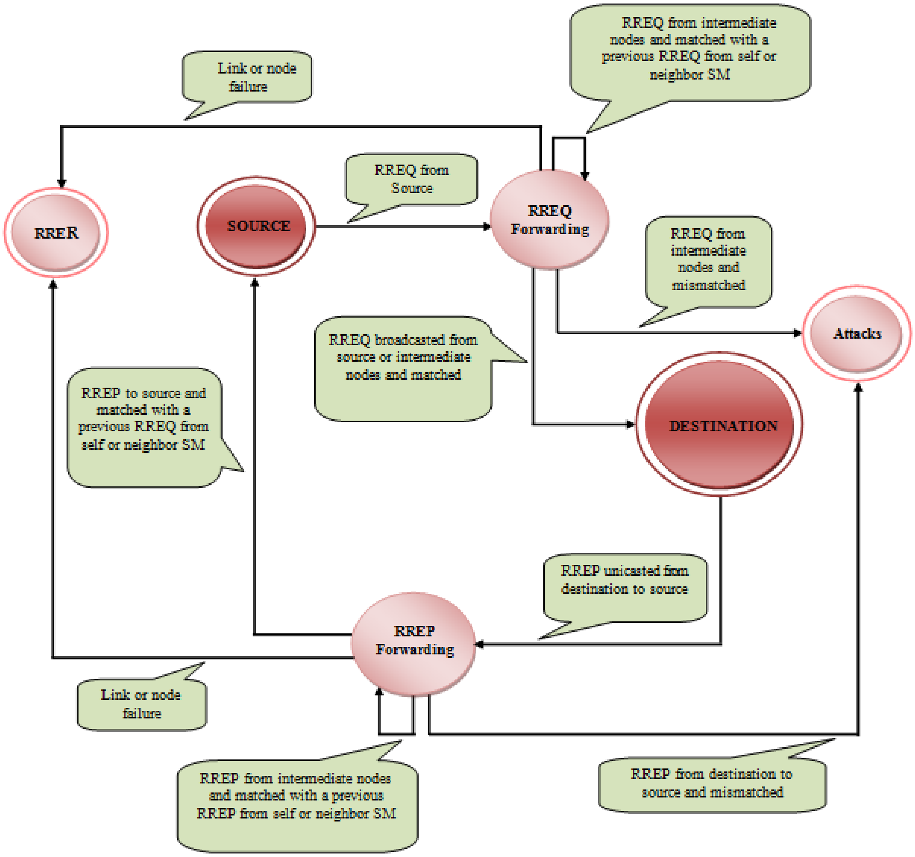

RREQ from intermediate nodes and matched with a previous RREQ from self or neighbor SM
Link or node failure
RREQ from Source
RRER
SOURCE
RREQ Forwarding
RREQ from intermediate nodes and mismatched
Attacks
RREQ broadcasted from source or intermediate nodes and matched
RREP to source and matched with a previous RREQ from self or neighbor SM
DESTINATION
RREP unicasted from destination to source
RREP Forwarding
Link or node failure
RREP from intermediate nodes and matched with a previous RREP from self or neighbor SM
RREP from destination to source and mismatched

Fig. 5. FSM to detect unmatched RREQ and RREP messages

5.2 Working Steps of Finite State Machine

Each sensor monitor uses a finite state machine (FSM) for detecting unauthorized RREQ and RREP messages. Figure 5 describes that type of state machine.

1. RREQ message is broadcasted from the source and go to RREQ Forwarding state.
2. When a packet move from one node to another node than sensor monitor sniff this packet and match with the sensor table.
3. If sensor monitor is able to find out that this packet is already registered then it allows the packet to move further.
4. If packet is not matched then the sensor monitor asks its neighbor monitors. Neighbor monitors search it in their sensor table. If the monitor get the details of

the packet then the monitor update its sensor table and allows the packet to go ahead. But if the detail of the packet is not found then monitor declared this is an incorrect forwarding. Then it goes to Attacks state.

5. If it is matched with the previous record then go to DESTINATION.
6. Reply message is unicasted from destination and go to RREP Forwarding state.
7. Step 2, 3, 4 are also same for RREP packet.
8. If any link or node is nor responding at the time of RREQ forwarding and RREP forwarding.

6 Conclusion and Future Work

In this paper we concentrated on Power Aware AODV (POW-AODV), which is an enhancement of AODV routing protocol to improve the quality of service for wireless sensor network. In normal case it gives far better result than that of traditional AODV. But POW-AODV is not enough concern about unauthorized access. As a result performance of POW-AODV is poor than that of traditional AODV if any malicious attacks will happen. In our paper we state that POW-AODV can hamper by what type of attack. Also try to introduce a detection mechanism system for this protocol. This mechanism system will able to find out the unauthorized access.

In our paper we state the possible vulnerabilities but a number of vulnerabilities can happen. In future we will study the different types of vulnerabilities and try to improve the detection mechanism system for these type of attacks. Further simulation needs to be done to show that POW-AODV with this detection mechanism performs well than the original at the time of unauthorized access.

References

1. Stankovic, J.: Wireless Sensor Networks. In: Handbook of Real-Time and Embedded Systems. CRC (2007)
2. Lewis, F.L.: Wireless sensor networks. In: Cook, D.J., Das, S.K. (eds.) Smart Environments: Technologies, Protocols and Applications. Wiley, New York (2004)
3. Patwardhan, A., Parker, J., Joshi, A., Karygiannis, A., Iorga, M.: Secure Routing and Intrusion Detection in Ad Hoc Networks. In: Third IEEE International Conference on Pervasive Computing and Communications (April 2005)
4. On-demand distance vector (AODV) routing, RFC 3561 (2003)
5. Roesch, M.: Snort-lightweight intrusion detection for networks. In: LISA 1999: Proceedings of the 13th USENIX System Administration Conference, pp. 229–238. USENIX Association (1999)
6. Chandol, V., Banerjee, A., Kumar, V.: Anomaly detection: A survey. ACM Computing Survey 41(3), 1–58 (2009)
7. Tseng, C.-Y., Balasubramanyam, P., Ko, C., Limprasittiporn, R., Rowe, J., Levitt, K.: A specification-based intrusion detection system for AODV. In: Proceedings of the 1st ACM Workshop on Security of ad hoc and Sensor Networks, pp. 125–134 (2003)

8. Alshanyour, A.M., Baroudi, U.: Bypass AODV: improving performance of ad hoc on-demand distance vector (AODV) routing protocol in wireless ad hoc networks. In: Proceedings of the 1st International Conference on Ambient Media and Systems, pp. 17:1–17:8 (2008)
9. Liu, J., Li, F.-M.: An Improvement of AODV Protocol Based on Reliable Delivery in Mobile Ad Hoc Networks. In: Proceedings of the 2009 Fifth International Conference on Information Assurance and Security, vol. 01, pp. 507–510 (2009)
10. Chen, D., Wang, X.: AODV with lower routing overhead. In: Proceedings of the 5th International Conference on Wireless Communications, Networking and Mobile Computing, pp. 2649–2652 (April 2009); Michalewicz, Z.: Genetic Algorithms + Data Structures = Evolution Programs, 3rd edn. Springer, Heidelberg (1996)
11. Sethi, S., Udgata, S.K.: SRMAODV: a scalable reliable MAODV for MANET. In: Proceedings of the International Conference and Workshop on Emerging Trends in Technology, pp. 368–373 (2010)
12. Bharathi, V., Poongkuzhali, T.: A Performance Enhancement of an Optimized Power Reactive Routing based on AODV Protocol for Mobile AD-HOC Network. International Journal of Technology and Engineering System 2(1), 39–45 (2011)
13. Tan, C.-W., Bose, S.K.: Modifying AODV for Efficient Power-Aware Routing in MANETs. In: IEEE TENCON, pp. 1–6 (2005)

English to Hindi Machine Translator Using GMT and RBMT Approach

Ambrish Srivastav[1] and Nitin Hambir[2]

[1] Assistant Professor, Dept. of Comp. Sci. and Engg.,
Swami Vivekanand College of Engineering,
Indore, India
a.srivastav30@gmail.com
[2] Assistant Professor, Dept. of Comp. Sci. and Engg.,
Acropolis Institute of Technology and Research,
Indore, India
nitin.hambir@gmail.com

Abstract. This research focuses on development of Grammar mapping technique and Rules Based Machine Translation (RBMT) approach for English to Hindi machine translator. Development of a machine translation (MT) system typically demands a large volume of computational resources. Rule based MT systems require extraction of syntactic and semantic knowledge in the form of rules. In this research I have to focus on grammar matching rule of both type of language (Source language and Target language).

Keywords: Machine translator (MT), Grammar mapping technique(GMT), Rules based machine translator(RBMT), Subject verb object(SVO), Subject object verb (SOV).

1 Introduction

Machine Translation (MT) is the process of translating text units of one language (source language) into a second language (target language) by using computers. In our approach, we focused on Grammar mapping of source language (English) to target language (Hindi). Under Grammar Mapping technique (GMT) we defined English to Hindi translation rules according to tense of sentences. We also focused on meaning of same word in English have different meaning in Hindi according to their gender. For instance, in English a verb *going* is used for both masculine and feminine noun but in Hindi translation *going* word translate according to gender of noun (For masculine noun- jaa raha and feminine noun- jaa rahi).

The need for MT is greatly felt in the modern age due to globalization of information, where global information base needs to be accessed from different parts of the world. Although most of this information is available online, the major difficulty in dealing with this information is that its language is primarily English. Starting from science, technology, education to manuals of gadgets, commercial advertisements, everywhere predominant presence of English as the medium of communication can be

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 219–225.
springerlink.com

easily observed. This world, however, is multi-lingual, where different languages are spoken in different regions. This necessitates the development of good MT systems for translating these works into other languages so that a larger population can access, retrieve and understand them.

2 Problem Definitions and Proposed Solution

English is a highly fixed order language with undeveloped morphology. In English language structure of sentence is Subject + Verb + Object (SVO). Hindi language is morphologically rich and relatively free word order. In Hindi language default sentence structure is Subject + Object + Verb (SOV). In addition, there are many stylistic differences. In translation process we have to organize target words of sentence according to their grammar. There are many differences in between English and Hindi which makes translation process difficult.

Machine Translation presents many challenges, of which the biggest is the ambiguity of English language. MT systems have to deal with ambiguity and various other Natural Language phenomena. In addition, the translation divergence between the source and target language makes MT a bigger challenge. This is particularly true of widely divergent languages such as English and Hindi languages.

2.1 Example Based Machine Translation (EBMT)

Example based machine translation is one such response against traditional models of translation. Like Statistical MT, it relies on large corpora and tries somewhat to reject traditional linguistic notions (although this does not restrict them entirely from using the said notions to improve their output).

2.2 Statistical- (or Corpus-) Based Machine Translation (SBMT)

Statistical translation models are trained on a sentence-aligned translation corpus, which is based on n-gram modeling, and probability distribution of the occurrence of a source-target language pair in a very large corpus.

2.3 Rule-Based Machine Translation (RBMT)

Here large number of rules is used for analysis and representation of the meaning of the source language texts, and the generation of equivalent target language texts.

2.4 Hybrid MT

In this approach, we merge two or more different machine translation approach for getting a correct sense of target language. For the solution of better output we fallow hybrid approach of Grammar mapping technique and Rules Based Machine Translation.

In a grammar mapping technique we define a grammar mapping rules for both language (source language and target language). Under the mapping technique we define a translation rules for source language to target language.

Ex- I am going. <English sentence>

मै जा रहा हूँ <हिन्दी अनुवाद>

Here grammar of English sentence is S_E →<noun><aux><verb>
and the corresponding grammar of Hindi sentence is
S_H→<noun><verb><aux>.

Under this grammar mapping technique we define a all English to Hindi translation rules.

For getting quality output of translator, it is necessary to create a strong database. In a Hindi language meaning of maximum word depends on its previous or next word or sense of sentence. So it is necessary to write a all meaning of word in a database.

Ex-- meaning of English word "You" is तुम, तुम्हें, तुमको, तुमने etc .

3 System Description

In a grammar mapping technique we define a grammar mapping rules for source to target language. Here we define a grammar mapping rules for one word, two word, and three word sentences.

3.1.1 One Word Sentences

One word sentences work like a dictionary. In one word sentences part of speech of source to target language not changed. There are some examples of one word sentences.

Come	→ <verb>	→ आना
Intelligent	→ <adjective>	→ बुद्धिमान
Was	→ <aux verb>	→ था
There	→ <adverb>	→ वहाँ
And	→ <conjunction>	→ और

3.1.2 Two Word Sentences

In a two word sentences grammar mapping rules and there example are –

S_E → < noun || pronoun> < verb >
S_H → < noun || pronoun > < verb >
I play (English sentence)
मै खेलता हूँ (हिन्दी वाक्य)

Alignment ((1,1),(2,2))

S_E → < adj > < noun >

S_H → < adj > < noun >

Beautiful color (English sentence)

सुन्दर रंग (हिन्दी वाक्य)

Alignment ((1,1),(2,2))

S_E → < aux > < noun || pronoun >

S_H → < noun || pronoun > < aux >

Is Ram (English sentence)

राम है (हिन्दी वाक्य)

Alignment ((1,2),(2,1))

S_E → < verb > < adverb >

S_H → < adverb > < verb >

Go there (English sentence)

वहाँ जाओ (हिन्दी वाक्य)

Alignment ((1,2),(2,1))

3.1.3 Three Word Sentences

In a three word sentences grammar mapping rules and there example are –

S_E → < noun || pronoun > < aux > < verb >

S_H → < noun || pronoun > < verb > < aux >

Ram is going (English sentence)

राम जा रहा है (हिन्दी वाक्य)

Alignment ((1,1),(2,3),(3,2))

S_E → < noun || pronoun > < verb > < adverb >

S_H → < noun || pronoun > < adverb > < verb >

You go there (English sentence)

तुम वहाँ जाओ (हिन्दी वाक्य)

Alignment ((1,1),(2,3),(3,2))

S_E → < aux > < pronoun > < noun >

S_H → < kya > < pronoun > < noun > < aux>

Is he ram (English sentence)

क्या वह राम है (हिन्दी वाक्य)

Alignment ((0,1),(1,4),(2,2),(3,3))

S_E → <aux > < noun || pronoun > < verb >

S_H → <kya > < noun || pronoun > < verb> < aux>

Is he going (English sentence)

क्या वह जा रहा है (हिन्दी वाक्य)

Alignment ((0,1),(1,4),(2,2),(3,3))

S_E → < wh > < aux > < noun || pronoun >

S_H → < noun || pronoun > < wh > < aux>

Who are you (English sentence)

तुम कौन हो (हिन्दी वाक्य)

Alignment ((1,2),(2,3),(3,1))

S_E → < please > < verb > < adverb >

S_H → < please > < adverb > < verb>

Please go there (English sentence)

कृप्या उधर जाओ (हिन्दी वाक्य)

Alignment ((1,1),(2,3),(3,2))

Under this grammar mapping technique, define English to Hindi translation rules.

4 System Architecture

The system architecture, as shown in figure 1, has the following stages through which the source text is passed.

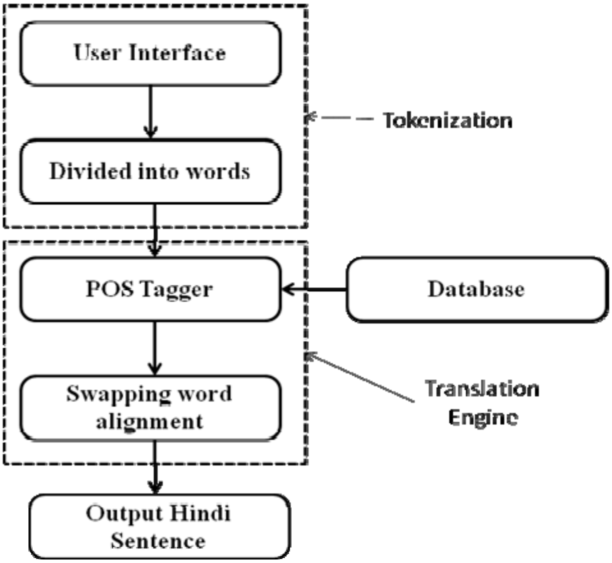


Fig. 1. System Architecture of translator

4.1 Tokenization

Tokenization consist user interface which takes English sentence from user and tokens generator that divide given sentence into words. Tokens are separated by using break characters like space, comma, question mark etc.

4.2 Translation Engine

The translation engine takes tokens from Tokenization and it is responsible for two task first one it is find tag of each English word and finds the match of a given token in Hindi Language. For this Translation Engine uses lexical resources. Second one it is align the words of Hindi language according to Hindi grammar structure.

4.3 Implementation Steps

1. Firstly write an English sentence in a Text Field. After pressing OK button, English sentence decompose into a token by StringTokenzier.
2. Every token firstly get its tag (part of speech) from the dictionary table of database.
3. According to position of English word tag we define a grammar mapping and English grammar rules for English to Hindi machine translator. Every English token gets its Hindi meaning from different table of database.
   Ex. If in an English sentence "going" word come then translator gets its Hindi meaning from table "verb4th".
4. We define a grammar mapping and corresponding English to Hindi translation rules for one word, two words three words sentences.

Ex. One word sentences: -

One word sentences behave like a dictionary. We get a Hindi meaning and part of speech tag of English word from the dictionary table.

Beautiful,

Intelligent,

Going etc are one word sentences.

Ex. Two word sentences:-

Ram comes	< noun > < verb >
He was	< pronoun > < aux verb >
Go there	< verb > < adverb >

Ex. Three word sentence:-

I am going <pronoun><aux verb> <verb>

Ram is intelligent <noun><aux verb> < adjective >

Who are you <wh><aux verb><pronoun>

5. In a database I write a every possible Hindi meaning of English word.

Ex. Going	(जा रहा, जा रही, जा रहे)
I	(मै, मुझे)

Could have (सकता था, सकती था,
सकते थे)

6. I use Unicode string for writing Hindi words.

Ex. For Hindi word “ जा रहा ” I write “\u091c\u093e \u0930\u0939\u093e”.

5 Results

We implemented the English to Hindi machine translator in which we trying to remove a problem of Google translator. We create a translator for three words sentence and find out the correct result compare to Google translator.

6 Conclusions

This translator translates the one word, two word, and three word sentences of English language into a corresponding Hindi language sentences. For a better output it is necessary to increase the database and translation grammar mapping rules.

For the translation of more then three word sentences, we create a grammar mapping rules for grater then three word sentence.

For the large sentence, we search the conjunction word from the English sentence. Decompose a sentence into more sentences from a conjunction word.

Translate an every sentence into a corresponding Hindi sentence and merge all sentences it is necessary to define a large number of grammar mapping rules for the source language to target language for the better output and also create a large database. There are more then 1.75Lac word in a English dictionary. So it is necessary that store all meaning of a particular English word in our database. For a good translator, database must be a very strong.

References

[1] Martin, J.: Chapter 20 Natural language Generation and Chapter 21 Machine Translation. In: Speech and Language Processing (2005) (Fourth Indian Reprint)

[2] Probst, K.: Learning Transfer Rules for Machine Translation with Limited Data, August 15. Language Technologies Institute School of Computer Science Carnegie Mellon University Pittsburgh, Pennsylvania (2005)

[3] Gupta, D.: Contributions to English to Hindi Machine Translation using Example-based approach, New Delhi-110016, India. Department of Mathematics Indian Institute of Technology Delhi Hauzkhas (January 2005)

[4] Word Alignment in English-Hindi Parallel Corpus Using Recency Vector Approach: Some Studies Chatterjee, N., Dept. of Mathematics, Indian Institute of echnology Delhi Hauz Khas, New Delhi, Agrawal, S., Department of Mathematics, Indian Institute of Technology Kharagpur, West Bengal, INDIA - 721302

[5] Chang, J.-S., Su, K.-Y.: Corpus-Based Statistics-Oriented (CBSO) Machine Translation. Researches in Taiwan. Behavior Design Corporation 2F, No. 5, Industrial East Road IV, Science- Based Industrial Park, Hsinchu, Taiwan 30077, R.O.C

Plugging DHCP Security Holes Using S-DHCP

Amit Kumar Srivastava[1] and Arun Kumar Misra[2]

[1] M.Tech(Information Secuirty), Computer Science and Engineering Department
Motilal Nehru National Institute of Technology, Allahabad
amitksrivastava170405@gmail.com
[2] Professor, Computer Science and Engineering Department
Motilal Nehru National Institute of Technology, Allahabad
akm@mnnit.ac.in

Abstract. Dynamic Host Configuration Protocol(DHCP) allows the terminals to have IP address for their introduction in to the network. In this paper, a secure DHCP system is proposed, not only for user authentication but also for addressing the other security issues like confidentiality, integrity and privacy. Proposed S-DHCP, makes use of the users legitimate right for an IP address such that legitimate user on a particular terminal can exercise it's right to get allocated and use an IP address from S-DHCP. Thus S-DHCP protects unauthenticated and unauthorized access of confidential informationat one hand which keeping their integrity intact in a private mode from within the network to the outside world over Internet. Simulation and preventing the fabrication of IP addresses and MAC addresses leading to unauthorized access control is difficult and our effort is to incorporate session-ID for user authentication without any modification on DHCP client to produce the desired result.This paper attempts to introduce proposed S-DHCP, its principle, to plug vulnerabilities and other efforts on optimization over the design of protocols in a secured environment.

Keywords: Access Control, Attribute Certificate, Authentication, DHCP, Identity Certificate X.509, User Authentication.

1 Introduction

Dynamic Host Configuration Protocol (DHCP) takes care of dynamic attribution of Internet addresses in both manual and in automated modes. If the nodes are mobile in the network, such allocation of IP addresses is typically governed by the rules of Mobile and Adhoc network (MANET). MANET does not allow static attribution of address. It is, therefore, required to be adapted to dynamic addressing by a suitable protocol. DHCP allows the terminals to have IP address for their introduction into the network. Originally DHCP server did not perform any authentication of clients and as such any intruder can avail the facilities of DHCP server as a legitimate client. The intruder in this way can change the MAC address or any other identification parameter used for the purpose. This may initiate denial of service attack at ease.DHCP server authenticates the

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 227–238.
springerlink.com

terminal through its MAC address rather than the client. A situation may arise when a host leaves the network and does not return its IP address to the DHCP server and it simply becomes a zombie.

Mobile Ad-hoc Network is equally vulnerable to malicious internet host like hosts in the wired network. The protocols have to perform in truly heterogeneous wireless environments where mobile nodes may enter or leave across different networks. Such movements of mobile nodes cause change in the allocated IP address. Such changes may not be broadcasted at once when they enter or leave and a node becomes unreachable. Such vulnerability makes the node self generated Denial-of-Service (DOS) attacked node. The problems are typically addressed by location manager using DNS dynamic updates.Authentication and authorization in DNS dynamic updates are addressed using zone security servers.Authentications can be done by using DNSSEC SIG (0) or through TSIG [Wellington, 2000].

In redirection attack, the attacker updates its own IP address against the domain name of another node. Thus, the attacker's IP address will be sent back in response to the queries against the compromised domain name.DHCP is authorized to make updates on behalf of other nodes. The IP address, acquired from DHCP, is updated for a corresponding entry in DNS.

Komori and Satio in 2002[6] suggested a technique for a concept of session between the DHCP server and the user which is not purely user authentication oriented. DHCPs legally valid MAC addresses are related to register terminals that can use legitimately obtained IP address. The attackers deceive DHCP to obtain a legitimate IP address followed by subsequent hacking. In contrast, our proposed S-DHCP supports user authentication and other security issues.

2 System Environment

The main purpose of S-DHCP is to protect unauthenticated and unauthorized access of information keeping their integrity from within the network and over internet.So the system environment of S-DHCP has been decided based upon these basic objectives.

2.1 System Architecture of S-DHCP

S-DHCP system is composed of following units that are interrelated by various actions (Fig.1).

Gateway: It is based on the configured router. The router having NAT functionality is capable of IP filtering performed by software or configured hardware.

DHCP server: The DHCP server, having server side programming capabilities of allocating an IP address to the DHCP client. DHCP server follows design outline of RFC 2131 and RFC 2132. S-DHCP is built upon this server by add-on security features.

DHCP client: It is actually client side programs that can request the allocation of an IP address from the DHCP server. These clients are required to be authenticated by the proposed authentication mechanism before accessing any of the resources of the network.

Authentication Server: The user is authenticated by the software of the server of a given IP address.

Authentication Client: This is client side software capable of communicating with the authentication server.

Confirmation Server: The confirmation server confirm the Authentication client on the basis of successful authentication.

Certificate Issue Server : The server is capable of auto-generating the digital certificate for DHCP server as well as confirmation server (if requested)

Database Server: The database of S-DHCP is stored in database server.

The S-DHCP client consists of DHCP client (not secured) as authenticated client (secured).

2.2 Requirement of S-DHCP

The requirement of S-DHCP in order to establish a robust DHCP system are as follows (Fig.2)

a. Legitimate users of S-DHCP should register their user ID and password with the Authentication server in order to have an IP address.
b. There is a provision to change the password for the existing registered users and registration of log on ID and password for the new user of the network.
c. The password is required to be stored in a hashed password file in S-DHCP system.
d. S-DHCP client becomes activated only when S-DHCP is installed in registered authenticated clients node.
e. The local area network is typically separated from the Internet or outside network by the gateway. If an unauthorized user uses an IP address, it is considered illegal access to the system.
f. S-DHCP exists by its software installed in the client and server of DHCP system.
g. The DHCP server allocates a local IP address in the private network to the authorized users. Having an illegitimate local IP address will not allow to communicate with the outside network through the Gateway.

3 Designed Behaviour of S-DHCP

The network structure of S-DHCP as presented in Fig.1 has been elaborated in system sequence diagram presented in Fig. 2. The designed behaviour is achieved through the steps Fig.3 and Fig.4. The legitimate user initiates the action of S-DHCP by registering the user ID and the password with S-DHCP server. Typical connection mechanism of the user node is through a NIC (Network Interface Card) and through the DHCP client. The user obtains an IP address from the DHCP server. This authentication is actually the Lease Phase of the S-DHCP design. If the S-DHCP client becomes in active due to timeout or any other reason, then it will enter into the restart phase again. It will require an IP address before next timeout. The timeout is usually gets fixed for confirmation and is called confirmation time.

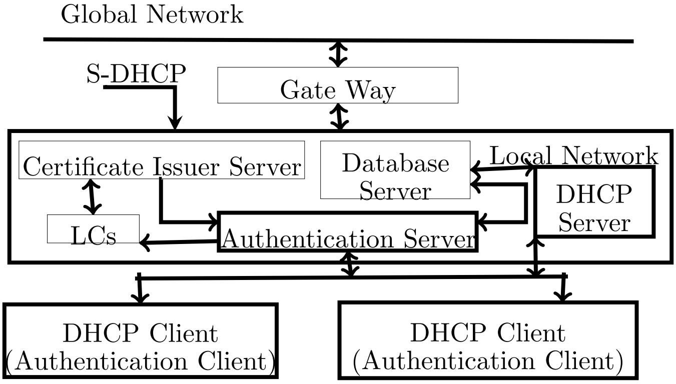

Global Network
S-DHCP
Gate Way
Certificate Issuer Server
Database Server
Local Network
DHCP Server
LCs
Authentication Server
DHCP Client (Authentication Client)
DHCP Client (Authentication Client)

Fig. 1. Block Diagram of Architecture

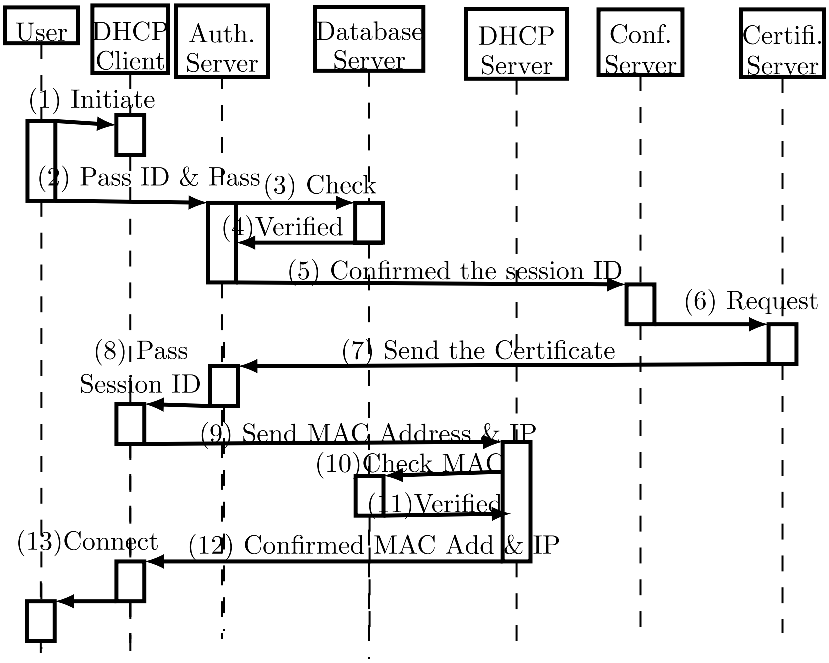


Fig. 2. Sequence Diagram for S-DHCP

It may be designed to be duration of 2-5 minutes. Within this confirmation time the authentication client will get reconfigured and will try to make the communication between the Authentication client and Authentication server. This phase is turned as Re-authentication phase. In this phase, an IP address is required again and typically Lease phase is repeated. If S-DHCPs client system is not currently going through the time out or any other system interruption process, the session is permitted to be established with in this confirmation time.

After the completion of Lease Phase and Re-Authentication phase successfully, the Confirmation server confirm the session with the Authentication client. The approval of session is known as confirmation Phase. It may happen that the user decides to release the IP address due to his leaving of the network or for any other reason. If the user releases the IP address the Release Phase is initiated.

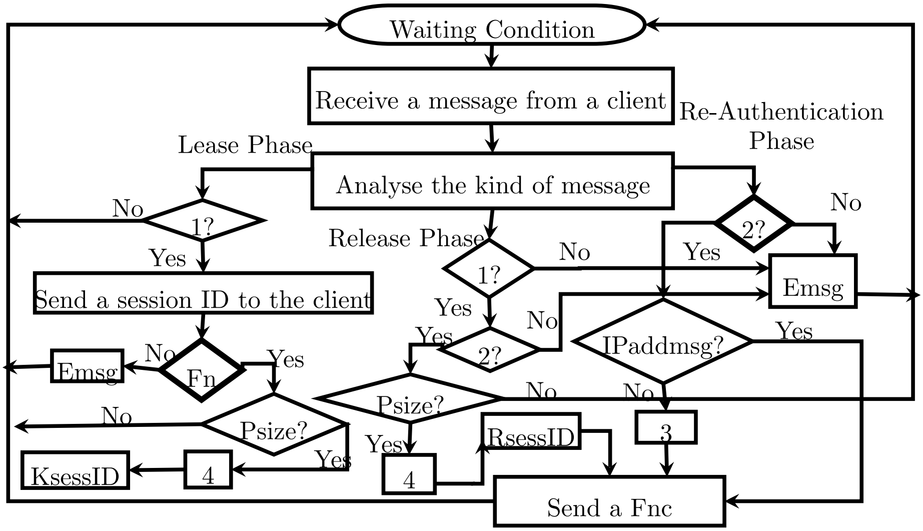


Fig. 3. The Designed behaviour of the Authentication Server

Description of Fig. 3 varibles are:
1=does the ID and the password of user exist?
2= does a session ID exist in the database?
3=rewrite the IP address in the Database to the IPaddress with in the msg and setup the routing table of the router again.
4= Seq.No. stored in the database.
Emsg= send an error message to the client.
Fn= was authentication finished normally?
Fnc=normality finishing message to the client.
IPaddmsg= does the IP address within the msg align with the database?
KsessID= the setup of gateway is rewritten and the the session ID is kept in the Database.
Psize= Paket Size(=56byte)
RsessID=the setup of gateway is rewritten and the session ID is removed from the database.

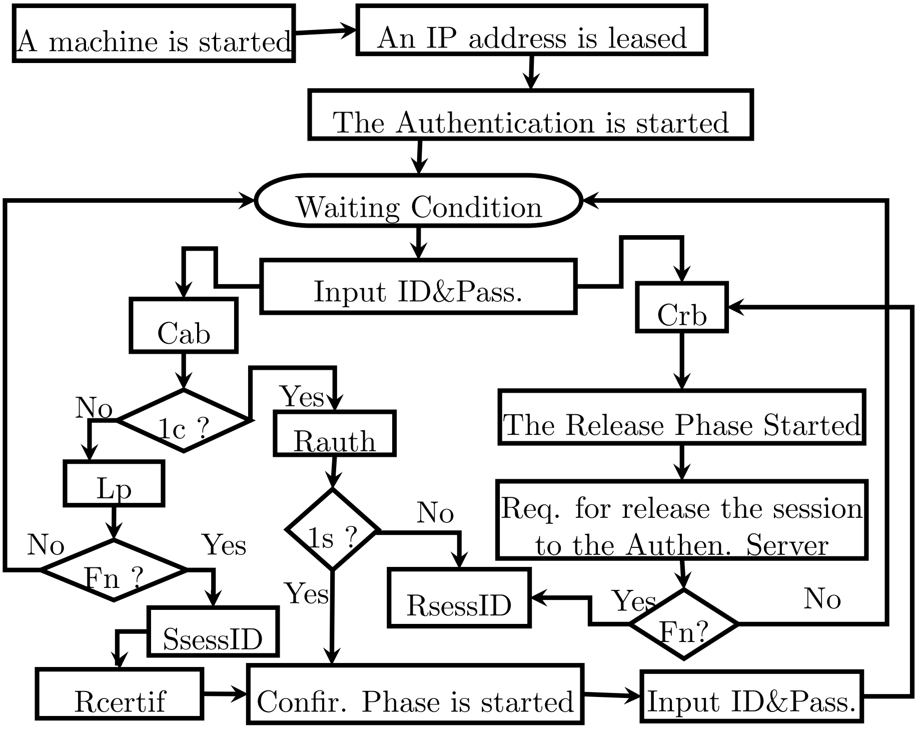


Fig. 4. The Designed behaviour of the Authentication Client

Description of Fig 4 varibles are: Cab= Click the authentication button. Crb= Click the release button. 1c=Does the Authen. client have the session ID. Lp= The lease phase is started. Rauth= The Re-Authen.Phase is started. Fn= Is Finished Normally? 1s= Does the Authen. Server have the session ID? SsessID= Save the session ID to the local file. RsessID= Remove the session ID from the local file. Rcertif= Req. the certificate.

3.1 Phase-Wise Protocol Formulation of S-DHCP

S-DHCP Server goes through four defined phases, namely, the Lease Phase, the Release Phase, the Re-Authentication phase and the Confirmation phase.

3.1.1 Terms of Communication of S-DHCP

Table 1 gives description and their corresponding message codes and Table 2 shows the sizes of the variables used in communication protocol.The sizes of variables are required to be decided in such a way that the overhead becomes minimum. For example, Nonce is a string representing a random number that remains non-existent before it is actually generated. A nonce is dependent on algorithms of nonce generation

Table 1. Description of Variables with Corresponding Message Code

Code	Meaning	Code	Meaning
M_l	Message for Lease Phase	ID_{Lc}	Client Identifier
M_{lerr}	Error message for Lease Phase	N_c	Client Nonce
M_{rel}	Message for release phase	T_1	Time Stamp by client
M_{relerr}	Error for release message	$MAC_{add(Lc)}$	MAC address for Client
M_{reauth}	Message for re-authentication phase	sessID	Session ID
$M_{reautherr}$	Error message for re-authentication phase	$MAC_{add(Ls)}$	MAC address for authentication server
M_{conf}	Message for confirmation phase	ID_{Ls}	Server Identifier
$M_{conferr}$	Error message for confirmation phase	T_2	Time stamp by server
$H(M_l)$	Hash value message for lease phase	N_{Cs}	Nonce of confirmation server
$H(M_{lerr})$	Hash value error message for release phase	L_c	Client
$H(M_{rel})$	Hash value message for release phase	L_s	Server
$H(M_{relerr})$	Hash value error message for release phase	L_{Cs}	Confirmation server
$H(M_{reauth})$	Hash value message for authentication phase	C_s	Certificate
$H(M_{reautherr})$	Hash value error message for authentication	IP	IP address
$H(M_{conf})$	Hash value message for confirmation phase	$H(M_{conferr})$	Hash value for error message of confirmation phase

Table 2. Size of Variables of Message of Each Phase

Variables	Size	Variable	Size
Message for each phase	1	Session ID	4
Hash value of Each phase message	16	MAC address of the Authentication server	16
Identifier of a User	4	IP address	4
Nonce of the Authentication client	8	Identifier of Authentication server	4
Time stamp by Authentication client	1	Time stamp by Authentication Server	1
MAC address of the Authentication client	16	Nonce of the Confirmation server	8

A nonce usually is obtained from a range of pseudo random number generators. This range of numbers should be large enough for securely purpose but an eye should be kept on the application as well so that unnecessarily it should not increase the overhead for no reason. A trade off between the range of nonces and its selection for a purpose is to be mode for optional utilization of network resources in a wireless medium. Moreover, nonce are generated fresh in each phase, every time it is required. The selection of size to accommodate a specified range must not slow down the transmission in a wireless (or medium). The terms like Epass denote encryption of the message M with the key pass.

3.1.2 Lease Phase of S-DHCP

This phase of proposed communication protocol allows sharing of Session ID (SessID). The Server and Client (L_c andL_s respectively) authenticate each other and confirm the legitimacy of use in a secured environment. The security requirement initiates a process that can be expressed in the following manner:

$(1)L_c \rightarrow L_s : E_{pass}(M_l||H(M_l)||N_c||ID_{Lc}||T_1)$
$(2)L_s \rightarrow L_c : E_{pass}(M_l||H(M_l)||ID_{Lc}||MAC_{add(Ls)}||sessID$
$||N_c||T_1||T_2)$
$(3)L_c \rightarrow L_s : E_{pass}((M_l||H(M_l))||ID_{Lc}||MAC_{add(Lc)}||sessID||IP)$
$(4)L_s \rightarrow L_c : E_{pass}((M_l||H(M_l))||ID_{Lc}||sessID)$
$OrE_{pass}((M_{lerr}||H(M_{lerr}))||ID_{Lc}||sessID)$

3.1.3 Release Phase of S-DHCP

The purpose of the phase of the proposed Communication protocol of S-DHCP is to release the session between Authentication Client and Authentication Server. The Authentication Client and the Authentication Server authenticate each other with Session ID (SessID) and confirms the legitimacy of use in a secured environment. The procedural phase is depicted between the Authentication Client and Authentication Server as in following manner.

$(1)L_c \rightarrow L_s : E_{pass}((M_{rel}||H(M_{rel}))||ID_{Lc}||N_c||sessID||T_1)$
$(2)L_s \rightarrow L_c : E_{pass}((M_{rel}||H(M_{rel}))||ID_{Lc}||N_c||T_1||ID_{Ls})$
Or
$E_{pass}((M_{relerr}||H(M_{relerr}))||ID_{Lc}||ID_{Ls}||N_c||T_1)$

The step 1 is a demand of the release of the Session between Authentication Client and the Authentication Server. SessID specifies the session to be released. If the release is successful, both the Authentication Client and Server clear the earlier state and enters into the other state of logout (from login).

3.1.4 Re-authentication Phase of S-DHCP

The main purpose of the Re-Authentication phase of the proposed communication protocol of S-DHCP is for Authentication Client to confirm with Authentication Server in order to continue the former Session with the same session ID.

Again, the Authentication Client and the Authentication Server authenticate each other with Session ID (SessID) and confirm the legitimacy of use of the formal session in a secured environment. The procedure between Authentication Client and Authentication Server has been depicted in following manner.

$(1)L_c \rightarrow L_s : E_{pass}((M_{reauth}||H(M_{reauth}))||N_c||ID_{Lc}||sessID||IP||T_1)$
$(2)L_s \rightarrow L_c : E_{pass}((M_{reauth}||H(M_{reauth}))||ID_{Ls}||N_c||sessID||T_1)$
Or
$E_{pass}((M_{reautherr}||H(M_{reautherr}))||LD_{Ls}||N_c||sessID||T_1)$

The step 1 demands confirmation of the continuation of the former session with the same Session ID (SessID). If the continuation of the Session can be confirmed, the authentication Client continues the session in the second step.

3.1.5 Confirmation Phase of S-DHCP

The Confirmation phase of the proposed communication protocol of S-DHCP confirms authentication by Authentication client. The Authentication Client and Confirmation Server authenticate each other with the Session ID (SessID). The process confirms the legitimacy of use of the Session in a secured environment. The procedure between the Authentication Client and confirmation server presented as follows:

$(1)L_{Cs} \rightarrow L_c : E_{pass}((M_{conf}||H(M_{conf})]||sessID||N_{Cs}||T_3$
$(2)L_c \rightarrow L_{Cs} : E_{pass}((M_{conf}||H(M_{conf}))||sessID||N_{Cs}||N_c||T_1||T_3$
Or
$E_{pass}((M_{conferr}||H(M_{conferr}))||sessID||N_{Cs}||N_c||T_1||T_3$

The step1 depicts the exchange of message for continuation of a present Session Confirm The confirmation session confirms the continuation of the Session by confirming in step 2.

3.1.6 Privacy Enhancement of S-DHCP

The issues related to privacy and integrity of communication (message data) have not been addressed before. However, the proposed communication protocol may be designed in the way again described by the four phases to take care of privacy and integrity of data and hence of the user in a wireless environment. This designed protocol is likely to plug a sizeable chunk of vulnerabilities experienced by DHCP servers.

The security of S-DHCP is further enhanced by not allowing an illegitimate user detected by Detection Server and will be prevented from communicating with the outside network. The related sequential communication are depicted through the protocol presentation.

1.Lease Phase:
$(1)L_c \rightarrow L_s : [E_{pass}(M_l||H(M_l)]||E_{uk}(ID_{Lc})||N_c||T_1$
$(2)L_s \rightarrow L_c : E_{pass}((M_l||H(M_l))||E_{uk}(ID_{Lc})||MAC_{add(Ls)}||sessID||N_c||$
$T_1||T_2||C_s)$
$(3)L_c \rightarrow L_s : E_{pass}((M_l||H(M_l))||E_{uk}(ID_{Lc})||MAC_{add(Lc)}||sessID||IP||C_s)$

$(4) L_s \rightarrow L_c : E_{pass}(M_l||H(M_l))||E_{uk}(ID_{Lc})||sessID||C_s$
Or
$E_{pass}(M_{lerr}||H(M_{lerr}))||E_{uk}(ID_{Lc})||sessID||C_s$

2.Release Phase Protocol:
$(1) L_c \rightarrow L_s : E_{pass}(M_{rel}||H(M_{rel}))||E_{uk}(ID_{Lc})||sessID||N_c||T_1$
$(2) L_s \rightarrow L_c : E_{pass}(M_{rel}||H(M_{rel}))||E_{uk}(ID_{Lc})||N_c||T_1||ID_{Ls}||C_s$
Or
$E_{pass}(M_{relerr}||H(M_{relerr}))||E_{uk}(ID_{Lc}))||N_c||T_1||ID_{Ls}||C_s$

3. Re-Authentication Phase Protocol
$(1) L_c \rightarrow L_s : E_{pass}(M_{reauth}||H(M_{reauth}))||E_{uk}(ID_{Lc})||sessID||N_c||IP||T_1$
$(2) L_s \rightarrow L_c : E_{pass}(M_{reauth}||H(M_{reauth}))||ID_{Ls}||N_c||sessID||T_1||C_s$
$Or E_{pass}(M_{reautherr}||H(M_{reautherr}))||LD_{Ls}||N_c||sessID||T_1||C_s$

4.Confirmation Phase Protocol
$(1) L_{Cs} \rightarrow L_c : E_{pass}(M_{conf}||H(M_{conf}))||sessID||N_{Cs}||C_s$
$(2) L_c \rightarrow L_{Cs} : E_{pass}(M_{conf}||H(M_{conf}))||sessID||N_{Cs}||N_c$
$Or E_{pass}(M_{conferr}||H(M_{conferr}))||sessID||N_{Cs}||N_c$

4 Privacy Enhancement Of S-DHCP

The privacy of the user is ensured in a number of circumstances such as

1. An illegitimate user spoofs a local IP address that is no more effective This type of simulation may arise when user authentication remains incomplete. The gateway is designed to filter local IP address and therefore, the illegitimate user cannot connect to outside network. The detection server promptly detects the user. The Detection server keeps an eye over all the IP addresses within packet header that flows through the network.
2. An illegitimate user may spoof a legitimate local IP address having the legitimate user status as login. This situation is very similar to hacking of a session of a legitimate user and consumption of illegal resources and loss of privacy of the legitimate user. If there is an IP address conflict between two or more users, only one of the users (legitimate one) is able to communicate with the outside network through the Gateway. The legitimate user can be detected by the Detection Server and confirmed by the Confirmation Server through the session ID along with time stamp or matching of MAC address by Detection Server extracted from packets flowing through the network. Communication may then be terminated in a natural way. The introduction of Detection Server and using its functionality as a double check is to check the vulnerability related to IP address spoofing and the same results in privacy enhancement of the user.
3. An illegitimate user may use a legitimate local IP address which the legitimate user is actually not using or moved outside the network.The Detection

Server can detects an illegitimate user by the mechanism available in the Detection server for checking the MAC address with local IP address.

4. An illegitimate user can cause harm to privacy of a legitimate user by spoofing a legitimate local IP address and forging a legitimate corresponding MAC address while the legitimate user is actually no more connected to the network or moved outside the network.

The other things (local IP address, the corresponding MAC) remaining the same, the illegitimate user can be detected by the Detection Server the session ID (SessID) and corresponding timestamp. This set of session ID and timestamp behaves like an unique identifier and detection becomes full proof enhancing privacy of user in a considerable way.

5 Conclusion

The proposed S-DHCP system addresses the issues related to user authentication successfully. The system, however, remains extensible in the sense that it is also able to take care of the issues like confidentiality, privacy and integrity of the message of the user. At the same time it protects the privacy of the user. The system not only stores the local IP address and corresponding MAC address for registration but also derives further detection mechanism through session ID and time stamp storage that behaves like an unique identifier. The local IP address thus gets checked doubly and protects the legitimate user. It blocks the communication of an illegitimate user by blocking the connection of local IP address at the Gateway level.

6 Limitation of the Proposed S-DHCP and Future Direction of Work

The proposed communication protocol is specially designed for an wireless environment, and extending it further for a different application, say, sensor network may require parallel effort in programming.The proposed communication protocol S-DHCP may be made domain specific. The increase of robustness has been found directly proportional to computational overhead. Hence experiments over various domain-specific trade-off may be conducted.

References

1. Wellington, B.: RFC 3007, Secure Domain Name System(DNS) Dynamic updates, Internet Engineering Task Force (November 2000)
2. Ramaswamy, C., Scoot, R.: Secure Domain Name System(DNS) Deploying Guide, Draft National Institute of Standards and Technology. NIST Special Publication 800-81, April 11 (2005)
3. Robert, J.: Dynamic DNS and Location Tracking Risks and Benefits, Craic comuting Technical report (2006)

4. Yasmin, S., Yousaf, M., Qayyum, A.: Security Issues Related with DNS Dynamic Updates for Mobile Nodes: A Survey. In: FIT 2010, Islamabad, Pakistan, December 21-23 (2010)
5. Perkins, C., Luo, K.: Using DHCP with computers that move. Wireless Networks (March 1995)
6. Komori, T., Satio, T.: The Secure DHCP systems with User Authentication. In: Proceedings of the 27th Annual IEEE Conference on Local Computer Networks, LCN 2002, November 6-8, pp. 123–131 (2000)
7. Droms, R., Arbaugh, W.: Authentication for DHCP Messages, RFC 3118, IETF (June 2001)
8. Drach, S.: DHCP option for open Groupś User Authentication Protocol, RFC 2485, IETF (January 1999)
9. Kent, S., Atkinson, R.: Security Architecture for the Internet Protocol, RFC 2401, IETF (November 1998)
10. Demerjian, J., Serhrouchni, A.: EDHCP: Extended Dynamic Host Configuration protocol (2004), doi:0-7803-8482-2/04/IEEE
11. Jonsson, J., Kaliski, B.: Public-Key cryptography Standards (PKCS) 1: RSA Cryptography Specification Version 2.1, RFC 3447, IETF (February 2003)
12. ITU-T Recommendations(a) X.509, Information technology Open System Interconnection The Directory: Authentication Frameworks (1997)
13. ITU-T Recommendations (b) X.509, Information technology Open System Interconnection The Directory: Public Key and attribute Certificate Frameworks (2000)
14. Ishibashi, H., Yamai, N., Abe, K., Ohnishi, K., Matsuura: A Protection method against Unauthorized access and/or Address Spoofing for open network Access Systems. Journal of Information Processing Society of Japan 40(12), 4335–4361 (1999)

Efficient Cryptography Technique on Perturbed Data in Distributed Environment

Nishant Goswami[1], Tarulata Chauhan[1], and Nishant Doshi[2]

[1] Gujarat Universisty,Ahmedabad, Gujarat, India
nishant_j_goswami@yahoo.com, taruchauhan114@gmail.com
[2] S V National Institute of Technology, India
doshinikki2004@gmail.com

Abstract. Data mining is a method through which we can search for a large pattern in huge database system. Now with the increasing growth of technology, the data requirements and amount of data will drastically increasing. Therefore, the data mining uses new methods for pattern matching which can be used for decision making. The organizations are stores data in bulk. Therefore, when a particular query is given by user, the amount of important or secure data can also be revealed as an answer of a query. This can harm to reputation of an organization. Therefore, privacy can concern to the above issue that not reveals any such kind of information about data provider and vice versa. Therefore, data needs to be modified without losing the data integrity. This paper outlines a method that achieve confidentiality from client and owner side which relatively less size of cipher text through mediator.

Keywords: Data mining, Cryptography, data perturbation, privacy, sensitive data.

1 Introduction

Data mining is one of the useful fields that connects different major areas like Artificial intelligence, databases etc. To investigate the unidentified data pattern from huge database, data mining can be act as dominant tool as contended by authors in [14-15, 21, 24]. In [31-32] authors said that organization depend on data mining, gives better throughput to their customers.[22] shows an example which uses hospital record for collecting large data for patients. With the increasing use of technologies like internet, networking, hardware and software the amount of data with different organizations is collected in huge, which also include the sensitive data also. It may be possible that the answer of query issued by customer, can revealed the important data regarding health care, finance , security etc. The common tendency of people is to hide the sensitive information. In real scenario like hospital records, the analyst requires the records for more than one hospital as it will provide mutual benefits to the hospitals. In this example each hospital want to share the data but neither of them want to share the data of their patients or nor disclose it. In this situation, use of the privacy preserving data mining will enhance the integrity of data [20].

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 239–244.
springerlink.com

The rest of the paper is organized as follows. The section 2 and 3 we have given background study or literature study. In section 4, we discussed the proposed system. Related work and future expansion are given in section 5and References are at the end.

2 Background Study

In general, there are two main approaches for the given problem one is use data transformation based approach and another is the cryptographic based approach. In firstapproach, the sensitive data is modified in such a way that it maintains its sensitive information. There are various data modification techniques as given in [1-5, 8, 11]. In cryptography technique data is modified using encryption techniques. In this one the communicating parties uses secured multiparty protocols as given in [6-7,17-19] which is not released any information to the third party. The basic techniques used were secure sum, secure size, set union etc.[27] suggest that in presence of adversary, which get some leakage data or gain some access to sensitive data. So in order to prevent this we require the third party that called the "trusted party". All the parties send their query or data to the trusted party and vice versa. [28] Suggests that any cryptography technique do not reveal any data with presence of the trusted party. [25-26] suggests the privacy preserving rules.

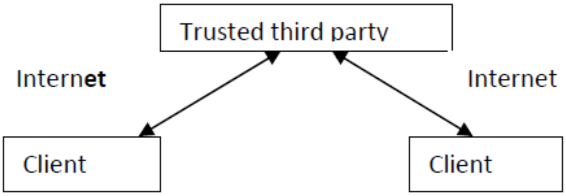


Fig. 1. System using trusted party [29]

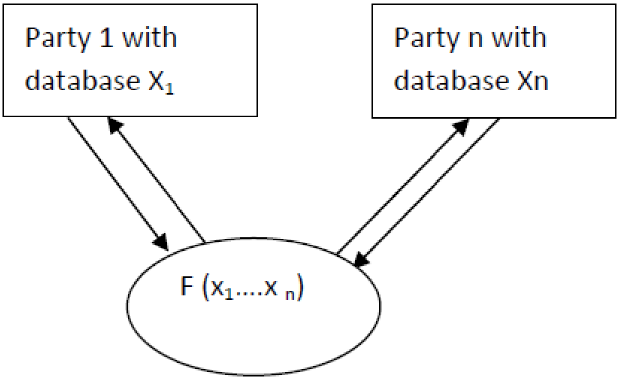


Fig. 2. Secure multiparty communication [29]

As shown in figure 1 there are clients who are on the internet and they are communicating with the trusted server through the communication link. So in real scenario all the data which communicate between client and server must be encrypted otherwise an adversary can gain the view or access or modify the data. Now consider the figure 2 in which there are n party each of which had their database and they will merge their data using function F and get the perturbed data. The data sent by each data provider do not contain any sensitive information. We can enhance our scheme by incorporating features as given in [9-10].

3 Overview of Randomization Perturbation Technique

In this approach the privacy of data is maintained by perturbed data[12,13,15,23] with randomization algorithm approach. [16] Suggest adding noise in this method. The Gaussian distribution technique is used for this one. As shown in figure 3 first the Gaussian algorithm and using the random variables applied to data then different conditions for data integrity will be checked than the noise will be given to data and lastly the perturbed data is ready for communications. Figure 4 represent the framework which use to share data using the perturbation and encryption techniques.

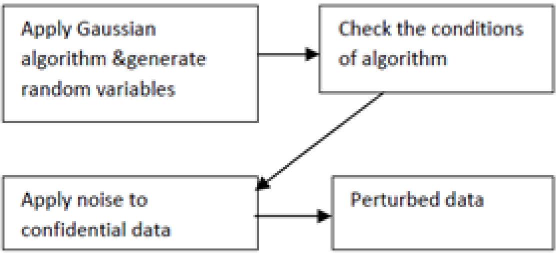

Apply Gaussian algorithm &generate random variables
Check the conditions of algorithm
Apply noise to confidential data
Perturbed data

Fig. 3. Perturbation technique [29]

4 Proposed System

The problems with previous [29] approach is, with increasing amount of data, the channel overhead is also increased for communication channel. Assume that scenario in which if there are 2^32 data providers than each of which will generate N different keys assume that N is sufficiently large. So the total overhead of channel becomes too large. We proposed the approach we can reduce the N keys to one key only for mediator and still maintain the integrity that neither client nor the data providers knows about each other. Here we assume the same model as in [29]. Therefore, the steps of the proposed algorithm are as follow. Here each client contain its *PKc* (Public key) and *SKc*(secret key) same as each mediator contains *PKm* and *SKm*. Data providers do not need any keys but they know the *PKm*.

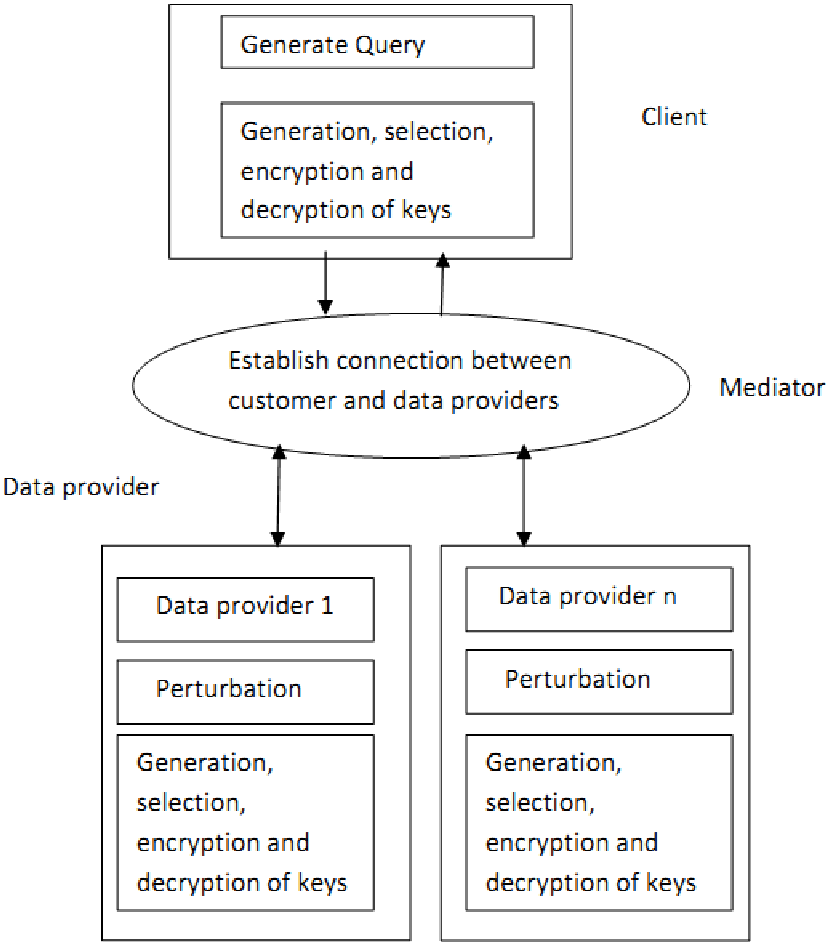

Generate Query
Generation, selection, encryption and decryption of keys
Client
Establish connection between customer and data providers
Mediator
Data provider
Data provider 1
Perturbation
Generation, selection, encryption and decryption of keys
Data provider n
Perturbation
Generation, selection, encryption and decryption of keys

Fig. 4. Framework that represent sharing of data using encryption and perturbation method[29]

1. Client *c* sends the query to mediator to get the data.
2. Mediator maintains the table in which a random generated number (or sequential number) *Rc* is associated with each incoming client request. Now mediator send query of client with random number to all data providers.
3. The data provider which satisfies the client requirements sends the perturbed data M with *Rc* and whole encrypted under pubic key *PKm* to mediator.
4. The mediator decrypts the data using own secret key *SKm* and checks the corresponding client for *Rc* and sends data back to client which encrypted under public key of client *c* e.g. *PKc*.
5. Client decrypts the data using its private key *SKc* and gets the required perturbed data..

5 Conclusion

This paper gives the proposed algorithm to reduce the network overheads between clients and data providers and still maintaining the privacy of data providers and clients with each other. Therefore, the main aim of this paper is to compare with [29] and suggest a new method that reduces the network overheads. Therefore, we had tried to avoid repeat of all data and concentrate on algorithm. The current open problem is if mediator is compromised than all the communication is compromised. We will also look at concept of [32] to enhance our paper.

References

1. Agrawal, R., Srikant, R.: Privacy Preserving Data Mining. In: The Proceedings of the ACM SIGMOD Conference (2000)
2. Muralidhar, K., Sarathi, R.: A General additive data perturbation method for data base security. Journal of Management Science 45(10), 1399–1415 (2002)
3. Agrawal, D., Aggarwal, C.C.: On the Design and Quantification of Privacy Preserving Data mining algorithms. In: ACM PODS Conference (2002)
4. Muralidhar, K., Sarathy, R.: Data Shuffling- a new masking approach for numerical data. Management Science (forthcoming 2006)
5. Iyengar, V.S.: Transforming data to satisfy privacy constraints. In: Proc. of SIGKDD 2002, Edmonton, Alberta, Canada (2002)
6. Lindell, Y., Pinkas, B.: Privacy Preserving Data Mining. In: Bellare, M. (ed.) CRYPTO 2000. LNCS, vol. 1880, pp. 36–54. Springer, Heidelberg (2000)
7. Yu, H., Vaidya, J., Jiang, X.: Privacy-Preserving SVM Classification on Vertically Partitioned Data. In: Ng, W.-K., Kitsuregawa, M., Li, J., Chang, K. (eds.) PAKDD 2006. LNCS (LNAI), vol. 3918, pp. 647–656. Springer, Heidelberg (2006)
8. Agarwal, D., Aggarwal, C.C.: On the design and quantification of privacy preserving data mining algorithms. In: Proceedings of the 20th Symposium on Principles of Database Systems, Santabarbara, California, USA (May 2001)
9. Karandikar, P., Deshpande, S.: Preserving Privacy in Data Mining using Data Distortion Approach. International Journal of Computer Engineering Science 1(2) (2011) ISSN: 2250-3439
10. Ravi Kumar, G., Ramachandra, G.A., Sunitha, G.: An Evolutionary Algorithm for Mining Association Rules Using Boolean Approach. International Journal of Computer Engineering Science 1(3) (2011) ISSN: 2250-3439
11. Kantarcioglu, M., Clifton, C.: Privacy-preserving distributed mining of association rules on horizontally partitioned data. In: The ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery (DMKD 2002), Madison, pp. 24–31 (June 2002)
12. Muralidhar, K., Sarathy, R., Parsa, R.A.: A general additive perturbation method for database security. Management Science 45(10), 1399–1415 (1999)
13. Agrawal, R., Evfimievski, A., Srikant, R.: Information sharing across private databases. In: Proc. of ACM SIGMOD (2003)
14. Han, J., Kamber, M.: Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers (2000)
15. Kargupta, H., Datta, S., Wang, Q., Sivakumar, K.: On the privacy preserving properties of random data perturbation techniques. In: Proc. of 3rd IEEE Int. Conf. on Data Mining, Washington, DC, USA, pp. 99–106 (2003)

16. Muralidhar, K., Parsa, R., Sarathy, R.: A general additive data perturbation method for database security. Management Science 19, 1399–1415 (1999)
17. Pinkas, B.: Cryptographic techniques for privacy preserving data mining. SIGKDD Explorations, 12–19 (2002)
18. Evfimievski, A.: Randomization in privacy preserving data mining. ACM SIGKDD Explorations Newsletter 4, 43–48 (2002)
19. Vaidya, J., Clifton, C.: Privacy-Preserving Data Mining: Why, How and When. IEEE Security and Privacy 2, 19–27 (2004)
20. Clifton, C., Kantarcioglu, M., Vaidya, J., Lin, X., Zhu, M.Y.: Tools for privacy preserving distributed data mining. SIGKDD Explor. News., 28–34 (2002)
21. Weiss, G.M.: Data Mining in Telecommunications. In: Data Mining and Knowledge Discovery Handbook, A Complete Guide for Practitioners and Researchers, pp. 1189–1201. Kluwer Academic Publishers (2005)
22. Kargupta, H., Datta, S., Wang, Q., Sivakumar, K.: On the privacy preserving properties of random data perturbation techniques. In: Proceedings of the 3rd IEEE International Conference on Data Mining, Melbourne, Florida, November 19-22, pp. 99–106 (2003)
23. Muralidhar, K., Parsa, R., Sarathy, R.: A General Additive Data Perturbation Method for database Security. Management, 1399–1415 (1999)
24. Liu, L., Kantarcioglu, M., Thuraisingham, B.: The applicability of the perturbation based privacypreserving data mining for real-world data. Data & Knowledge Engineering 65, 5–21 (2008)
25. Evfimievski, A., Srikant, R., Agrawal, R., Gehrke, J.: Privacy preserving mining of association rules. In: Proceedings of 8th ACM SIGKDD International Conference on Knowledge Discovery Data Mining (July 2002)
26. Lindell, Y., Pinkas, B.: Privacy Preserving Data Mining. In: Bellare, M. (ed.) CRYPTO 2000. LNCS, vol. 1880, pp. 36–54. Springer, Heidelberg (2000)
27. Vaidya, J., Clifton, C.: Privacy preserving association rule mining in vertically partitioned data. In: Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, July 23-26 (2002)
28. Kargupta, H., Datta, S., Wang, Q., Sivakumar, K.: On the privacy preserving properties of random data perturbation techniques. In: Proc. of Intl. Conf. on Data Mining, ICDM (2003)
29. Kamakshi, P., VinayaBabu, A.: Preserving Privacy and Sharing the Data in Distributed Environment using Cryptographic Technique on Perturbed data. Journal of Computing 2(4) (April 2010) ISSN 2151-9617
30. Agarwal, R., Srikant, R.: Privacy preserving data mining. In: Procseedings of the 19th ACM SIGMOD Conference on Management of Data, Dallas, Texas, USA (May 2000)
31. Canny, J.: Collaborative filtering with privacy. In: IEEE Symposium on Security and Privacy, Oakland, pp. 45–57 (May 2002)
32. Jothimani, K., Antony SelvadossThanmani, S.: MS: Multiple Segments with Combinatorial Approach for Mining Frequent Item sets Over Data Streams. International Journal of Computer Engineering Science 2(2) (2012) ISSN : 2250-3439

A Novel Triangle Centroid Authentication Protocol for Cloud Environment

K. Anitha Kumari[1], G.Sudha Sadasivam[2], Bhandari Chetna[1], and S. Rubika[1]

[1] Department of Information Technology
[2] Department of Computer Science and Engineering, PSG College of Technology
Coimbatore, India
{anitha.psgsoft,chetna.bhandari,rubika.optim}@gmail.com
sudhasadhasivam@yahoo.com

Abstract. Cloud computing allows the use of Internet-based services to support business processes and rental of IT-services on a utility-like basis. Cloud computing is a concept implemented to decipher the daily computing needs of users for hardware, software and other resources dynamically. Server consolidation poses risks for data privacy and cloud security. Authentication remains a significant challenge in cloud. The proposed approach used a triangle centroid based authentication protocol. In traditional authentication protocol a single server stores the sensitive user credentials, like username and password. When such a server is compromised, a large number of user passwords, will be exposed. Our proposed approach uses a dual authentication protocol in order to improve the authentication service in cloud environment. The protocol utilizes the fundamental concept of triangle centroid and strengthening parameters derived from it to perform user authentication. In the proposed protocol median angles and the prime numbers representing the intercepts of triangle sides will be used for authentication. During the registration process, the password given by user is transformed to represent the centroid and strengthening parameters are derived using it. Whenever a user logs in, based on his password centroid is recalculated and these strengthening parameters stored are used to authenticate the user. The entire authentication protocol is hosted as a service in the cloud environment to authenticate the web services. Authentication security is ensured due to the splitting of the strengthening parameters asymmetrically.

Keywords: Dual authentication, authentication protocol, triangle parameters, centroid, cloud computing, cloud security.

1 Introduction

'Cloud computing' is the newly-minted buzz term to describe the next stage in the Internet's evolution, providing the means through which everything can be delivered to the computer user as a service wherever and whenever needed. As online access to services becomes ubiquitous and the cloud access model gains momentum, authentication is increasingly becoming a focal point for security professionals. The major issue involved with maintenance of bank accounts, health records, corporate

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 245–254.
springerlink.com

intellectual property and politically sensitive information is establishment of user identity [1].

The dynamic nature of the cloud environment necessitates the establishment of user identities before an entity can join the cloud. Because cloud computing should involve a large amount of entities, such as users and resources from different sources, the authentication is important and complicated. Authentication is the process of proving identity, typically through credentials, such as a user name and password. In the cloud this also encompasses authentication against varying identity stores. A user in this case could be a person, another application, or a service; all should be required to authenticate. Many enterprise applications require that users authenticate before allowing access. Authorization, the process of granting access to requested resources, is pointless without suitable authentication. Both the cloud provider and the enterprises must consider the challenges associated with credential management and implement cost effective solution that reduce the risk appropriately [2].Password authentication is considered as one of the simplest and most convenient authentication mechanisms [3].But password authentication protocols are subject to replay, password guessing and stolen-verifier attacks as described below [4].

(1) Replay attack: A replay attack is an offensive action in which an adversary impersonates or deceives another legitimate participant via the reuse of information obtained in a protocol.

(2) Guessing attack: A guessing attack involves an adversary simply (randomly or systematically) trying passwords, one at a time, in hope that the correct password is set up. Ensuring passwords selected from an adequately large space can resist exhaustive password searches. However, the majority of the users choose pass-words from a small subset of the full password space. Such weak passwords with low entropy are easily guessed by means of the dictionary attack [5].

(3) Stolen-verifier attack: In the majority of the applications, the server stores verifiers of users' passwords (e.g., hashed passwords) instead of the clear text of passwords. In stolen-verifier attack, the adversary who steals the password-verifier from the server can use it directly to masquerade as a legitimate user in a user authentication execution [5].

The architecture for authentication systems fall under 4 categories as shown in Figure. In single-server model (Fig. 1) a single server maintains a database of user passwords. Most of the existing authentication systems follow this single-server model. The main drawback of single server is the single point of vulnerability. It leads to offline dictionary attacks against the user password database.

The second type is the plain multi-server model depicted in Fig. 2, the server side comprises of multiple servers to overcome the single point of vulnerability; the servers are equally exposed to users and a user has to communicate in parallel with several or all servers for authentication. The main problem with the plain multi-server model is the demand on communication bandwidth and the need for synchronization at the user side since a user has to engage in simultaneous communications with multiple servers.

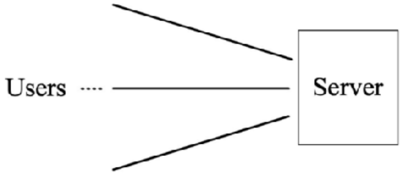


Fig. 1. Single Server Model

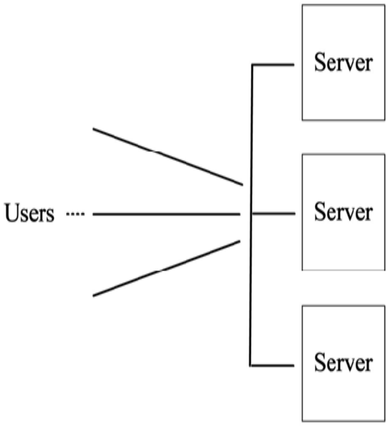


Fig. 2. Plain MultiServer Model

The third type is the gateway augmented multi-server model shown in Fig. 3. A gateway is positioned as a relaying point between users and servers and a user only needs to contact the gateway. Apparently, the introduction of the gateway removes the demand of simultaneous communications by a user with multiple servers as in the plain multi-server model. However, the gateway introduces an additional layer in the architecture, which appears "redundant" since the purpose of the gateway is simply to relay messages between users and servers. It does not in any way involve in service provision, authentication, and other security enforcements. Gateways also reduce system reliability.

The fourth type is the two-server model (Fig. 4).It comprises of two servers at the server side, one of which is a public server exposing itself to users and the other is a back-end server staying behind the scene. Users contact only the public server, but the two servers work together to authenticate users. We propose to use two server model for

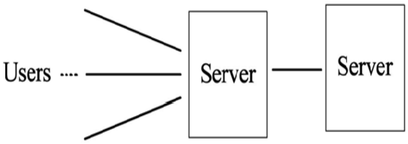


Fig. 3. Gateway augmented multiserver model

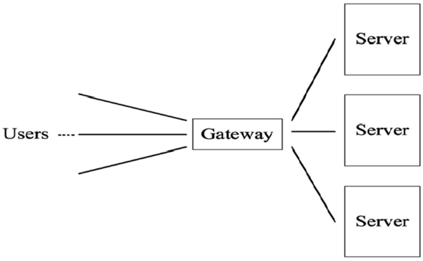


Fig. 4. Two-Server Model

our authentication framework. From a security point of view, servers in the multi-server models are equally exposed to outside attackers, while in the two-server model, only the public server faces such a problem. These clearly improve the server side security and in turn the overall system security in the two-server model. The two-server model eliminates drawbacks in the plain multi-server model (i.e., simultaneous communications between a user and multiple servers) and the gateway augmented multi-server model (i.e., redundancy) [12, 13, 14].It distributes user passwords and the authentication functionalities to two servers in order to eliminate a single point of vulnerability in the single-server model. As a result, the two-server model appears to be a sound model for practical applications [6, 11].

Our proposed triangle centroid approaches uses two server model for implementation of authentication protocol.

This paper proposes a dual authentication protocol which utilizes dual servers for authentication to enhance the cloud security. The significance of the protocol is the usage of the fundamental concepts and basic elements of the triangle centroid for authentication. With these triangle parameters, the user credential is interpreted and then stored in two servers which provide solid security for authentication protocol. The dual authentication protocol gives authentication to the cloud user if and only if both the servers are mutually involved in the authentication mechanism. It is not possible to obtain the password by hacking a single server. The triangle centroid protocol offers effective security against the attacks like replay attack, guessing attack and stolen-verifier attack as the user authentication is a combined mechanism of two servers. The remaining of the paper is organized as follows: Section 2 deals with some of the existing research works and Section 3 is constituted by the proposed dual authentication protocol. Section 4 discusses about the implementation approach. Section 5 is about discussion and results and Section 6 concludes the paper. Section 7 discusses future enhancement.

2 Related Work

Lishan Kang [7] has proposed an Identity-Based Authentication (IBA) scheme over traditional mutual authentication. In cloud storage sharing, mutual authentication between users and between user and Cloud environment is critical in ensuring data security. However, traditional mutual authentication using public-key operation unleashes cloud storage system load, computation and communication overhead and reduces scalability. An IBA scheme has short key size, is identity-based and non-interactive. This scheme divides the sharing users between domain. In the domain global master key is shared to exercise mutual authentication. By the analysis of performance, this scheme improves the computational and communicational efficiency over two times. This scheme is enabled by an emerging cryptographic technique from the bilinear pairing and its security can be assured by the Bilinear Diffie-Hellman Problem (BDHP). In IBA scheme, the master key of some do-main becomes the bottleneck of Cloud Storage System's security. Once the master key of some domain is leaked, the domain's security will be wrecked. In addition, if a user wants to share another user's data, they must be in the same domain [7].

Zhidong Shen [8] have proposed Trusted Computing Platform (TCP) to aid the process of authentication in cloud computing. The TCP is based on the Trusted Platform Mod-ule (TPM). The TPM is a logic independent hardware. It can resist the attacks from both software and the hardware. The TPM contains a private master key to protect for other information stored in cloud computing system. Because the hardware certificate is stored in TPM it is hard to attack it. So TPM provides the trust root for users. Since the users have full information about their identity, the cloud computing system can use some mechanism to trace the users and get their origin. In TCP the user's identity is proved by user's personal key and this mechanism is integrated in the hardware, such as the BIOS and TPM. So it is very hard to deceive a user-id. Each site in the cloud computing system will record the visitor's information. By using the TCP mechanism in cloud computing, the trace of participants can be known by the cloud computing trace mechanism. The TCP provides cloud computing a secure base for achieving trusted computing. Integration of hardware modules with cloud computing system is a challenging research issue [8].

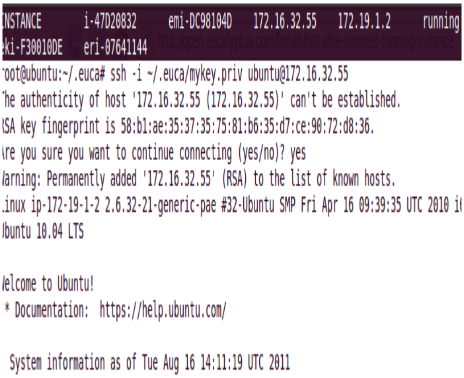
When organizations begin to utilize applications in the cloud, authenticating users in a trustworthy and manageable manner becomes an additional challenge. Our proposed work on a dual authentication protocol utilizes dual servers for authentication to enhance the cloud security. The significance of this protocol is the usage of the fundamental concepts and basic elements of the triangle, namely triangle centroid to authenticate.

3 Proposed Methodology

The proposed approach is based on a triangle centroid authentication protocol. In traditional authentication protocol a single server stores the sensitive user credentials, like username and password. When such a server is compromised, a large number of user passwords, will be exposed. Our proposed approach uses a dual authentication protocol in order to improve the authentication service in cloud environment. In the proposed protocol, angles made by centroid to the three sides of triangle and the centroid itself will be stored, using which authentication will be performed. There are 2 phases-registration phase and authentication phase. During the registration process, the password given by the user is transformed to represent the centroid and derivation of strengthening parameters is made. Whenever a user logs in, based on his password centroid is recalculated and the strengthening parameters which are stored is used to authenticate the user. The entire authentication protocol is hosted as a service in the cloud environment to authenticate the user accessing the web services. So this type of authentication provides security due to asymmetrically splitting up the strengthening parameters

4 Implementation

The proposed approach is implemented as a service over private cloud. Private cloud has been setup using Eucalyptus [9]. Eucalyptus is an open-source software platform that implements IaaS-style cloud computing using the existing Linux-based



```
NSTANCE        i-47D20832     emi-DC98104D    172.16.32.55    172.19.1.2      running
ki-F30010DE    eri-07641144   http://open.eucalyptus.com/forum/not-able-connect-running-instance
oot@ubuntu:~/.euca# ssh -i ~/.euca/mykey.priv ubuntu@172.16.32.55
he authenticity of host '172.16.32.55 (172.16.32.55)' can't be established.
SA key fingerprint is 58:b1:ae:35:37:35:75:81:b6:35:d7:ce:90:72:d8:36.
re you sure you want to continue connecting (yes/no)? yes
arning: Permanently added '172.16.32.55' (RSA) to the list of known hosts.
inux ip-172-19-1-2 2.6.32-21-generic-pae #32-Ubuntu SMP Fri Apr 16 09:39:35 UTC 2010 i
buntu 10.04 LTS

elcome to Ubuntu!
* Documentation:  https://help.ubuntu.com/

 System information as of Tue Aug 16 14:11:19 UTC 2011
```

Fig. 5. Login to VM Instance

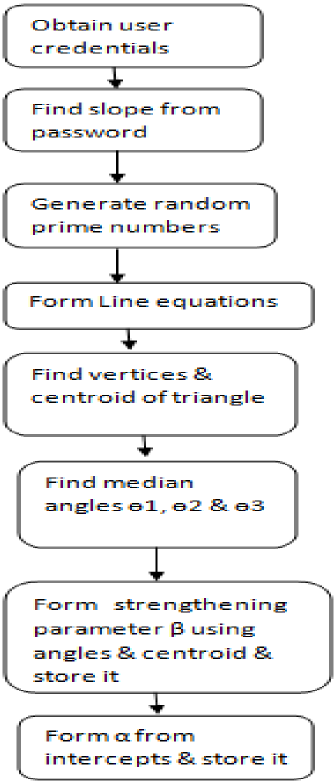

Obtain user credentials
Find slope from password
Generate random prime numbers
Form Line equations
Find vertices & centroid of triangle
Find median angles θ1, θ2 & θ3
Form strengthening parameter β using angles & centroid & store it
Form α from intercepts & store it

Fig. 6. User Registration Process

infrastructure found in the modern data center. Eucalyptus works with most of the currently available Linux distributions including Ubuntu Red Hat Enterprise Linux (RHEL), CentOS, SUSE Linux Enterprise Server (SLES), openSUSE, Debian and Fedora. Similarly, Eucalyptus can use a variety of virtualization technologies including VMware, Xen, and KVM to implement the cloud abstractions. Eucalyptus Enterprise Edition is built upon the open source core platform and a suite of additional products and features that allow Enterprises and Service Providers to implement a portable, scalable and high performing private cloud solution [9]. Once the private cloud is started a VM (virtual machine) instance can be started and logged in as shown in fig 5. Using Apache Tomcat server as the web server for deploying the web service and MySQL as the backend DBMS.

4.1 Authentication Protocol

The registration process and authentication process as shown in Fig 6 and Fig 7 is implemented as an authentication service. These phases are as in Fig. 8. Forward centroid algorithm includes calculating strengthening parameters from password entity and using dual servers for its storage which will later be used for authenticating users by reverse centroid algorithm.

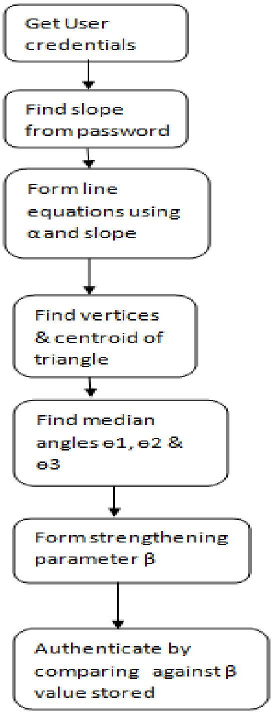

Get User credentials
Find slope from password
Form line equations using α and slope
Find vertices & centroid of triangle
Find median angles θ1, θ2 & θ3
Form strengthening parameter β
Authenticate by comparing against β value stored

Fig. 7. User Authentication Process

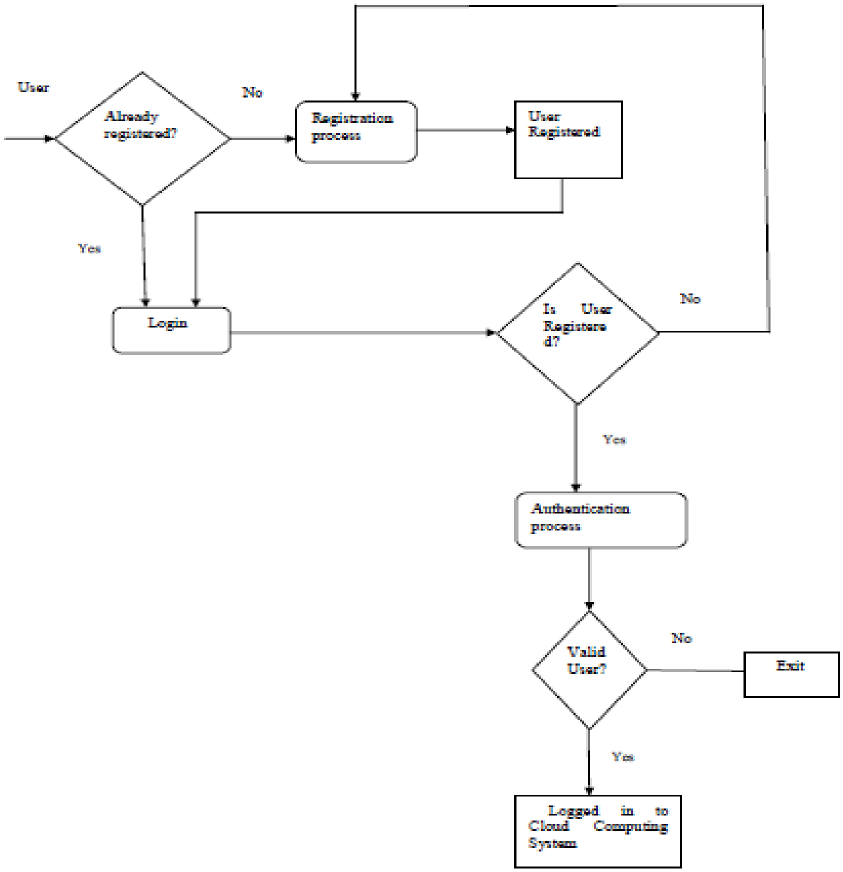


Fig. 8. Flowchart for Registration & Authentication Phase

4.2 Hosting Authentication Protocol as a Service in Cloud

Authentication protocol is hosted as a service in VM instances once they are up and running.VM instances running in different nodes makes possible dual server implementation. Web Service can be accessed from client machine using VM instance public IP address and web service context path. The client is authenticated to use the web service using the proposed authentication protocol.

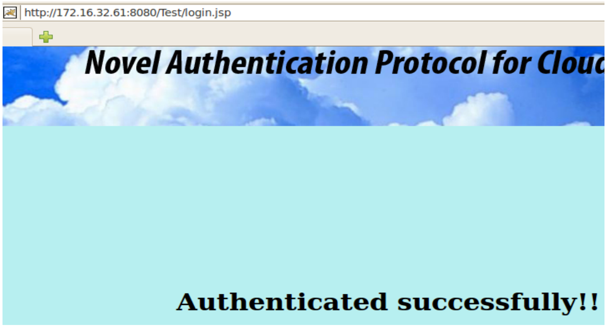
5 Discussion and Results

The common threats faced by authentication process in cloud environment and the countermeasures in triangle centroid protocol to overcome these attacks are as follows.

Replay attack: Usually replay attack is called as ‘man in the middle’ attack. Adversary stays in between the user and the server and hacks the user credentials when the user contacts server. To overcome this, the user has to change the credential randomly. But it is less probable to do that. Our protocol is robust when the replay attack happens in between the two servers as the credentials are interpreted and alienated into two parts.

Guessing attack: Guessing attack is nothing but the adversaries just contacts the servers by randomly guessed credentials. The effective possibility to overcome this attack is to choose the password by maximum possible characters, so that the probability of guessing the correct password can be reduced. As the proposed work uses random generation of prime numbers for the representation of intercepts of the sides of the triangle, it is more difficult to guess the password.

Stolen-verifier attack: Instead of storing the original password, the server is normally storing the verifier of the password. If the password attacker steals the verifier from the server, then it will masquerade as the legitimate user. This does not happen in any two server protocol, as the password is alienated into two modules. Hence, we can justify that our protocol is also more robust against the attack, as the password is interpreted and then alienated into two modules and stored in the two servers. We have completed forward centroid algorithm and reverse centroid algorithm of the authentication protocol with dual servers along with hosting this as a service in cloud and we can access it as shown in Fig 9.


http://172.16.32.61:8080/Test/login.jsp
Novel Authentication Protocol for Clouc

Authenticated successfully!!

Fig. 9. Accessing Authentication service

6 Conclusion

Triangle centroid based authentication protocol, enhances the cloud security as authentication mechanism utilizes two servers for authentication. As the servers maintain the interpreted and distinct form of user credentials, there is very less chance to reveal the user credentials to the adversary. Moreover, the protocol utilizes the fundamental properties of the triangle.

The triangle centroid parameters make the cloud more secure as the alienated passwords are derived from these parameters. This simple triangle centroid concept utilization in the authentication protocol introduces revolutionary idea in the authentication mechanism as well as in cloud environment. We have completed registration phase of the authentication protocol along with building private cloud and are proceeding with authentication phase and hosting as a service in cloud.

7 Future Enchancement

Multiple centers of triangle can be considered as entities for passwords and relationships between them can be used for authentication. This protocol can also be extended for multidimensional shapes.

References

1. Chow, R., Jakobsson, M., Masuoka, R.: Authentication in the Clouds: A Framework and its Application to Mobile Users. ACM (2010) 978-1-4503-0089-6/10/10
2. Cloud Security Alliance, Domain 12: Guidance for Identity & Access Management V2.1 (April 2010)
3. Yeh, H.-T., Sun, H.-M., Hwang, T.: Efficient Three- Party Authentication and Key Agreement Protocols Resistant to Password Guessing Attacks. Journal of Information Science and Engineering 19(6), 1059–1070 (2003)
4. Lin, C.L., Hwang, T.: A password authentication scheme with secure password updating. Computer & Security 22(1), 68–72 (2003)
5. Yoon, E.-J., Ryu, E.-K., Yoo, K.-Y.: Attacks and Solutions of Yang et al.'s Protected Password Changing Scheme. Informatica 16(2), 285–294 (2005)
6. Yang, Y., Deng, R.H., Bao, F.: A Practical Password-Based Two-Server Authentication and Key Exchange System. IEEE
7. Kang, L., Zhang, X.: Identity-Based Authentication in Cloud Storage Sharing. In: 2010 International Conference on Multimedia Information Networking and Security (2010)
8. Shen, Z., Tong, Q.: The Security of Cloud Computing System enabled by Trusted Computing Technology. In: 2010 2nd International Conference on Signal Processing Systems (2010)
9. Johnson, D., Murari, K., et al.: Eucalyptus Beginner's Guide–UEC edition (Ubuntu Server 10.04 - Lucid Lynx) v1.0, May 25 (2010)
10. Ruckmani, V., Sudha Sadasivam, G.: A novel trigon-based dual authentication protocol for enhancing security in grid environment. (IJCSIS) International Journal of Computer Science and Information Security 6(3) (2009)
11. Yang, Y., Bao, F.: Enabling Use of Single Password Over Multiple Servers in Two-Server Model. In: 2010 10th IEEE International Conference on Computer and Information Technology, CIT 2010 (2010)
12. Yang, D., Yang, B.: A Novel Two-Server Password Authentication Scheme with Provable Security. In: 2010 10th IEEE International Conference on Computer and Information Technology, CIT 2010 (2010)
13. Lee, J.H., Lee, D.H.: Secure and Efficient Password-based Authenticated Key Exchange Protocol for Two-Server Architecture. In: 2007 International Conference on Convergence Information Technology (2007)
14. Yang, Y., Bao, F.: Enabling Use of Single Password Over Multiple Servers in Two-Server Model. In: 2010 10th IEEE International Conference on Computer and Information Technology, CIT 2010 (2010)

Security and Availability of Data in the Cloud

Ahtesham Akhtar Patel, S. Jaya Nirmala, and S. Mary Saira Bhanu

Department of Computer Science and Engineering,
National Institute of Technology, Tiruchirappalli, Tamilnadu, India
ahtesham.patel@gmail.com, {sjaya,msb}@nitt.edu

Abstract. With advances in the field of cloud computing many computing resources and/or services are being provided to the end user on a pay-as-you-use basis. Data outsourcing is a new paradigm in which a third party provides storage services. This is more cost effective for the user as there is no need of purchasing expensive hardware and software for data storage. The user is also relieved from administrative activities of software upgrades and maintenance. The private data of the enterprises can be stored at secure and reliable sites, and ubiquitously made available to the user on demand. CIA (Confidentiality, Integrity and Availability) are the challenging issues associated with data storage management with/without data outsourcing. This paper proposes a method of data storage which achieves CIA using data dispersal. The practicality of this proposal is demonstrated by implementing it in a private cloud setup using the OpenStack cloud framework.

Keywords: data security, cloud, information dispersal.

1 Introduction

Cloud Computing can be defined as the shifting of computing resources like processing power, network and storage resources from desktops and local servers to large data centers hosted by companies like Amazon, Google, Microsoft etc. These resources are provided to a user or company on highly scalable, elastic and pay as you use basis. It reduces the administrative and maintenance cost of IT organizations. From an individual's perspective Cloud Computing is a revolutionary concept as it removes the obstacles created due to lack of finance and resources thus enabling easy large scale deployment an application.

Computing resources provided by cloud vendors can be categorized as computing power, network resources (bandwidth, IP addresses etc.) and storage. The services provided by the cloud can be Software as a Service (SaaS), Platform as a Service (PaaS), Infrastructure as a Service (IaaS) or it can be Anything-as-a-Service (XaaS).

Several free and reliable online storage services available to the users are Apple iCloud, Microsoft SkyDrive, Google Drive, Amazon S3, Dropbox and Box. As the use of these services becomes widespread, security of the out sourced user data becomes an important research topic.

The parameters that are taken into consideration for data security are Confidentiality, Integrity, Availability and Performance. [1]- [3] have emphasized the importance

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 255–261.
springerlink.com

of ensuring remote data integrity. The problem of outsourcing data faces the following obstacles:

1.1 Storage of Data at an Untrusted Host

Though a service provider gives the guarantee of protecting the privacy of user data, the reality is that the data may be physically located in some country and subject to the local rules and regulations. For example, according to the USA PATRIOT Act the government can access data being hosted by a third party without the permission or knowledge of the user or company using the hosting services [4]. In the recent Megaupload trial regarding the fate of digital files belonging to some 60 million global users there is a possibility that the court will allow Carpathia Hosting, the company that has maintained the servers at its own expense since Megaupload was taken down, to delete the information on them or possibly sell off the servers [5]. Moreover most cloud computing vendors give users little control over their stored data. Under such circumstances it becomes paramount for a user or company to ensure the confidentiality of the data before it is moved off premise.

1.2 Availability of Data

Data availability and durability are vital for cloud storage providers, as data loss or unavailability can be damaging to the business. This is usually achieved by replicating the data without the knowledge of the user. Regardless of the precautions taken by a storage service provider, there were major cloud outages recently. VMware Cloud Foundary was down on 25th and 26th April 2011 and Microsoft Azure was out on 28th February 2012 [6] [7]. Amazon's S3 cloud storage service replicates data across "regions" and "availability zones" so that data and applications can be available even in the face of a disaster affecting an entire location. The user should carefully understand the details of the replication scheme. Though Amazon guarantees that an application using multiple availability zones will not suffer any down time, Amazon web services outage on 21st April 2011 took down several online sites like Reddit, HotSuite, FourSquare and Quora [8]. In view of these events, it is essential for a user or company to ensure the availability of the data without being dependant on the storage service provider. The proposed method overcomes these obstacles by data dispersal and message authentication code.

The rest of this paper is organized as follows. Section 2 discusses the existing work. Section 3 describes the proposed solution for secure data storage. The experimental setup is presented in section 4 and conclusion in section 5.

2 Existing Work

To provide confidentiality while out sourcing of data to the service providers, data encryption has traditionally been used. Sion [9] emphasizes that a system that employs encryption for data outsourcing is secure if it ensures *correctness, confidentiality and data access privacy.* He also discusses about how these components are inter-related.

Outsourcing by encrypting the whole data is prohibitive in terms of performance and it does not solve the problem of availability. Moreover encryption techniques are secure based on the present computing power but with the advances being made in computing speeds this may not be always true. For example, distributed.net and Electronic Frontier Foundation joined hands in January 1999, to break a DES key, which was previously thought to be unbreakable, in 22 hours and 15 minutes. Additionally, secure maintenance of keys used for encryption becomes important. Also, to execute queries on the outsourced data, the data needs to be encrypted and decrypted. This takes more time and hence the query response time becomes larger. Homomorphic encryption [10] suggests executing queries over encrypted data, but due to its dependence on the public key cryptosystem, it becomes impractical to implement it.

The concept of Private Information Retrieval (PIR) was first discussed in [11]. Private information retrieval protocols intend to hide the queries performed by the user on a public database, stored on a set of servers. By providing the privacy of user queries the PIR protocols tend to hide the user's intensions from the Service provider. The idea of PIR has been extended to Symmetric Private Information Retrieval (SPIR), in which the privacy of user data is the main concern. In [12], Sion and Carburnar have extensively discussed about the practical infeasibility of implementing the single-server computational PIR protocol.

To overcome these problems, it is proposed to use Information Dispersal techniques. Two popular information dispersal techniques are the Shamir's Secret Sharing method and Rabin's Information Dispersal Algorithm (IDA). In Shamir's[13] secret sharing algorithm, a file F to be outsourced is split into n parts $F_1, F_2, F_3 \ldots F_n$, such that each file F_i, $i \leq n$, is padded with redundant information to make its size same as of F. The file F can be retrieved if k out of n pieces is available. Shamir calls this as *threshold (k,n).* The drawback of Shamir's approach in a pay-per-use cloud computing model is that the amount of storage required is increased by n times.

Rather, Rabin [14] suggested splitting a secret S into n pieces such that a person can obtain the secret only if $k < n$ of these pieces are available, where k is the threshold. Here, each secret S_i, $i \leq n$, is of size $|S|/k$, where $|S|$ is the size of the secret. The total sizes of all the secrets are $(n/k)*|S|$. Thus, with the Rabin's IDA the storage complexity is reduced.

Hence, we propose to use the Rabin's IDA to solve the problem of secure data outsourcing to an untrusted host in a cloud.

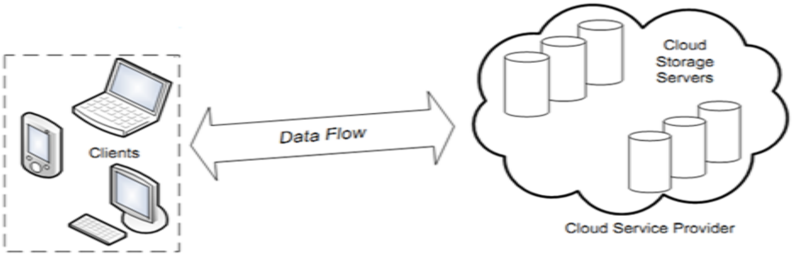


Fig. 1. Cloud data storage architecture

3 Proposed Solution

The cloud data storage architecture used in this work is based on the model proposed by Qian Wang et al [15], as shown in figure-1. The different entities are the Client and Cloud Storage Server.

Client: The end user who has large amount of data to store in the cloud and relies on the service provider for maintenance. This can either be an individual user or a large organization.

Cloud Storage Server: An entity, which is managed by a Cloud Service Provider, has significant storage space and computation resource to maintain client's data.

Figure-2 and figure-3 illustrate the information dispersal and recovery processes used in the proposed method.

Let k be the threshold value and n be the number of slices. The data to be stored is arranged in terms of the data matrix D of size *k x t* where t varies according to the size of the data. C is the secret Vandermonde Matrix of size *n x k*. The matrix M of size *n x t* is computed as

$$M= C * D \tag{1}$$

Each of the *n* rows of *M* represents a slice. This modified data is stored at multiple data centres such that none of them have access to $s < k\text{-}1$ slices.

Data retrieval can be achieved by obtaining any *k* of the *n* slices and applying the reverse IDA. Consider *M'* to be the *k x t* matrix formed by obtaining the *k slices* of data stored in the cloud and *C'* to be the *k x k* matrix obtained by selecting the corresponding rows of C. Then the data matrix D can be retrieved as:

$$D=C'^{-1} * M' \tag{2}$$

Even with the loss of *(n-k)* slices, the data can be reproduced thus ensuring availability. A message authentication code can be applied to the test data before dispersal to achieve integrity.

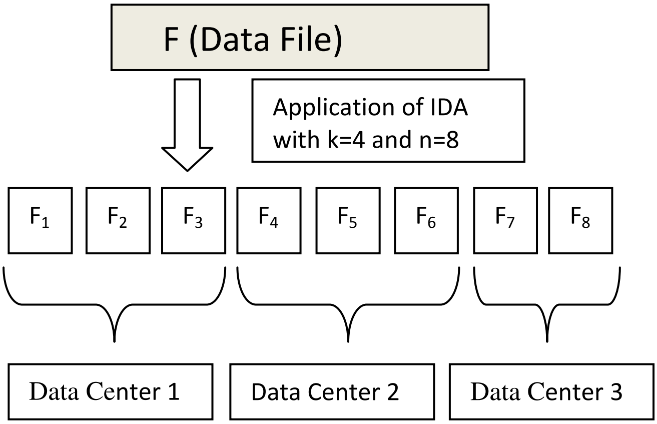


Fig. 2. Information Dispersal

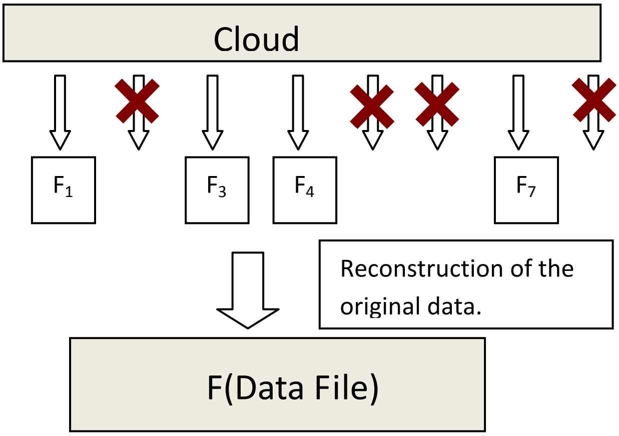


Fig. 3. Information Recovery

4 Experimental Setup

To demonstrate the effectiveness of this method the OpenStack cloud framework was used. OpenStack is an IaaS cloud computing project. OpenStack is gaining importance in the academia as well as in the industry. Major cloud players like AMD, Intel, HP, Linux and Cisco have joined the OpenStack project [16].

Linux machines were used as compute nodes, controller and client. The controller node consisted of the nova-api, network controller, scheduler and the RabbitMQ asynchronous messaging server. The compute nodes had the compute controller component installed in them. The virtual machines were instantiated and run in these compute nodes. Six virtual machines were instantiated, each one representing a cloud storage server. *IDA(n,k)* was applied to the data file at the client side. These files were randomly placed at the storage servers using the SCP (secure copy) protocol in such a way that no server had *k* files. The FlatManager network configuration was used for the setup of the OpenStack cloud.

The Employees sample database developed by Patrick Crews and Giuseppe Maxia was used as the test data. This large database has six separate tables and a total of four million records [17].

IDA and MAC (Message Authentication Code) were implemented using Crypto++. It is a free and open source C++ class library of cryptographic algorithms and schemes written by Wei Dai [18]. AES (Advanced Encryption Standard) is the commonly used encryption technique for large size of data. The performance comparison of AES and IDA was done using Crypto++ library. It was observed that AES encryption and decryption using CTR (Counter) mode and a 16 byte key was faster than splitting and combining operations of IDA.

Modern processors have special hardware support for AES resulting in faster encryption and decryption but AES does not provide the guarantee of availability which is more important in a cloud environment. Integrity was achieved using the SHA1 algorithm. It has a block size of 64 bytes and message digest size of 20 bytes.

We were able to reconstruct the whole data successfully even when *(n-k)* slices of data were unavailable.

Figure-4 represents the time taken to split a 94 MB file into 10 pieces for different threshold values and Figure-5 represents the time taken to combine them. Overhead is calculated as *((combined size of split files/original file size) * 100)* for different values of n and k.

It was observed that a considerable decrease in the dispersal time as the overhead decreases. The Recovery time remains almost constant with a maximum variation of 0.4 seconds with the best recovery time at the threshold value of eight. The overhead at this point was 25%.

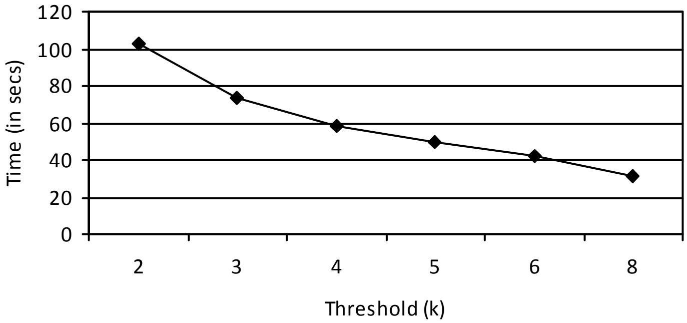


Fig. 4. Information Dispersal for 94 MB file with n=10

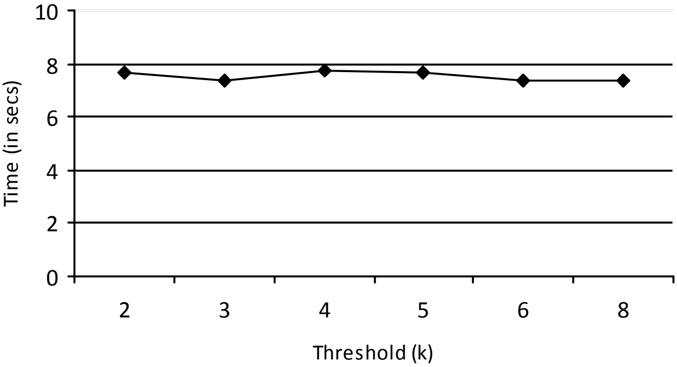


Fig. 5. Information Recovery for 94 MB file with n=10

5 Conclusion

Proposed method of secure data outsourcing provides the CIA of security along with comparable performance, as compared to traditional encryption techniques.

The security of this method is not bounded by the computational capabilities of present hardware. Further, secure storage of the encryption keys is done away with. Taking the above points into consideration we conclude that, owing to the distributed nature of the cloud, information dispersal of data is the more optimal approach.

References

1. Juels, A., Burton, J., Kaliski, S.: PORs: Proofs of Retrievability for Large Files. In: Proc. of CCS 2007, pp. 584–597 (2007)
2. Shacham, H., Waters, B.: Compact Proofs of Retrievability. In: Pieprzyk, J. (ed.) ASIACRYPT 2008. LNCS, vol. 5350, pp. 90–107. Springer, Heidelberg (2008)
3. Bowers, K.D., Juels, A., Oprea, A.: Proofs of Retrievability: Theory and Implementation. In: Cryptology ePrint Archive, Report 2008/175 (2008), `http://eprint.iacr.org/`
4. Wikipedia: Patriot ACT (2012), `http://en.wikipedia.org/wiki/Patriot_Act` (cited 2012)
5. Rauf, D.S.: Politico: Megaupload data: Hosting company chafes at maintenance, `http://www.politico.com/news/stories/0312/74354.html` (cited 2012)
6. Tanke, D.: Cloud Foundary: Analysis of April 25, 26 (2011) Downtime, `http://support.cloudfoundry.com/entries/20067876-analysis-of-april-25-and-26-2011-downtime` (cited 2011)
7. Laing, B.: Windows Azure: Summary of Windows Azure Service Disruption on February 29 (2012), `http://blogs.msdn.com/b/windowsazure/archive/2012/03/09/summary-of-windows-azure-service-disruption-on-feb-29th-2012.aspx` (cited 2012)
8. Amazon Web Services team: Summary of the Amazon EC2 and Amazon RDS Service Disruption in the US East Region, `http://aws.amazon.com/message/65648/` (cited 2011)
9. Sion, R.: Secure data outsourcing. In: Proc. of the VLDB Conf., pp. 1431–1432 (2007)
10. Ge, T., Zdonik, S.B.: Answering aggregation queries in a secure system model. In: Proc. of the VLDB Conf., pp. 519–530 (2007)
11. Chor, B., Goldreich, O., Kushilevitz, E., Sudan, M.: Private information retrieval. Journal of the ACM 45(6), 965–982 (1998)
12. Sion, R., Carbunar, B.: On the computational practicality of private information retrieval. In: Proc. of the Networks and Distributed Systems Security (2007)
13. Shamir, A.: How to share a secret. Commun. ACM 22(11), 612–613 (1979)
14. Rabin, M.O.: Efficient dispersal of information for security, load balancing and fault tolerance. Journal of The ACM 36(2), 335–348 (1989)
15. Wang, C., Wang, Q., Ren, K., Lou, W.: Ensuring data storage security in Cloud Computing. In: 17th International Workshop on Quality of Service, IWQoS, vol. 186, pp. 1–9 (2009)
16. `http://www.openstack.org`
17. `https://launchpad.net/test-db/`
18. `http://www.cryptopp.com/`

A Novel Power Balanced Encryption Scheme for Secure Information Exchange in Wireless Sensor Networks

Shanta Mandal and Rituparna Chaki

Department of Computer Sc. and Engineering
West Bengal University of Technology
Salt lake, Kolkata, India
{shanta.mandal.ghosh,rituchaki}@gmail.com

Abstract. A lot of research on wireless sensor network is focused on field of performance, security and energy. Public key cryptography suffers from high computational complexity and overhead, when Symmetric key schemes can be utilized more efficiently in order to provide more security. This paper proposes a more appropriate cryptography scheme for wireless sensor networks in respect of low energy consumption. The proposed security scheme overcomes the limitations of both public-key and symmetric-key protocols. The scheme is utilizes pre-distributed keys to implement data confidentiality service. Special attention has been given to data authenticity. The proposed scheme is suitable for data centric routing using direct diffusion protocols in wireless sensor networks.

Keywords: Wireless Sensor Network, Public Key Cryptography, Symmetric Key Cryptography, Elliptic Curve Cryptography, Scalable Encryption Algorithm, RC5.

1 Introduction

Wireless sensor networks are widely used indifferent fields. Energy awareness is an essential design issue in wireless sensor networks (WSN). A wireless sensor network consists of spatially distributed autonomous sensors to monitor physical or environmental conditions, such as temperature, sound, vibration, pressure, motion or pollutants and to cooperatively pass their data through the network to a main location. Security is a key consideration when deploying Wireless Sensor Networks. The energy efficiency is one of the key concerns in WSN. Sensor networks are deployed in a hostile environment, security becomes extremely important as these networks are prone to different types of malicious attacks. To provide security, communication transactions should be encrypted and authenticated. Symmetric key scheme is more appropriate cryptography (SKC) for wireless sensor networks due to its low energy consumption and simple hardware requirements, but most of them cannot provide sufficient security level (e.g. integrity, confidentiality, and authentication) as public key approach (PKC) does.

Cryptographic primitives are the basis of security solutions and the most frequently executed security operations in sensor networks. Symmetric algorithms, both parties share the same key for encryption and decryption. To provide privacy, this key needs

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 263–271.
springerlink.com

to be kept secret. Once somebody else gets to know the key, it is not safe anymore. Symmetric algorithms have the advantage of not consuming too much computing power. A public key cryptography algorithm uses two different keys for encryption and decryption. The key used for decryption kept secret (Private) whereas the encryption key can be distributed openly (Public).Encryption algorithms and their use are essential part of the secure transmission of information. There are extensive studies on using symmetric-key cryptography to achieve various aspects of security in sensor networks [5, 11].The symmetric key function is used to guarantee secure communications between in-network nodes while the public key function is used to guarantee a secure data delivery between the source node and the sink node.

This paper describes a new hybrid approach that combines the advantages of the well-known PKC and SKC schemes in wireless sensor networks. Symmetric-key and public-key are the main key-based tools used insecurity implementation. It is suitable for wireless sensor networks that incorporate data centric routing protocols. We have calculated the computational and communication overheads in terms of energy consumption in the new scheme.

Rest of the paper is organized as follows: section 2 deals with the state of the art studies in this field, section 3 presents the proposed framework, section 4 includes the simulation results, followed by conclusion in section 5.

2 Related Work

The main challenge in sensory networks is how to secure communications between sensor nodes and how to set up secret keys between communicating nodes. In this section, explain how different security schemes can be implemented in Direct Diffusion (DD) protocol. In DD protocol an interest travels between three different nodes the sink node, intermediate node, and source node. Therefore, the schema show how each of these node implements the security schemes and how much energy is consumed to run such implementation within the node. We assume that a node uses the first radio model for sending and receiving data [3, 2]. This problem is known as the key agreement problem which has been handled via two security mechanisms: Public Key Cryptography (PKC) and Symmetric Key Cryptography (SKC). In direct diffusion protocol used both public and private key. Mohammad AL-Rousan, A. Rjoub and Ahmad Baset used Directed Diffusion (DD) protocol that uses Elliptic Curve Cryptography (ECC) public key and RC5 symmetric key. But the novel proposed scheme uses ECC public key and scalable encryption algorithm (SEA) symmetric key and it is most suitable for wireless sensor networks that incorporate data centric routing protocols.

PKC is preferred for security purpose as it provides security services for the system under consideration including confidentiality, integrity, authentication, and non repudiation [12]. The public key and symmetric key approaches maintain all features of the Directed Diffusion (DD) protocol. Directed Diffusion routing protocol has been developed in data-centric routing [7, 4]. In data-centric routing, the sink sends queries to certain regions and waits for data from the sensors located in the selected regions. Data centric routing has proven to be a good scheme for minimizing communication overhead and energy consumption by using in-network aggregation. The DD protocol has several advantages. First, there is no need for a node to have a global or a local address since all communications occurs between neighboring nodes. Second, it is highly energy efficient since the node does not have to maintain global information about network topology. Finally, individual nodes can do aggregation and caching, in addition to sensing. For network deployment to implement a secure Directed Diffusion using the hybrid security scheme it Store Public key, Symmetric key, and hash function codes in each node and also each node, select and save a randomly private key and keep the associated public key at the sink. It Save a public key of the sink at each node and the same common symmetric key in all sensor nodes.

The main drawback of PKC is that it suffers from high computational complexity and overhead so PKC schemes must be improved to their high complexity and high memory overheads. Rivest-Shamir-Adelman algorithm (RSA) [1] and (ECC) [8] are amongst well-known public key algorithms used in security systems. It is shown that ECC is more efficient than RSA in terms of memory requirements because it requires much lower key size than RSA to achieve the same security level. So the author uses ECC as a public key cryptography. The experimental result of executing the ECC with 160-bit key size and 1024-bit message size [9] shows that the execution time of the ECC on 8-bit ATMEL microprocessor with 8 MHz clock rate is 0.81s.

The main idea In SKC techniques is that the secret keys are pre-distributed among sensors before their deployment [10] and SKC schemes, must be utilized more efficiently in order to provide more security satisfaction. The paper uses RC5 as a symmetric key, which has substantial overhead associated with its implementation. The symmetric key encryption does not guarantee authenticity or the integrity. So we uses secure hashing algorithm. W Stalling Cryptography and Network Security discusses the implementation of SHA and showed that an m-bit message is processed by SHA [(m+65)/512] times. It has been found in [6] that the total execution time of SHA-1 using a 512-bit message is 0.007777 seconds.

3 Proposed Work

In this section we elaborate the proposed schema used for sensor networks. The study of previous works show that the use of RC5 often leads to high energy consumption. In this paper, a novel method for key generation using Scalable Encryption Algorithm (SEA) is proposed in order to reduce the energy consumption of the network. The hash function is used for this purpose, albeit with additional (h) bits to be sent along with the original data packet and here we proposed to use SEA [14] as a symmetric key.

The proposed scheme uses ECC public key and SEA symmetric key and is suitable for wireless sensor networks that incorporate data centric routing protocols. The scheme involves three sub-cases, (i) Source_node_centric, (ii) intermediate_node_centric (iii) sink_node_centric

3.1 Case (i) Source_Node_Centric

This is the case when the data is encrypted/decrypted by the source node itself. Data is encrypted by using both the symmetric and public key techniques. The symmetric key encryption is coupled with hash function to guarantee the authenticity and integrity. The final encrypted message thus have an additional h bits associated with it. This m+h bit data packet is then sent to intermediate node.

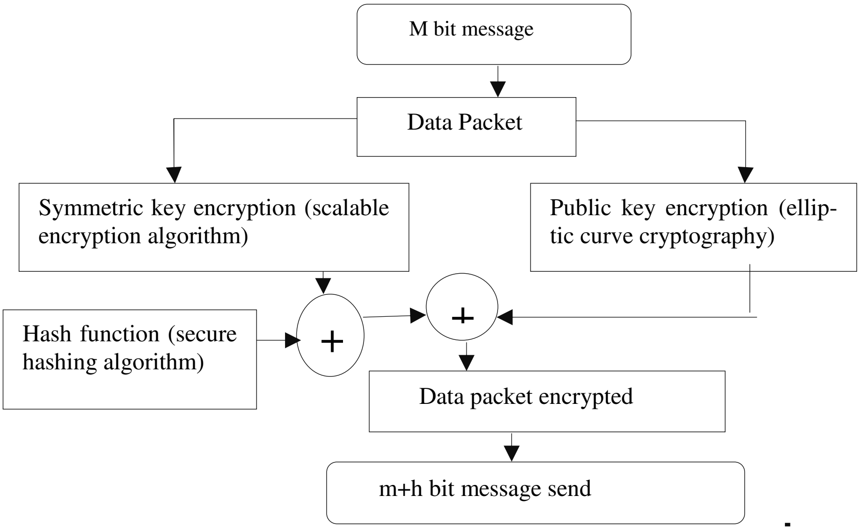

M bit message
Data Packet
Symmetric key encryption (scalable encryption algorithm)
Public key encryption (elliptic curve cryptography)
Hash function (secure hashing algorithm)
+
+
Data packet encrypted
m+h bit message send

Fig. 1. Data encryption by the source node

3.2 Case (ii) Intermediate_Node_Centric

This part deals with the steps executed by each intermediate node after receiving the encrypted. Our proposed scheme use direct diffusion protocol, a well known data centric routing protocol. In Data-Centric routing in-network aggregation of data is used to yield energy efficient dissemination. In data aggregation, whenever similar data happens to meet at a branching node in the tree, the copies of similar data are replaced by a single message it is known as data aggregation. It is important feature of DD protocol. For intermediate nodes, each node does not need to encrypt the part of the packet that is encrypted by the source/sink node using the public key, it rather needs to decrypt and encrypt the aggregation data. This is done by using the scalable encryption algorithm (SEA) and SHA, as shown in the Figure 2. Suppose there are two messages M1 and M2, such that M1=M2 and the encryption and decryption keys are the same. This implies that the cipher of both M1 and M2 are equal so the node will only check if the encrypted data already exists in the data cache. The proposed techniques first decrypt the message and checks for its existence in cache. The message is then encrypted using hashing algorithm and scalable encryption algorithm. The encrypted message is then sent to next step.

3.3 Case (iii) Sink_Node_Centric

This section describes the steps executed by sink node after receiving the encrypted message. The sink node decrypts the received data using the ECC and SEA algorithm.

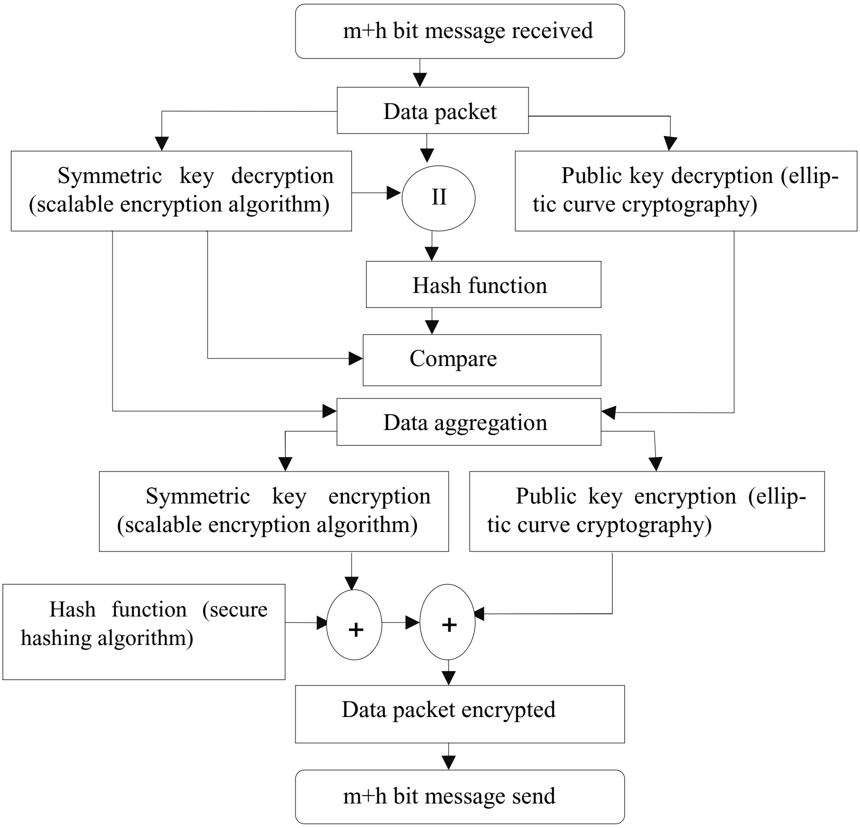

m+h bit message received
Data packet
Symmetric key decryption (scalable encryption algorithm)
II
Public key decryption (elliptic curve cryptography)
Hash function
Compare
Data aggregation
Symmetric key encryption (scalable encryption algorithm)
Public key encryption (elliptic curve cryptography)
Hash function (secure hashing algorithm)
+
+
Data packet encrypted
m+h bit message send

Fig. 2. Data encryption by intermediate node

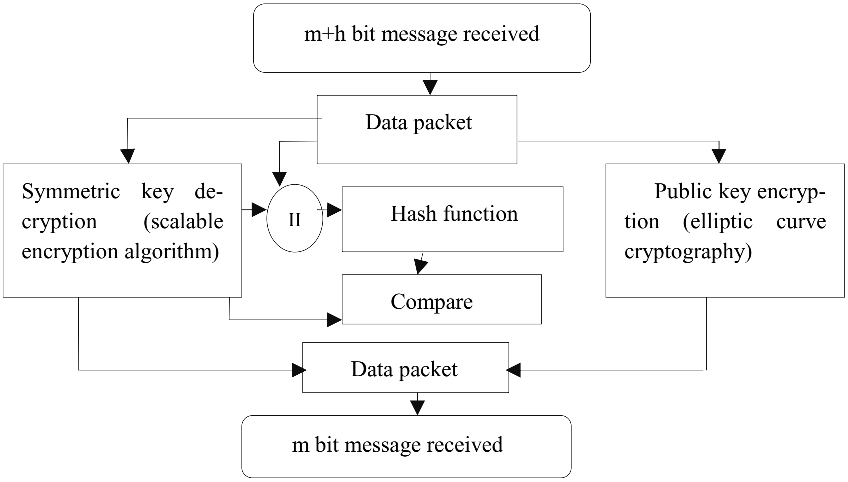

m+h bit message received
Data packet
Symmetric key decryption (scalable encryption algorithm)
II
Hash function
Public key encryption (elliptic curve cryptography)
Compare
Data packet
m bit message received

Fig. 3. Data encryption by sink node

4 Result Discussion

This section presents a comparative analysis of the proposed technique against the techniques as presented in [13]. We now analyze the energy consumption of the symmetric key for encryption/decryption process. It is assumed that the proposed scheme is executed on Atmega 128 16MHz 8-bit architecture AVR instruction set. The network size is taken as n= 10. In our discussion we show and compare the energy consumptions at the sink, source, and intermediate node for all security schemes under consideration. The message of size is taken as 1024.

Figure 4 compares the energy consumed by the source node in the hybrid scheme vs the schemes proposed in [13]. The energy consumption by source node in the proposed scheme is only 86 micro joules, whereas the scheme presented in [13] shows the minimum energy consumption to be 96 micro joules. The reason for improvement is attributed to the fact that RC5 has been replaced by SEA.

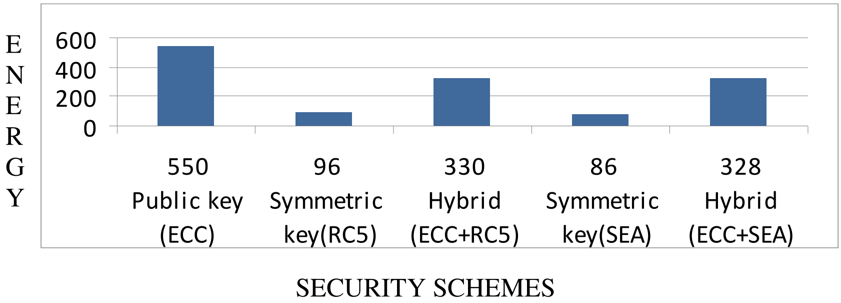


Fig. 4. Energy Consumption at Source Node

Figure 5 compares the energy consumed by the intermediate node in the hybrid scheme vs the schemes proposed in [13]. The energy consumption by intermediate node in the proposed scheme is only 173 micro joules, whereas the scheme presented in [13] shows the minimum energy consumption to be 192 micro joules. The reason for improvement is attributed to the fact that RC5 has been replaced by SEA.

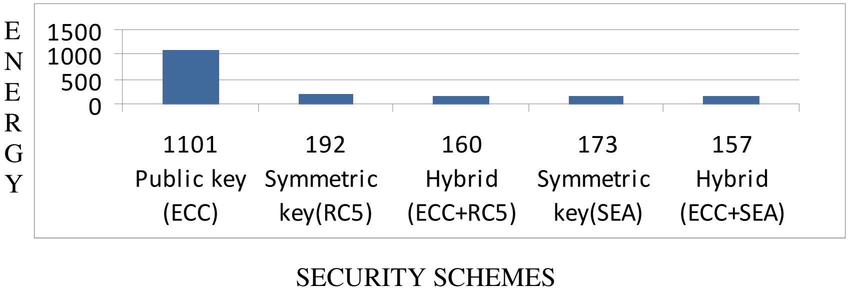


Fig. 5. Energy Consumption at intermediate Node

Figure 6 compares the energy consumed by the sink node in the hybrid scheme vs the schemes proposed in [13]. The energy consumption by sink node in the proposed scheme is only 76 micro joules, whereas the scheme presented in [13] shows the minimum energy consumption to be 96 micro joules. The reason for improvement is attributed to the fact that RC5 has been replaced by SEA.

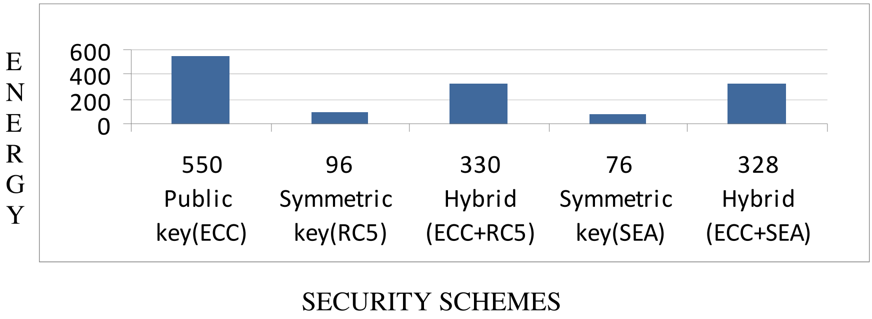


Fig. 6. Energy Consumption at sink Node

Figure 7 shows the average energy consumption of the overall node. The approach shows that if we are using SEA instead of RC5 [13] (Rivest Cipher 5) then the energy consumption will be reduced a lot. The energy consumption by all nodes in the proposed scheme is only 1900 micro joules and 2226 micro joules, whereas the scheme presented in [13] shows the minimum energy consumption to be 2113 micro joules and 2264 micro joules.

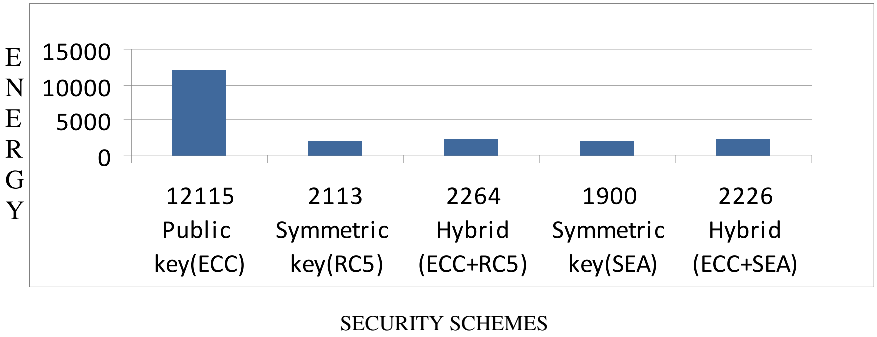


Fig. 7. Overall Energy Consumption

5 Impact of Network Parameters

It is obvious that the novel technique based on the parameter network size n. Network size is the number of the sensor nodes in the network. The size of the network has a direct effect on the number of intermediate nodes between the source and sink. Figure 8 shows comparative result analysis of the proposed technique using SEA against the techniques RC5.

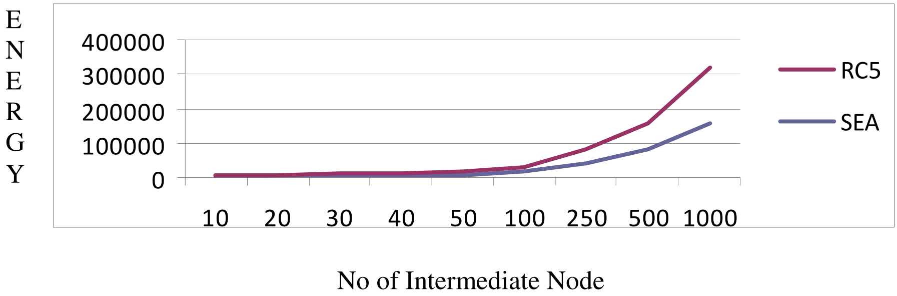


Fig. 8. Impact of network size (intermediate node)

Figure 9 shows that the overall energy consumption for all schemes when increase the network size. The proposed hybrid scheme shows better performance to use SEA instead of RC5 in small and larger sensor networks.

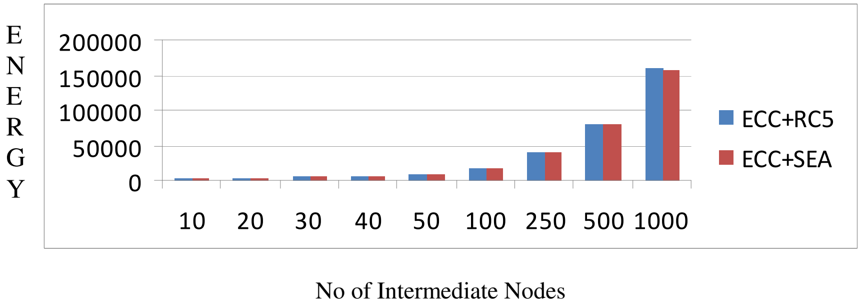


Fig. 9. Impact of network size (all node)

6 Conclusion

The state of the art studies that most of the existing encryption technique scheme consume lot of energy. This paper proposes a more appropriate cryptography scheme for wireless sensor networks in respect of low energy consumption is Modified Symmetric Key Cryptography algorithm. Here we are using Scalable Encryption algorithm instead of RC5. Our scheme is suitable for Data Centric routing using DD protocol which is good scheme for minimizing the energy consumption. PKC suffers from high computational complexity and overhead, when SKC schemes can be utilized more efficiently in order to provide more security. The proposed security scheme overcomes the limitations of both public-key and symmetric-key protocols. Proposed scheme uses secure hashing algorithm and this will incur additional (h) bits to be sent along with the original data packet. The symmetric key function is used to guarantee secure communications between the nodes in a network while the public key function is used to guarantee a secure data delivery between the source node and the sink node.

References

1. Rivest, R.L., Shamir, A., Adleman, L.A.: A method for obtaining digital signatures and public-key crypto systems. Communications of the ACM 21(2), 120–126 (1998)
2. Heinzelman, W., Kulik, J., Balakrishnan, H.: Adaptive Protocols for Information Dissemination in Wireless Sensor Networks. In: Proceedings of the 5th ACM/IEEE Mobicom, pp. 174–185 (August 1999)
3. Heinzelman, W., Chandrakasan, A., Balakrishnan, H.: Energy-Efficient Communication Protocol for Wireless Micro sensor Networks. In: Proceedings of the 33rd Annual Hawaii International Conference on System Sciences (HICSS 33), pp. 1–10 (2000)
4. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: Wireless sensor networks: a survey. Computer Networks 38(4), 393–402 (2001)
5. Chan, H., Perrig, A., Song, D.: Random key pre distribution schemes for sensor networks. In: Proceedings of the IEEE Symposium on Research in Security and Privacy, pp. 197–213 (2003)
6. Ganesan, P., Venugopalan, R., Peddabachagari, P., Dean, A., Mueller, F., Sichitiu, M.: Analyzing and modeling encryption overhead for sensor network nodes. In: Proceedings of WSNA 2003, pp. 151–159 (September 2003)
7. Di Pietro, R., Mancini, L.V., Law, Y.W., Etalle, S., Havinga, P.: LKHW: A directed diffusion-based secure multicast scheme for wireless sensor networks. In: Proceedings of the First International Workshop on Wireless Security and Privacy (WiSPr 2003), pp. 397–406 (2003)
8. Gaubatz, G., Kaps, J.-P., Sunar, B.: Public Key Cryptography in Sensor Networks—Revisited. In: Castelluccia, C., Hartenstein, H., Paar, C., Westhoff, D. (eds.) ESAS 2004. LNCS, vol. 3313, pp. 2–18. Springer, Heidelberg (2005)
9. Gura, N., Patel, A., Wander, A., Eberle, H., Shantz, S.C.: Comparing Elliptic Curve Cryptography and RSA on 8-bit CPUs. In: Joye, M., Quisquater, J.-J. (eds.) CHES 2004. LNCS, vol. 3156, pp. 119–132. Springer, Heidelberg (2004)
10. Lee, J.-Y., Stinson, D.R.: Deterministic Key Predistribution Schemes for Distributed Sensor Networks. In: Handschuh, H., Hasan, M.A. (eds.) SAC 2004. LNCS, vol. 3357, pp. 294–307. Springer, Heidelberg (2004)
11. Cheng, Y., Agrawal, D.P.: Efficient pair wise key establishment and management in static wireless sensor networks. In: Proceedings of the Second IEEE International Conference on Mobile ad hoc and Sensor Systems, Washington, DC (2005)
12. Law, Y., Doumen, J., Hartel, P.: Survey and benchmark of Block Cipher for Wireless Sensor Neworks. ACM Transactions on Sensor Networks 2(1), 65–93 (2006)
13. AL-Rousan, M., Rjoub, A., Baset, A.: A Low-Energy Security Algorithm for Exchanging Information in Wireless Sensor Networks. Journal of Information Assurance and Security 4, 48–59 (2009)
14. Standaert, F.-X., Piret, G., Gershenfeld, N., Quisquater, J.-J.: SEA a Scalable Encryption Algorithm for Small Embedded Applications

A Cryptographic Approach towards Black Hole Attack Detection

Pratima Sarkar and Rituparna Chaki

Department Of Computer Science and Engineering
West Bengal University of Technology
Salt lake, Kolkata, India
{psmoon2,rituchaki}@gmail.com

Abstract. A MANET is an infrastructure less network that consists of mobile nodes that communicate with each other over wireless links. In the absence of a fixed infrastructure, nodes have to cooperate with each other in order to provide the necessary network functionality for discovery and maintaining path. One of the principal routing protocols used in Ad hoc networks is AODV protocol. The security of the AODV protocol is threaded by a particular type of attack called 'Black Hole' attack. In this attack a malicious node advertises itself such a way that it has the shortest path to the destination node. This paper gives solution for co-operative Black Hole Attack at the time of route discovery phase and gives solution without monitoring other nodes. Monitoring technique increases network traffic and also cause of power loss. This paper uses RSA security Algorithm to identify Reply packet comes from destination only.

Keywords: Ad hoc Networks Routing, Protocols, AODV, Black Hole Attack.

1 Introduction

Wireless networks can be basically either infrastructure based networks or infrastructure less networks. The infrastructure based networks uses fixed base stations, which are responsible for coordinating communication between the mobile hosts (nodes). The ad hoc networks falls under the class of infrastructure less networks, where the mobile nodes communicate with each other without any fixed infrastructure between them. MANET is composed of many mobile devices with wireless interfaces. MANET has fully decentralized topology. It changes dynamically as nodes move and the nodes reorganize themselves to enable communications. In MANET, each mobile device can be treated as a node. The source node can send packets to the destination node by other nodes in MANET. Therefore, the routing protocol must ensure both connectivity and security to achieve the network stability. Since the memory, the bandwidth, and the power saving are very important in MANET. While these characteristics are essential for the flexibility of MANETs, they introduce specific security concerns that are absent or less severe in wired networks. MANETs are vulnerable to various types of attacks. These include passive eavesdropping, active interfering, impersonation, and denial-of-service. Intrusion prevention measures such as strong authentication and redundant transmission can be used to improve the security of an

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 273–278.
springerlink.com

ad hoc network. However, these techniques can address only a subset of the threats. Moreover, they are costly to implement. The dynamic nature of ad hoc networks requires that prevention techniques should be complemented by detection techniques, which monitor security status of the network and identify malicious behaviour.

One of the most widely used routing protocols in MANETs is the ad hoc on-demand distance vector (AODV) routing protocol [2]. It is a source initiated on-demand routing protocol. However, AODV is vulnerable to the well-known black hole attack. AODV's basic working principle contains two key components: route discovery phase and route maintenance phase. This paper studies black hole attack occurs in route discovery phase. In this paper, a mechanism is proposed to identify multiple black hole nodes cooperating as a group in an ad hoc network.

Rest of the paper is organized as follows: section 2 deals with the state of the art studies in this field, section 3 presents the proposed solution, section 4 includes the result description, followed by conclusion in section 5.

2 Related Work

In the intruder attack, the attacker must realize the routing protocol mechanism to fake the network. Researcher understands the routing protocol mechanism to protect the network as well. The researchers use different types of intrusion detection mechanisms on different routing protocols to defend against same attack or different types of attacks. Such type of few papers discussed below that gives solution for Black Hole Attack.

[1] In this paper the authors proposed Source Intruder Detection (SID) security routing mechanism that detects the attacker when an intermediate node sends the RREP packet. In SID security routing mechanism, when the source nod receives a RREP packet from the intermediate node, the source node sends a Further Route request packet to the next hop through a new route to verify that it has a route to the intermediate node, which they sent back the RREP packet and announced that they have a route to the destination. As soon as the next hop receives FRREQ packet, it sends a Further Route Reply packet to the source node. The source node checks the FRREP Packet information and if the next hop node has routes to the destination nod and intermediate node, the source node establishes the route. That means source node monitors other nodes in the network.

[3] In this paper, an approach is proposed to combat the Black hole attack. In this approach any node uses number rules to inference about honesty of reply's sender. Activities of a node in a network show its honesty. To participate in data transfer process, a node must demonstrate its honesty. Early of simulation, all nodes are able to transfer data; therefore they have enough time to show its truth. If a node is the first receiver of a RREP packet, it forwards packets to source and judge the replier. The activities of a node are logged by its neighbors. These neighbors are requested to send their opinion about a node. When a node collects all opinions of neighbors, it decides if the replier is a malicious node. The decisions are made based on number rules.

[4] This paper proposes a secure and efficient MANET routing protocol, the SAODV protocol which aims to solve black hole attack. In SAODV after route discovery phase, when the source node in MANET receives a RREP, source node will deposit the RREP in its routing table, and immediately sends another packet SRREQ to the destination node along the opposite direction route of RREP received. This SSREQ contains a random number generated by the source node. When receives two SRREQ or more from different routing paths, the destination node firstly deposits them to local routing table, and compares the content of SRREQ whether contains a same random number. If number is same then destination node sends SRREP that also contains a random number to the source node along opposite direction path. When source node receives two or more SRREP, it compares random numbers generated by the destination node if same then it sends its data packet to the destination.

[6] This paper gives solution for two problems those are "Do Not Forward RREQ Messages" and "Do Not Forward DATA Messages" This paper has defined two roles: the checking node and the checked node. If a node broadcasts a RREQ message and monitors its neighbors whether have forwarded the RREQ message or not. This node is called the RREQ checking node. The monitored node is called the RREQ checked node. After broadcasting a RREQ message to its neighbors, the RREQ checking node monitors its neighbors and records those neighbors have rebroadcast the same RREQ message. After waiting for period of time, the RREQ checking node will look over the routing table to examine which nodes do not forward the RREQ message. These nodes that do not forward the RREQ message is identified selfish nodes. In this paper all activities of other node are monitored by checking node.

[7] In this paper, proposed mechanism for removing a cooperative black hole attack is presented. The mechanism modifies the AODV protocol by introducing two concepts, (i) data routing information (DRI) table and (ii) cross checking. In the proposed scheme, two bits of additional information are sent by the nodes that respond to the RREQ message of a source node during route discovery process. Each node maintains an additional data routing information (DRI) table. In the DRI table, the bit 1 stands for 'true' and the bit 0 stands for 'false'. The first bit 'From' stands for the information on routing data packet from the node, while the second bit 'Through' stands for information on routing data packet through the node. Fields are used for checking reliability of a node. In cross checking phase, the intermediate node that generates the RREP has to provide information regarding its next hop node and its data routing information entry for that next hope node. On receiving the RREP message from intermediate node, source node will check its own table to see whether intermediate nodes are its reliable node. If source node has used intermediate node before for routing data packets, then intermediate node is a reliable node for source node.

[8] This paper resolves the problem of [1] paper that is resending FRREQ packet from the source node towards the next hop and waiting for FRREP packet from the next hop means increasing in routing overhead packets between the source and next hop node. In this paper instead of source node previous node monitors the next node by using same packets FRREQ and FRREP packets. The proposed LID security routing mechanism takes into consideration that the previous node is

trusted and there is no group attack in the network; which means that if the intermediate node is suspected, then it performs a single attack on the network.

3 Proposed Solution

It is observed from the state of the art studies that most of the existing intrusion detection schemes suffer from communication overhead, due to the frequent monitoring involved. This may prove to be dangerous in case of high traffic. In the following section, a novel technique for intrusion detection is proposed to overcome the monitoring overhead problem.

Above discussed papers uses different procedures to monitor intermediate nodes. The monitoring nodes may be either source node or previous node or the checking node itself. The proposed solutions have been observed to generate extra overhead leading to power loss and increase of traffic. This paper tries to solve this problem and find solution path in such a way that confirms that RREP packet only comes from the destination node. This solution need not monitor intermediate nodes after establishing the path.

The proposed solution uses the following assumptions:.

- Each node contains a same set of prime numbers with increasing order.
- RREQ and RREP packet contains a counter that increases with each hop.
- It uses AODV algorithm with some additional features.

3.1 Logic Description

The proposed solution appends some fields to the RREP and RREQ messages. The RREP message format is {HP_count, S_INFO, original_message}. The RREQ message format is {HP_count, R_INFO, E, original_message}. The HP_count is the total number of hops needed to send packets between source and destination. E holds an integer value which is used by the RSA algorithm for encryption/decryption. When RREQ and RREP packets are sent by source or destination, this packets use its counter to counts the number of intermediate node between source and destination node. The HP_count value is used to select two prime numbers for encrypting data. S_INFO (value of INFO field in the RREQ packet) field contains some integer value sent by source node which is used for confirming that reply comes from destination node. R_INFO (value of INFO field in the RREP packet) contains encrypted value that is sent by destination node.

Algorithm

Step 1: Source node sends RREQ to its neighbor nodes.
Step 2: On receipt of RREQ packet, destination node, it uses HP_count field value to find out prime number which will be chosen from the set of prime numbers for implementing RSA algorithm. For example, if hop count value is 10 then two prime numbers will be chosen those are 10th and 11th prime number.

Step 3: Then destination node adds value of S_INFO field of RREQ with 10th prime number and selects an integer E i.e 1<E< multiplication of prime numbers i.e N.
Step 4: Then added value is encrypted by implementing RSA algorithm using N, E then encrypted value is being kept in R_INFO field.
Step 5: Destination node sends value of E and R_INFO via same path to the source node so that value of HP_ count (10) in RREP packet is same as RREQ packet.
Step 6: On receiving packet, the source node uses HP_count value and it retrieves prime number from the set of prime numbers having same sequence number.
Step 7: By using value of E and prime numbers it decrypts R_INFO of RREP packet and subtract value of 10th prime number and checks that it is equal to the value of S_INFO of RREQ packet. If value is equal then only it sends packet via that path.

4 Result Description

This section presents a comparative analysis of the proposed technique against the techniques as presented in [8] which contains effect of number of nodes in the network overhead.

4.1 Network Overhead

Network over head is calculated by the number of control packets exchanged by all nodes before sending data. In order to study the effect of the number of nodes in network overhead by proposed algorithm and LID security routing mechanisms over AODV routing protocol, the combination of 20, 40, 60 and 80 nodes are used for simulation and simulation time is 15 minutes. The effect number of nodes in network overhead is shown below.

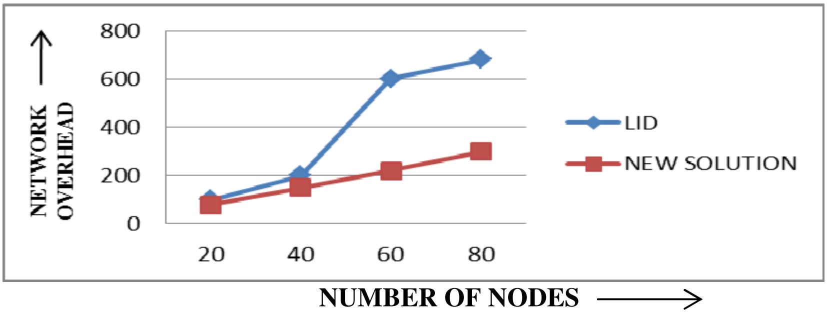


Fig. 1. Effect of number of nodes in network overhead

5 Conclusion

In MANET, the selfish nodes with some misbehavior are common, one type of such attack is Black Hole Attack in AODV routing protocol. In this attack selfish node gives false reply to source node. If this reply comes before destination node reply then it creates problem. The state of the art study shows that most of the existing black hole attack detection schemes suffer from network overhead. The proposed paper aims to reduce the network overhead. Most of the previous paper uses monitoring to solve this problem but this paper proposes a solution without monitoring. This paper uses RSA algorithm with AODV to confirm that reply comes from destination node and its number of control packet same as AODV. As number of control packet is less, overhead is minimized.

References

1. Deng, H., Li, W., Agrawal, D.: Routing security in ad hoc networks. IEEE Communications Magazine 40(10), 70–75 (2002)
2. Perkins, C., Belding-Royer, E., Das, S.: Ad-hoc on-demand distance vector (AODV) routing. Internet Draft, RFC 3561 (July 2003)
3. Medadian, M., Yektaie, M.H., Rahmani, A.M.: Combat with Black Hole Attack in AODV routing protocol in MANET. IEEE (2009)
4. Lu, S., Li, L., Lam, K.-Y., Jia, L.: SAODV: A MANET Routing Protocol that can Withstand Black Hole Attack. In: International Conference on Computational Intelligence and Security (2009)
5. Saini, A., Kumar, H.: Effect Of Black Hole Attack On AODV Routing Protocol In MANET. IJCST 1(2) (December 2010)
6. Wu, L.-W., Yu, R.-F.: A Threshold-Based Method for Selfish Nodes Detection in MANET, pp. 875–882. IEEE (2010)
7. Sen, J., Koilakonda, S., Ukil, A.: A Mechanism for Detection of Cooperative Black Hole Attack in Mobile Ad Hoc Networks. In: Second International Conference on Intelligent Systems, Modelling and Simulation (2011)
8. Abdelhaq, M., Serhan, S., Alsaqour, R., Hassan, R.: A Local Intrusion Detection Routing Security over MANET Network. In: International Conference on Electrical Engineering and Informatics, Bandung, Indonesia, July 17-19 (2011)

Connecting Entropy-Based Detection Methods and Entropy to Detect Covert Timing Channels

Bukke Devendra Naik, Sarath Chandra Boddukolu, Pothula Sujatha,
and P. Dhavachelvan

Department of Computer Science, Pondicherry University, Pondicherry, India
{devendrabukke,yours.sarath,spothula,
dhavachelvan}@gmail.com

Abstract. In this paper an entropy-based approach for detecting the covert timing channels is proposed. The detection of covert timing channels is the challenging task over the internet. Ordinary things such as existence of a file or time used for computation, have been the medium through which covert channel communicates. Covert timing channels are not easy to detect because these media are so numerous and frequently used. Different approaches are implemented to detect various covert timing channels. Existing techniques are efficient but have to adopt more than one approach. Applying more than one approach to detect the covert timing channels is the risk process. In this paper, only one approach is used by this efficiency is improved while applying this proposed technique improvements to be made for proposed entropy and corrected conditional entropy in detecting covert timing channels. An entropy-based approach is sensitive to the current covert timing channels.

Keywords: covert timing channel, network security, detection tests, overt channel.

1 Introduction

Lately, Internet has been used for every activity of our daily lives and it plays a vital role in our daily lives. People send sensitive information through emails, posts etc, that sensitive information should not revealed to the third party. In contrast a network covert channel leaks the information across the network, by violating security policies [2]. The covert channels are of two types, namely covert storage channels and covert timing channels. One of the differences between covert storage channels and covert timing channels is that covert timing channels are essentially memory less, whereas covert storage channels are not. In covert storage channel data is written to and read from sections of network packets not intended for data transmission. With a time channel while the sender sends the data the receiver have to receive immediately else the data will lost. But comparatively timing channels are more intricate to detect. So the detection of the covert timing channels over the Internet (0vert channel) is the challenging research. The definition of the covert timing channel is "the computer security attack which is capable of communicating the transforming information objects in the opposition of security policies. In simple manner leaking of the information

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 279–288.
springerlink.com

into the network, which is difficult to detect. While the overt channel is the communication path designed to transmit the information in the authorized way. In the original design of the communication channel design covert channel is not a part and is not authorized to access information in the communication channel [17] because of high variation in legitimate network traffic. The regularity of the timing channel can be used to differentiate the channel and measure their efficiencies. The timing or ordering of network is tampers by the covert timing channel over the Internet for mystery information transfer. Even though information is transforming secretly, the users are affected by the Internet threats due to the leakage of the covert timing channel. In the same way detecting covert timing is a well-known challenging task in the security community. The covert channel is relegated based on scenarios, noise and information flows. The covert timing channel first proposed in [21], where the sender either transmits or stays silent in a specific interval time.

Covert channel is used for information transmission, but is not designed nor intended for communication [20]. For detecting the covert timing channels there are some tests and approaches like statistical tests [1], [2], [3], but are not accurate and robust in capturing a covert timing channel and border detection, but are over sensitive to the high variation of network traffic. Huskamp's dissertation **[**13**]** has encountered the first systematic capacity analysis of timing channels**.** Quantized pump [14], which is easy to assemble, easy to analyse, has a provable upper bound on the covert channel bandwidth, and has better performance characteristics than the Pump. Later entropy-based approach is suggested to detect the covert timing channels. It is more difficult to detect the covert timing channels, due to the creation of the covert timing channel is created without pretending the entropy of the original process. Thus for detecting the covert timing channels entropy and conditional entropy approaches are used but for finite samples, the exact entropy rate of a process cannot be measured and must be forecasted [4]. Thus, we forecast the entropy rate with the corrected conditional entropy, a measure used on biological processes [5]. The corrected conditional entropy is designed to be accurate with limited data, which makes it excellent for small samples of network data. To evaluate our new entropy-based approach, we conduct a series of experiments to validate whether our approach is capable of differentiating covert traffic from legitimate traffic. We perform the fine-binned estimation of entropy and the coarse-binned estimation of corrected conditional entropy for both covert and legitimate samples. We then determine false-positive and true-positive rates for both types of estimations. Our experimental results show that the combination of entropy and corrected conditional entropy is very effective in detecting covert timing channels.

In this paper we propose the combination of the both entropy-based detection methods and entropy to detect the covert timing channels in the easiest way. The entropy relates to covert timing channel capacity.

The remainder of this paper is structured as follows: Section 2 covers related work in covert timing channels and their detection schemes. Section 3 describes covert timing channels in brief of different types of covert timing channels. Section 4 proposed method describes the basics of detection and implementation details. Section 5 has the experimental results for the proposed method. Finally section 6 is conclusion.

2 Background and Related Work

In recent years, researchers had suggested many different solutions to detect, interrupt, and take out covert traffic, but due to disruption system performance takes down. In normal way to mark difference between the covert traffic and the legitimate traffic, and the statistical methods are used and then detects the covert timing. In the statistical methods first order and second order or higher order statistics are used to describe the shape of the traffic.

In the normal way compare to the covert storage channel the detection of the covert timing channels are hard. The communication may be one way or two ways (duplexers). The following section follows the overview of the covert timing channels, and detection tests.

a) Ack filter

In one way communication (duplex and half-duplex) systems also covert timing channels are exist. Here low security network will communicate to high security network are communicate each other with some conditions by using ACK filter [15] as a gateway. In these communication lower can forward data to higher but in response to the lower higher can sends only acknowledgements [15]. But intruders can observing the timing of acknowledge from higher to the lower and they can send acknowledges to lower as they are higher. The other disadvantage is it is for one way communication.

b) Buffer the information in the transmission process

In these method of technique the data from lower to high will send , and it pass through a gateway and it sends the acknowledge to lower and forwards data to higher, then if higher sends acknowledge , the data from buffer will be deleted otherwise again it will sends the data [15]. Here we have to maintain a buffer and there is also a chance that an intruder may act as lower, the data will send from other way, other than through gateway. It is also one way communication.

c) Observing the packet delays in the communication

The covert timing channels, observes the packet inter-arrival time to encode covert messages. In the network, packets delays method it will be easy to investigates the channel capacity of Inter-based timing channels [19] and will be help for detecting covert timing channels based on how close a source comes to achieving that capacity of a channel.

Covert channels are three categories storage, Legitimate and Covert timing channel. In these detection method the statistical analysis of the Inter-packet delays does a good job for classifying between regular network traffic and traffic that is communicating through modulation of the inter-packet delays. Here only statistical analysis gives good result compare to the other analysis. In this method of detection, two methods are introduced. One is based on information theory concepts. It requires a thorough understanding of the network situation and a highly skilled attacker. The second method is based on the simple idea that if an even number of input symbols is used. Here the number of packets at the mean delay is very low compared to the maximum number of packets or any given delay. These are highly sensitive to channels noise, rendering them useless in practice. So the encoding packet delays in network also have some disadvantages.

d) Pelt and Attempt in Time

In this method of acting controlling the transmission time between the consecutive payload packets in overt network communication, the covert timing channels aim to transmit the hidden messages. Overt network communication is the communication path with in a computer system or network designed for authorized transfer of data. The overt communication is the authorized communication while the covert communication is not authorized. The covert timing is designed systematically which is statistically undetectable by shape and regularity tests and is robust against disruptions caused by active adversaries and/or noise in the network. Here by encoding the message, robustness is achieved using a spreading code scheme. But in this method by using a model-based modulation scheme estimate any legitimate traffic distribution is fulfilled undetectable.

3 Different Covert Timing Channels

The objective of the proposed scheme is to select the most relevant features using statistical characteristics of the subband coefficients, thus reduce the dimensionality of feature set and increase the accuracy of detection. In this paper, the first four normalized moments of high frequency, low frequency subband coefficients and structural similarity measure of medium frequency subband coefficients are taken as the feature set. With these five features, a three layer back propagation neural network is trained for further classification. The block diagram of the proposed model is given in figure.1. The following sub sections briefly explain contour let transformation how the feature set is extracted from images and how the classification is done.

Covert timing channel is one of the vulnerability. It conveys the information by modulating some aspects of system behaviour over time, so that the program receiving the information can observe system behaviour and infer protected information. The main purpose of covert timing channel is to leaking information in the network. These are of different types and are as follows.

a) Two Types of Covert Timing Channels

There are two types of covert timing channels: active and passive. Active covert timing channel generates additional traffic (as shown in the fig 1), while passive refers to covert timing channels that manipulate the timing of existing traffic. As shown in the fig1: the normal traffic is mentioned with green colour and additional traffic is mentioned with the maroon red colour. Depending upon the types the covert timing channels are explained as follows.

b) Internet Protocol Covert Timing channel (IPCTC)

The IPCTC investigates a number of design issues. A scenario where IPCTC can be used is illustrated in Fig. 1. In this scenario, a machine is compromised and the defensive perimeter represented as a perimeter firewall or intrusion detection system monitors communication with the outside.

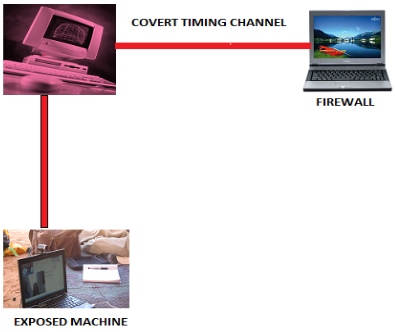


Fig. 1. IPCTC / TRCTC / MBCTC scenario

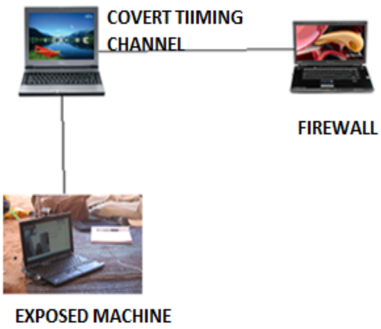


Fig. 2. passive or jitterbug

c) Model-Based Covert Timing Channels (MBCTC)

The MBCTC fits a sample of legitimate traffic to several models, such as Exponential, and selects the model with the best fit. It uses the inverse distribution function and cumulative distribution function for the selected models as encoding and decoding functions.

d) Time-Replay covert Timing Channel (TRCTC)

The TRCTC uses a sample of legitimate traffic Sin as input and replays Sin to transmit information. Sin is partitioned into two equal bins S0 and S1 by a value tcutoff. This will creates legitimate network traffic.

e) Swing Bug

It is a keyboard that slowly leaks the information over the network and is a passive covert timing channel. A scenario where swing bug or Jitterbug can be used is illustrated in Fig. 2. It belongs to class of inline interaction mechanisms [18]. In the trusted environment it is positioned at input devices and transmits the data covertly to affect externally observable network traffic by perturbing the timing of input events. There by leaks the sensitive data without compromising the host or its software and captures passwords through small variations in the precise times at which keyboard events are delivered to the host.

f) other covert timing channels

A simple binary covert timing channel based on the arimoto-Blahut algorithm [6] [7]. Combinatorics–based scheme, called cloak (consider as storage and timing channels), to transmit information in the order of packets within different flows. Covert timing channel based on the timing of TCP bursts [8].

g) Network Covert Timing channel (NCTC)

It is the attack in which attackers are used to communicate with the compromised hosts particularly in distributed denial of service (DoS) attacks [9].

4 Proposed Method

In this section, we have explained basics of the method later the combination of entropy-based detection methods and entropy i.e., the actual proposed method.

a) ***Entropy***

Entropy is the quantitative measure of disorder in a system. Entropy is what the equations define it to be. The word entropy came out from the thermodynamics. The rate of entropy is the average entropy per random variable, can be used as a measure of complexity or regularity [19]. Now we can explain the entropy in the form of equation. The entropy H, of a discrete random variable X is a measure of the amount of uncertainty associated with the value of X.

Suppose one transmits 1000 bits (0s and 1s). if these bits are known ahead of transmission logic dictates that no information has been transmitted. Here we are using to be a certain value with absolute probability. If however, each is equally and independently likely to be 0 or 1, 1000 bits have been transmitted. Between these two extremes, information can be qualified as follows. If X is the set of all messages {x1,x2,……,xn} that X can be, and p(x) is the probability of x given some x € X, then the entropy of X is defined.

$$H(x) = Ex\,[I\,(x)\,] = -\sum_{x \,€\, X} P\,(\,x\,)\log P(\,x\,)$$

Ix is the self information, which is the entropy contribution of an individual message, and Ex is the expected value.

The special case of information entropy for a random variable with two comes is the binary entropy function, usually taken to the logarithmic base 2:

$$Hb\,(\,P\,) = -\,P \log 2\, P - (\,1\text{-}\,P\,) \log 2\,(\,1 - P\,)$$

Here we are using joint entropy of two discrete random variables X and Y is merely the entropy of their pairing: (X,Y). This implies that if X and Y are independent, then their joint entropy is the sum of their individual entropies.

For example , if (X,Y) represents the position of a chess piece X the row and Y then columns then the joint entropy of the row of the piece and the column of the piece will be the entropy of the position of the piece.

$$H(X,Y) = EX,Y\,[-\log p\,(x,\,y)] = -\sum p(x,\,y)\log p\,(x,\,y)$$

For sequence of infinite length, the conditional entropy is the entropy rate. So for finite length samples it is unable to measure.

b) ***Corrected Conditional Entropy (CCE)***

For the finite samples the exact entropy rate cannot be measured, so they must be estimated. Based on the histograms the probability density functions can be replaced with empirical probability density functions. By using specific strategy the data is binned into Q bins. The empirical probability density functions are determined by the proportions of bin sequence number patterns.

In sympathetic outflow the regularity can be measured by the corrected conditional entropy. It is a new method for measuring the regularity of a process over short data sequences is reported. It is designed to decrease in relation to the regularity of the process, but it is able to increase when no robust statistic can be performed as a result of a limited amount of available samples.

In the conditional entropy the length m pattern can be predicted by the length m-1pattern and the length m and m-1patterns cancel out. But in corrected conditional entropy there is no need to fix the length m.

The conditional entropy of a random variable X, Y is the average conditional entropy over Y.

$$H(X/Y) = EY\,[H(X/y)] = -\sum_{y\text{€}Y} P(y) \sum_{x\text{€}X} p(x/y)$$

$$\log p(x/y) = -\sum_{x,y} P(x,y) \log p(x,y)/p(y)$$

Because entropy can be conditioned on a random variable or on that random variable being certain value, care should be taken not to confuse these two definitions of the conditional entropy, the former of which is in more common use. A basic property of this form of conditional entropy is that,

$$H(X/Y) = H(X,Y) - H(Y)$$

c) ***Contribution***

In this paper we are proposing the connection between the entropy-based detection methods and entropy as it relates to covert timing channel capacity. Here we explain entropy and entropy-based detection and in the implementation we can show the performance.

5 Sample Experimental Results

The covert timing channels are very serious threatens in the networking. After the experimental results is also found that detecting covert timing channels is more hard compare to the covert storage channels. In this paper it has been proved that the single proposed method will be able to detect all the covert timing channels. Compare to the other methods the proposed method will perform effectively in detecting the timing channels. The entropy is a measure of uncertainty, so it is used in the detection of timing channels and performance of this method is effective compare to the previous methods as follows.

The detection rates of IP Covert Timing Channel (IPCTC), Time-Replay Covert Timing Channel (TRCTC), Model-Based Covert Timing Channel (MBCTC) with true positive values are listed in the table 1. Here KSTEST for IPCTC is 1.00, TRCTC is .01 and MBCTC is .03. Regularity test for three timing channels .54, .04, .02, EN test is 1.00, .02, .55 and CCE is 1.00 for IPCTC and TRCTC and for MBCTC is .95. By observing the results it can understand that efficiency is increasing and by the proposed method also the efficiency is increased.

Table 1. Detection rates of IPCTC, TRCTC, MBCTC with HTTP TEST

	HTTP-TEST	IPCTC	TRCTC	MBCTC
TEST	False positive	True positive	True positive	True positive
KSTEST	.00	1.00	.01	.03
Regularity	.01	.54	.04	.02
EN	.01	1.00	.02	.55
CCE	.01	1.00	1.00	.95

In the fig 3 three timing channels are detected with different detection methods. We are mentioning with different colour for each test.

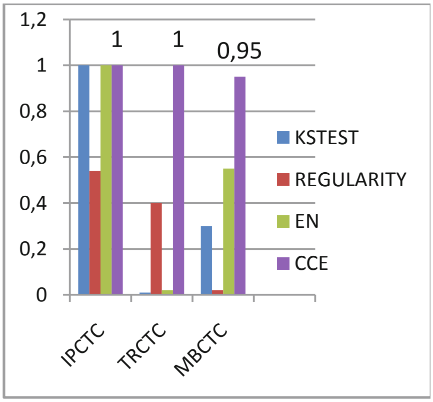


Fig. 3. Entropy Detection Method

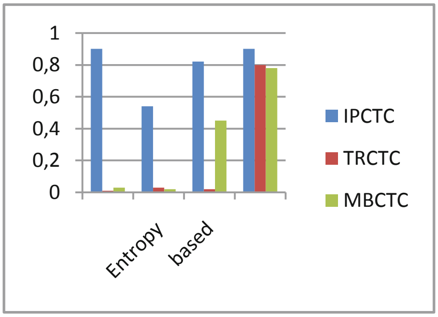


Fig. 4. Entropy-Based Detection Methods

In the proposed method we are increasing the efficiency of different covert timing channels detection. The results are as in the table 2 and by the experimental results it is observed that our proposed method is efficient compare to the existing methods.

In table 2, by taking some samples we conducted test in detecting different covert timing channels. Here we classified the data in to bins and there by test the data and by observing the obtained results it is conform that proposed method is efficient in detecting the different covert timing channels.

Table 2. Proposed method results

	IPCTC	TRCTC	MBCTC
TEST	True positive	True positive	True positive
Entropy-based detection method	0.9	.01	.03
	.54	.03	.02
	1.00	.02	.45
	0.9	0.8	0.78

By obtaining sample experimental results and observing the results we conforming that the proposed method is efficient in detecting the different covert timing channels. In fig 3 and fig 4 it is observed that maximum value is 1 and 0.9 respectively. Here with in short period timing channels are detecting.

6 Conclusion

In this paper we going to implement the connection between entropy-based detection methods and entropy as both are related to capable of detecting covert timing channels. The possibility of covert channels cannot be completely eliminated, although it can be significantly reduced by careful design and analysis. In previous there are many detection methods, in that some are failed to detect some covert timing channels and some are restricted to detect some timing channels. But by using the corrected conditional entropy test and entropy test is able to detect covert timing channels with abnormal regularity and abnormal shape respectively. Here we are using connection between entropy-based detection methods and entropy to detect the covert timing channels. The covert timing channels are more difficult to detect compare to the covert storage channel. There are some other methods to detect the covert timing channels as it is the high variation in legitimate network traffic.

References

1. Cabuk, S.: Network Covert Channels: Design, Analysis, Detection and Elimination. Purdue Uni. (December 2006)
2. Cabuk, S., Brodley, C., Shields, C.: IP Covert Timing Channels: Design and Detection. In: Proc. ACM Conf. Computer and Common. Security (October 2004)
3. Shah, G.: Keyboards and Covert Channels. In: Proc. USENIX Security Symp. (July/August 2006)
4. Cloak: A Ten-Fold Way for Reliable Covert Communication. In: Luo, X. (ed.) Proc. European Symp. Research in Computer Security (September 2007)
5. Porta, Baselli, Liberati: Measuring Regularity by Means of a Corrected Conditional Entropy in Sympathetic Outflow. Biological Cybernetics (January 1998)
6. Arimoto, S.: An Algorithm for Computing the capacity of Arbitrary Discrete Memory less Channels. Proc. IEEE Trans. Information Theory (January 1972)
7. Blahut, R.E.: Computation of Channel Capacity and Rate-Distortion Functions. IEEE Trans. Information Theory (July 1972)
8. Luo, X., Chan, E.W.W., Chang, R.K.C.: TCP Covert Timing Channels: Design and Detection. In: Proc. IEEE Int'l Conf. Dependable Systems and Networks (June 2008)
9. Henry, P.A.: Covert channels provided hackers the opportunity and the means for the current distributed denial of service attacks. Technical report (2000)
10. Wang, X., Chen, S., Jajodia, S.: Tracking Anonymous Peer-to-Peer VoIP Calls on the Internet. In: Proc. ACM Conf. Computer and Comm. Security (November 2005)
11. Peng, P., Ning, P., Reeves, D.: On the Secrecy of Timing-Based Active Watermarking Trace-Back Techniques. In: Proc. IEEE Symp. Security and Privacy (May 2006)
12. Moddemeijer, R.: On Estimation of Entropy and Mutual Information of Continuous Distributions. Signal Processing (1989)

13. Huskamp, J.C.: Covert communication channels in timesharing systems, Ph.D. thesis. Univ. of Califomia, Berkeley, CA (1978); also tech. rep. UCB-CS-78-02 and Electron. Res. Lab. Memo. No. ERLM78/ 37
14. Ogurtsov, N., Orman, H., Schroeppel, R., O'Malley, S., Spatscheck, O.: Experimental Results of Covert Channel Limitation in One-way Communication Systems. Department of Computer Science University of Arizona Tucson, AZ 85721, nicko,ho,rcs,sean,spatsch @cs.arizona.edu
15. Ogurtsov, N., Orman, H., Schroeppel, R., O'Malley, S., Spatscheck, O.: Experimental Results of Covert Channel Limitation in One-way Communication Systems Department of Computer Science University of Arizona Tucson, AZ 85721 nicko,ho,rcs,sean,spatsch @cs.arizona.edu
16. A guide to understand covert channels analysis of trusted systems. Virgil Gligor. Technical Report NCSC_TG_030, National Computer Security Center, Ft. George G.Meade, Maryland, U.S.A (November 1993)
17. Shah, G.: Keywords and Covert Channels. Andres Molina and Matt blaze. University of Pennsylvania
18. Gianvechio, S., Wang, H.: An Entropy-Based Approach to Detect Covert Timing Channels. IEEE Trans. Dependable and Secure Computing 8(6) (November 2011)
19. Berk, V., Giani, A., Cybenko, G.: Detection of Covert Channel Encoding in Network Packet Delays. Dept. of Comp. Sci. (November 2005)
20. Liua, Y., Ghosal, D., Katzenbeisser, S.: Hide and Seek in Time- Robust Covert Timing Channels. Dept. of Electrical and Comp. Sci., University of California, Davis, USA
21. Padlipsky, M., Snow, D., Karger, P.: Limitations of end-to-end encryption in secure computer networks. Tech. Rep. ESD TR-78-158, Mitre Corporation (1978)
22. Lampson, B.W.: A note on the confinement problem. Proc. of the Communications of the ACM (16), 10 (1973)

Route and Load Aware Channel Assignment Algorithm for Multichannel and Multi Radio Vehicular Ad-Hoc Networks

Jagadeesh Kakarla and S. Siva Sathya

School of Engineering and Technology, Department of Computer Science
Pondicherry University, Pondicherry, India
{jagadeesh0826,ssivasathya}@gmail.com

Abstract. Vehicular Ad-hoc Network (VANET) is a special type of Mobile Adhoc Network (MANET) technology that recently gained lot of prominence in urban areas. VANETs consider a combination of Road Side Units (RSUs) and vehicles with wireless radios. One primary issue in VANET is Medium Access Control (MAC), which aims to utilize the radio spectrum efficiently, to resolve potential contention and collision among vehicles for using the medium. Contention reduces the performance of single channel MAC protocols and the workload distribution among the vehicles is also not well defined. Hence Multi channel Multi radio architecture is useful to provide better quality of services (QoS). But here multi channel interference is a major problem to assign channel to a particular vehicle. The proposed work describes a cross layered approach with distributed channel assignment algorithm. The multipath routing is carried out with the help of a distributed channel assignment algorithm to assign channels to each vehicle by finding out each channel load. This route and channel load aware algorithm provides better quality of service in VANET. Simulation results reveal that, Weighted Cumulative Expected Transmission Time (WCETT) routing metric in multipath routing algorithm integration with multi channel and multi radio outperforms AODV and DSR in various QOS parameters.

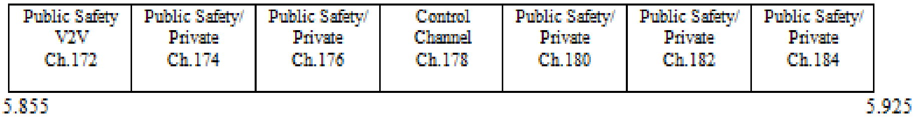
1 Introduction

Vehicular Ad hoc Network (VANET) is a special type of Mobile Ad hoc Network (MANET) which is an emerging technology and recently gained lot of prominence in urban areas. In VANET each vehicle acts as a router to exchange data between nodes in the network. This type of networks supports vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (VIC) communication. Such networks are used in traffic control applications, safety applications, driver assistance and location based services. The main challenge in VANET is high mobility [10] with the constraint of road topology, initially low market penetration ratio, unbounded network size and infrastructure support that differentiate it from MANET.

Based on IEEE 802.11p [1], VANET uses Dedicated Short Range Communication (DSRC) for improving the drivers comfort and safety. The U.S.Federal Communication Commission (FCC) has allocated 75MHz of spectrum at 5.9GHz exclusively for

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 289–298.
springerlink.com

vehicular communications. The overall spectrum [5] is divided into seven 10 MHz wide channels as described in figure 1. The Channel 178 is assigned as Control Channel (CCH) for announcements and short data messages. Channel 172 and Channel 184 are used for safety application in V2V communication. The remaining channels are designated as Service Channels (SCHs), where additional data transfer takes place. DSRC allows a data payload communication capability of 3, 4, 5, 6, 9, 12, 18, 24, and 27 Mbps. Using the optional 20 MHz channels, it allows data payload capabilities up to 54 Mbps.



Public Safety V2V Ch.172	Public Safety/ Private Ch.174	Public Safety/ Private Ch.176	Control Channel Ch.178	Public Safety/ Private Ch.180	Public Safety/ Private Ch.182	Public Safety/ Private Ch.184

Fig. 1. 5.9GHz DSRC spectrum

The rest of the paper is organized as follows. Section 2 discusses related work, section 3 defines problem definition, section 4 describes network model, section 5 discusses our proposed work, section 6 deals with simulation results and section 7 concludes the paper.

2 Related Works

Multi channel protocols contain two major components, channel assignment and media access control. An efficient channel assignment method can reduce radio interference among concurrent transmissions within a single channel, and make medium access control easier. Channel assignment plays an important factor in multi channel communication. Channel assignment methods are of three types as static, dynamic and hybrid.

2.1 Static Channel Assignment

In static channel assignment, channel selection process is executed at the beginning of system deployment, or very infrequently during runtime. For example, G. Zhou [9] proposed an Even Selection channel assignment scheme, nodes first exchange their IDs among two communication hops to reduce hidden and exposed terminal problems. In next phase nodes begin to choose channels in the increasing order of their ID. One drawback of this scheme is it doesn't use traffic information of a node for channel assignment and assumes traffic is evenly distributed on each node. But this is often not true in reality. To overcome this problem Yafeng Wu [8] proposed a Traffic-Aware Channel Assignment in Wireless Sensor Networks. In this scheme, nodes exchange their IDs and their traffic weights among two communication hops to gather information about its two-hop neighbors. In the next phase nodes take channel decisions in the decreasing order of their traffic weight, with the smallest node ID used as a tie breaker. If a node has the greatest traffic weight among its two communication hops, it chooses the channel with the least load among available channels, and then beacons the channel choice within two hops. After receiving this beacon, nodes update the load of the corresponding channel. One disadvantage for this scheme is it assumes that the traffic data rate does not change in future. One major drawback of all static channel assignment schemes is the channels for all the radios of a node are fixed and degrades the performance of entire network.

2.2 Dynamic Channel Assignment

In dynamic channel assignment scheme, a distributed or centralized mechanism assigns channels dynamically to nodes according to the traffic information in the network. For example, X. Chen [7] proposed a Cluster Channel Assignment (CCA) approach to maximize the aggregate throughput while minimizing the interference and exploiting spatial reuse and local dynamic switching of the channels. OzlemDurmazIncel [6] proposed a schedule-based multichannel Mac protocol provides an interference and collision free parallel transmissions on different channels. Later Xiaodong Wang proposed [2] a multi-channel real-time communication protocol. It is a flow-based channel allocation strategy and partitions the network based on many-to-one data flows. One potential pitfall of all these schemes is they can change the network topology. Topology changes can lead to sub-optimal routing and even network partitioning in case of node failures.

2.3 Hybrid Channel Assignment

In hybrid channel assignment scheme, the channels for a subset of radios are fixed and for the remaining radios channels may vary dynamically during communication. The hybrid multichannel protocol (HMCP) [3] ensures connectivity between nodes by allowing one of the two wireless interfaces to switch across channels as required. The other interface remains fixed on a channel as long as the channel is perceived to be good. In this scheme a node can potentially transmit and receive simultaneously, if the channels on which they transmit and receive are different and control messages are exchanged to communicate the channel information between the nodes. Later Yang Yuwang [4] designed a Hybrid Multi-channel MAC protocol with Virtual Mechanism and Power Control for Wireless Sensor Networks. This protocol combines dynamic channel assignment, quasi-reservation mechanism, a virtual MAC frame mechanism to support larger network layer packets, a multi-channel virtual carrier sensing mechanism to estimate idle or busy channels effectively, and has an intelligent power control which adjusts the transmission power levels automatically according to the distance among network nodes, therefore it reduces the energy consumption and increases the life of the entire network.

3 Problem Definition

Given a vehicular Adhoc network with multiple radio interfaces, multiple paths from source to destination the problem is to assign one or multiple channels to each node, such that the number of different channels assigned to a node is not more than the number of radios on the node. The objective of the channel assignment problem for

VANET is to maximize the throughput and minimize the delay between a source node and a destination node.

Formally, the problem of hybrid channel assignment for a multi-radio VANET over a set of N nodes, given a set of K channels and set of multiple paths from source to destination is to compute a function f: N → P (K), to maximize the throughput and minimize the delay between a given source and destination pair (src, dst).

4 Network Model

A cluster is a group of nodes that can communicate without disconnection and that identify themselves to be part of a cluster. These nodes select a cluster head to coordinate the communication among them. In VANET due to high mobility of vehicles and lack of fixed infrastructure cluster formation is a dynamic, tangled process and utilizes lot of bandwidth for transferring control packets to form the cluster. So in our proposed model we form fixed clusters with help of Road Side Units (RSU) and select one RSU as a cluster head based on highest-degree algorithm. Each cluster head contains a list of RSU's and list of vehicles within the cluster. Reactive protocol is used in vehicle to vehicle communication and a proactive protocol is used between vehicle and RSU. The proposed algorithm assumes the VANET architecture as shown in figure-2. We use some infrastructure network with the help of road side units (RSUs). RSUs are equipped with minimum number of interfaces for providing multi channel assignment. The channels will be assigned based on the available channels in the network.

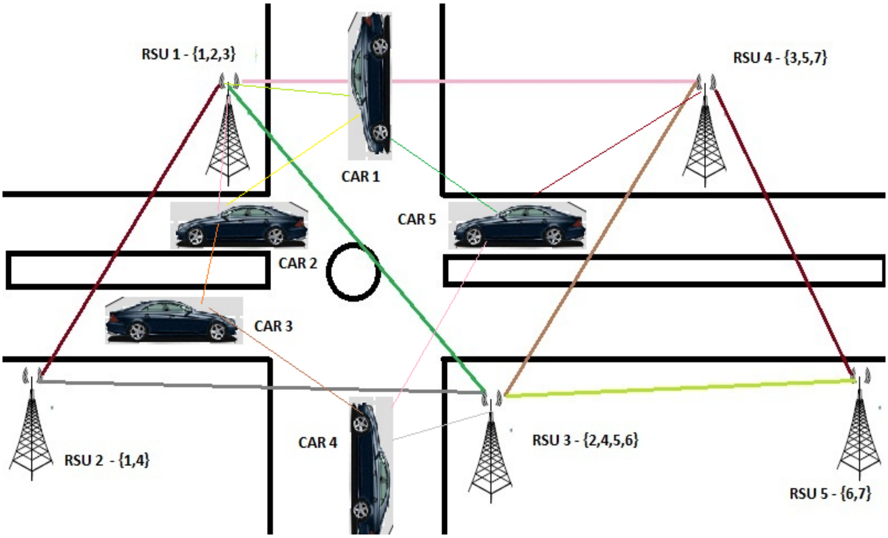


Fig. 2. VANETs architecture for multi-radio and multi channel

The network considered is purely an undirected graph with V set of nodes and E set of edges. We denote this as G (V, E). Each edge in set E has interference channels. Let I(E) be the set of interference channels with edge E, that are identified to reduce the interference in the network. In figure -2, different channels are indicated with different colors. Common color is used to indicate adjacent vehicles/RSUs.

5 Proposed Work

In this proposal, load aware channel assignment algorithm is discussed followed by multi-path routing in the network.

5.1 Load Aware Channel Assignment Algorithm

Let N- represents set of all nodes, R- represents set of road side unit radios, C-set of channels

1. Assign a default channel to all nodes to break the circular dependency between nodes workload and channel assignment, because the channel assignment depends on nodes workload and nodes workload depends on the channel assignment.

   For all Node i that belongs to a node in set of nodes N ( ) has default Channel assignment.

   So, $S_i^1 \leftarrow$ 1 means if a node i has C channels, channel 1 is assigned as a default channel and is used for all nodes. Binary values are used to represent channel assignment. If it is 1 channel is assigned, otherwise it is not assigned.
2. After assignment, that channel information is stored in a temporary variable. The Channel assignment information is useful for further processing.
3. Compute weight factor for each and every channel which is available in network. This weight factor is required to assign channel during transmission.
4. **Weight factor calculations**

   Calculation of weight factor Z is based on the available packet delivery ratio of that channel and duty D is number of times channel used in a time period.

$$Z_i^k = \frac{R_i^k L_i^k}{1 - \prod_{j<i}(1 - R_i^k(1 - link\ loss\ probability(i,j))} \tag{1}$$

   R_i^k denotes channel k is assigned to node i.

   L_i^kprovides channel assignment information of a link at node i.
5. Compute throughput for assigned channel for node S and its weight factor is Z and

   Throughput =

$$\sum_{s\epsilon S}(R_d^k \sum_{i\epsilon N} T_i^k\ (1 - linklossprobability(i,j)) \tag{2}$$

T_i^kProvides all channels normalized affective transmission rates at node i.

6. While throughput values are grater then zero, do
7. While assigned channel c is less than the number of channels C at i do
8. Maximum weight factor is assigned for channel C at node i
9. If i is not equal to destination (d) then
10. Minimum weight factor assigned for channel C at i to eliminate the starvation problem.
11. Assign default channel for node.
12. End if
13. Again compute weight factor at D
14. End while
15. Compute throughput1 for channel Z at node D
16. Throughput = Throughput1 - Throughput
17. If throughput is greater than 0 then
18. Assign channel between the S and D.
19. end

5.2 Multi-path Routing Algorithm

Multipath routing is useful to find various paths from source to destination for reducing packet loss due to frequent route breakdowns. One major problem of multiple path routing is the large routing overhead generated during route search and maintenance. But it is necessary to provide better QoS in VANET. In Route discovery process the Route Request (RReq) message includes a route record which specifies the sequence of nodes traversed by the message. Intermediate nodes do not keep a route cache, so they can't reply to RReq's. This allows the destination to receive all the routes so that it can select the disjoint paths.

Let d(s) – represent alternative paths form source to destination, v(s, d ,l) – represent all possible paths between the source and destination, q(l) – gives the queue length.

Calculate the positive routing fraction of each path by using above parameters and Greedy approach.

Calculate the positive routing fraction of each path by using above parameters and Greedy approach.

$$\beta_s P_{sd} + \sum_{l=1}^{L} V_{sd}^{l} \times ql = \max(\beta_s P_{sd}) + \sum_{l=1}^{L} V_{sd}^{l} \times ql \triangleq q_{s,max} \tag{3}$$

$$then \sum_{l=1}^{L} V_{sd}^{l} \times ql > q_{s,max} \tag{4}$$

In this $V_{sd}^{l} istheroutingmatrixand\beta_s isthepositivenumberforeachs.$

Step 1: P_{sd} denotes positive routing fraction of all paths from source to destination

$$\overrightarrow{P_{sd}} = [P_{s1}, P_{s2}, \ldots., P_{sd(s)}]$$

If path partial fraction between any source node and destination is greater than or equal to 0 and the assigned summation of all paths between the source and destination is equal to 1, for all source nodes then

$$\boldsymbol{if} P_{sd} \geq 0 \; and \sum\nolimits_{d=1}^{d(s)} P_{sd} = 1, \forall s \; \boldsymbol{then} \tag{5}$$

Find the maximum path partial fraction between source and destination with the help of V_{sd}^{l},ql,

$$\max_{\overrightarrow{P_s}} = \frac{-\beta_s}{2} \sum\nolimits_{d=1}^{d(s)} (P_{sd})^2 - \sum\nolimits_{d=1}^{d(s)} P_{sd} \times \sum_{l=1}^{L} V_{sd}^{l} \times ql(t) \tag{6}$$

Step 2: calculate the next queue length at intermediate nodes using below equation

$$ql(t+1) = ql(t) + \sum\nolimits_{d=1}^{d(s)} V_{sd}^{l} P_{sd}(t) - \sum_{c=1}^{C} y_l^c(t) \tag{7}$$

Step 3: Repeat the step 2 to improve the efficiency of finding the queue length. The above algorithm is useful to find out the best possible path between source and destination.

6 Simulation Results

We have implemented our proposed algorithm in NS2, which has been highly validated by the networking research community.

Table 1. Simulation Parameters

Parameters	Value
MAC Layer	IEEE 802.11
Number of RSU + Number of nodes	4 + 6, 8+ 12, 12+18, 16+24, 20+30
Number of Radios and Channels	3,3
Simulation Area	600*600, 1000*1000 etc...
Simulation Duration	200 sec
Traffic Flow	TCP ,CBR
Mobility Pattern	Random wave point
Packet Size	512, 1024 etc...
Transmission range	100, 250 m
Node mobility speed	0 – 50 m/sec

End-to-End Delay: It is defined as the averaged time required for transferring a data packet from source to destination.

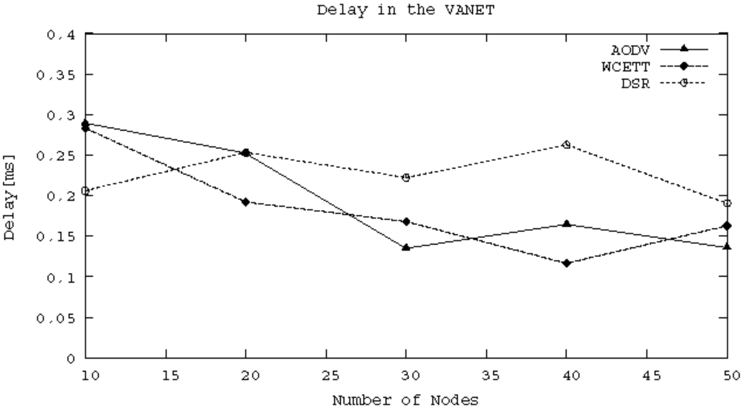


Fig. 3. Packet end-to-end delay

In figure 3, we have observed AODV and DSR have more end-to-end delay when compared to WCETT.

Packet Loss Ratio: The number of packets that were lost due to high mobility in VANET and MAC layer collisions.

In figure 4, we have observed that packet loss ratio is high in AODV and DSR when compared to WCETT because of less contention for channel in our algorithm and high throughput.

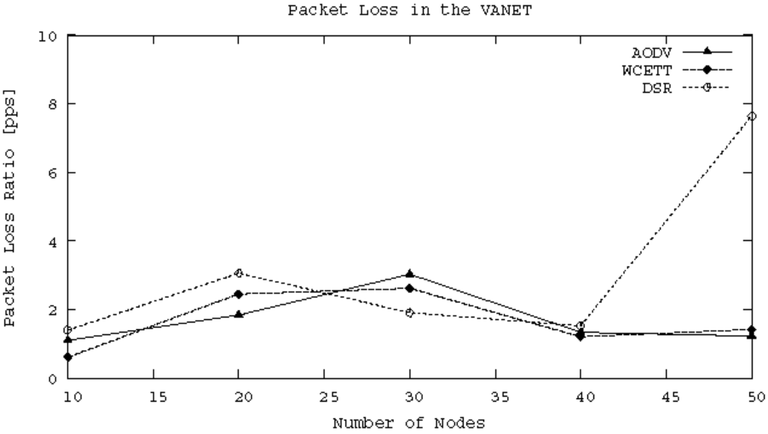


Fig. 4. Packet loss ratio

Throughput: Throughput denotes how much data can be transferred from source to destination in a given amount of time.

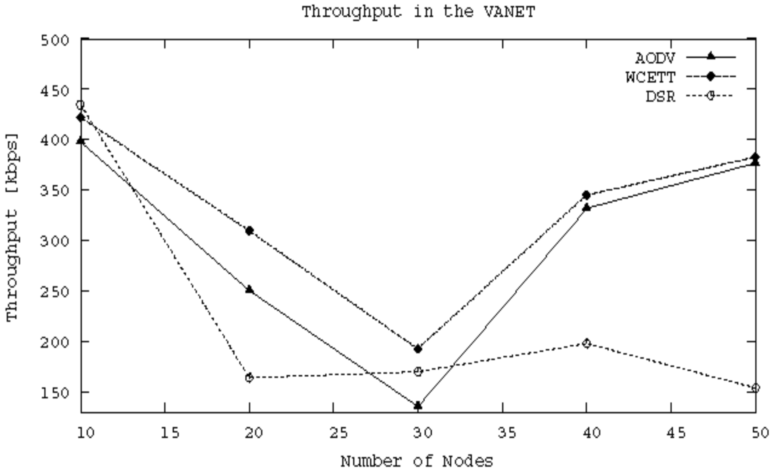


Fig. 5. Throughput in Kbps

In figure 5, we have observed Throughput is high in WCETT when compared to AODV and DSR because channel interference is less and channel assignment for a node is based on channel weight factor and duties.

Routing Overhead: The total number of control packets is transferred to route the data packets. It is defined as a ratio of control packets to the data packets.

In figure 6, we have observed routing overhead is high in WCETT when compared to AODV and DSR

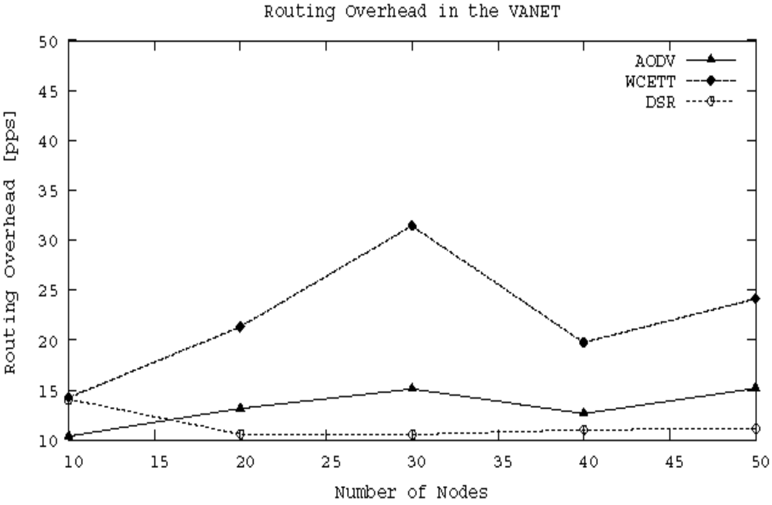


Fig. 6. Routing overhead

7 Conclusions

In this paper we investigated the performance of vehicle safety applications with the help of Cross layered Multi channel, Multi radio and Multi path routing algorithm. The study proceeded through two phases. In the first phase, we calculated the best path among multiple paths from source to destination by using Path partial fraction and Queue length at intermediate nodes. In the second phase assign the channels to each node among the available channels by calculating weight factor and duty of each channel. The simulation results of the algorithm show that the proposed method has improved the packet delivery ratio and throughput substantially. In future, to reduce the routing overhead we have to consider more parameters which will be our next course of research.

References

[1] Liu, K., Guo, J., Lu, N., Liu, F.: RAMC: A RSU-Assisted Multi-channel Coor-dination MAC Protocol for VANET. In: GLOBECOM Workshops. IEEE (2009)

[2] Wang, X., Wang, X., Fu, X., Xing, G., Jha, N.: Flow-Based Real-Time Communication in Multi-Channel Wireless Sensor Networks. In: Roedig, U., Sreenan, C.J. (eds.) EWSN 2009. LNCS, vol. 5432, pp. 33–52. Springer, Heidelberg (2009)

[3] Kyasanur, P., Vaidya, N.: Routing and link-layer protocols for multi-channel multi-interface ad hoc wireless networks. In: ACM MC2R (2006)

[4] Yang, Y., Ju, Y., Jin, B., Yu, J., Sun, Y., Yang, J.: A Hybrid Multi-channel MAC protocol with Virtual Mechanism and Power Control for Wireless Sensor Networks 3, 144, http://www.ubicc.org

[5] Kakarla, J., Siva Sathya, S.: A Survey and Qualitative Analysis of Multi-channel Mac Protocols for VANET. International Journal of Computer Applications 38(6) (January 2012)

[6] DurmazIncel, O., Jansen, P., Mullender, S.: MC-LMAC: A Multi-Channel MAC Protocol for Wireless Sensor Networks. In: Proceedings of IEEE INFOCOM (2008)

[7] Makram, S.A., Gunes, M., Kchiche, A., Krebs, M.: Dynamic Channel Assignment for Wireless Mesh Networks Using Clustering. In: Proceedings of ICN, pp. 539–544 (2008)

[8] Wu, Y., Keally, M., Zhou, G., Mao, W.: Traffic-Aware Channel Assignment in Wireless Sensor Networks. In: Liu, B., Bestavros, A., Du, D.-Z., Wang, J. (eds.) WASA 2009. LNCS, vol. 5682, pp. 479–488. Springer, Heidelberg (2009)

[9] Zhou, G., He, T., Stankovic, Abdelzaher, T.F.: Radio Interference Detection in Wireless Sensor Networks. In: IEEE INFOCOM (2005)

[10] Kakarla, J., Siva Sathya, S., Laxmi, B.G., Ramesh Babu, B.: A Survey on Routing Protocols and its Issues in VANET. International Journal of Computer Applications 28(4) (August 2011)

Performance Analysis of TCP & UDP in Co-located Variable Bandwidth Environment Sharing Same Transmission Links

Mayank Kumar Goyal, Yatendra Kumar Verma, Paras Bassi,
and Paurush Kumar Misra

Deptt. of CSE/IT, JIIT University, Noida, UP, India
{mayankrkgit,yatendra54,paras123,misrapaurush}@gmail.com,
http://www.jiit.ac.in

Abstract. Various communication protocols can be used simultaneously in a networking environment. This paper address the question that how much bandwidth is used by Transmission Control Protocol (TCP) and User Datagram Protocol (UDP) when they share the same link in transport layer and which protocol consumes more bandwidth than the other. A set of simple experiments has been conducted to find the effect of constant bit rate UDP traffic on adaptive TCP and vice versa. For that, four types of TCP which are TCP Tahoe, TCP Reno, TCP NewReno and TCP Vegas are used with UDP in variable bandwidth environment. From there, we are going to differentiate them in terms of bandwidth usage and define how it works and describes several effects that occurred when they work together.

Keywords: TCP, UDP, NS2, Tahoe, Reno, NewReno, Vegas, RTT.

1 Introduction

Transmission Control Protocol (TCP) is the predominant Internet protocol & carries approximately 90% of the internet traffic. One key point of TCP is its congestion control which regulates the millions of flows of the Internet to provide both efficiency and fairness. In wired networks, the available bandwidth for best-effort traffic is usually constant and is set to the link capacity most of the time. However, major changes are foreseen as many Internet providers (ISP) are beginning to deploy Quality of Service (QoS) features with reservation-like or priority like mechanisms in their networks to guarantee a given QoS level.

1.1 TCP Tahoe

In TCP Tahoe, the sender maintains a congestion window (CWD) that limits the no. of packets that it is allowed to send into the network without waiting for acknowledgment. The congestion control is a two-phase control mechanism: a slow-start phase and a congestion avoidance phase. The TCP sender dynamically increases/decreases the congestion window size according to the congestion level, which is

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 299–306.
springerlink.com

conjectured by packet losses. Switching from the first phase to second phase depends on the slow-start threshold (ssthres). Slow starts suggest that the sender set the congestion window to 1 and then for each ACK received it increase the CWD by 1. So in the first round trip time (RTT) it sends one packet, in the second it sends two and in the third it sends four. Thus it increases exponentially until it loses a packet which is a sign of congestion. The important thing is that TCP Tahoe detects packet losses by timeouts [6].

1.2 TCP RENO

TCP Reno adds some intelligence to TCP Tahoe so that the packets which are lost are detected earlier and the pipeline is not vacant every time a packet is lost. Reno requires that receiving immediate acknowledgement whenever a segment is received [1]. The logic behind this is that whenever it receive a duplicate acknowledgment, then his duplicate acknowledgment could have been received if the next segment in sequence expected, has been delayed in the network and the segments reached there out of order or else that the packet is lost. So Reno suggests an algorithm called 'Fast Re-Transmit'**.** Whenever it receives three duplicate ACK's, it takes it as a sign that the segment was lost, so it retransmits the segment without waiting for timeout [7]. Another modification that RENO makes is in that after a packet loss, it does not reduce the congestion window to 1. However, performance of TCP Reno suffers in case of multiple packet losses within a window of data [9].

1.3 TCP NEW-RENO

Like Reno, New-Reno also enters into fast-retransmit when it receives multiple duplicate packets, it doesn't exit fast-recovery until all the data which was out standing at the time it entered fast recovery is acknowledged. Thus it overcomes the problem faced by Reno of reducing the CWD multiples times.

1.4 TCP VEGAS

TCP Vegas detects congestion at an incipient stage based on increasing Round-Trip Time (RTT) values of the packets in the connection unlike other flavors like Reno, NewReno, etc., which detect congestion only after it has actually happened via packet drops.

1.5 User Datagram Protocol (UDP)

UDP is a simple, connectionless transport protocol (Stevens, 1994) .UDP does just about as little as a transport protocol can. Aside from the multiplexing/demultiplexing function and some light error checking, it adds nothing to IP. In fact, if the application developer chooses UDP instead of TCP, then the application is talking almost directly with IP. UDP takes messages from application process, attaches source and destination port number fields for the multiplexing/demultiplexing service, adds two other fields of minor importance, and passes the resulting "segment" to the network layer. The network layer encapsulates the segment into an IP datagram and then makes a best effort attempt to deliver the segment to the receiving host. With UDP there is no handshaking between sending and receiving transport layer entities before sending a segment. For this reason, UDP is said to be connectionless [5].

2 Design of Simulation Model

To test the performance effect of the TCP and UDP in term of bandwidth usage, the simulation was carried out with 10 nodes and by sending two applications (FTP and CBR) to represent the protocols TCP and UDP using Network Simulator 2 (NS2). The simulations have been carried out on different versions of TCP (Tahoe, Reno, NewReno and Vegas) and the rates of UDP were set to 200 Kb and 4Mb.

3 Results

Two scenarios with all TCP versions once with 200 Kb rates of UDP and other with 4Mb rate of UDP were simulated.

3.1 TCP Tahoe with 200Kb Rate of CBR

As shown in figure 1, TCP Tahoe consumed more bandwidth than UDP when the CBR is small (200Kb) and the maximum value of usage bandwidth is (2.2*10^6).

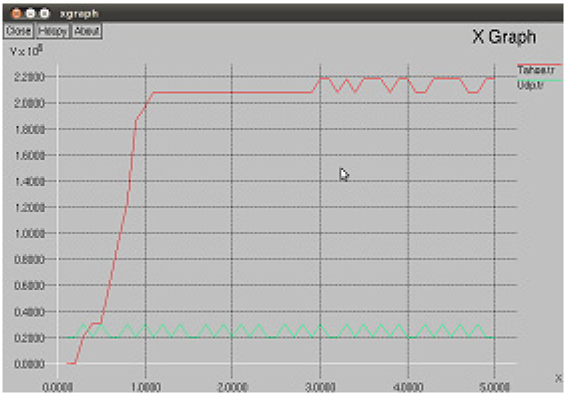


Fig. 1. The bandwidth usage of TCP Tahoe with 200Kb of UDP

3.2 TCP Tahoe with 4Mb Rate of CBR

As shown in figure 2 the bandwidth used by TCP Tahoe is less than bandwidth usage by UDP when the CBR is large (4Mb), and the UDP is started with big value of bandwidth usage (2.75*10^6) while the TCP started with zero value.

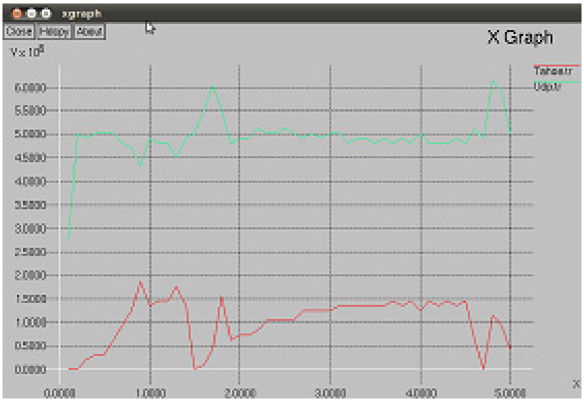


Fig. 2. The bandwidth usage of TCP Tahoe with 4Mb of UDP

3.3 TCP Reno with 200Kb Rate of CBR

As shown in figure 3, TCP Reno consumed more bandwidth than UDP when the CBR is small (200Kb) and the maximum value of usage bandwidth is (2.2*10^6) which is same as earlier in TCP Tahoe , which tells that there is no major difference in bandwidth usage in comparison to TCP Tahoe. It doesn't reduce the congestion window too much prematurely.

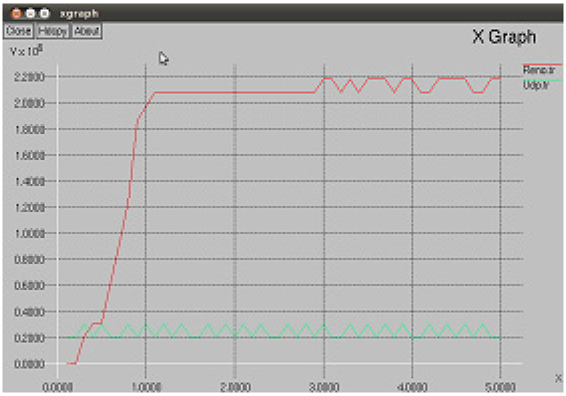


Fig. 3. The bandwidth usage of TCP Reno with 200Kb of UDP

3.4 TCP Reno with 4Mb Rate of CBR

As shown in figure 4 the bandwidth usage by TCP Reno is less than bandwidth usage by UDP when the CBR is large (4Mb), and UDP started with big value of bandwidth usage (2.75*10^6) while the TCP started with zero value.

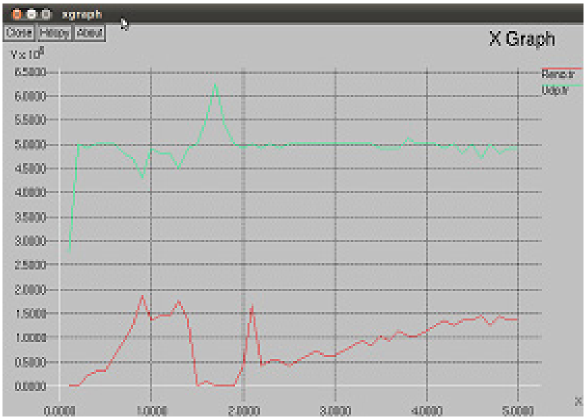


Fig. 4. The bandwidth usage of TCP Reno with 4Mb of UDP

3.5 TCP NewReno with 200Kb Rate of CBR

As shown in figure 5 TCP NewReno consumed more bandwidth than UDP when the CBR is small (200Kb) and the maximum value of usage bandwidth is (2.2*10^6) which is same as earlier in TCP Tahoe and TCP Reno. Fig 3.1, 3.3, 3.5 represents almost identical drawings for TCP Tahoe, TCP Reno and TCP NewReno. This fact represents that when TCP Tahoe, TCP Reno & TCP NewReno are used with UDP sharing same links, the value of bandwidth for UDP in all simulations is almost same.

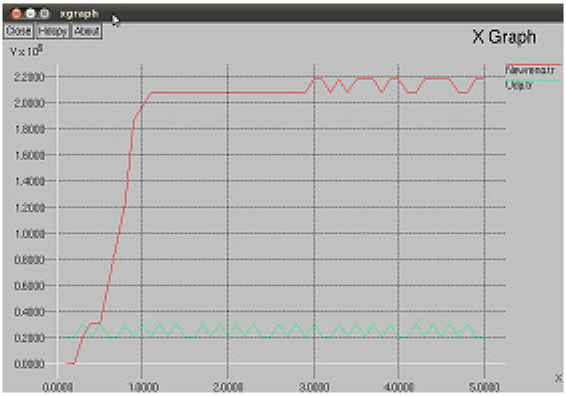


Fig. 5. The bandwidth usage of TCP NewReno with 200Kb of UDP

3.6 TCP NewReno with 4Mb RATE of CBR

As shown in figure 6 the bandwidth usage by TCP NewReno is less than bandwidth usage by UDP when the CBR is large (4Mb), and UDP started with big value of bandwidth usage (2.75*10^6) .

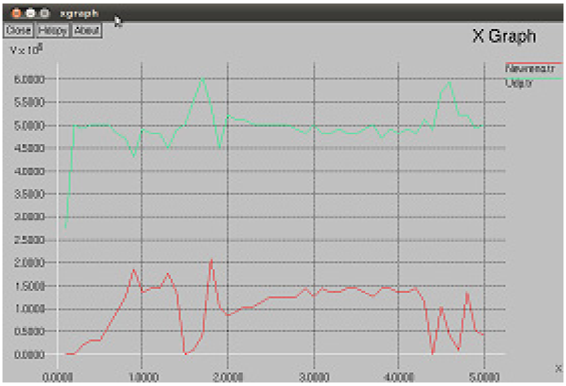


Fig. 6. The bandwidth usage of TCP NewReno with 4Mb of UDP

3.7 TCP Vegas with 200Kb Rate of CBR

As shown in figure 7 TCP NewReno is consumed bandwidth more than UDP because the CBR is small (200Kb) and the maximum value of usage bandwidth was (2.2*10^6).

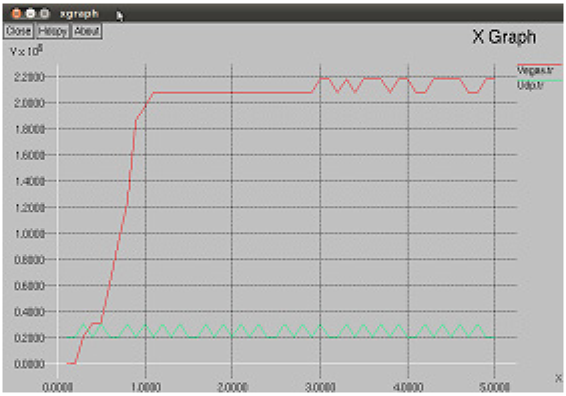


Fig. 7. The bandwidth usage of TCP Vegas with 200Kb of UDP

3.8 TCP Vegas with 4Mb Rate of CBR

As shown in figure 8 the bandwidth usage by 4Mb rate of UDP is more than bandwidth usage by TCP Vegas, but in this scenario the maximum value of bandwidth usage by TCP Vegas which is (5.15*10^6) is less among previous scenarios.

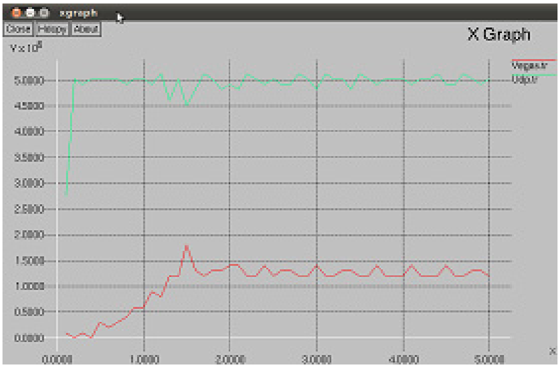


Fig. 8. The bandwidth usage of TCP Vegas with 4Mb of UDP

4 Conclusion

The result shows that the average bandwidth available to different versions of TCP congestion control mechanism are not the same. But for this network simulation environment, we prefer to adopt TCP Vegas because the average bandwidth is less than other TCP versions. Regarding the bandwidth share of TCP and UDP, we agree that UDP traffic is suppressing the bandwidth share of TCP traffic. To prevent that, the UDP rate should be decreased if the UDP rate is higher than TCP rate. Based on the graphs, there is no difference between all versions of TCP except for TCP Vegas which the average bandwidth of the TCP Vegas is the least among all the TCP versions. We also notice that the rate of UDP has the possibility to affect TCP send rates, depending on the UDP's rate. More higher the UDP's rate, more suppressed the TCP traffic will be at the shared bandwidth.

References

[1] Floyd, S., Fall, K.: Simulation based comparisons of Tahoe, Reno and SACK TCP. ACM Computer Communication Review 26(3), 5–21 (1996)

[2] Brakmo, L.S., Peterson, L.L.: TCP Vegas: End to End Congestion Avoidance on a Global Internet. IEEE Journal on Selected Areas in Communication 13, 1465–1490 (1995)

[3] Holland, G., Vaidya, N.: Analysis of TCP performance over mobile ad hoc networks. In: Proc. ACM/IEEE Int. Conf. on Mobile Computing, Seattle, WA, USA, pp. 219–230 (September 1999)

[4] Shang, Y., Hadjitheodosiou, M.: TCP splitting protocol for broadband l satellite network. In: Proc. 23rd IEEE Digital Avionics Syst. Conf., Salt Lake City, UT. 2, pp. 11.C.3-1–11.C.3-9 (October 2005)

[5] Zhu, J., Roy, S., Kim, J.H.: Performance modeling of TCP al-satellite hybrid networks. IEEE/ACM Trans. Netw. 14(4), 753–766 (2006)

[6] Zeng, W.G., Trajkovic, L.J.: TCP packet control for wireless networks. In: Proc. IEEE Int. Conf. on Wireless and Mobile Computing, Networking and Communications (WiMob 2005), Montreal, Canada, vol. 2, pp. 196–203 (August 2005)

[7] Omueti, M., Trajkovic, L.J.: M-TCP+: using disconnection feedback to improve performance of TCP in wired/wireless networks. In: Proc. SPECTS, San Diego, CA, USA, vol. 2, pp. 443–450 (July 2007)

[8] Casetti, C., Gerla, M., Mascolo, S., Sanadidi, M.Y., Wang, R.: TCP Westwood: end-to-end congestion control for wired/wireless networks. Wireless Netw. 8(5), 467–479 (2002)

[9] Lee, H., Lee, S., Choi, Y.: The influence of the large bandwidth-delay product on TCP Reno, New Reno and SACK. In: Proc. Information Networking Conference, Oita, Japan, pp. 327–334 (2007)

Personalised High Quality Search with in a Web Site: No User Profiling

L.K. Joshila Grace[1], V. Maheswari[2], and Dhinaharan Nagamalai[3]

[1] Research Scholar, Department of Computer Science and Engineering
[2] Professor and Head, Department of Computer Applications,
[1,2] Sathyabama University, Jeppiaar Nagar, Chennai, India
[3] Network Architect and Security Expert, Wireilla Net Solutions PTY Ltd, Australia
joshilagracejebin@gmail.com

Abstract. The main objective of this work is to make a web site developer to understand the efficiency of the web site. the custom now among the web site users is to select the web site which is satisfying the needs of the user. the needs of each user will be different. the user may perform a professional oriented search or a non professional oriented search. whichever may be the criteria the user must be satisfied when a web site is opened for his search. considering a single web site either an entertaining web site or a professional oriented web site. the web site must be efficient in its own way. thus analysing the various parameters the web site can be improvised in its area of speciality. the world wide web contains large amount of information. gaining knowledge over the information and also over the web user is the only way to extract effective information while mining.

Keywords: Search Key, Utility, Web Warehouse.

1 Introduction

Considering a single web site all the visitors' details are being extracted from the log file. These log files are resided in the server side. They are extracted as raw data and then converted into a user understandable data. A web site that has to track about the detail a tracking code must be inserted in the web site coding. This enables to listen each and every activity the user is performing. A user who ever enters the web site is tracked about all the activities they perform. The various user performed activities are moving the mouse, clicking the mouse, moving to a different link in the same web site, download various contents or exit out of the web site. The main reason why a particular user exits out of the web site is for a similar search key may have different views for different user. All their views may not be satisfied during the first time of search. For this the mind of the user has to be read. For this purpose the system has to be trained in such a way that according to the first click done by the user the system must be able to know the main target of the user for doing such a word search.

Let us consider an example of a word search. Let the word to be searched given by the user is "tree". Some may look for the cultivation of the tree, some may want to know the various types of trees present in a particular area. A professional may need

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 307–316.
springerlink.com

information regarding the various trees structure to organise the data and others may need the tree to just to have a look at its importance in growing. Therefore we could conclude that a single word will not be helpful for learning the mind of the user.

The proposed algorithm NUP (No user Profile) gives the knowledge to know the necessity of the user during the search. This requires just the search key from the user to proceed with knowledge extraction, ranking and ordering.

The second section is about knowledge management, third section is about the existing related work, fourth to tenth sections is about the proposed method, eleventh section gives conclusion and twelfth section provides the enhancement needed in the proposed work

2 Knowledge Management

In most of the web site we can find the search segment in order to search for a suitable link for the user in the web site. Once a search key word is being typed and searched by the user immediately the links corresponding to the search key is being displayed. The user those who currently start surfing through the web site is said to be handling a session S1. The knowledge that has to be gained in this session is done in two segments while mining. The knowledge from the data that is present in the data ware house or the web warehouse has to be extracted. This knowledge is helpful for knowing about the contents of the information present on the whole. This information is obtained when the data or the link is created and are added to the warehouse. The knowledge gaining process is done even before the mining is done. This knowledge will not change unless the data are modified by the owner of the data. The entire knowledge that is being acquired in one session is helpful for further processing of data.

The next stage of knowledge is extracted from the user while they are performing the mining process. This is dynamically performed. This information dynamically changes.

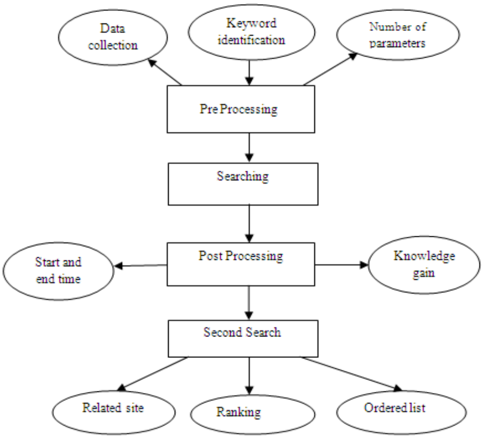


Fig. 1. Web mining subtasks

The Fig. 2 gives the clear outline of the procedure done in the proposed system. The stepwise procedure is discussed in the forth coming sections.

3 Related Work

Web Usage Mining (WUM) is the process of extracting knowledge from Web user’s access data by exploiting Data Mining technologies[1].

A partitioning method was one of the earliest clustering methods to be used in Web usage mining[1]. Incremental algorithm produces high quality clusters[2].

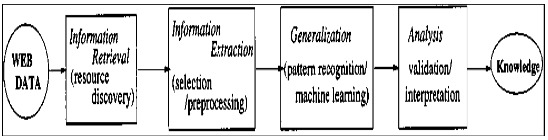

WEB DATA
Information Retrieval (resource discovery)
Information Extraction (selection /preprocessing)
Generalization (pattern recognition/ machine learning)
Analysis validation/ interpretation
Knowledge

Fig. 2. Web mining subtasks

The overall web usage mining process can be divided into Four interdependent stages as shown in Fig 2 are Data collection, Pre Processing, Pattern Discovery and Pattern Analysis.

Almost all web sites have searching function, and the search engines under use are confronting the following troubles [6]

3.1 Over Abundance

If the there are duplicate data present in the web site. Duplicates are said to be replication of data present in the web site. There may be more links corresponding to a single search key.

3.2 Limited Coverage

All the web site cannot satisfy all the customer needs. There can be limitation in the amount of data present in the web site.

3.3 Limited Query

Most search engines access on simple keyword-based searching, retrieve or order pages based on popularity of pages.

3.4 Limited Customization

Query results are determined only by query itself, often dependent on the background and knowledge of the user. The focus of this paper is to provide an overall view of how to use frequent pattern mining techniques for discovering different types of patterns in a web log database.

4 Ordering of Data

Just by the single word the intention of the user cannot be read. In this each set of links has the topic of the link as well as the hint about what the particular link is providing. The frequently used keywords would form the search keys of a particular data. This eliminates the commonly used words like the, is, was, about etc., So whenever the search matches the search keys immediately the listing of the link are done in the normal order. This would not give the preference to the intention of the user. Since there is no information about the user is provided. When all the links with a short line of description about the link are displayed, 95% of the user's would read the description before they follow through the link. By analysing the first click done on the link the system would learn about the information the user is looking for. According to the example the links regarding the cultivation, the information about the types, the data about the structuring of information or data are all formed in to separate groups.

The next time user searches for the same search key he/she would be able to get the most related information in the first set of options. This would increase the effectiveness of the mining process.

5 Essential Fields

There are two tables used in this entire process. The first table would contain the information about the web site. The second table would contain the information of the user.

The information regarding the web site that are present in the table are

5.1 Title of the Web Page

This is the main heading of the web page. This would be provided by the data provider.[3] Successful knowledge discovery requires a sufficient document collection.

5.2 Web Address of All the Links

The data should be present in a particular address. This address is provided so that whenever a selection of the particular option is made the particular web address is linked.

5.3 Hint about Each Link

This gives a small description about the particular link. While a search is made the list of option along with the hint is listed. This hint should not exceed 100 characters.

5.4 List of Key Word

Each link would have a list of key word by which they can be identified. The list of key word which would help to identify the particular link whenever searched. The total number of parameters provided in the site is entered.

The second table will have the details of the web site user.

5.5 Web Site User

Each web site user does not need to waste time in registering. Therefore they are identified by unique session ID. Each user enters in to the web site a new session is started and they are identified through out only by the session id.

5.6 Start Time

This gains the information about what time the user has traversed through the link. The format of the time is retrieved as “hh;mm;ss”.

5.7 End Time

This gains the information about what time the user has moved out from the link. The format of the time is same as that of the start time. The format of the time is retrieved as “hh;mm;ss”.

5.8 Traversed Path

The path traversed by the user is tracked. This field gets an entry only if the user moves through the path provided by the web page. The time duration spent on those paths are also identified.

5.9 Scrolls and Clicks

As the user traverse through the links the click and the scrolls that is being made in each link is also identified. The increment process is done for each scroll and clicks done throughout the surfing.

This table is being generated for each user separately. This table is in existence only till the user under a session is surfing through the same web site. Once the user exits immediately the table is destroyed. If two users enters the web site at a time two table under a different session ID is generated.

If the links provided by the page are used it is assumed as the desired web page. It is assumed as the page is being read by the user before moved to the next link in the page.

By list of parameters provided it can find the efficiency of the web page. This specifies the details it can provide about the topic. The quality of the web page is determined by the quantity of information it provide.

The time is mainly used in order to know the time spent by the user. Since the user may go through the description and have made an attempt to traverse through the link. But the page would not have the desired information. Then user may attempt to move back in a few seconds or minutes. Therefore a threshold is being set. Whenever the time period exceeds the threshold value then it is decide by the system as the desired site by the user. Therefore in the next time of search done for the same search key in the next instant would give a different set of options. These options would be ordered in such a way that the most related sites are present in first few options.

By just analysing the scrolls and clicks the utilisation of the page can be identified. The user may have opened the web page but would not have done any activities. This is also considered as the web page is not utilised.

6 Temporary Buffering

In the case of user profiling based web search the user profiles would have a separate table. Each time the user enters in and the activities done by the user is analysed until they exits out. This information is saved in a separate memory space for further analysis. These information helps the system to know what the user is actually interested in. Therefore every time the user makes a log in, the options are provided according to the knowledge gained while traversed in the site previously in the same user session.

The main disadvantage regarding the user profile are

- The profiling would occupy large memory space.
- The user has to remember the username and password each time they log in.

[5]User profiles can be categorized into three groups: interviewing, semi-interviewing and no interviewing. Current web information gathering systems attempt to satisfy user requirements by capturing their information needs. For this purpose, user profiles are created for user background knowledge description. Generate a topic profile for the user based on the discovered categories contained in the sub-trees.

To avoid all these drawbacks in the proposed method there is no separate memory wasted for storing the user information. There is just a temporary storage done about the user who is doing the latest search operation. Once the user does the first search and spends time greater than the threshold value immediately the entry is made regarding the search key and the area the user is interested in. This information is helpful while the search is done for the same search key at the next instant.

Thus just a temporary storage is done. When the time elapses for the threshold set for this storage. Then immediately the information of the previous search becomes unavailable to the system. Only the access time and traversed path is being identified. Which are deleted after the session of search is over.

7 Setting the Threshold

As it is discussed above if the user opening a web site is not doing any activity in the web page that has to be accounted. To identify wether the user is interacting with the web site efficiently the number of clicks and scrolls are being noted in the table. There is a necessity to set threshold values for various reasons.

- To identify a user is interacting with the web site
- To identify whether the customer needs are satisfied.

There are two threshold values generated. For each 3 minutes the scrolls and click column is verified. If there is no increment of values in the scroll and click column for more than 3 times then the session is expired. Then the user is continuing surfing in the same web site then a new session is generated.

The next threshold value would be generated to compare with the time duration that the user had spent after opening the link.

$ET-ST > TV1$

Where ET => End time

ST => Start time

TV1 => Threshold value

The difference between the start time and the end time would provide the time spent by the user on the particular link. If the time spent is lesser than the threshold it is considered as the site not interested by the user. At this point of time there would not occur any reordering of the options instead the same set of options are shown with the title and the description. The total knowledge is retrieved only by the activities of the user.

Once the condition is satisfied then if the user moves back to look for more options then at that instant the options are re ordered and the sites falling under the group which the user is interested is displayed in the first few options. Even though the user was not provided with a wider range of options to his area of interest the second time of search at that instant would provide with large related site information.

Since there is just a temporary storage done here, there is a threshold set even for this storage to become unavailable. This threshold is compared with the end time [ET] and next search done for the same search key.

$BST - ET < TV2$

` Where BST => Back Search Time

ET => End time

TV2 => Threshold value 2

The difference between the end time and the back search time would give the time gap of the next search done for the same search key . If the threshold is larger than the time gap then the machine analyses that the next search for the same search key is done immediately. If the threshold value is smaller than the time gap then the next immediate search is not done by the same user. This is the analysis done by the system.

By analysing these threshold values the knowledge is being generated according to the user and the ranking of the related links are done.

8 Ranking and Reordering

When it comes to the point of reordering, they are done only based on the rank that is being generated. The ranking is done only after identifying the user desired web page. The desired web page is identified by the previous equations that are discussed. Once that is identified the entire related webpage is considered. The total number of parameters provided by the developer are given in the extracted parameter field. The page which has more number of parameters takes up the first rank

Once this ranking is performed they display the result from the first rank to the last rank during the next search operation done by the user. This helps to identify the site that provide wide range of details about a particular topic.

9 Training the System

Here the system is trained in such a way that it should analyse two factors.

9.1 Users Utility of the Web Page

How much time user spends time on the particular link is analysed. This helps to find the most interesting web site. During this analysis the system would be able to find which is the area of interest of the user. With out getting the user profile information the area of interest is found. Once this area of interest for the particular search key is found and the list of search option is re ordered. The group of site falling under the area of interest of the user is given the first priority and the listing is re generated.

Now only the time duration the page is been viewed but also the how effectively it has been utilised is analysed. This is done by the clicks and scrolls column.

9.2 Users Next Idea of Searching

This analysis is done in order to know whether the same user wants the listing for the search key. To analyse this point the time gap is considered that is between the first search completion and the next search. According to this information the system may conclude whether the user is the same user. Do they need still more options for the same search key. The time gap should be too small so that no other user is getting the search result. This information helps in making the temporary buffer to be unavailable to the system. Minimal wastage of memory space.

These two factors help the system to gain knowledge over the entire process done through out mining.

[4]The two techniques of knowledge discovery can be applied for fully meeting the analyst needs: association rules and sequential patterns. Association rules mining provides the end user with correlations among references to various pages and sequential patterns can be used to determine temporal relationships among pages. Furthermore, mining sequences with time constraints allows a more flexible handling of the visitor transactions.

A number of clustering approaches have been proposed, all of which use web server logs to generate a model of user actions that is then grouped with a clustering algorithm.

10 Steps Involved

The various steps done in these type of mining are

10.1. Pre Processing

These two activities are done before the search process

- The data that are present are sorted to the respective key words
- Then they are made into groups according to the area they fall in spite of similar search key.
- Total number of parameters provided by each link is identified

10.2 Searching

This is done by the user once the user types the search key

- They found match with the data search key words
- All the matched data are listed with their respective link address, title and the hint.

10.3 Post Processing

After the search result is displayed the following activities are carried out

- The start time and the end time are noted.
- They compared with the threshold and the corresponding activity is carried out as given in the previous sections.

10.4 Second Time Search for Same Search Key

When the second time of search done for the same key word the time threshold is analysed

- Only the related sites are extracted
- Ranking based on the number of parameters is done.

Ordering the link according to the rank and displayed.

11 Conclusion

The proposed method is much effective than that of the existing machine learning and the user profiling method. An efficient personalised search is made in the system is done by this method. The user's area of interest is given the highest priority. Which inturn provides a efficient set of options for the web search made by the user.

12 Future Enhancement

The proposed method is implemented with a limited amount of data and information. This can be implemented in more than one web site and tested. With increase in the set of information in the web this type of personalisation is very much essential.

References

[1] Dixit, D., Gadge, J.: A New Approach for Clustering of Navigation Patterns of Online Users. International Journal of Engineering Science and Technology 2(6), 1670–1676 (2010)

[2] Maheswara Rao, V.V.R., Valli Kumari, V., Raju, K.V.S.V.N.: An Advanced Optimal Web Intelligent Model for Mining Web User Usage Behavior Using Genetic Algorithm. In: Proc. of Int. Conf. on Advances in Computer Science (2010)

[3] Pierre, J.M.: Mining Knowledge From Text Collections Using Automatically Generated Metadata. Interwoven, Inc. (2002)

[4] Masseglia, F., Poncelet, P., Teisseire, M.: Using Data Mining Techniques on Web Access Logs to Dynamically Improve Hypertext Structure. ACM Sigweb Newsletter 8(3) (October)

[5] Tao, X., Li, Y., Zhong, N.: A Personalized Ontology Model for Web Information Gathering. IEEE Transactions on Knowledge and Data Engineering 23(4) (April 2011)

[6] Tao, X., Li, Y., Zhong, N.: Constraint Based Frequent Pattern Mining for Generalized Query Templates From Web Log. International Journal of Engineering, Science and Technology 2(11), 17–33 (2010)

Usability Evaluation Using Specialized Heuristics with Qualitative Indicators for Intrusion Detection System

Tulsidas Patil[1], Ganesh Bhutkar[2], and Noshir Tarapore[2]

[1] Research Scholar, [2] Assistant Professor
Department of Computer Engineering,
Vishwakarma Institute of Technology,
Pune - 411 037, India

Abstract. Network security promises protection of valuable and accessible network resources from viruses, trojans, keyloggers, hijackers and unauthorized access. One of the important subsets of the network security tools is Intrusion Detection System (IDS). It is found that current IDS systems are not easy to use. As a result user has difficulties in judging the quality of the output, i.e. getting efficient alarm and severity level for detected intrusions, information about the detected intrusions as per layers, ports and visitors. Also the problems in installing and configuring the system go unnoticed. Therefore, the usability evaluation is extremely vital to help users in efficient interaction and enhance usage of IDS system. In this paper a specialized set of heuristics combined with objectively defined usability indicators are proposed for the usability evaluation of IDS systems. This study presents the evaluation of specialized set of heuristics based on usability problems and design deficiencies commonly prevalent among IDS systems. This set of specialized heuristics can be used to evaluate various IDS systems or other network security tools with diversified platforms and features.

Keywords: Usability, Intrusion Detection System, Human Computer Interaction, Usability Heuristics, Usability Indicators, Network Security.

1 Introduction

From the evolution of internet, people have been facing challenges of the network security. To face security challenges, network users utilize various tools such as firewall, antivirus software, ethereal, nmap, nessus, and Intrusion Detection System (IDS) [12]. But according to report of US-CERT the rate of incident in year 2010 is nearly six times as compare to year 2005. So, the focus of work is moving towards the usability in security. Among these tools, IDS plays vital role in addressing the issues of network security as it is designed to provide a timely identification of malicious activities and to support effective response to the real-time attacks. But user often fails to get all these functional advantages from the IDS as usability of IDS comes into picture. Users complain about operations and maintenance of IDS [13]. There are two main problems regarding the state of art and state of practice in IDS. The first

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 317–328.
springerlink.com

problem is about the underlying technique that is used in detecting attacks and the second problem is about user interaction to know and quickly respond to the detected attacks [7, 14].

This paper is a one step towards the improvement of usability in the network security tools. In this paper, we used well known heuristic evaluation method for the usability evaluation of IDS. The concept of qualitative indicator is also very much helpful to rate the importance of heuristics.

1.1 Usability Heuristics

Usability heuristics are used as guidelines to evaluate the usability of system under consideration. These heuristics are derived from the user as well as expert's experiences and observations of the IDS systems. With these heuristics user can evaluate the usability of various IDS systems. According to the importance and its impact on the usability of IDS, the formation of heuristics was done from the observations. Usability evaluation is done by various methods such as cognitive walkthrough, formal usability inspection, heuristic evaluation or pluralistic walkthrough [5]. For usability evaluation of IDS, a heuristic evaluation method was selected where heuristics are specially designed for the IDS. The usability problems and design deficiencies commonly prevalent among all three IDS systems were identified and list down as observations. A comprehensive list of relevant problems and observations, which contribute to usability of IDS systems, was prepared. With IDS and usability expert the only observations which are applicable for the usability of IDS are considered as usability heuristics which are further divided into six groups. These six groups are elaborated in next section - Introduction to heuristics. The qualitative usability indicators [4] were identified to measure the compliance. Instead of applying the 1-5 Likert scale uniformly across all parameters, we have chosen an indicator based evaluation method. Some heuristic indicators are checked in term of their absence or presence and some are elaborated in terms of their qualitative attributes.

1.2 Intrusion Detection System

The notion of IDS came from the seminal paper by James Anderson in 1980 [7]. Intrusion detection is the process of monitoring the events occurring in a computer system or network and analyzing them for signs of intrusions, defined as attempts to compromise the confidentiality, integrity, availability to bypass the security mechanisms of a computer or network. Intrusions can be caused by attackers accessing the systems from the internet, authorized users of the systems who attempt to gain additional privileges for which they are not authorized and authorized users who misuse the privileges given them. IDS is a software that automate this monitoring and analysis process[1].

1.3 Users of IDS Systems

Traditionally IDS is used only by the administrators but because of its advantage over other network security tools IDS users are continuously increasing from network administrator to the daily computer users who wants to have a full picture of the traffic going through one's PC or LAN segment. These users can be mainly classified as LAN administrators, security professionals and network programmers.

- LAN Administrators – The LAN administrator provides support and management of the local area network within a company. This management involves a number of functions that have to do with regular maintenance of the network, overseeing enhancements and upgrades to the local area network. As part of the regular maintenance of the network, the LAN administrator will monitor the daily activity on the network, ensuring that the resources of the company are utilized in ways that are within the standards set for employee usage.
- Security Professionals – Security professionals have broad understanding of a growing list of topics and technologies, including identity management, strong authentication, biometrics, anti-virus, intrusion detection, anti-spyware, firewalling and encryption.
- Network Programmers – Network programmers are the persons which are deal with the design of network. Based upon the traffic on the network and occurrence of intrusions, network programmer designs the networks and program according to traffic.

2 Related Work

It is observed that Even with strong (financial) incentives, users tend to ignore security indicators, such as absence or invalidity of SSL certificates [17]. Andrew Zhou has studied a IDS system and proposed a set of 6 heuristics for the usability improvement of IDS [1]. In the set of six heuristics 4 heuristics from Nielson's heuristics and 2 additional heuristics – 'Display of information'

and 'Information navigation'; exclusively for IDS. These set of six heuristics are useful but not sufficient as it consider only IDS systems. In another study, a heuristic evaluation of 3 touch screen ventilator systems has been performed using qualitative indicators [4]. We have used the concept of a comprehensive set of usability heuristics and well-defined usability indicators in our study with IDS. The next study has highlighted various challenges while using IDS such as considerations for deployment, configuration of security settings, availability of information about log storage in IDS [6, 13] and requirement of additional software for better operations. These challenges have propelled us to arrive at some vital usability heuristics in our study. Masone and Smith [3] discusses the problems related to the secure email applications based on digital certificates, certification authorities, and public key infrastructures. Similarly, Peter Mell has discussed issues in testing of IDS [12]. These issues have guided us in designing heuristics such as type of IDS output provided, sorting of detected events and provision of attack signatures for IDS in our study.

We have come across several usability evaluation studies which are carried out using Nielsen's heuristics. Ficarra has performed a heuristic evaluation of multimedia products [3] and another study was focused on heuristic evaluation of paper-based web pages [9]. These studies have helped us to get better insights into a method of heuristic evaluation.

3 Research Methodology

It involves following activities during research:

3.1 Selection and Study of IDS Systems

To design heuristics for IDS systems, we had chosen five IDS systems out of which three IDS systems namely Snort, KFSensor and Easyspy are used for the observation and study purpose and two IDS systems namely Sax-2 and X-ray along with Snort are used for evaluation.

For the study of IDS systems, we installed IDS systems on three different operating systems likewise Snort on Linux, KFSensor on Microsoft Windows 7 and Easyspy on Microsoft Windows XP. These three IDS systems are widely used and readily available. Operating System (OS) platform supported by IDS is an important aspect in selection of IDS. During the study of IDS systems we have noted down different aspects of IDS systems such as usability issues during installation, configuration, operation and maintenance of IDS systems. All major and differentiating observations providing useful insights into IDS usability were noted down.

3.2 Use of Card Sorting Method

To classify heuristics we use the card sorting method which is a useful method to determine and understand about how items can be classified [11]. In our research, we have used a set of index cards for classification of heuristics. Each card has an entry of heuristic written on it along with a related identification number. These cards are then provided to IDS experts for classification into suitable categories. The output of this process has provided 6 well-defined categories such as Installation, Output, Event, Customization, Help and Miscellaneous Heuristics. These classified heuristics are discussed in next section.

4 Introduction to Heuristics

We proposed a set of 35 usability heuristics for usability evaluation of IDS. To design heuristics we considered not only the suggestion from IDS experts but also the experience of naïve user. So the final set of specialized heuristics contains the heuristics useful for both naïve user as well as the IDS experts. By considering the inputs from IDS experts, experience from naïve user and suggestion by usability experts the final draft of specialized set of 35 heuristics for the usability evaluation of IDS systems prepared. These heuristics are discussed ahead in details.

4.1 Installation Heuristics

The heuristics under this group are useful to select the IDS system for the network. This group is divided into two categories like pre-installation and installation. The heuristics under pre-installation group is helpful when user selects the IDS system among the others. Before installation of IDS systems user must have sufficient information about IDS systems like whether the chosen system is applicable for OS platform, the type of installation. The other category of heuristics is suitable during installation process. The heuristics under this category is required at the time of installation of IDS systems.

Table 1. Installation heuristics

Pre-Installation Heuristics		
H1	Operating system platform supporting the IDS	Windows (4) Linux (3) FreeBSD (2) Other (1)
H2	Type of IDS available	Freeware (4) Open source(2) Proprietary (0)
H3	Type of installation supported in IDS	Offline (2) Online (1)
Installation Heuristics		
H4	Provide Graphical User Interface (GUI) for installation of IDS	Provided (3) Not provided (0)
H5	Require installation of additional software for IDS	Not required (3) Available at vendor's website (2) Online (1) Not available (0)
H6	Allow user to customize installation of IDS	Allowed (2) Not allowed (0)
H7	Provide facility to select database during installation	Provided (1) Not provided (0)

4.2 Output Heuristics

This group of usability heuristics contains the heuristics which are related to the ou tput provided by IDS. One of the most important part of IDS systems is the output. So usability of IDS systems is the most critical section. It is observed that users fail to understand the output of IDS systems as in output IDS systems provides unrelated information also it contains too many technical specifications which are not require to user. Use of table and charts in the output of IDS systems makes it more quickly and rapidly understandable.

It is always better if the output of IDS contains information about the number of attacks and detailed information for each attack such as its type, time and/or severity. A set of heuristics for output is depicted in table 2.

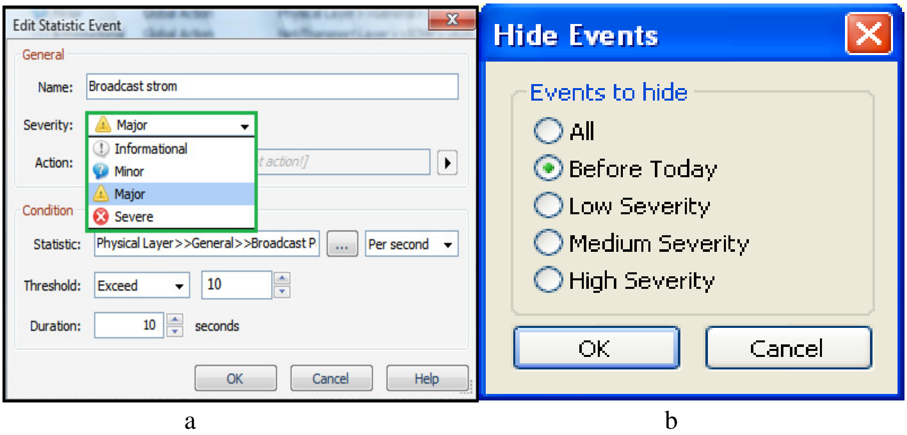
Table 2. Output heuristics

Output Heuristics		
H8	Understanding of the output provided by IDS	Easy to understand (4) Difficult to understand (2) Not understood (0)
H9	Customization of GUI for the output of IDS	Allowed (3) Not allowed (2) GUI not provided (0)
H10	The type of IDS output provided	As per ports (2) As per layers (2) As per visitors (2) Not provided (0)
H11	Provide printable output report	Provided in PDF format (2) Provided in Excel Format (2) Provided in other format (1) Not provided (0)
H12	The additional information available apart from number of attacks in IDS output	Type of intrusion (2) Severity of intrusion (2) Time of intrusion (2) Number of intrusion attempts (2) Not available (0)
H13	Provide tables and charts to represent output information	Charts provided (2) Tables provided (1) Not provided (0)

4.3 Event Heuristics

We have proposed 7 heuristics related with usability of events in IDS as shown in table 3. IDS system continuously detects events from the network. So with in less time IDS systems detects large number of events. To get the useful information from the events, these event heuristics are helpful. IDS systems should provide facility to categories events in ascending as well as in descending order. Also color codes as per the severity of intrusion makes the good difference among the detected events. For hiding the events, IDS should provide customization through display options like ‘before today’ and ‘as per the severity level. Assignment of severity level to each event is a vital feature in IDS can be observed in figure 1.

With the help of severity levels, user can not only differentiate events, but also assign appropriate actions to the events. These actions may include an execution of a suitable file, audio alert, email alert or flashing an icon. Users find it difficult to search event(s) based on a particular criteria and so, there is a need of advance search option.



Edit Statistic Event

General

Name: Broadcast strom

Severity: ⚠ Major

Condition

Statistic: Physical Layer>>General>>Broadcast P

Per second

Threshold: Exceed 10

Duration: 10 seconds

Hide Events

Events to hide

○ All

● Before Today

○ Low Severity

○ Medium Severity

○ High Severity

Fig. 1. Sample screen shots of events in Easyspy and KFSensor a) severity level b) events to hide

Table 3. Event heuristics

Event Heuristics		
H14	Amount of false-positive and false-negative events	Negligible (4) Marginal (3) Higher (0) Don't know (2)
H15	Type of intelligence technique used in IDS	Rule based (3) Anomaly based (3) Honey pot based (2) Other (1)
H16	Categories provided for customizing the event alerts in IDS	Execution of file (4) Flashing icon (3) Mobile messaging (3) Audio alert (2) Mobile messaging (2) E-mail alert (1) Not provided (0)
H17	Number of severity levels provided in IDS	3-5 levels provided (3) Levels out of range (2) Not provided (0)
H18	Availability of search option for events	Advanced search available (2) Search available (1) Not available (0)
H19	Provide color codes for the events in IDS	Appropriate (3-5 colors) (2) Confusing (1) Not provided (0)
H20	Sorting of detected events in IDS	Ascending order (1) Descending order (1) Not possible (0)
H21	Categories provided for loading/hiding events	Before today (1) Low severity (1) Medium severity (1) High severity (1) Not provided (0)

4.4 Customization Heuristics

Customization heuristics reduces the effort of re-installing the IDS systems. Also customization heuristics reduces the redundancy in information and makes the interaction of user with IDS systems more efficient. IDS systems detect events with the help of signatures. So to reduce false negative and false positive alarm rate support for customized signature is necessary.

A set of heuristics for customization is depicted in table 4.

Table 4. Customization heuristics

Customization Heuristics		
H22	Availability of wizard help option in IDS	Whenever necessary (3) At the end (2) Not available (0)
H23	Provide support for customized signatures	Provided (2) Not provided (0)
H24	Provide facility to set updation policy for IDS	Auto update (3) Manually update (2) Not provided (0)
H25	Provide facility for customize activation time of IDS.	Auto-started (1) Need to start explicitly (0)

4.5 Help Heuristics

The heuristics under this category is required when user came across core ny networking terminologies of IDS. Also when some errors / warnings and related messages appears on the screen. IDS systems have to provide online help for understanding terminologies and error-handling. IDS also stores logs for the detected events, so it should provide the path of the log file for its users. A set of help heuristics is depicted in table 5.

Table 5. Help heuristics

Help Heuristics		
H26	Provide help about the networking terminology in IDS	Help is provided in software(3) At vendor's website (2) Not provided (0)
H27	Provide information about the log file in IDS	Provided with path of log file (2) Provided without path of log file(1) Not provided (0)
H28	Provide appropriate error/warning message	Provided along with help (4) Provided without help (2) Not provided (0)
H29	Provide help for icons in IDS	As a tool tip (2) In user manual document (1) Not provided (0)
H30	Provide multilingual support	Provided (1) Not provided (0)

4.6 Miscellaneous Heuristics

With the help of card sorting method all 35 heuristics are categorized into 5 categories and the remaining heuristics are group into miscellaneous category. The heuristics under this category is important for the performance as well as usability of IDS. To keep the IDS always updated, provision for attack signatures is needed. Users find convenience in getting the information about attack signatures through notification than downloading it from the vendor's website. An active IDS should not degrade the system performance. Also the time required to provide new signatures for attack detection should be less than a day. A set of miscellaneous heuristics is depicted in table 6.

Table 6. Miscellaneous heuristics

Miscellaneous Heuristics		
H31	Provision for attack signatures for IDS	Through notification (4) At vendor's website (2) Not provided (0)
H32	Time required to provide signature for new vulnerability	Within 24 Hrs (4) Within a week (2) More than a week (0) Don't know (2)
H33	Provide previous and next options at every screen	Provided (2) Not provided (0)
H34	Provide information about scalability of IDS	Provided in help option (3) Provided in user manual (2) Provided at vendor's website (1) Not provided (0)
H35	Effect of IDS on system performance	Not affected (2) Marginally affected (1) Severely affected (0)

5 Evaluation

We have evaluated the usability of IDS systems using the heuristics and usability indicators with following objectives:

1. Measure the usability and overall efficacy of IDS systems in terms of usability index.
2. Study the reliability of the heuristics by involving two more IDS experts to carry out the evaluation of additional two IDS systems.

This heuristic evaluation has been carried out by two more Usability Evaluators (UE) with author. In the following table UE1 is referred to author where as UE2 and UE3 are for other two IDS expert involved in the heuristic evaluation of IDS systems. For the heuristic evaluation we used two more IDS systems with Snort IDS namely Sax-2 and X-ray.

The usability evaluators have adequate understanding of Human Computer Interaction (HCI). They were sensitized about the proposed heuristics, fundamentals about network security and the usability evaluation of IDS systems. Before the evaluation, Evaluators have hands free session on the all three IDS systems. All the evaluators got the IDS systems on internet after they installed it on the allocated computer system. Before they installed IDS systems, the list of all 35 heuristics are provided to them and ask to give score as per their observation and experience with the IDS systems. Their queries about the heuristics and related evaluation were discussed and then they carried out individually the heuristic evaluation of the IDS systems provided to them. The total scores of usability evaluations by all three usability evaluators are consolidated in table 7.

Table 7. Heuristic evaluation of IDS by three evaluators

Heuristic Categories	**Max. Score**	**Usability Evaluators**	**Scores for IDS Systems**		
			IDS-1	**IDS-2**	**IDS-3**
Pre-Installation	**19**	**UE1**	**16**	**05**	**06**
		UE2	16	06	06
		UE3	16	06	06
Installation	**12**	**UE1**	**04**	**05**	**04**
		UE2	01	05	04
		UE3	01	05	04
Output	**29**	**UE1**	**14**	**20**	**09**
		UE2	14	21	13
		UE3	14	20	13
Events	**33**	**UE1**	**08**	**20**	**07**
		UE2	07	20	07
		UE3	08	17	07
Customization	**11**	**UE1**	**04**	**08**	**05**
		UE2	04	08	05
		UE3	04	08	05
Help	**13**	**UE1**	**03**	**06**	**02**
		UE2	03	06	02
		UE3	03	06	02
Miscellaneous	**21**	**UE1**	**09**	**04**	**03**
		UE2	10	07	04
		UE3	10	08	04
Total	**138**	**UE1**	**58**	**68**	**36**
		UE2	55	73	41
		UE3	56	70	41

6 Validation

The figure 2 shows a graph for a comparison of usability evaluation of three IDS systems under evaluation by three usability evaluators (UE1, 2 and 3). The usability evaluation by other usability evaluators – UE1 differs from UE3 by -2.16 % for IDS-1, 2.17 for IDS-2 and -2.63 for IDS-3 IDS.

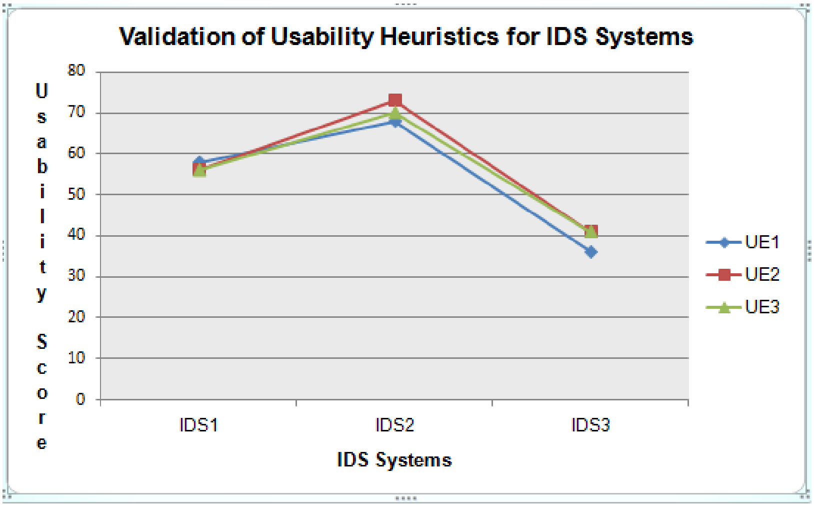


Fig. 2. Usability evaluation resuts of three IDS systems with specialized set of heuristics by three evaluators

Above figure depicts a graph representing closeness in usability evaluation of three IDS systems by three evaluators with the set of specialized heuristics for IDS. The above graph shows the heuristics are very appropriate and gives similar results in the evaluation process by different evaluators.

7 Conclusion

A study of IDS systems and the outcomes of their usability evaluation using specialized set of heuristics with indicators show that there are many usability issues as well as design deficiencies, which needs to be addressed. The specialized set of heuristics categorized into relevant groups and the objectively defined usability indicators ensure better understanding and efficiency among IDS systems to make them more user-friendly and humanized. This process helps in better understanding and usage of IDS systems by maximum possible users including novice users.

In future, this study can be extended to evaluation of other network security tools.

Acknowledgement. We express our gratitude to IDS experts – Axel Kloth, CTO and Vice President Engineering, Parimics Inc., USA and Hal Flynn, Vulnerability Analyst, Symantec Corporation, USA for their timely inputs and feedback during this research work.

References

1. Zhou, A., Blustein, J., Heywood, N.-Z.: Improving Intrusion Detection System through Heuristic Evaluation. In: IEEE CCECE 2004-CCGEI, Niagara Falls, pp. 1641–1644 (2004)
2. Abran, A., Khelifi, A., Suryn, W.: Usability Meaning and Interpretation in ISO Standards. Software Quality Journal, 325–338 (2003)
3. Masone, C., Smith, S.: Towards Usefully Secure Email. IEEE Technology and Society Magazine 26(1), 25–34 (2007)
4. Ficarra, F.: Evaluation of Multimedia Components. In: IEEE International Conference, Ottawa, Ont., Canada, pp. 557–564 (1997)
5. Bhutkar, D.K.G., Karmarkar, S.: Usability Heuristics and Qualitative Indicators for the Usability Evaluation of Touch Screen Ventilator Systems. In: Katre, D., Orngreen, R., Yammiyavar, P., Clemmensen, T. (eds.) HWID 2009. IFIP AICT, vol. 136, pp. 83–97. Springer, Heidelberg (2010)
6. Neilsen, J., Molich, R.: Heuristic Evaluation of User Interfaces. In: ACM SIGCHI Conference on Human Factors in Computing Systems: Empowering People, pp. 249–256 (1990) ISBN: 0-201-50932-6
7. McHugh, J., Christie, A., Allen, J.: The Role of Intrusion Detection Systems. IEEE Software, 42–51 (2000)
8. Anderson, J.: Computer Security Threat Monitoring and Surveillance, pp. 1–56 (February 1980)
9. Baker, K., Greenber, S., Gutwin, C.: Empirical Development of Heuristic Evaluation Methodology for Shared Workspace Groupware. In: ACM CSCW, New Orleans, Louisiana, USA, pp. 96–105 (2002)
10. Ahmed, M., Pal, R., Hossain, M., Bikas, A.N., Hasan, K.: A Comparative Study on Currently Existing Intrusion Detection Systems. In: IACSIT-SC, pp. 151–154. IEEE (2009)
11. Allen, M., Currie, L., Bakken, S., Patel, V.: Heuristic Evaluation of Paper-based Web Pages A Simplified Inspection Usability Methodology. Journal of Biomedical Informatics, 412–423 (2006)
12. Neumann: Audit Trail Analysis and Usage Collection and Processing. Technical Report Project 5910. SRI International (1985)
13. Nurmuliani, N., Zowghi, D., Williams, S.: Using Card Sorting Technique to Classify Requirements Change. In: IEEE International Requirement Engineering Conference, Kyoto, Japan, pp. 240–248 (2004)
14. Dhanjani, N., Clarke, J.: Network Security Tools. O'reily Media Inc. (2005)
15. Mell, P., Hu, V.: An Overview of Issus in Testing Intrusion Detection System. In: DARPA (2002)
16. Werlinger, R., Hawkey, K., Muldner, K., Jaferian, P.: The Challeges of Using an Intrusion Detection System: Is It Worth the Effort. In: SOUPS, Pittsburg, PA, USA (2008)
17. SANS Institute: Intrusion Detection System: Definition, Need and Challenges (2001)
18. Schecter, S.E., Dhamija, R., Ozment, A., Fischer, I.: The Emperor's New Security Indicators: An evaluation of website authentication and the effect of role playing on usability studies. In: IEEE Symp. on Security and Privacy, Oakland, CA, USA, May 20-23 (2007)

Analysis and Synchronization of the Hyperchaotic Yujun Systems via Sliding Mode Control

Sundarapandian Vaidyanathan

R & D Centre, Vel Tech Dr. RR & Dr. SR Technical University
Avadi-Alamathi Road, Avadi, Chennai-600 062, India
sundarvtu@gmail.com
http://www.vel-tech.org/

Abstract. In this paper, we deploy sliding mode control (SMC) to derive new results for the global chaos synchronization of identical hyperchaotic Yujun systems (2010). The synchronization results derived in this paper are established using the Lyapunov stability theory. Numerical simulations have been provided to illustrate the sliding mode control results derived in this paper for the complete synchronization of identical hyperchaotic Yujun systems.

Keywords: Sliding mode control, chaos synchronization, hyperchaos, hyperchaotic Yujun system.

1 Introduction

Chaotic systems are nonlinear dynamical systems that are characterized by sensitive dependence on initial conditions and by having evolution through phase space that appears to be quite random. This sensitive dependence on initial conditions is commonly called as the *butterfly effect* [1].

Chaos theory has been applied to a variety of fields including physical systems [2], chemical systems [3], ecological systems [4], secure communications ([5]-[7]) etc.

Since the pioneering work by Pecora and Carroll ([8], 1990), chaos synchronization problem has been studied extensively in the literature. In the last two decades, various control schemes have been developed and successfully applied for the chaos synchronization such as PC method [8], OGY method [9], active control method ([10]-[13]), adaptive control method ([14]-[17]), time-delay feedback method [18], backstepping design method ([19]-[20]), sampled-data feedback synchronization method ([21]-[22]) etc.

In this paper, we adopt the *master-slave* formalism of the chaos synchronization approaches. If we call a particular chaotic system as the *master* system and another chaotic system as the *slave* system, then the goal of the global chaos synchronization is to use the output of the master system to control the slave system so that the states of the slave system track asymptotically the states of the master system. In other words, global chaos synchronization is achieved when the difference of the states of master and slave systems converge to zero asymptotically with time.

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 329–337.
springerlink.com

In this paper, we derive new results based on the sliding mode control ([23]-[25]) for the global chaos synchronization of identical hyperchaotic Yujun systems ([26], Yujun *et al.* 2010).

The sliding mode control is often adopted in robust control theory due to its inherent advantages of easy system realization, fast response and good transient performance. The sliding mode control results are also insensitive to parameter uncertainties and external disturbances.

This paper has been organized as follows. In Section 2, we describe the problem statement and our methodology using sliding mode control. In Section 3, we describe an analysis of the hyperchaotic Yujun system (2010). In Section 4, we discuss the sliding mode controller design for the global chaos synchronization of identical hyperchaotic Yujun systems (2010). Section 5 contains the conclusions of this paper.

2 Problem Statement and Our Methodology Using Sliding Mode Control

2.1 Problem Statement

Consider the chaotic system described by

$$\dot{x} = Ax + f(x) \tag{1}$$

where $x \in \mathbb{R}^n$ is the state of the system, A is the $n \times n$ matrix of the system parameters and $f : \mathbb{R}^n \to \mathbb{R}^n$ is the nonlinear part of the system. We take the system (1) as the *master* system.

As the *slave* system, we consider the following chaotic system described by the dynamics

$$\dot{y} = Ay + f(y) + u \tag{2}$$

where $y \in \mathbb{R}^n$ is the state of the system and $u \in \mathbb{R}^m$ is the controller of the slave system.

If we define the *synchronization error* e as

$$e = y - x, \tag{3}$$

then the error dynamics is obtained as

$$\dot{e} = Ae + \eta(x, y) + u, \quad \text{where} \quad \eta(x, y) = f(y) - f(x) \tag{4}$$

The objective of the global chaos synchronization problem is to find a controller u such that

$$\lim_{t \to \infty} \|e(t)\| = 0 \quad \text{for all initial conditions} \quad e(0) \in \mathbb{R}^n \tag{5}$$

2.2 Our Methodology

First, we define the control u as

$$u(t) = -\eta(x, y) + Bv(t) \tag{6}$$

where B is a constant gain vector selected such that (A, B) is controllable.

Substituting (6) into (4), the error dynamics becomes

$$\dot{e} = Ae + Bv \tag{7}$$

which is a linear time-invariant control system with single input v.

Thus, we have shown that the original global chaos synchronization problem is equivalent to the problem of stabilizing the zero solution $e = 0$ of the linear system (7) by a suitable choice of the sliding mode control.

In the sliding mode control, we define the variable

$$s(e) = Ce = c_1 e_1 + c_2 e_2 + \cdots + c_n e_n \tag{8}$$

where $C = [\, c_1 \quad c_2 \quad \cdots \quad c_n \,]$ is a constant vector to be determined.

In the sliding mode control, we constrain the motion of the system (7) to the sliding manifold defined by

$$S = \{x \in \mathbb{R}^n \mid s(e) = 0\} = \{x \in \mathbb{R}^n \mid c_1 e_1 + c_2 e_2 + \cdots + c_n e_n = 0\}$$

which is required to be invariant under the flow of the error dynamics (7).

When in sliding manifold S, the system (7) satisfies the following conditions:

$$s(e) = 0 \tag{9}$$

which is the defining equation for the manifold S and

$$\dot{s}(e) = 0 \tag{10}$$

which is the necessary condition for the state trajectory $e(t)$ of the system (7) to stay on the sliding manifold S.

Using (7) and (8), the equation (10) can be rewritten as

$$\dot{s}(e) = C\,[Ae + Bv] = 0 \tag{11}$$

Solving (11), we obtain the equivalent control law given by

$$v_{\text{eq}}(t) = -(CB)^{-1} CAe(t) \tag{12}$$

where C is chosen such that $CB \neq 0$.

Substituting (12) into the error dynamics (7), we get the closed-loop dynamics as

$$\dot{e} = [I - B(CB)^{-1}C]Ae \tag{13}$$

where C is chosen such that the system matrix $[I - B(CB)^{-1}C]A$ is Hurwitz.

Then the system dynamics(13) is globally asymptotically stable.

To design the sliding mode controller for the linear time-invariant system (7), we use the constant plus proportional rate reaching law

$$\dot{s} = -q\text{sgn}(s) - ks \tag{14}$$

where sgn$(\cdot)$ denotes the sign function and the gains $q > 0, k > 0$ are determined such that the sliding condition is satisfied and sliding motion will occur.

From equations (11) and (14), we obtain the control $v(t)$ as

$$v(t) = -(CB)^{-1}[C(kI + A)e + q\text{sgn}(s)] \tag{15}$$

Theorem 1. *The master system (1) and the slave system (2) are globally and asymptotically synchronized for all initial conditions $x(0), y(0) \in \mathbb{R}^n$ by the feedback control law*

$$u(t) = -\eta(x, y) + Bv(t) \tag{16}$$

where $v(t)$ is defined by (15) and B is a column vector such that (A, B) is controllable. Also, the sliding mode gains k, q are positive.

Proof. First, we note that substituting (16) and (15) into the error dynamics (7), we obtain the closed-loop dynamics as

$$\dot{e} = Ae - B(CB)^{-1}[C(kI + A)e + q\text{sgn}(s)] \tag{17}$$

To prove that the error dynamics (17) is globally asymptotically stable, we consider the candidate Lyapunov function defined by the equation

$$V(e) = \frac{1}{2}s^2(e) \tag{18}$$

which is a positive definite function on $\mathbb{R}^n$.

Differentiating V along the trajectories of (17) or the equivalent dynamics (14), we obtain

$$\dot{V}(e) = s(e)\dot{s}(e) = -ks^2 - q\text{sgn}(s) \tag{19}$$

which is a negative definite function on $\mathbb{R}^n$.

Thus, by Lyapunov stability theory [27], it is immediate that the error dynamics (17) is globally asymptotically stable for all initial conditions $e(0) \in \mathbb{R}^n$.

This completes the proof. □

3 Analysis of the Hyperchaotic Yujun System

The 4-D Yujun dynamics is described by

$$\begin{aligned} \dot{x}_1 &= a(x_2 - x_1) + x_2x_3 \\ \dot{x}_2 &= cx_1 - x_2 - x_1x_3 + x_4 \\ \dot{x}_3 &= x_1x_2 - bx_3 \\ \dot{x}_4 &= -x_1x_3 + rx_4 \end{aligned} \tag{20}$$

where x_1, x_2, x_3, x_4 are the states and a, b, c, r are constant, positive parameters of the system.

It has been shown by Yujun *et al.* [26] that the system (20) exhibits hyperchaotic behaviour when the parameter values are taken as

$$a = 35, \quad b = \frac{8}{3}, \quad c = 55, \quad 0.41 < r \leq 3 \tag{21}$$

When $r = 15$, the system (20) has the Lyapunov exponents

$$\lambda_1 = 1.4944, \quad \lambda_2 = 0.5012, \quad \lambda_3 = 0, \quad \lambda_4 = -38.9264$$

Since the system (20) has two positive Lyapunov exponents *viz.* λ_1 and λ_2, it is hyperchaotic.

The phase portrait of the hyperchaotic Yujun system is depicted in Figure 1.

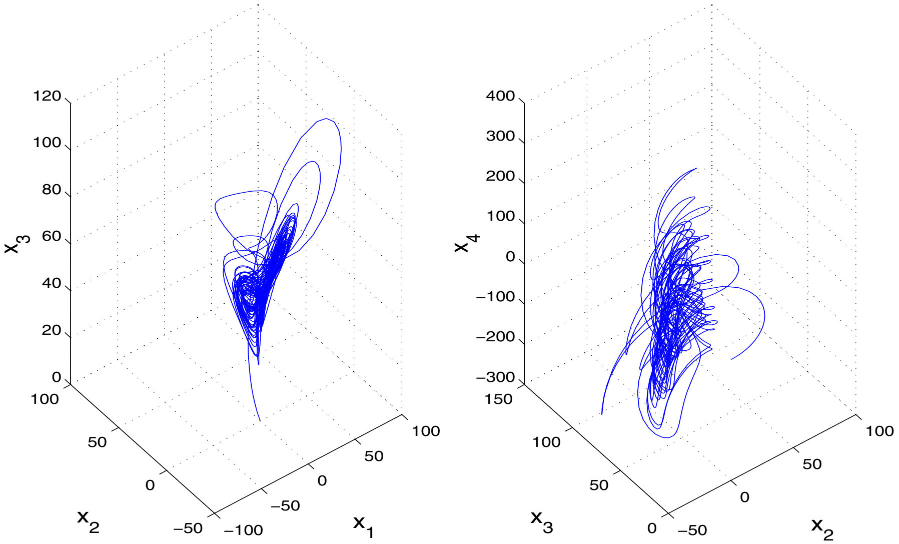


Fig. 1. State Portrait of the Hyperchaotic Lorenz System

4 Global Chaos Synchronization of the Identical Hyperchaotic Yujun Systems

4.1 Main Results

In this section, we apply the sliding mode control results derived in Section 2 for the global chaos synchronization of identical hyperchaotic Yujun systems ([26], 2010).

Thus, the master system is described by the hyperchaotic Yujun dynamics

$$\begin{aligned} \dot{x}_1 &= a(x_2 - x_1) + x_2 x_3 \\ \dot{x}_2 &= c x_1 - x_2 - x_1 x_3 + x_4 \\ \dot{x}_3 &= x_1 x_2 - b x_3 \\ \dot{x}_4 &= -x_1 x_3 + r x_4 \end{aligned} \tag{22}$$

where x_1, x_2, x_3, x_4 are the states of the system and a, b, c, r are the constant, positive parameters of the system.

The slave system is also described by the hyperchaotic Lorenz dynamics

$$\begin{aligned} \dot{y}_1 &= a(y_2 - y_1) + y_2 y_3 + u_1 \\ \dot{y}_2 &= c y_1 - y_2 - y_1 y_3 + y_4 + u_2 \\ \dot{y}_3 &= y_1 y_2 - b y_3 + u_3 \\ \dot{y}_4 &= -y_1 y_3 + r y_4 + u_4 \end{aligned} \tag{23}$$

where y_1, y_2, y_3, y_4 are the states of the system and u_1, u_2, u_3, u_4 are the controllers to be designed.

The chaos synchronization error e is defined by

$$e_i = y_i - x_i, \quad (i = 1, 2, 3, 4) \tag{24}$$

The error dynamics is easily obtained as

$$\begin{aligned} \dot{e}_1 &= a(e_2 - e_1) + y_2 y_3 - x_2 x_3 + u_1 \\ \dot{e}_2 &= c e_1 - e_2 + e_4 - y_1 y_3 + x_1 x_3 + u_2 \\ \dot{e}_3 &= -b e_3 + y_1 y_2 - x_1 x_2 + u_3 \\ \dot{e}_4 &= r e_4 - y_1 y_3 + x_1 x_3 + u_4 \end{aligned} \tag{25}$$

We can write the error dynamics (25) in the matrix notation as

$$\dot{e} = Ae + \eta(x, y) + u \tag{26}$$

where the associated matrices are

$$A = \begin{bmatrix} -a & a & 0 & 0 \\ c & -1 & 0 & 1 \\ 0 & 0 & -b & 0 \\ 0 & 0 & 0 & r \end{bmatrix}, \quad \eta(x, y) = \begin{bmatrix} y_2 y_3 - x_2 x_3 \\ -y_1 y_3 + x_1 x_3 \\ y_1 y_2 - x_1 x_2 \\ -y_1 y_3 + x_1 x_3 \end{bmatrix} \text{ and } u = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} \tag{27}$$

The sliding mode controller design is carried out as detailed in Section 2.

First, we set u as

$$u = -\eta(x, y) + Bv \tag{28}$$

where B is chosen such that (A, B) is controllable. We take B as

$$B = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \tag{29}$$

In the hyperchaotic case, the parameter values are

$$a = 35, \;\; b = 8/3, \;\;\; c = 55 \;\;\text{and}\;\;\; r = 1.5$$

The sliding mode variable is selected as

$$s = Ce = [\,-1 \quad -2 \qquad 0 \quad 1\,]\,e \tag{30}$$

which makes the sliding mode state equation asymptotically stable.

We choose the sliding mode gains as $k = 6$ and $q = 0.2$.

We remark that a large value of k can cause chattering and q must be chosen appropriately to speed up the time taken to reach the sliding manifold as well as to reduce the system chattering.

From equation (15), we can obtain $v(t)$ as

$$v(t) = -40.5e_1 - 22.5e_2 + 2.75e_4 + 0.1\,\text{sgn}(s) \tag{31}$$

Thus, the required sliding mode controller is obtained as

$$u(t) = -\eta(x, y) + Bv(t) \tag{32}$$

where $\eta(x, y)$, B and $v(t)$ are defined in equations (27), (29) and (31).

By Theorem 1, we obtain the following result.

Theorem 2. *The identical hyperchaotic Yujun systems (22) and (23) are globally and asymptotically synchronized for all initial conditions with the sliding mode controller* u *defined by (32).* □

4.2 Numerical Results

For the numerical simulations, the fourth-order Runge-Kutta method with time-step $h = 10^{-8}$ is used to solve the hyperchaotic Yujun systems (22) and (23) with the sliding mode controller u given by (32) using MATLAB.

For the hyperchaotic Lorenz systems, the parameter values are taken as

$$a = 35, \quad b = 8/3, \quad c = 55, \quad r = 1.5$$

The sliding mode gains are chosen as $k = 6$ and $q = 0.2$.

The initial values of the master system (22) are taken as

$$x_1(0) = 2, \;\; x_2(0) = 17, \;\; x_3(0) = 22, \;\; x_4(0) = 16$$

and the initial values of the slave system (23) are taken as

$$y_1(0) = 14, \;\; y_2(0) = 26, \;\; y_3(0) = 38, \;\; y_4(0) = 5$$

Figure 2 depicts the synchronization of the hyperchaotic Yujun systems (22) and (23).

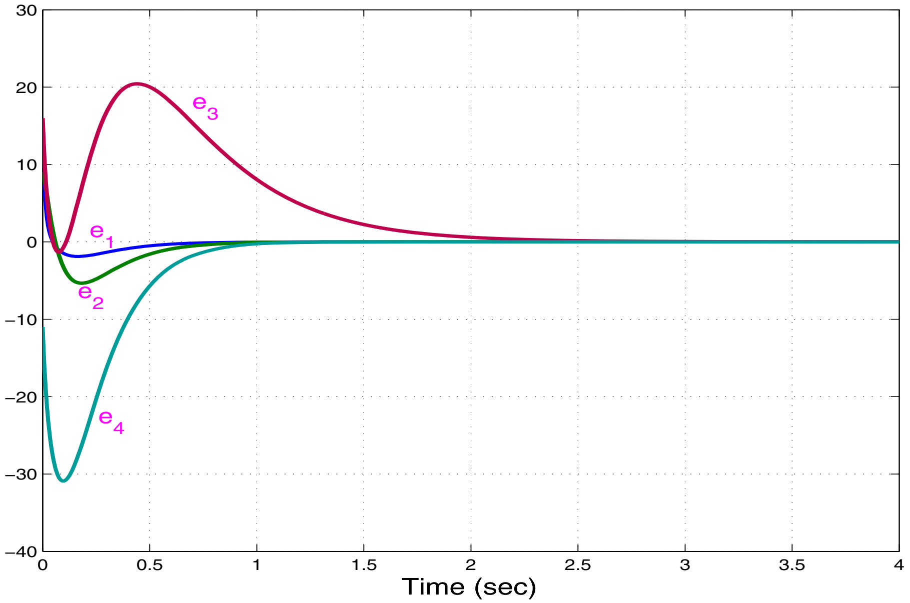


Fig. 2. Synchronization of the Identical Hyperchaotic Yujun Systems

5 Conclusions

In this paper, we have used sliding mode control (SMC) to achieve global chaos synchronization for the identical hyperchaotic Yujun systems (2010). Our synchronization results for the identical hyperchaotic Yujun systems have been established using the Lyapunov stability theory. Since the Lyapunov exponents are not required for these calculations, the sliding mode control method is very effective and convenient to achieve global chaos synchronization for identical hyperchaotic Yujun systems. Numerical simulations have been shown to demonstrate the effectiveness of the synchronization results derived in this paper using sliding mode control.

References

1. Alligood, K.T., Sauer, T., Yorke, J.A.: Chaos: An Introduction to Dynamical Systems. Springer, New York (1997)
2. Lakshmanan, M., Murali, K.: Chaos in Nonlinear Oscillators: Controlling and Synchronization. World Scientific, Singapore (1996)
3. Han, S.K., Kerrer, C., Kuramoto, Y.: Dephasing and burstling in coupled neural oscillators. Phys. Rev. Lett. 75, 3190–3193 (1995)
4. Blasius, B., Huppert, A., Stone, L.: Complex dynamics and phase synchronization in spatially extended ecological system. Nature 399, 354–359 (1999)
5. Kwok, H.S., Wallace, K., Tang, S., Man, K.F.: Online secure communication system using chaotic map. Internat. J. Bifurcat. Chaos 14, 285–292 (2004)

6. Kocarev, L., Parlitz, U.: General approach for chaos synchronization with applications to communications. Phys. Rev. Lett. 74, 5028–5030 (1995)
7. Murali, K., Lakshmanan, M.: Secure communication using a compound signal using sampled-data feedback. Applied Math. Mech. 11, 1309–1315 (2003)
8. Pecora, L.M., Carroll, T.L.: Synchronization in chaotic systems. Phys. Rev. Lett. 64, 821–824 (1990)
9. Ott, E., Grebogi, C., Yorke, J.A.: Controlling chaos. Phys. Rev. Lett. 64, 1196–1199 (1990)
10. Ho, M.C., Hung, Y.C.: Synchronization of two different chaotic systems using generalized active network. Phys. Lett. A 301, 421–428 (2002)
11. Huang, L., Feng, R., Wang, M.: Synchronization of chaotic systems via nonlinear control. Phys. Lett. A 320, 271–275 (2004)
12. Chen, H.K.: Global chaos synchronization of new chaotic systems via nonlinear control. Chaos, Solit. Frac. 23, 1245–1251 (2005)
13. Sundarapandian, V.: Global chaos synchronization of Shimizu-Morioka and Liu-Chen chaotic systems by active nonlinear control. Internat. J. Advances in Science and Technology 2(4), 11–20 (2011)
14. Chen, S.H., Lü, J.: Synchronization of an uncertain unified system via adaptive control. Chaos, Solit. Frac. 14, 643–647 (2002)
15. Lu, J., Han, X., Lü, J.: Adaptive feedback synchronization of a unified chaotic system. Phys. Lett. A 329, 327–333 (2004)
16. Samuel, B.: Adaptive synchronization between two different chaotic dynamical systems. Adaptive Commun. Nonlinear Sci. Num. Simul. 12, 976–985 (2007)
17. Sundarapandian, V.: Adaptive synchronization of hyperchaotic Lorenz and hyperchaotic Lü systems. Internat. J. Instrument. Control Sys. 1(1), 1–18 (2011)
18. Park, J.H., Kwon, O.M.: A novel criterion for delayed feedback control of time-delay chaotic systems. Chaos, Solit. Fract. 17, 709–716 (2003)
19. Wu, X., Lü, J.: Parameter identification and backstepping control of uncertain Lü system. Chaos, Solit. Fract. 18, 721–729 (2003)
20. Yu, Y.G., Zhang, S.C.: Adaptive backstepping synchronization of uncertain chaotic systems. Chaos, Solit. Fract. 27, 1369–1375 (2006)
21. Yang, T., Chua, L.O.: Control of chaos using sampled-data feedback control. Internat. J. Bifurcat. Chaos. 9, 215–219 (1999)
22. Zhao, J., Lu, J.: Using sampled-data feedback control and linear feedback synchronization in a new hyperchaotic system. Chaos, Solit. Fract. 35, 376–382 (2008)
23. Slotine, J.E., Sastry, S.S.: Tracking control of nonlinear systems using sliding surface with application to robotic manipulators. Internat. J. Control 38, 465–492 (1983)
24. Utkin, V.I.: Sliding mode control design principles and applications to electric drives. IEEE Trans. Industrial Electr. 40, 23–36 (1993)
25. Sundarapandian, V.: Global chaos synchronization of Pehlivan systems by sliding mode control. Internat. J. Comp. Sci. Eng. 3(5), 2163–2169 (2011)
26. Yujun, N., Xingyuan, W., Mingjun, W., Huaguang, Z.: A new hyperchaotic system and its circuit implementation. Commun. Nonlinear Sci. Numer. Simulat. 15, 3518–3524 (2010)
27. Hahn, W.: The Stability of Motion. Springer, New York (1967)

String Matching Technique Based on Hardware: A Comparative Analysis

Aakanksha Pandey and Nilay Khare

Department of Computer Science and Engineering, Maulana Azad National Institute of Technology, Bhopal, India
aakankshapandey7@gmail.com,
nilay.khare@yahoo.co.in

Abstract. Network Intrusion Detection Systems is one of the most effective way of providing security to those connected to the network, and the string matching algorithm is the heart of the intrusion detection system .IDS checks both packet header and payload in order to detect content-based security threats.Payload scan requires efficient string matching techniques, since each incoming packet must be compared against the hundreds of known attacks .Checking every byte of every packet to see if it matches one of a set of ten thousand strings becomes a computationally intensive task as network speeds grows up .For high speed networks it can be difficult to keep up with intrusion detection using purely software approach without affecting performance of the system intended for designed application. It is essential to use hardware systems for intrusion detection. A string matching algorithm is implemented in hardware with the focus on increasing throughput, and reasonable area cost while maintaining the configurability provided by the software IDSs .This paper consist a review of different string matching techniques implemented in FPGA for detecting malicious packet over the network.

Keywords: Field Programmable Gate Array, Network intrusion detection, String matching.

1 Introduction

Security is a big issue for all networks in today’s enterprise environment. Hackers and intruders have made many successful attempts to bring down high-profile Company, networks and web services. Many methods have been developed to secure the network infrastructure and communication over the Internet, among them the use of firewalls, encryption, IDS etc. Firewall is the way of blocking the outside traffic selectively and the requirement is that all the traffic should pass through the network firewall, encryption transform the data before sending into the network. Intrusion detection is one of the best techniques to identify malicious packet over the network. Intrusion detection is the act of detecting unwanted traf-fic on a network.

The main motivation of implementing the Intrusion Detection System into the hardware instead of software is the performance gap and their dissimilar execution paradigm. Since the speed of the network is increasing rapidly performance of

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 339–347.
springerlink.com

sequential execution in software limits the throughput but hardware can use other technique like pipelining etc for parallel processing .Under these condition there is a need of hardware implementation. Field Programmable Gate Array(FPGA) is suitable for the implementation because it provide the flexibility.

The main concern of this paper is to review different string matching tech-niques like Content Addressable Memory (CAM) ,Finite Automata(FA) and hash-ing for intrusion detection.

The remainder of this paper is organized as follows section 2 contains the review of different string matching algorithm, in section 3 a brief discussion about the results and comparison and section 4 is the conclusion of the paper.

2 Related Work

Different techniques have been implemented in hardware for string matching. The efficient technique will be one which is more efficient, fast and having limited implementation area.

2.1 CAM (Content Addressable Memory) Based Matching

CAM implementation uses discrete comparators for pattern matching so it has several advantages: (i) it is simple and regular, (ii) it allows for fine grain pipelining and high operating frequencies, and (iii) it is straightforward to use multiple comparators in order to process multiple input bytes per cycle. This is the mostly used pattern matching technique [7,8] in which distinct comparator are used for pattern matching. In this section we have presented the basic CAM then the Decoded CAM (DCAM) architecture given by Ioannis Sourdis et al[8].

2.1.1 Basic CAM

The input stream to be matched is inserted into the shift register and the individual entries are compared using the comparator. So for n length string, n comparator is used.

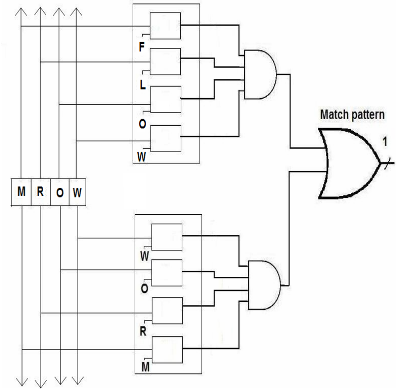


Fig. 1. CAM based matching [7]

The above CAM(fig 1) has been made for two pattern wolf and worm .The incoming data is compared against these patterns using comparator and OR operation is performed at the end for getting result. Its main drawback is the high area cost. To remedy this cost, they had suggested sharing the character comparators for strings with "similarities" and the architecture is called optimized CAM.

2.1.2 Optimized CAM

In this optimized technique sharing a comparator with similar characters(Fig 8). In this optimized CAM six comparator is sufficient. This approach achieves not only the sharing of the equality logic for character O and W, but also transforms the 8-bit wide shift register used in basic CAM into possibly multiple single bit shift register. Hence, the advantage is the potential for area savings.

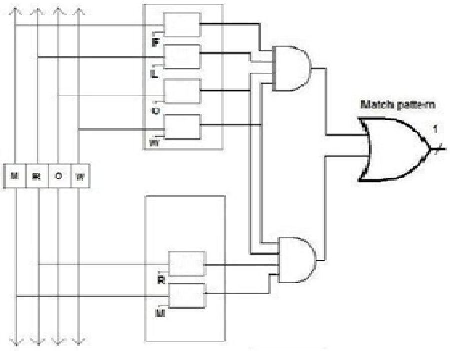


Fig. 2. Optimized CAM based matching [7]

2.1.3 Decoded CAM (DCAM)

The motivation of this architecture is to use a centralized comparator(Decoder) which reduce the Decoded CAM cost with high extent and in this architecture instead of using a n bit shift register using a single bit shift register.

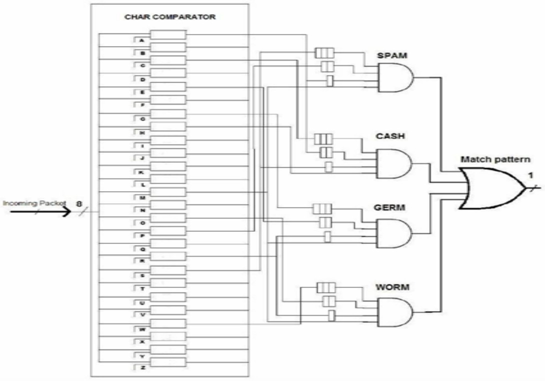


Fig. 3. Decoded Cam for pattern matching[7]

In the DCAM implementation we use partitioning to achieve better performance and area density. In terms of performance, a limiting factor to the scaling of an implementation to a large number of search patterns is the fan out and the length of the interconnections.

2.1.4 Pipelined DCAM

Parallelism in DCAM architecture (fig 4) is used to increase the processing throughput .Parallelism with factor p provide the p copies of the central compartor

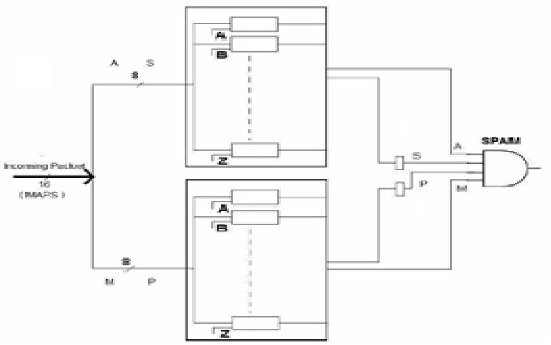


Fig. 4. Pipelined Decoded Cam for pattern matching[7]

The processing speed increases with the cost of area.In CAM based approach the data is stored into the memory so operation is performed in a fast way.

2.2 Finite Automata Based Pattern Matching

Below we informally review the regular expression ,NFA and DFA for string matching.Regular expression are generated for every string in the rule set and nondeterministic / deterministic finite automata is generated that examines the one byte input at a time.

2.2.1 Regular Expression

A regular expression is used for pattern matching that matches one or more string of characters. Regular expression are having different metacharacters like kleen closure *, positive closure +, concatenation | , alteration etc. Regular expressions are combined using these metacharacters. Suppose r1 and r2 are two regular expression then r1|r2 can be used for any string matched by r1 or r2 , r1* can be used for the string having one or more r1 etc.suppose the regular expression (wh(oisater)*) can be used to match who, what , whois ,where etc.

2.2.2 Regular Expression Matching Using Non-deterministic Finite Automata

Sidhu and Prasanna[12] have mapped NFA into FPGA. A NFA is constructed from the given regular expression .The NFA can have more than one active state at any point of time and all the active state needs to be processed for any given input symbol to be read.

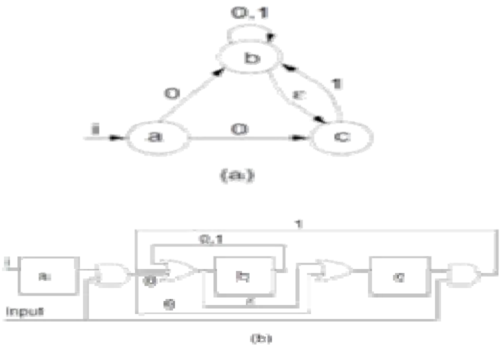


Fig. 5. (a)Simple NFA and(b) its logic structure[12]

The above fig 5 shows the simple NFA for the regular expression and its logic structure. Three flip flops are used for the three different states and the flip flop stores a 1 that signifies the current state or active state .The output is 1 only when the flip flop stores a 1 and the input character matches the character stored inside the comparator.

2.2.3 Regular Expression Matching Using Deterministic Finite Automata

Babu Karuppiah et al[1] have presented the string matching algorithm using DFA.
The DFA constructed from the NFA by using thompson or any other rule then it is used for the pattern matching .the essential property of DFA in that at any point of time it has only one active state so DFA is faster than NFA.fig 6 shows a simple DFA for the regular expression ((a|b)*(c|d)).

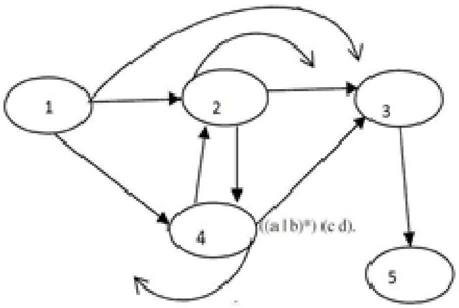


Fig. 6. DFA for regular expression ((a|b)*)(cd)[1]

DFA are usually efficient when the expression is not repetitive and becomes larger than NFA.when expression is repetitive .In general DFA may require up to 2n states compared to an equivalent NFA with only n states that leads to state explosion.

2.3 Pattern Matching Using Hash Function

The Hash function is a one to one direct mapping function. It hashes the incoming data and determines a unique memory location that contains the possibly matching pattern and a comparison is made between incoming data and its particular memory output to determine the match. But generating a perfect hash function for large incoming data bytes with no collision is difficult and time consuming process. So it can be made faster by finding sub hashes for substrings. So for this purpose Dhanapriya et al [7]have designed word split hash algorithm[WHSA]in which the string is split in to smaller substrings and sub hash is calculated for the first substring and checked for matching of known pattern. If not matched then that whole string is detected to be free from virus and need not find hash value for the rest of the bytes of a string. If the initial substring matches then there is some known pattern starting with those characters and the process is continued.

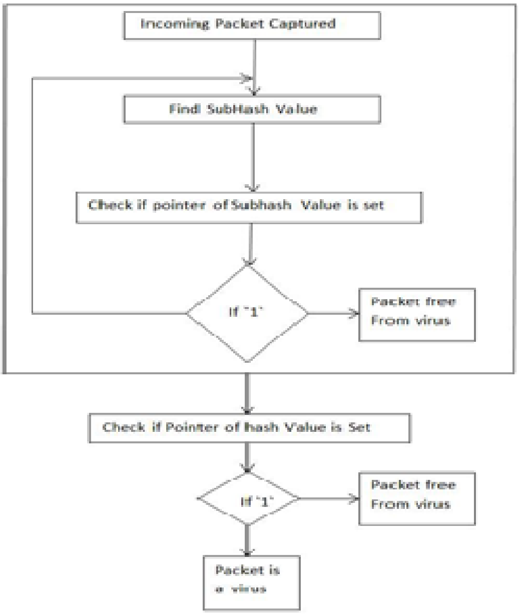


Fig. 7. Word split pattern matching algorithm [7]

Hence virus can be eliminated initially at the first Step itself or second step or finally.

3 Results and Comparison

After surveying the different string matching techniques it is necessary to make comparison of different approaches and their outcomes. These approaches will help in increasing the efficiency and to improve the performance of the string matching to a significant level.

3.1 Comparison of CAM and DCAM

Here presents the comparison of improvement of the DCAM (Decoded CAM) architecture with the earlier discrete comparator CAM design.

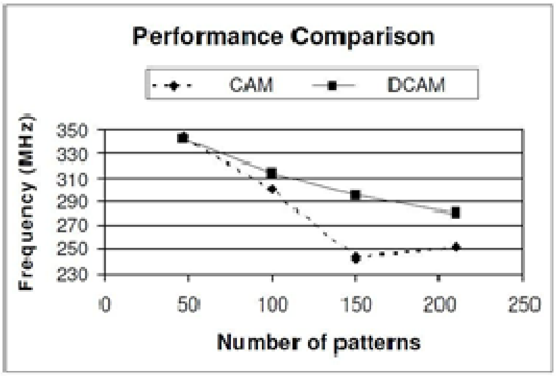


Fig. 8. The Performance comparison between the Discrete Comparator CAM and the DCAM architectures[8]

Figure 9 plots the cost of the designs again in terms of logic cells per search pattern character. It is clear that the DCAM architecture results in drastically smaller designs for the largest rule set, the DCAM area cost is about 4 logic cells per character, while the cost of our earlier design almost 20 logic cells per character.

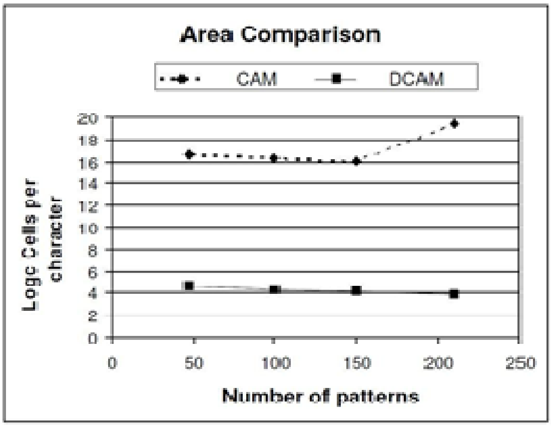
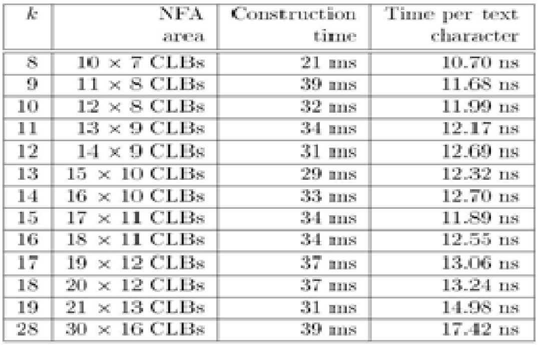


Fig. 9. The Area cost comparison between the Discrete Comparator CAM and the DCAM architectures[8]

3.2 Comparison of NFA and CFA Pattern Matching

From sidhu et al[12]of NFA implementation The performance result for the NFA construction time, NFA size and time required to process a text character for (a|b)*a(a|b)k ,k ranging from 8 to 28 as follows.



k	NFA area	Construction time	Time per text character
8	10 × 7 CLBs	21 ms	10.70 ns
9	11 × 8 CLBs	39 ms	11.68 ns
10	12 × 8 CLBs	32 ms	11.99 ns
11	13 × 9 CLBs	34 ms	12.17 ns
12	14 × 9 CLBs	31 ms	12.69 ns
13	15 × 10 CLBs	29 ms	12.32 ns
14	16 × 10 CLBs	33 ms	12.70 ns
15	17 × 11 CLBs	34 ms	11.89 ns
16	18 × 11 CLBs	34 ms	12.55 ns
17	19 × 12 CLBs	37 ms	13.06 ns
18	20 × 12 CLBs	37 ms	13.24 ns
19	21 × 13 CLBs	31 ms	14.98 ns
28	30 × 16 CLBs	39 ms	17.42 ns

Fig. 10. The NFA area,NFA constructiontime and time per text character for the FPGA implementation[12]

DFA are usually efficient when the expression is not repetitive and becomes larger than NFA when expression is repetitive .In general DFA may require up to 2n states compared to an equivalent NFA with only n states that leads to state explosion .This becomes very hard to implement in hardware when resource are not sufficient[1].

3.3 Comparison of Normal Hash and Word Split Hash

The graph is drawn in Fig.11 with execution time against network traffic, for normal hash and word split hash with the assumption that the probability of virus in the incoming data is low.

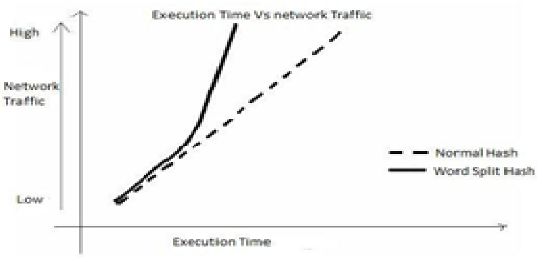


Fig. 11. Execution time vs network traffic for normal hash and WSHA

Execution time difference is low between the two methods in low network traffic since the packet frequency is less the time difference is also not significant. But in high traffic network word split hash has huge difference in the execution time over normal hash.

4 Conclusions

In this paper we have seen existing hardware based technique for pattern matching, the background of network intrusion detection system and also we have seen that string matching as the major performance bottleneck in intrusion detection systems. This document also surveys the various efficient method for string matching with current state of the string matching implementation for intrusion detection using FPGA with the goal of increasing throughput ,speed and reduce the area further work can be done in this field to achieve these goals .

References

1. Babu Karuppiah, A., Rajaram, S.: Deterministic Finite Automata for Pattern Matching in FPGA for intrusion Detection. In: International Conference on Computer and Electrical Technoogy, ICCCET 2011, March 18-19 (2011)
2. Nakahara, H., Sasao, T., Matsuura, M.: A Regular Expression Matching Using Non-Deterministic Finite Automata. IEEE (2010)
3. Bonesana, I., Paolieri, M., Santambrogio, M.D.: An adaptable FPGA based system for regular expression Matching. IEEE (2008)
4. Aldwairi, M., Conte, T., Franzon, P.: Configurable string Matching Hardware for Speeding up Intrusion detection. ACM SIGARCH Computer Architecture News 33(1) (March 2005)
5. Tummala, A.K., Patel, P.: Distributed IDS using Reconfigurable Hardware. IEEE (2007)
6. Le, H., Prasanna, V.K.: Ming Hsieh Department of Electrical Engineering University of Southern California Los Angeles, CA 90089, USA A Memory-Efficient and Modular Approach for String Matching on FPGAs (2010)
7. Dhanapriya, M., Vasanthanayaki, C.: Hardware Based Pattern Matching Technique for Packet Inspection of High Speed Network. In: International Conference on Control, Automation, Communication and Energy Consevation 2009, June 4-6 (2009)
8. Sourdis, I., Pnevmatikatos, D.N., Vassiladis, S.: Scalable Multigigabit Pattern Matching for Packet Inspection. In: Proc. IEEE Symp. Field Program. Custom Comput. (February 2008)
9. Hutchings, B.L., Franklin, R., Carver, D.: Scalable hardware implementation usonf Finite Automata. Department of Electrical and Computer Engin.
10. Bloom, B.: Space/Time Tradeoffs in Hash Coding with Allowance Errors. Comm., ACM 13(7), 422–426 (1970)
11. Hasan, J., Cadambi, S., Jakkula, V., Chakradhar, S.: Chisel: A Storage-efficient, Collision-free Hash-based Network Processing Architecture. In: 33rd International Symposium on Computer Architecture, pp. 203–215
12. Sidhu, R., Prasanna, V.K.: Fast Regular Expression Matching using FPGAs. In: 9th Annual Symposium IEEE (2001)

Impact of Blackhole and Rushing Attack on the Location -Based Routing Protocol for Wireless Sensor Networks

R. Shyamala[1] and S. Valli[2]

[1] Assistant Professor, Department of Information Technology,
University College of Engineering Tindivanam., Melpakkam 604 001
vasuchaaru@gmail.com
[2] Associate Professor, Department of Computer Science and Engineering,
Anna University, Chennai-25
valli@annauniv.edu

Abstract. Wireless sensor networks have millions of sensors, which cooperate with one on other in home automation, military surveillance, entity tracking systems and several other applications. In these networks, multicast is the basic routing service for efficient data broadcasting for task assignment, targeted queries and code updates. The sensor nodes have low computational capabilities, and are highly resource-constrained. So, the multicast routing protocols of wireless sensor networks are prone to various routing attacks, such as black hole, rushing, wormhole and denial of service attacks. The objective of this paper is to study the effects of the black hole and rushing attack on the location based Geographic multicast routing (GMR) protocol. The NS-2 based simulation is used in analyzing the black hole and rushing attacks on the GMR. The black hole and rushing attack degrades the network performance by 26% and 18% respectively.

Keywords: Wireless sensor networks, geographic multicast routing, black hole attack, rushing attack, joules and throughput.

1 Introduction

A wireless sensor network (WSN) consists of sensor nodes, which are simple processing devices. The sensor nodes have the capability of sensing parameters like temperature, humidity and heat. The sensor nodes (Eric et al, 2009) communicate with each other, using wireless radio devices and form a WSN. The WSN is dynamic and has a continuous changing network topology, which makes routing difficult. Bandwidth and power limitations are the important resource constraints.

The authors (Kai et al, 2010), classify the attacks on WSN as active and passive attacks. The monitoring of and listening to the communication channel by unauthorized attackers, are known as passive attacks. The attack against privacy is passive in nature. Some of the more common attacks against sensor privacy are monitoring and eavesdropping, traffic analysis and camouflage adversaries. If the unauthorized attackers monitor, listen to and modify the data stream in the communication channel,

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 349–359.
springerlink.com

then the attack is an active attack. Routing attacks such as spoofing, replay, selective forwarding, sinkhole, sybil, wormhole and HELLO flood are active attacks. Denial of service attacks such as neglect and greed, misdirection and black hole are also active in nature.

(Hoang et al, 2008) in their study, classify the rushing, black hole, neighbor and jelly fish as severe routing attacks. The impact of these routing attacks was studied by varying the number of senders and receivers. From the results it is shown, that the rushing attack causes more damage to the routing, irrespective of the number of senders and receivers. In the black hole attack, if the attacker is closer to the destination, heavy damage is caused to the network. According to (Hoang et al, 2008) a large mesh MANET has negligible damage from any type of routing attacks. (Kannhavong et al, 2007) have handled flooding, black hole, link withholding, link spoofing, replay, wormhole and colluding misrelay attacks on Mobile ad-hoc network(MANET) routing protocols. (Avinash et al, 2010) used a non-cooperative game theory to identify black hole nodes and proposed a new Ad hoc On-demand Distance Vector (AODV) routing protocol for Mobile Ad hoc network (MANET). (Jorge et al, 2010) used watchdog and Bayesian filters to detect a black hole attack by means of which malicious nodes are identified in MANET. (Anoosha et al, 2011) identified black holes using honey pot agents. This roaming software agent performs a network tour and identifies the malicious node through route request advertisements and maintains intrusion logs. (Kai et al, 2010) proposed an energy efficient, denial of service (DoS) and flooding attack resistant routing protocol, using ant colony optimization. In this algorithm, every node has a trust value. Faithful forwarding nodes are selected based on the remaining energy and trust value. (Guoxing et al, 2010) proposed a trust aware secure multi hop routing protocol for the WSN. The trust values are calculated by exploiting the replay of routing information, by which all the malicious nodes are dropped from routing decisions. In the previous work (Shyamala et al, 2009) a TESLA based secure route discovery is suggested for MAODV. The effect of sybil and wormhole attack (Shyamala et al, 2011) in GMR is investigated for a WSN. It is found that the wormhole attack does more damage than the sybil attack on the routing procedure.

This paper simulates the black hole and rushing attack in Geographic Multicast Routing (GMR). The simulation was carried out using NS-2 and the network performance is studied with and without the black hole and rushing attack in the WSN.

The rest of this paper is organized as follows. Section 2 describes the Geographic Multicast Routing protocol (GMR). Section 3 describes the rushing attack and section 4 the black hole attack. Section 5 describes the simulation environment and analyses the performance of the network in the presence and absence of the black hole and rushing attack and section 6 concludes the work.

2 Geographic Multicast Routing Protocol

Depending on the network structure, routing in WSNs can be divided into flat-based routing, hierarchical-based routing and location-based routing. Sensor protocols for information via negotiation (SPIN), directed diffusion and rumor routing are some of the flat-based routing algorithms. Low energy adaptive cluster hierarchy (LEACH), leach centralized (LEACH-C), power efficient gathering in sensor information system (PEGASIS) are the hierarchical routing protocols.

(Sancez et al, 2007) proposed an energy efficient routing protocol for WSN, called the Geographic Multicast Routing Protocol (GMR), which is one of the location-based protocols. The GMR protocol calculates the position of the sensor nodes from the Global Positioning System (GPS) (Lianming et al 2008) or it can use the virtual co-ordinates. Each sensor node communicates its position to its neighbors, using periodic beacons. The GMR forms a multicast tree to send a data packet, from a source to multiple destinations, using a single broad cast transmission.

In the GMR (Sancez et al, 2007), each forwarding node selects a subset of its neighbors in the direction of the destination as relay nodes, based on the cost-over-progress ratio. The cost is equal to the number of selected neighbors. Progress is the reduction of the remaining distances to the destinations. The cost-over-progress metric is explained with respect to Fig. 1. The remote source node S multicasts the message M to a set of destinations $\{D_1, D_2, D_3, D_4, D_5\}$. The forwarding node C receives the message M from the source S and uses its neighbors A_1 and A_2 as the relay nodes. In the GMR, the multicasting task could be given to one neighbor, or it could be handled by several neighbors. Each neighbor could address a set of destinations.

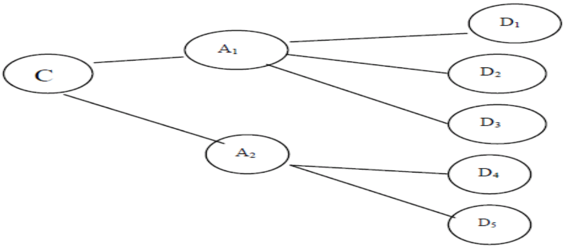


Fig. 1. Neighbor Selection

$$T_1= |CD_1|+|CD_2|+|CD_3|+|CD_4|+|CD_5| \tag{1}$$

$$T_2=|A_1D_1|+|A_1D_2|+|A_1D_3|+|A_2D_4|+|A_2D_5| \tag{2}$$

$$P_i = 2 / (T_1 - T_2) \tag{3}$$

C_{ID} $A_{1(Id)}, \{D_{1(Id)}, D_{2(id)}, D_{3(id)}\}$ $A_{2(Id)}, \{D_{4(Id)}, D_{5(id)}\}$

Fig. 2. Header Format

From node C the total distance for multicasting is T_1 as given in equation (1). Hence, node C applies the greedy partitioning algorithm, and selects A_1 as the relay node responsible for D_1, D_2 and D_3. Node A_2 is chosen as the relay node for D_4 and D_5. For the next level of the multicast tree, a new total distance T_2 is calculated as given in equation (2). The progress is the difference between T_1 and T_2 as given in equation (3). The cost-over-progress ratio (P_i) for the new forwarding set $\{A_{1,}\ A_2\}$ is $2/(T_1 - T_2)_.$ Node C informs its neighbors that they are selected as the relay nodes through the header, as given in Figure 2. The GMR adds this header to the data message.

In Fig. 2, the first field is the node namely, node C, which applies the greedy partitioning algorithm. The next field is the first relay node A_1 and the set of destinations it has to handle $\{D_{1,}\ D_2\ ,D_3\}$. The third field is the second relay node, A_2 and the set of destinations $\{D_{4,}\ D_5\}$. Thus, the sender broadcasts a single message and it reaches the destination by selective forwarding. Hence, the energy and bandwidth consumption are minimized.

3 Rushing Attack in GMR

The rushing attack (Hoang et al, 2008) is a kind of denial of service attack. When the source node floods the network with route discovery packets to find routes to the destinations, each intermediate node processes only the first non-duplicate packet, and discards the other duplicate packets that arrive at a later time. A rushing attacker exploits this duplicate suppression mechanism by quickly forwarding the route discovery packets, in order to gain access to the forwarding group.

This paper studies the rushing attack in terms of its effect on the operation of the GMR. In this study the GMR is implemented in the source node(C) which initiates a data message to 20 destinations. The malicious nodes are uniformly distributed throughout the network. The cost-over-progress ratio is calculated. In this simulation, the rushing attack is introduced by setting the data packets processing delay time to 10 ms for all the good nodes. For the rushing nodes (M) the processing delay time is set to zero. Therefore, node M_i is chosen as the forwarding node by the greedy partitioning algorithm of the GMR (Sancez et al, 2007). The malicious node M_i is chosen as the relay node, since it has the best cost-over-progress ratio. From this experimentation, it has been found that the introduced malicious nodes will be selected as the forwarding nodes. Fig. 3 is the pseudo code of the rushing attack in the GMR.

```
RUSHING(M: Set of malicious node)
Begin
for all nodes do
  set bestCOP = 0;
end for;
M={$M_1$,$M_2$,$M_3$,...$M_n$}, Where, $M_i$ = Malicious Node i
for i=1 to n do
          Set Processing Delay as 0 for $M_i$    // Malicious nodes
end for
for i =1 to n do
          Set Processing Delay as 10 ms for $N_i$// Normal nodes
end for;
```

Fig. 3. Pseudo Code For A Rushing Attack

```
// node C receives a multicast message from source node S
If (GMR_neighbour_ID == ID of node C) then
  Get the neighbor list (A)
for i=0 to k do // k neighbors ($A_i$) of C
for j=0 to m do // m destination ($D_j$)
CurrentDistance(i,j) - = distance (C,$A_i$ )+ distance($A_i$, $D_j$);
end for //for j;
Progress($P_i$) = Min(CurrentDistance(i,j) );
Calculate Cost($A_i$) = Packet_arrival_time(C,$A_i$)
newCOP=Cost($A_i$) / Progess($P_i$)
if COP($A_i$) > newCOP($A_i$)
  bestCOP($A_i$)=newCOP($A_i$)
end if
end for
//for i;
else
drop PKT;
end if;
```

Fig. 3. *(continued)*

4 Black Hole Attack in the GMR

When all the messages are redirected to a specific node, it is defined as black hole attack (Santhosh et al, 2007). The node could be a malicious node. The traffic migrates into that malicious node. The node would not exist after a black hole attack. A black hole attack has two stages. In the first stage, the black hole exploits the routing protocol to advertise itself as having a valid route to the destination, even though the route is spurious. In the second stage, the node consumes the intercepted packets and suddenly disappears.

This paper implements the black hole attack in the GMR protocol, using the pseudo code given in Fig. 4. A set of malicious nodes(M) with the processing delay of 0 ms is launched, and the normal nodes are set with a processing delay of 10 ms. The black hole node advertises its ID and location information to its one hop neighbor by a beacon message. Then, GMR partition algorithm is executed. Since the black hole nodes have less processing delay and hence the best cost-over-progress ratio(COP), they are selected as the relay nodes. In our implementation only 6 nodes were selected as the forwarding nodes in the first iteration. So, the loop is repeated until all the 10 malicious nodes are selected as the forwarding nodes in the multicast tree. After 100 ms of simulation time, the malicious node starts dropping the packets. When 200ms is reached the energy is set to zero. So, the black hole node, disappears from multicast tree, as shown in Fig. 5. Fig. 6 is the data header format.

```
BLACKHOLE( M: set of Malicious Node )
repeat
        RUSHING(M);
Until all malicious nodes are selected as forwarding nodes.
if (Simulation time =100ms) then
        M drops PKT.
else
        if (Simulation time =200ms) then
        for i = 1 to n do
        Set energy of $M_i$ as 0;
        end for;
        end if;
end if;
```

Fig. 4. Pseudo Code For The Black Hole Attack

S_{ID} M_{ID},{$D_{1(Id)}$, $D_{2(Id)}$, $D_{3(Id)}$, $D_{4(Id)}$, $D_{5(Id)}$ }

Fig. 5. Black hole Header Format

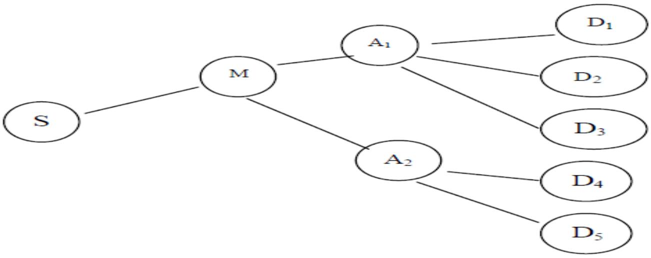


Fig. 6. Black Hole Attack

5 Simulation Environment

To evaluate the effectiveness of the proposed attacks, the GMR is simulated using NS-2. The goal of the evaluation is to test the effectiveness of the black hole and the rushing attack variations under normal conditions. The size of the data payload is 512 bytes. This simulation considers 200 sensor nodes. Nodes 11-200 are simple sensor nodes, and nodes 1 to 10 are the malicious nodes. Table 1 shows the simulation parameters. Table 2 is the obtained mean values of the network performance under no attack, the black hole attack and the rushing attack. The number of malicious nodes was varied from 2 to 10. The network performance is evaluated, using the packet delivery ratio (PDR), network throughput (NTh), packet drop ratio (PDrR) and energy loss metrics, in the presence of the black hole and rushing attack.

Table 1. Simulation Parameters

Examined Protocol	GMR	Movement model	Static
Simulator	NS-2	Initial energy	5J
Simulation time	250 Sec	RxPower	1.75mW
Simulation area	1000m x 1000m	TxPower	1.75mW
Number of sensor nodes	200	SensePower	1.75mW
Number of base stations	1	IdlePower	1.75µW
Number of malicious nodes	1-10	Transmission range	250m

5.1 Performance Analysis

The performance of the network is studied, by analyzing the packet delivery ratio(PDR), network throughput(NTh), packet drop ratio(PDrR), and energy loss for ten malicious nodes and the values are shown in figures 7 to 10.

5.2 Packet Delivery Ratio(PDR)

The packet delivery ratio is defined as the ratio of the total number of data packets received by the destination node to the number of data packets sent by the source node as given in equation (4).

Fig. 7 represents the packet delivery ratio measured for the GMR protocol in the presence of ten malicious nodes. The packet delivery ratio dramatically decreases in the presence of malicious node in the network. The mean packet delivery ratio calculated for 200 nodes is 80% when there is no attack. When 10 malicious nodes are introduced, the mean packet delivery ratio decreases to 53 % for black hole attack. In the case of the rushing attack, the mean PDR decreases to 69% because of fast message forwarding.

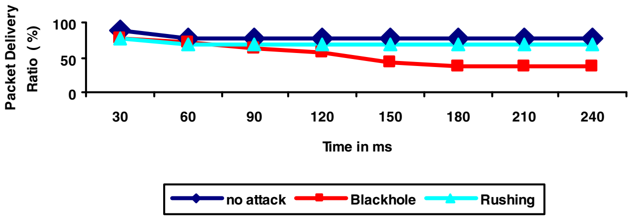


Fig. 7. Packet Delivery Ratio for 10 Malicious Nodes

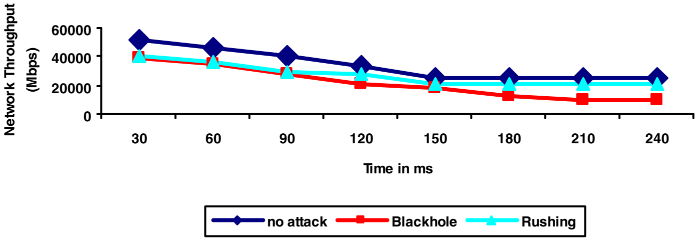


Fig. 8. Network Throughput for 10 Malicious Nodes

5.3 Network throughput (NTh)

The network throughput (NTh) represents the numbers of data packets generated by the source node to the number of data packets received in the destination, as given in formula (5). In Fig. 8, is the throughput values of the network is 67%, when there is no attack. In the rushing attack, where 10 malicious agent is launched at 100 ms, and it floods the data packets to all its neighbors. As a result, the mean throughput is reduced to 53%. In the case of the black hole attack, the malicious node, which is activated at 100ms, starts dropping the packets. Hence, the throughput regularly drops by 10%, and the mean throughput decreases to 42% for the black hole attack. In Fig. 8, the throughput is seen to be high, in the absence of an attack.

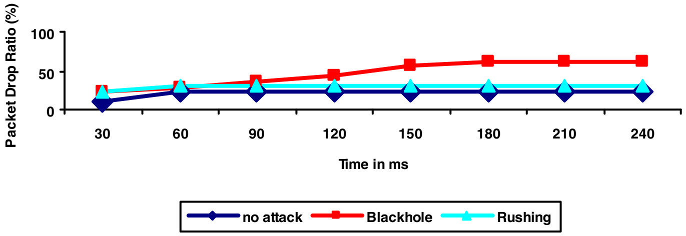


Fig. 9. Packet Drop Ratio for 10 Malicious Nodes

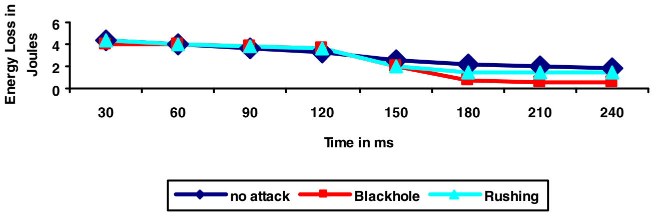


Fig. 10. Energy Loss for 10 Malicious Nodes

$$\text{Packet Delivery Ratio (PDR)} = \frac{\sum \text{ of packets received by the destination node}}{\sum \text{of packets sent by the source node}} \text{......(4)}$$

$$\text{Network Throughput (NTh)} = \begin{array}{l} \sum \text{ of packets generated by the source node} \\ \sum \text{of packets received at the destination} \end{array} \text{......(5)}$$

$$\text{Packet Drop Ratio (PDrR)} = \begin{array}{l} \sum \text{ of packets dropped by the network} \\ \sum \text{of packets generated by the network} \end{array} \text{......(6)}$$

5.4 Packet Drop Ratio (PDrR)

The packet drop ratio is the average number of packets dropped by the network to the number of packets generated by the network as given in equation (6). Fig. 9 shows the packet drop ratio in the case of the rushing and black hole attacks for ten malicious nodes. In the rushing attack, there is a packet loss between 50ms to 120 ms because, the malicious node floods the data packets to all its neighbors in the next 70 ms. The rushing attack has a uniform packet loss after 120 ms. The packets dropped in the network are more in the black hole than in the rushing attack.

5.5 Energy Loss (EL)

From Fig. 10 it is seen that the energy loss is uniform in the case of no attack. The network drops its energy by two joules from 60 ms to 150 ms for the rushing attack. In the next 30 ms the rushing attack lost one joule, and the black hole attack lost two joules in the next 50 ms. At 210 ms of simulation, the energy loss drops to 0.5 joules, because of the sudden disappearance of the black hole nodes.

Table 2. Mean Values Of The Black Hole Attack And Rushing Attacks With 10 Malicious Nodes

Time	PDR (%)	Network Throughput (Mbps)	Packet Drop Ratio (%)	End to End delay (Time)	Energy Loss (Joules)
NO ATTACK:					
30	90	51200	10	3.8	4.3
60	78	46080	22	3.8	4
90	78	40960	22	3.9	3.9
120	78	33280	22	3.9	3.3
150	78	25600	22	4.2	2.5
180	78	25600	22	4.2	2.2
210	78	25600	22	4.2	2
240	78	25600	22	4.2	1.8
Mean	79.5	34240	20.5	4.025	3.05
BLACK HOLE ATTACK:					
30	76	38400	24	4.8	4
60	72	34816	28	4.8	4.2
90	64	28160	36	5	3.9
120	56	20480	44	5	3.2
150	44	17920	56	5	2
180	38	12800	62	5.5	0.8
210	38	10240	62	5.5	0.5
240	38	10240	62	5.5	0.5
Mean	53.25	21632	46.75	5.138	2.25
RUSHING ATTACK:					
30	76	40960	24	4.5	4.3
60	70	36864	30	4.5	4.2
90	68	29696	32	4.6	3.8
120	68	28160	32	4.8	3.1
150	68	21504	32	5.2	2.3
180	68	20480	32	5.2	1.8
210	68	20480	32	5.2	1.5
240	68	20480	32	5.2	1.5
Mean	69.25	27328	30.75	4.9	2.9

From the performance metrics it is understood, that the black hole is a severe routing attack for WSNs.

6 Conclusion

With rapid developments in the WSN environment, the services based on the WSN have increased. In this paper, the effects of the black hole and rushing attacks on the GMR protocol have been studied. The packet delivery ratio, throughput, end-to-end delay and energy loss have been evaluated. There is a reduction in the packet delivery ratio, throughput and end to end delay as observed from the graphs. In the black hole attack, all network traffic is redirected to a specific node, the malicious node causing serious damage to the GMR protocol. In the rushing attack, because of a lengthier transmission queue in each node, the performance of the network is degraded. The prevention of the black hole and rushing attacks in the GMR for the WSN is still considered to be a challenging task, and it will be handled in the future work.

References

[1] Prathapani, A., Santhanam, L., Agrawal, D.P.: Detection of blackhole attack in a Wireless Mesh Network using intelligent honeypot agents. The Journal of Supercomputing, 1–28 (2010) (online First)

[2] Krishnan, A., Manjunath, A., Reddy, G.J.: Retracted: A New Protocol to Secure AODV in Mobile AdHoc Networks. Communications in Computer and Information Science 133(5), 378–389 (2011)

[3] Sabbah, E., Kang, K.-D.: Security in Wireless Sensor Networks Computer Communications and Networks. In: Guide to WSN, 1st edn., pp. 491–512 (2009)

[4] Zhan, G., Shi, W., Deng, J.: TARF: A Trust-Aware Routing Framework for Wireless Sensor Networks. In: Silva, J.S., Krishnamachari, B., Boavida, F. (eds.) EWSN 2010. LNCS, vol. 5970, pp. 65–80. Springer, Heidelberg (2010)

[5] Nguyen, H.L., Nguyen, U.T.: A study of different types of attacks on multicast in mobile ad hoc networks. Journal on Adhoc Networks 6, 32–46 (2008)

[6] Hortelano, J., Calafate, C.T., Cano, J.C., de Leoni, M., Manzoni, P., Mecella, M.: Black-Hole Attacks in P2P Mobile Networks Discovered through Bayesian Filters. In: Meersman, R., Dillon, T., Herrero, P. (eds.) OTM 2010. LNCS, vol. 6428, pp. 543–552. Springer, Heidelberg (2010)

[7] Lin, K., Lai, C.-F., Liu, X., Guan, X.: Energy Efficiency Routing with Node Compromised Resistance in Wireless Sensor Networks, Mobile Networks and Applications, pp. 1–15 (2010) (online First^{TM})

[8] Xing, K., Srinivasan, S.S.R., Jose, M., Rivera, M., Li, J., Cheng, X.: Attacks and Countermeasures in Sensor Networks: A Survey. Network Security, 251–272 (2010)

[9] Xu, L., Deng, Z., Ren, W., Sch, H.W.: A Location Algorithm Integrating GPS and WSN in Pervasive Computing. Pervasive Computing and Applications, 461–466 (2008)

[10] NS-2, `http://www.isi.edu/nnam/ns/`

[11] Kannhavong, P., Nakayama, H., Nemoto, Y., Kato, N.: A Survey of Routing Attacks In Mobile Ad Hoc Networks. IEEE Wireless Comm. 14, 85–91 (2007)

[12] Sanchez, Ruiz, J.A., Stojmnenovic, P.M.: GMR: Geographic Multicast Routing for Wireless Sensor Networks. Journal on Computer Comm., 2519–2531 (2007)

[13] Kurosawa, S., Nakayama, H., Kato, N., Jamalipour, A., Nemoto, Y.: Detecting Black hole Attack on AODV based Mobile Ad-hoc networks by Dynamic Learning Method. International Journal of Network Sec. 5, 338–346 (2007)

[14] Shyamala, R., Valli, S.: Secure route discovery in MAODV for Wireless Sensor Networks. UbiCC Journal 4, 775–783 (2009)

[15] Ramachandran, S., Shanmugan, V.: Impact of Sybil and Wormhole Attacks in Location Based Geographic Multicast Routing Protocol for Wireless Sensor Networks. Journal of Computer Science 7, 973–979 (2011)

Analysis of Fractional Frequency Reuse (FFR) over Classical Reuse Scheme in 4G (LTE) Cellular Network

Chandra Thapa[1] and C. Chandrasekhar[2]

[1] SV College of Engineering & Technology, M. Tech II (DECS)
R.V.S Nagar, Chittoor-517127, A.P. India
Chandra2thapa@gmail.com
[2] SV College of Engineering & Technology, Head ECE Department
R.V.S Nagar, Chittoor-517127, A.P. India
hodece@svcetedu.org

Abstract. With the advent of fourth generation (4G) cellular networks like Long Term Evolution (LTE), there is always difficulty for proper frequency planning as it is targeting aggressive spectrum reuse (frequency reuse 1) to achieve high system capacity (rate and throughput). At same time, we face signal degradation at cell edge users due to interference by co-channel cells. There are various ways to manage interference by co-channel cells. This paper is focus on comparison of Fractional Frequency Reuse (FFR) over classical reuse scheme on basis of probability of acceptance rate with respect to Rate Threshold. And in order to allocate resources in Cells, Signal to Interference plus Noise Ratio (SINR) proportional resource allocation strategy is deployed. We observe that FFR provides better acceptance rate as well as improved coverage for cell edge users in LTE environment.

1 Introduction

LTE is accepted worldwide as the Long Term Evolution perspective for today's 2G and 3G networks based on WCDMA/HSPA, GSM/EDGE, TD-SCDMA and CDMA 2000 technology. The 4th Generation (4G) of wireless mobile systems is characterized by Long Term Evolution (LTE) [1] and WiMAX [2] technology which evolved with higher data rates and improved quality of service even for cell edge users. It is expected that LTE will be deployed in a reuse one configuration, in which all frequency resources are available to use in each cell. LTE has adopted Orthogonal Frequency Division Multiple Access (OFDMA) multiple access technique for Downlink and Single Carrier Frequency Division Multiple Access (SC-FDMA) for Uplink, other detail specification can be found in reference [3]. Release 7 on December 2007 by 3rd Generation Partnership Project (3GPP) contained first work on LTE [3]. Then afterward lots of works have been added in LTE and it is consider as recent hot research topic.

Now considering LTE cellular network, when a frequency reuse of 1 is supported, i.e. all cells will operate on same frequency channels to maximize the spectral efficiency (number of channels per cell is increased), the inter-cell interference is major concern. There will be less effect to the users near to Base station but cell edge users may suffer degradation in connection. This can be address by using reuse ratio of 3 (classical

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 361–367.
springerlink.com

frequency planning) i.e. dividing total spectrum band into 3 sub bands and allocate only one sub band to a given cell, so that adjacent cells use different frequency bands. Meanwhile co-channel interference is reduced but at expense of decrease in efficiency in terms of coverage and capacity.

Analyzing above scenario a mix frequency reuse 1 and 3 schemes can be used to avoid interference at cell edges. Here, the total frequency band is divided into two sub bands: a frequency reuse 1 sub band is allocated to users at cell centers of all cells, and a frequency reuse of 3 sub bands is allocated to cell edge users [4]. This decreased the interference but also decreases the data rates as full frequency band is not used by this method. For implementation if this new idea there are two different methods: A static approach where a user is assigned a bandwidth depending on its position (path loss), and another is dynamic approach where the frequency assignment is done on the basis of position as well as cell loads. This above present concept is of FFR technique and our paper focus on static approach.

There are various frequency allocation schemes like OFDMA, SC-FDMA, Partial isolation, classical frequency planning, Fractional frequency planning [5], here we basically focus on two of them which are as follows:

1.1 Classical Frequency Planning

This is simplest scheme to allocate frequencies in a cellular network by using reuse factor of 1 (Fig.1) which leads to high peak data rates. However, in this case, higher interference is observed on cell edges. The classical interference management is done by using reuse ratio 3 (Fig.2), by using this interference is low but large capacity loss because only one third of resources are used in each cell.

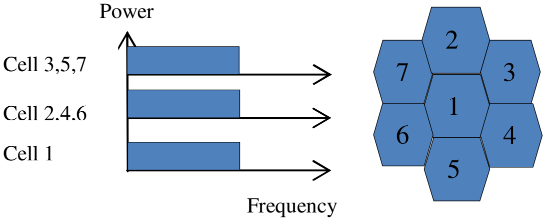


Fig. 1. Classical Frequencies Reuse with reuse ratio 1.

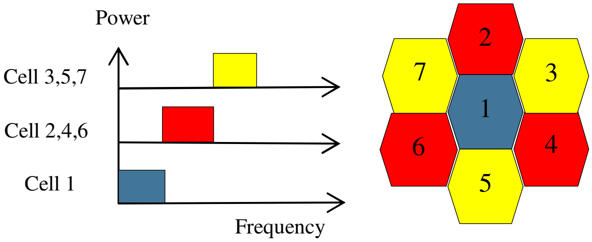


Fig. 2. Classical Frequencies Reuse with reuse ratio 3.

1.2 Fractional Frequency Planning

The basic idea of FFR is to partition the cell's bandwidth so that (i) cell-edge users of adjacent cells do not interfere with each other and (ii) interference received by (and created by) cell-interior users is reduced, while (iii) using more total spectrum than classical frequency reuse [6]. The use of FFR in cellular network is tradeoffs between improvement in rate and coverage for cell edge users and sum network throughput and spectral efficiency.

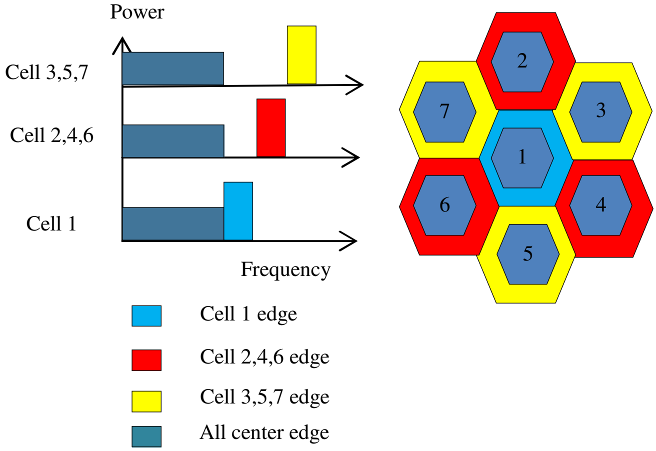


Fig. 3. FFR in LTE, Frequency Reuse factor for cell edge users is 3.

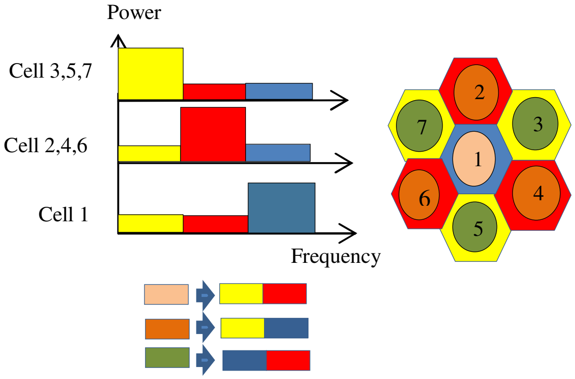


Fig. 4. Soft Frequency Reuse (SFR) with reuse ratio 3.

Among 3 major frequency reuse patterns, FFR is compromised between hard and soft frequency reuse. In Hard frequency reuse splits the system bandwidth into a number of distinct sub bands according to chosen reuse factor and it let neighboring cell transmit on different sub bands. FFR splits the given bandwidth into an inner and outer part. It allocates the inner part to the near users located near to BS in terms of path loss with reduced power applying frequency reuse factor of one i.e. the inner part is completely reused by all BSs (illustrated in Fig.3). The far users over cell edge, a fraction of the outer part of bandwidth are dedicated with the frequency reuse factor greater than one [7]. With soft frequency the overall bandwidth is shared by all base stations (i.e. reuse factor of one is applied) but for the transmission on each subcarrier, the BSs are restricted to a certain power bound [8].

There are two common FFR models: strict FFR and Soft Frequency Reuse (SFR) [6]. Strict FFR is modification of the traditional frequency reuse used extensively in multi-cellular networks (Fig.3 is example of strict FFR for reuse 3 at cell edge users) and they don't share the exterior sub bands to the inner frequency bands. Soft Frequency Reuse (SFR) employs the same cell-edge bandwidth partitioning strategy as Strict FFR, but the interior users are allowed to share sub-bands with edge users in other cells (illustrated in Fig.4). Thus, the shared sub bands by the interior users will be transmitted at lower power levels than the cell edge users [6]. SFR is more bandwidth efficient than strict FFR, it results more interference to both cell-interior and edge users [9]. Here in our paper we focus basically on Strict FFR type.

2 System Model

Here the system model is simple implementing Hexagon geometry. Each cell has given systematically integer label to indicate which frequencies are to be used in a frequency reuse scheme. The distribution of mobile are randomly scattered across the cell and stationary over the plane. Their intensity is λ, and distributed as per poisson distribution. Each mobile is communicating with nearest base-station. We assume that respective base station and respective user experience only Rayleigh fading with mean 1 and constant transmit power of 1/μ. Now, the received power at a typical node at a distance r from its base station is $hr^{-\alpha}$ where random variable h has exponential distribution with mean 1/μ and h ~ exp(μ). From above mentioned model, the SINR of the mobile user at a random distance of r from its associated base station is:

$$\text{SINR}=\frac{hr^{-\alpha}}{\sigma^2+I_r} \tag{1}$$

Here interference power which is sum of received power from all other base station other than the home base station treated as noise is:

$$I_r=\sum\nolimits_{i\in\phi/b_o}(g_iR_i^{-\alpha}) \tag{2}$$

Where,

‘g’ is statistical distribution and is fading value or value for fading, shadowing and any other desired random effect with mean (1/μ). When g is also exponential then simpler expression will result.

‘h’ is exponential random variable(h~ exp(μ)).

‘r’ is distance from mobile to its base station.

‘R’ is distance from the mobile to other stations on same reuse assignment.

‘ α ’ is path loss coefficient.

‘ σ^2 ’ is noise power.

And ‘i’ represents each of the mobiles which are interfering with the mobile whose SINR is being calculated. All above results are for single transmit and single receive antenna and similarly we consider that there is no same cell interference due to orthogonal multiple access (OFDMA) within a cell. The noise power is assumed to be additive and constant with value of σ^2 but no specific distribution is assumed [10]. The detail formulation of these equations is given in reference [10]. The coverage probability is the probability that a typical mobile user is able to achieve some threshold SINR, i.e. it is the complementary cumulative distribution function (CCDF). Mathematically, coverage probability is:

$$p_c(T,\lambda,\alpha) \cong P[SINR > T] \tag{3}$$

Where, T is target threshold SINR value.

The CDF gives P [SINR≤ T] so CCDF of SINR over the entire network is probability of coverage too. The achievable rate [10] shows $\tau \to \ln(1+SINR)$, i.e. Shannon bound. τ has unit nats/Hz (since log is base e and 1 bit =ln(2)=0.693nats).

3 Results

The following are the parameters considered during simulation:

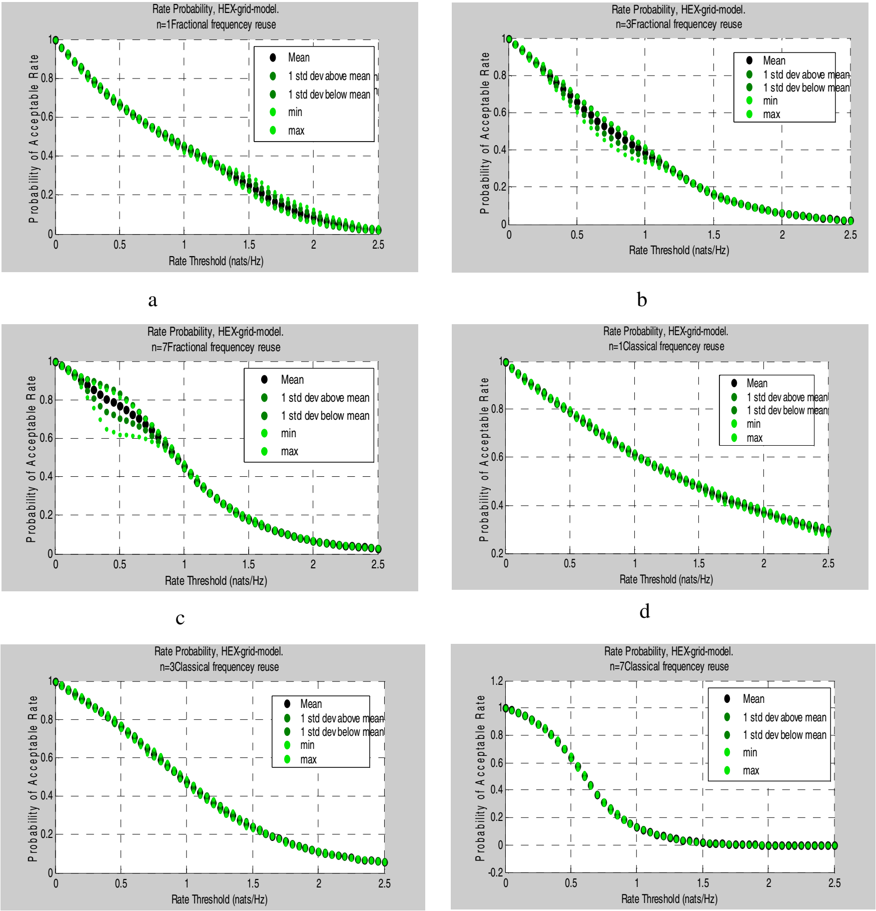
The intensity of user, λ=5

Path loss coefficient, α=4

Avg. SNR= 10dB

We have considered 10 cells with users and 15 cells as total cells under consideration during evaluation of parameters. For fractional frequency reuse we simulate with SINR threshold of 15 dB and 25% of power and bandwidth allocate to the center. This means if any mobile user has SINR ≥ 15dB will reside in center using frequency reuse 1 and other will be outside center using frequency reuse as per fractional frequency reuse scheme.

If we analyze above results, *fig.5a* and *fig.5d* depicts that we are getting higher probability of acceptance for respective rate under classical reuse and that is obvious. *Fig. 5b* and *fig.5e* has more or less similar outcome. But, for reuse 7, it is distinguished that FFR has better performance than classical in terms of acceptance probability and *fig. 5c* and *fig.5f* depict the output results. If we consider rate 1.5 nats/Hz then probability of acceptance are 20% for FFR assignment and nearly zero percentage for classical approach. Thus, this has 20 times better performance at this point.



a

b

c

d

e

f

Fig. 5. Graphs obtained from simulation for respective frequency reuse technique and ratio. Where, *fig a* is of FFR for reuse 1, *fig b* is of FFR for reuse 3, *fig c* is of FFR for reuse 7, *fig d* is of Classical Frequency Reuse for reuse 1, *fig e* is of Classical Frequency Reuse for reuse 3 and *fig f* is of Classical Frequency Reuse for reuse 7.

It is obvious that higher the reuse value there will be more SINR which results higher rate but meanwhile higher reuse means less bandwidth for each mobile so it points towards lower rate. Thus, "push and pull" activity is observed and results depend up on that factor which is dominant.

4 Conclusion

This paper presents comparison in between FFR and Classical frequency reuse scheme in 4G most specially in LTE environment (homogenous condition). These results shows, FFR provides better probability of acceptance rate for given rate threshold value considering "PUSH & PULL" effect. Thus, there should be taken care of SINR as well as bandwidth because both are important metrics. FFR balances the requirements of interference reduction and resource efficiency. In addition to this result, user friendly GUI is build so, further analysis at different values and ratios are made easier.

References

1. LTE-A, Requirements for Further Advancements for EUTRA, 3GPP TR 36.913 (2008)
2. 802.16m, Draft IEEE 802.16m Evaluation Methodology, IEEE802.16m-07/037 r1 (2007)
3. Agilent 3GPP Long Term Evolution: System Overview, Product Development and Test Challenges Application Note
4. 3GPP, R1-050507, Huawei, Soft frequency reuse scheme for UTRAN LTE (2005)
5. Elayoubi, S.-E., Ben Haddada, O., Fourestie, B.: Performance Evaluation of Frequency Planning Schemes in OFDMA-based Networks. IEEE Transaction on Wireless Communications 7(5) (May 2008)
6. Novlan, T.D., Ganti, R.K., Ghosh, A., Andrews, J.G.: Analytical Evaluation of Fractional Frequency Reuse for OFDMA Cellular Networks. AT & T Laboratories (January 23, 2011)
7. Ghaffar, R., Knopp, R.: Fractional Frequency Reuse and Interference Suppression for OFDMA Networks. Eurecom's research (2011)
8. Bonald, T., Borst, S.C., Proutiere, A.: Inter-cell scheduling in wireless data networks. In: Proceeding of European Wireless Conference (2005)
9. Doppler, K., Wijting, C., Valkealahti, K.: Interference aware scheduling for soft frequency reuse. In: Proc. IEEE Vehicular Technology Conf., Barcelona, pp. 1–5 (April 2009)
10. Andrews, J.G., Baccelli, F., Ganti, R.K.: A Tractable Approach to Coverage and Rate in Cellular Networks, February 22 (2011)

Temporary Parallel Route Recovery for Frequent Link Failure in VANET

B. Siva Kumar Reddy[1], M. Sakthi Ganesh[2], and P. Venkata Krishna[3]

[1] M.TECH-IT (Networking) School of Information Technology & Engineering (SITE)
[2] Assistant Professor, School of Information Technology & Engineering (SITE)
[3] Professor, School of Computing sciences & Engineering,
VIT University, Vellore
bskreddy64@gmail.com, sakthiganesh.m@vit.ac.in,
pvenkatakrishna@vit.ac.in.

Abstract. VANET is an upcoming technology to establish communication between the vehicles while travelling, which provides internet connectivity resulting in increased road safety, giving important alerts and accessing comforts while travelling. The VANET technology integrates WLAN, Cellular and Adhoc networks to achieve the continuous connectivity between the vehicles. Vehicle Ad-hoc Networks (VANETs) are systems that allow vehicles to communicate with each other. Wireless device can send information to nearby vehicles, and messages can be routed from one vehicle to another, so that the information can be spread throughout the city. In network there is a very frequent link failure due to high mobility of nodes from available network region. So this frequent link failure causes packets to not reach respective destination. The mechanism proposed here establishes a kind of parallel route discovery for real time application for packets to be delivered at destination by minimizing losses. The main goal is to establish parallel routes during link failures for real time application scenarios to deliver the data safely to destination. The parallel route recovery establishes temporary parallel path between the nodes when there is link failure. The node before the failure link buffers the packets, after establishing new parallel path it then forwards the buffered packets to the destination through newly established path.

Keywords: VANET, AODV, AOMDV, FROMR, buffer, parallel path.

1 Introduction

The main focus is to propose a new method for route recovery process that provides efficient routing when there is link failure and also to avoid congestion in network during link failures. This paper mainly deals with a temporary parallel route recovery mechanism during link failures and about the performance metrics. There are many problems in vehicular adhoc networks like frequent link failures due to the high mobility of nodes. Here the topology is high dynamic topology the vehicles will be moving with high speed, the topology formed will be always changing. The packet loss will be high, the packets does not reach destination due to frequent path breaks.

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 369–374.
springerlink.com

Generally in VANETS each and every node acts as a router in forwarding packets from source to destination. There is chance that the packets may not be delivered, that is they may be dropped because of some reasons. The lost data cannot be recovered back from the intermediate nodes during link failures, the lost packets can be retransmitted on request from source. To overcome the drawback we found a solution in proposed method. The temporary parallel path takes the better route from source to destination so the transmission delay can be minimized when there is route failure.

There are many routing protocols for adhoc networks like AODV, DSR, TORA and DSDV. Among all AODV is important on-demand routing protocol. AODV protocol establishes route for source node when there is data for transmitting. The AODV has phases like route discovery, route maintenance and data delivery. Simulation results for protocols like AODV, DSDV, TORA, DSR are found in many papers have summarized that AODV performs better when compare to all. AODV gives better performance in following metrics: packet delivery ratio, routing overhead, path optimization. The proposed system is the enhancement of AODV. In the proposed system "On demand temporary parallel route recovery for link failure" gives better performance when compared to AODV.

2 Related Work

In "On demand temporary parallel route recovery for frequent link failures in MANETS", [1] the author proposed a mechanism which propagates a parallel route discovery when there is a frequent link failures. The mechanism is proposed in order to save data losses during link failures. The packets are stored in buffer at nodes when there is link failure and buffered packets are delivered through new route established.

In "Enhancing AODV routing protocol using mobility parameters in VANET", [3] the authors proposed an enhanced routing protocol by enhancing AODV. They enhanced AODV protocol to make it adaptive to vehicular adhoc networks (VANETS). In their paper they took direction and position as important parameters in choosing next hop in route discovery phase. The main objective of the proposed protocol is to establish a new stable route in VANETS.

In "Design of Fast Restoration Multipath Routing in VANETs",[6] the authors proposed a multipath routing protocol for VANET and they named it as *Fast Restoration On-demand Multipath Routing* (FROMR), The FROMR protocol mainly focuses on rapidly establishing an alternate path if the original route is broken. In order to reduce the amount of control messages as well as increase the path robustness, The FROMR protocol divides the geographical region into squares of equal sizes called as grids. In each grid, the vehicle that is expected to stay for the longest duration is selected as the grid leader. Only grid leaders are responsible for route discovery, maintenance and restoration during link failures.

In "An Optimized Ad-hoc On-demand Multipath Distance Vector (AOMDV) Routing Protocol", [7] the author proposed an optimized AOMDV that solves the "route cutoff" problem in AOMDV by using a control packet RREP_ACK.

RREP_ACK control packet is defined in AODV, but in general it is ignored. Here they used it in the OAOMDV protocol. Even though the proposed routing scheme increases an additional routing packet, the simulations show that the routing overhead is decreased. The performance metrics such as packet loss, route discovery frequency, and average delay has been improved.

3 Proposed System

The proposed method is a new parallel route recovery concept for buffering the packets and then transmitting packets that are dropped by intermediate nodes during link failures. Delays that occur by the link failures are minimized by allocating buffers at the intermediate nodes. Each node that carries time sensitive critical data is allotted with buffer that holds the data that is dropped during link failures. As soon as the alternate route is recovered the data at intermediate nodes is transmitted through that link.

When the sender wants to send time critical data to destination and finds frequent link failures for it then it generates TRREQ (temporary route request) packet and broadcasts it to neighboring nodes. The TRREQ uses the fields like hop count, TPRREQ ID, destination IP address, source IP address destination sequence number and source sequence number. The hop count is number of hops require the source to the node handling the TPRREQ. Hop count increments by 1 till it finds the destination node. TPRREQ ID is a unique ID number which identifies request i.e., TPRREQ. If the TPRREQ ID in the packet matches with the TPRREQ ID in the nodes route entry table then the TPRREQ ID will be dropped if that node is not the destination node.

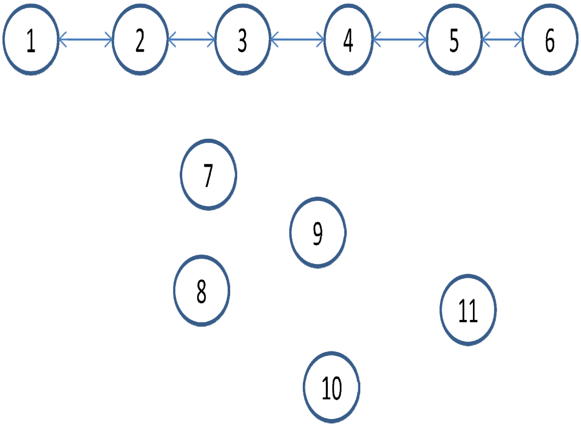


Fig. 1. Delivery of packets from node 1 to node 6 when there is no link failure

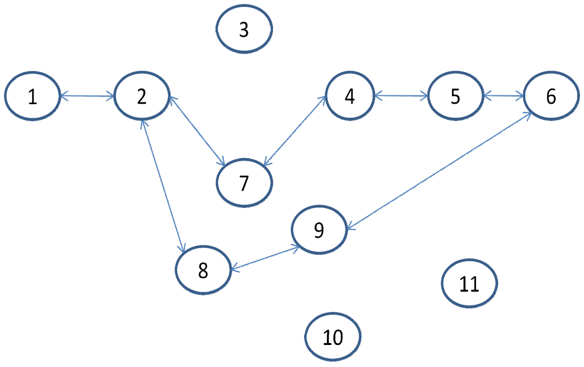


Fig. 2. Link failure at node 2 due to moving of node 3 away from transmission path

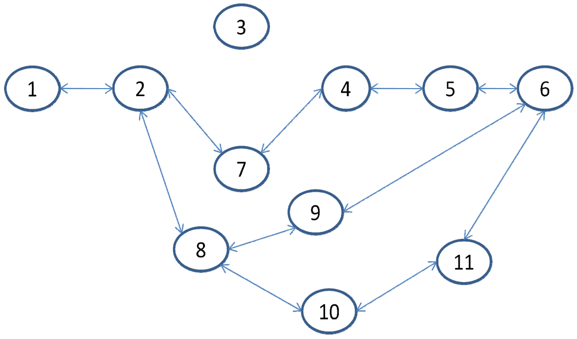


Fig. 3. Temporary parallel path for delivery of packets from node 1 to node 6 when there is link failure

Algorithm : Route Discovery

Route Request (RREQ) & Route Reply (RREP):

```
If packets are to be delivered from source to destination
Then
{
    Broadcast RREQ to all nodes N
    If (Node N$i$ = Destination D)
    {
        Send RREP to Source Node S,
    }
    Else
{
Broadcast RREQ to neighbor nodes
}
}
```

Route Maintenance

1. If there is link failure notify to source S by sending RERR message.
2. Establish new parallel path from source S to destination D.
3. Transfer the packets that are stored in buffer at intermediate during link failure.

When the destination node receives the TPRREQ packet, it prepares the TPRREP (reply packet) and increments its current destination number by 1 and forwards the TPRREP packets to the source through the nodes from which it received the TPRREQ packet at first. The source node waits for fixed amount of time for the TPRREP. If it does not arrive on time then it retransmits TPRREQ up to predefined number of times. If the response from destination is not arrived then source declares that the destination is not reachable. If the TPRREP is received then it allocates buffer to all the nodes that take data to destination in order to avoid data losses during link failures due to high mobility of nodes. So, if there is any link failure then the data is stored at the intermediate nodes and after creating a new temporary parallel path then the buffered data is transmitted.

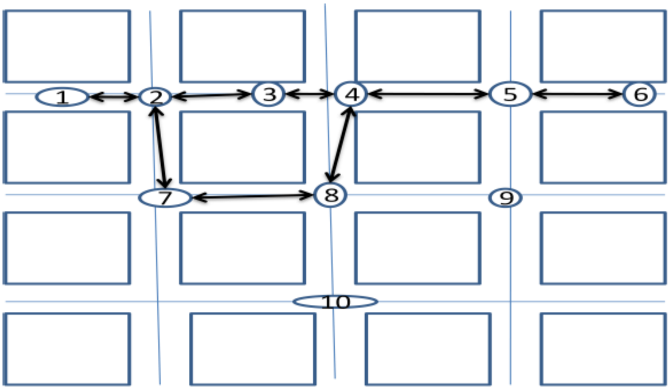


Fig. 4. Manhattan model

Temporary parallel route is calculated as soon as there is a link failure. When there is link failure at the time of transmission a route error (RERR) message is sent to the source node so that the source again retransmits the data. When there are link failures at intermediate nodes our temporary parallel route discovery helps us to find a new temporary parallel path to the destination to transmit the data that is buffered at the intermediate nodes.

We use the following metrics to evaluate the performance of the temporary route recovery mechanism packet delivery ratio, routing overhead and average delay.

4 Conclusion

When there is any link failure during packets transmission the temporary parallel route recovery scheme is introduced. We can get better results in packet delivery ratio, average delay time i.e., the difference between the packet receive time to packet sent time in temporary parallel route recovery scheme when compared to AODV. The routing overhead can be decreased because of maintenance of buffer space at nodes.

References

1. Kothari, A.D., Patel, A.R.: On Demand Temporary Parallel Route Recovery for Frequent Link Failure in Adhoc Networks. International Journal of Computer Applications (0975-8887) 11(11) (December 2010)
2. Perkins, C.E., Belding Royer, E.M., Das, S.: Ad Hoc on Demand Distance Vector (AODV) Routing. IETF Internet Draft, Draft-Ietfmanetaodv-10.txt (March 2002)
3. Abedi, O., Fathy, M., Taghiloo, J.: Enhancing AODV routing protocol using mobility parameters in VANET. IEEE (2008)
4. Kohli, S., Kaur, B., Bindra, S.: A comparative study of Routing Protocols in VANET. In: Proceedings of ISCET (2010) ISBN: 978-81-910304-0-21
5. Perkins, C., Belding-Royer, E., Das, S.: Ad hoc On- Demand Distance Vector routing. RFC 3561 (July 2003)
6. Wu, C.-S., Hu, S.-C., Hsu, C.-S.: Design of Fast Restoration Multipath Routing in VANETs. IEEE (2010)
7. Yuan, Y., Chen, H., Jia, M.: An Optimized Ad-hoc On-demand Multipath Distance Vector (AOMDV) Routing Protocol. In: Asia-Pacific Conference on Communications, Perth, Western Australia, October 3-5 (2005)

Analysis of MIMO Channel Characteristics in Indoor Environment Using Ray Tracing Simulator

Manjusha Karkhanis and Achala Deshmukh

S.C.O.E, Vadgaon Budruk, Pune
manjusha.karkhanis@gmail.com,
achala.deshmukh@gmail.com

Abstract. Multipath propagation is a fundamental requirement for the operation of Multiple Input Multiple Output wireless systems. By using antenna arrays at the transmitter and receiver sides in MIMO systems, very high channel capacity can be achieved, provided the environment has sufficient scattering. Inorder to predict static narrowband/wideband channel characteristics, the ray tracing algorithms are generally used. In this paper channel impulse response was taken from a ray tracing simulator (RPS). This paper investigates the effects on MIMO characteristics in an indoor environment. The effect on channel characteristics with fixed transmitter and moving receivers are studied.

Keywords: MIMO, ray tracing, channel parameters, Rayleigh distribution.

1 Introduction

The need of high speed data wireless systems is increasing day by day. Considering the fact that the frequency spectrum is a scare resource, the future systems should be characterized by enhanced spectral efficiently in order to increase the network capacity. With the expansion of indoor wireless LAN applications, the focus on Multiple Input Multiple Output (MIMO) systems research is growing. The MIMO systems have the advantage of significant bandwidth efficiency in broadband wireless applications. The MIMO systems overcome the drawback caused single antenna systems without additional energy of bandwidth consumption.

Research demonstrates that the effects of temporal variations caused by pedestrian or machinery movement in indoor MIMO channels are significant. The temporal variations affect channel capacity of indoor wireless systems. During propagation, radio waves are mainly affected by three different modes of physical phenomena: reflection, diffraction, and scattering. The paper discusses the effect on channel parameters in an indoor MIMO environment.

When multiple-input multiple-output (MIMO) systems are deployed in suitable rich scattering environments, a significant capacity gain can be observed due to the

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 375–383.
springerlink.com

assurance of multipath propagation [1]. MIMO systems are attractive in environments where multipath tends to create independent channels even with small antenna spacing. A MIMO system takes advantage of the spatial diversity that is obtained by spatially separated antennas in a dense multipath scattering environment. MIMO systems may be implemented in a number of different ways to obtain either a diversity gain to combat signal fading or to obtain a capacity gain. The time varying effects on the propagation channel within populated indoor environments depends on different line of sight (LOS) and no line of sight (NLOS) conditions, and is also related to the particular type of environment considered.

The section 2 of the paper describes the ray tracing methods. Section 3 describes the parameters of mobile fading channel and in section. 4 the simulation results and determination of characteristics based on simulation data are presented.

2 Ray Tracing

While designing an indoor or outdoor communication system there should an idea about the channel parameters that can be expected when the system is operating. We can get the real channel parameters of the system by doing measurements at the location where the system is to be deployed. There are three ways of doing those measurements: direct RF pulse system channel sounding, spread spectrum sliding corealtor channel sounding and frequency domain channel. By solving the Maxwell's equations with the building geometry as boundary conditions, it is theoretically possible to compute the propagation characteristics exactly.

Ray tracing is a powerful method to determine the radio channel characteristics. Especially, for broadband transmission and to determine the angle of incidence, which is becoming more important with the introduction of directional antennas, ray tracing provides useful and realistic results. Ray tracing methods are based on geometrical optics (GO) where the objects have dimensions that are much larger than the wave length and where electromagnetic waves are modelled as rays with flat wavefronts. Rays are followed until they hit an object, where a reflected/transmitted ray is initiated in the next reflection/ transmission depth. The direction of the new ray is determined by Snellius' law.

- Ray tracing has several advantages over doing actual measurements:
- No need for expensive equipment
- Potentially much faster
- No need to clear out the environment
- It is possible to send a real sharp delta pulse

There are some drawbacks compared to measurements:

- Model of the environment is an approximation
- Ray tracing is an approximation

- Computing resources are limited; if a very high accuracy is required (many receivers, low noise measurements), it is possible that the computer cannot finish the computation The various geometric optics based ray tracing techniques are as follows are i) Image model based ray tracing technique, ii) Ray launching based techniques, iii) Ray tube based tracing iv) Frustum ray tracing technique.

The channel impulse response can also be estimated by using ray tracing simulation software's like the I-Prop. Channel sounding by ray tracing simulation is very similar in the way that it is in fact simulating a direct RF pulse system: a transmit antenna sends out one very small pulse (which in this case could be an ideal Dirac delta function, δ (t)), it traces every ray (path) until it's power is below a certain noise level, and during the tracing, all the receiving antennas that the ray will penetrate, record the received pulses (time of arrival, amplitude and phase). For a path k from transmitter to receiver, the received power is

$$P_k = \frac{\alpha P_o}{l^s} \prod_i \sigma_i \tag{1}$$

with α is a function of the antenna patterns, wavelength and initial path direction, P_o is the transmitted power, l is the length of the unfolded path and σ_i is the transmission or reflection coefficient of the ith wall along the path.

3 Channel Parameters

A radio channel can be either static or dynamic. In a static channel, all the transmitters, receivers and all objects in the environment are standing still. While in a dynamic channel, at least one of them is moving. In this chapter, we will first discuss what will happen to the amplitude ($\hat{E}$), the phase (Φ) and the time of arrival (τ_A) when a continuous wave (unmodulated carrier) with frequency is transmitted in a static channel. Finally, we will briefly discuss the additional properties of a dynamic channel

3.1 Amplitude Fading

The received amplitude is of interest since it is directly related to the received power by $P_r = \frac{\hat{E}^2}{2Z_o}$ and the received power has influence on the average bit error rate.

Amplitude fading (or propagation path loss) is the decrease of the amplitude of a continuous wave. This decrease is caused by numerous effects, like the distance between the transmitter and receiver, the objects between them and interference with other waves. The signal is treated as a random process that gives signal strengths with certain probabilities and to analyse that process on a statistical basis.

3.2 Log Distance Model

The Friis free space model can be used if there is a real free space; that is, when the transmitter and receiver have a clear line of sight path between them. In reality, there is almost never a free space environment. Therefore, in the log-distance path loss model, the received power is:

$$P_{r,dBm} = P_{t,dBm} + G_s - 10_n \log d \tag{2}$$

where n=2 for a free space environment, but can be any positive real number for other environments. And $P_{r,dBm}$ and $P_{t,dBm}$ are the received and transmitted power respectively.

3.3 Log Normal Distribution

A plane electromagnetic wave in a dispersive medium decays exponentially. Before a radio wave reaches the receiver, it will have to travelled through different obstructions, like walls, doors and closets. Each obstruction has its own attenuation constant and thickness. If the i^th obstruction has attenuation constant α_i and thickness Δr_i, and a wave that enters this obstruction has amplitude E_{i-1}, then the amplitude of the wave after the obstruction is E_i:

$$E_t = E_{i-1} \exp(-\alpha_i \Delta r_i) \tag{3}$$

If a wave with amplitude E_o would travel through n obstructions, the amplitude would be [5]

$$E_n = E_o \exp(-\sum_{i=1}^{n} \alpha_i \Delta r_i) = E_o \exp(r) \text{ with } x = -\sum_{i=1}^{n} \alpha_i \Delta r_i \tag{4}$$

If the number of obstructions is large enough ($n = \infty$) in equation 4.15, the central limit theorem can be used to state the random variable x has a normal distribution $f_x(x)=p(x)$:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma_x}} \exp(-\frac{1}{2}(\frac{x-\mu_x}{\sigma_x})^2) \tag{5}$$

with μ_x is the mean of x and σ_x is its standard deviation.

The power loss caused by the obstructions is Y equation (6):

$$Y = P_{n,dB} - P_{0.dB} = 10\log_{10} \frac{E_n^2 / Z_o}{E_o^2 / Z_o} = 10\log_{10}\left(\frac{E_n}{E_o}\right)^2 = 20\log_{10}\frac{E_n}{E_o} = 20\log_{10} e \tag{6}$$

The power loss caused by a large amount of obstacles has a lognormal distribution (the power loss in dB has a normal distribution).

3.4 Rayleigh Distribution

The signal that is received on any location is usually the sum of scattered signals. Those signals are caused by reflections of the original transmitted signal by randomly placed obstructions. This phenomenon is called multipath propagation. It is reasonable to assume that the phases of the scattered waves have a uniform distribution from 0 to 2π rad and that the amplitudes and the phases are statistically independent from each other. Therefore, at a certain position, the waves will be in phase and produce a large amplitude (constructive interference) and at an other position, the waves will be out of phase and produce a small amplitude (destructive interference). If the original transmitted signal is an unmodulated carrier with frequency ω_0 and amplitude a, as formulated in eq (7)

$$s = ae^{j\omega_o t} \tag{7}$$

and if n scattered instances of this signal will reach the receiver, then the receiver signal will be the sum of n scaled and phase shifted instances of this signal:

$$s_r = \sum_{i=1}^{n} a_i e^{j(\omega_o t+\theta_i)} = e^{j\omega_o t}\sum_{i=1}^{n} a_i e^{j\theta_i} = e^{j\omega_o t} re^{j\theta} \tag{8}$$

Then the amplitude of the received signal is

$$r = \sqrt{x^2 + y^2} \tag{9}$$

The pdf of the amplitude of the received signal (r) is

$$f_r(r) = p(r) = \int_0^{2\pi} p(r,\theta) = \{ \begin{array}{ll} \frac{r}{\sigma_r^2}\exp(-\frac{r^2}{2\sigma_r^2}) & r \geq 0 \\ 0 & otherwise \end{array} \tag{10}$$

The cdf (cumulative distributive function) of r is the integral of equation 10 [5]:

$$F_r(R) = P(R) = prob(r \leq R) = \int_0^{R} p(r)dr = 1 - \exp(-\frac{R^2}{2\sigma_r^2}) \tag{11}$$

The cdf and pdf of the Rayleigh distribution can be plotted. The Rayleigh distribution is closely related to the central chi-square distribution with two degrees of freedom.

3.5 Time of Arrival

The time of arrival of a pulse depends on the distance that the pulse has to travel between transmitter and receiver, and on the objects that it has to penetrate on its path. A reflection against an object will cause a longer path length and therefore a longer delay. A penetration will also cause a longer delay, since the velocity of the pulse inside the object is less than the velocity of the pulse in vacuum (which is c, the speed of light).

The amount of reflections and penetrations are unknown and therefore, a deterministic approach for the time of arrival is not possible, and the time of arrival has to be analysed statistically. We can model the time of arrival of the ith received pulse as

$$t_i = t_o + \tau_i \, , \quad i \geq 0 \tag{13}$$

where t_o is the delay of the first pulse and τ_i is the excess delay for that pulse. The parameters like mean excess delay, rms delay spread and maximum excess delay are important.

4 Simulation

Ray tracing in an indoor environment was simulated by using I-Prop software. The parameters for the indoor environment were the room plan, thickness of the various walls, positions of transmitters and receivers.

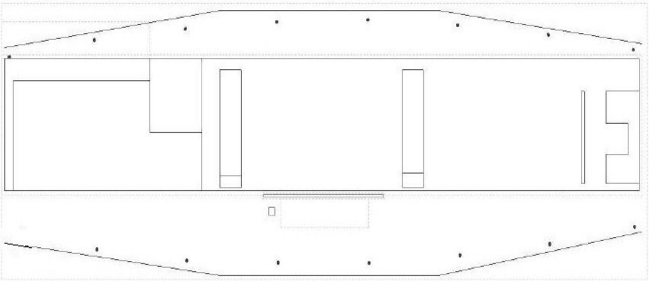


Fig. 1. Simulation Set – up

The parameters of the communication system and environment are as follows:

Table 1. Communication Parameters

Modulation Type	DPSK
Frequency	2.4GHz
Bit Rate [Max,Min]	[12.5Mbits/s,283bits/s]
Bit Error Rate	10^{-3}
Max Transmitted power	10 dBm
Max SNR	33 Db
Receiver Sensitivity Range	[-20,-22, … - 72dBm]
Maximum Speed of Mobile	20km/h
Transmit Antenna	Isotropic, loss less
Receiver Antenna	Isotropic, loss less
Small Scale Model	Rayleigh Distribution
Standard Deviation (for Gaussian)	0.69
Operating Radius	1 t 40m indoor
RMS delay spread	40ns
Coherence Distance	0.0212

All the antennas are assumed to be of fixed height. The transmitters were fixed in location while the locations of the receivers were changed while taking the ray tracing data. The ray tracing instances for 22 reflections were simulated and the data was imported to MATLAB®. The signal above the specified noise level was considered to be valid and processed further.

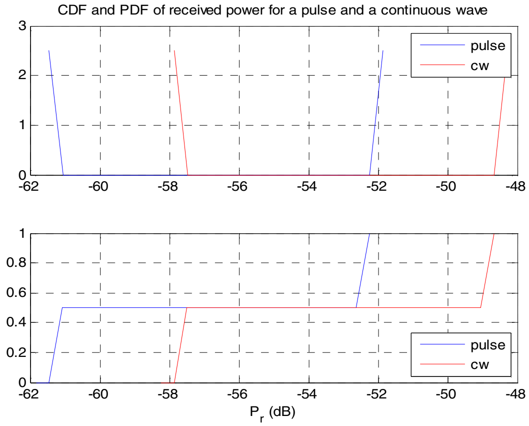


Fig. 2. CDF and PDF for received power

The CDF and PDF plots for the received power in the sample environment are plotted in Fig 2. The results obtained for the received power at different locations are plotted in Fig 3. It can be concluded from Fig 3 that as the distance increases there is a decrease in the received power. The validity of the models was found to be 66.6667% i.e the models are valid on locations were we do not receive 0 Watt power.

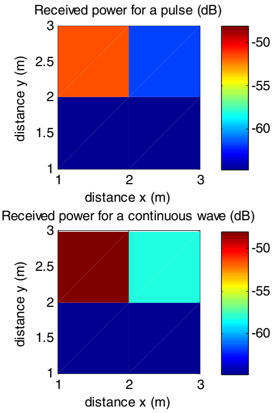


Fig. 3. Received power for a pulse and continuous wave

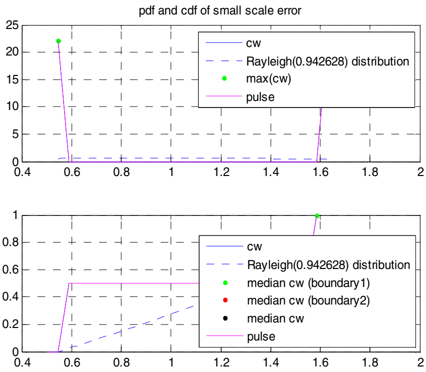


Fig. 4. PDF CDF and QQ plot for pulse and CW with only overall Rayleigh model

Considering small scale model (with Rayleigh distribution)

The overall values for Small scale – normal scale average, converted to amplitude-factors were obtained as:

	Std Deviation	Mean
Pulse	0.798474	1.08684
CW	0.798453	1.08685

The overall values for Small scale – normal scale average,

Small scale – normal scale average, amplitude-factors converted to power, overall values:

	Mean	Small scale path loss
Pulse	0.642714 dB	48.1238
CW	0.642545 dB	48.1254

The standard deviation of Gaussian distributions for Rayleigh was calculated as 0.942628.

5 Conclusion

The ray tracing simulators can be successfully used to specify parameters for a specific environment and to design a wireless communication system. The ray tracing software can be used over the ray tracing algorithms and the actual on site measurements. The channel parameters can be derived and it is seen that though the validity of the model is 66.667% the small scale error of the ensemble of all receivers still fits the Rayleigh distribution quite well (using a log-distance, log-normal model and averaging around every receiver).

References

[1] Suzuki, H.: Accurate and efficient prediction of coverage map in an office environment using frustum ray tracing and in-situ penetration loss measurement. In: Proc. IEEE Veh. Technol. Conf., pp. 236–240 (April 2003)

[2] Suzuki, H., Murray, B., Grancea, A., Shaw, R., Pathikulangara, J., Collings, I.B.: Real-time wideband MIMO demonstrator. In: Proc. 7th Int. Symp. Commun. Inf. Technol., pp. 284–289 (October 2007)

[3] Suzuki, H., Mohan, A.S.: Measurement and prediction of high spatial resolution indoor radio channel characteristic map. IEEE Trans. Veh. Technol. 49(4), 1321–1333 (2000)

[4] Ziri-Castro, K., Scalon, W.G., Evans, N.E.: Prediction of variation in MIMO channel capacity for the populated indoor environment using a radar cross-section-based pedestrian model. IEEE Trans. Wireless Commun. 4(3), 1186–1194 (2005)

[5] Designing wireless indoor radio systems with ray tracing simulators by Admar Schoonen

[6] Rapport, T.S.: Wireless communications principles and practice, 2nd edn. Pearson publication

[7] http://www.awe-communications.com

[8] http://i-prop.cz

Efficient Target Recovery in Wireless Sensor Network

Shailaja Patil, Ashish Gupta, and Mukesh Zaveri

Department of Computer Engineering,
Sardar Vallabhbhai National Institute of Technology, Surat, India
{p.shailaja,p10co991,mazaveri}@coed.svnit.ac.in

Abstract. In this paper, we present fast and efficient target recovery algorithm for a distributed wireless sensor network with dynamic clustering. As sensor nodes have limited power, the nodes performing frequent computation and communication have problem of battery exhaustion, causing failure in participation of tracking. Also, nodes may fail due to physical destruction. These reasons of node failure may result in loss of target during tracking. Therefore, we propose an efficient detection of lost target and recovery (DLTR) algorithm to recover lost target using the Kalman filter. From the simulation results, it is evident that, the proposed recovery algorithm outperforms existing algorithm in literature.

Keywords: Recovery, Tracking, Kalman filter, Wireless Sensor Network.

1 Introduction

Wireless Sensor Network (WSN) consists of small size sensor nodes capable of collecting, processing and transferring information. A sensor node consists of microcontroller, memories, a radio transceiver and sensor(s). All of these components are powered by battery. All the nodes are deployed randomly on the field and form network through self organization. Target tracking in WSN has been successfully implemented in applications like aqueous surveillance system for drinking water reservoir [1], wildlife habitat observation system [2], in-house person tracking [3] etc.

The target tracking suffers from loss of target due to depletion of nodes, which may occur in different situations like node failure, communication failure, localization error, prediction error etc. Nodes can fail due to various reasons such as exhaustion of battery, physically destroying the node by enemy in the battlefield. Similarly, during deployment, if nodes fall in water body like pond, these may fail if nodes are not water resistant. Due to this, a region gets developed inside the network, where target's movements are not observable. This may result in loss of target, as it goes undetected in the region of failed nodes. This could be a temporary loss as the target may still be present inside the network.

For tracking the target, nodes need to be always in active mode inside the network. However, in such case their battery may drain out early resulting in

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 385–394.
springerlink.com

node failure, consequently decreasing the network lifetime. The clustered architecture facilitates distributed operation of the network, and can handle this issue. In such an architecture, the Cluster Head (CH) is responsible for computation of future location of the target, selection of suitable nodes (nearer to the predicted location) to track the target. If the target enters into the next cluster, its information need to be handed over to the next CH. Similarly, the location computation of the target is also performed by CH in two different cases, 1. initially when it is detected in the network and 2. on its recovery. Thus, to perform all these tasks a lot of communication is required. If the CH is fixed as in static clustering, it is heavily loaded and probability of failure of CH is more than remaining nodes in the cluster. Thus, failing the CH prohibits participation of its corresponding cluster members in tracking. In dynamic clustering, one of the nodes with highest energy and connectivity can become the cluster head(CH) and can control the cluster members in tracking. The role of CH may be performed by every node of the network in rotation[4]. Hence we have employed a dynamic clustering in the algorithm. Our goal is to develop an energy efficient, and quick recovery mechanism, which can predict the next location of target with more accuracy, and prolong the network lifetime.

In our earlier work [5], we have used static clustering for recovery of lost target. In this paper, we propose a recovery mechanism using distributed network to find the lost target in WSN. As mentioned earlier, the algorithm uses a dynamic clustered architecture, where CHs are selected in rotation according to their energy, as described in LEACH [4]. We are using the energy model as explained in [6] for computing energy consumption.

The rest of paper is organized as follows: In Section 2, we present related work of target tracking and existing recovery algorithms described in literature. The proposed algorithm is presented in Section 3. The result and performance comparison with existing algorithm is presented in Section 4. Conclusion is presented in Section 5.

2 Related Work

In WSN tracking, three components are to be considered- 1. ranging technique 2. position estimation. 3. position tracking. We shall review literature for these components in this subsection. There are various ranging techniques mentioned in the literature such as Received Signal Strength(RSS) [7], Time of Arrival(TOA) or Time Difference of Arrival(TDOA)[8], Angle of Arrival(AOA) [9] etc. However, the RSS technique is preferred over remaining, as it does not need any additional hardware. The target tracking begins with detection of the target in the network. As soon as it intrudes the network, its current location need to be estimated. There are various techniques of position estimation like trilateration, triangulation, multilateration etc. [10]. We are using trilateration for position estimation, as it needs least number of nodes(3) to locate. In position tracking, there are several tracking mechanisms devised in literature based on underlying techniques such as tree based STUN[11], cluster based DELTA [12].

The prediction based tracking use various prediction techniques as explained in [13,14,15,16]. In spite of having an efficient tracking mechanism, it is badly affected in node depleted region as mentioned earlier. Few techniques have been devised to circumvent this issue. For example, a cluster based Distributed Tracking Protocol (DTP) is presented in [13], which deals with the recovery of the lost target. The protocol describes the level based recovery for capturing the lost target. The first level recovery is initiated when the detection of loss of target occurs. In this, previous target tracking sensor-triplet switches to high beam from normal beam. If target is recovered then the normal tracking mechanism is followed otherwise second level recovery is initiated. In second level of recovery, the nodes which are at a distance of p meters from the last node are activated otherwise N^{th} level of recovery is initiated. In the N^{th} level of recovery, a group of sensor nodes that are $(2N\text{-}3)p$ meters away are activated to detect the target. However, this method does not consider past information of object motion, which leads to participation of more number of nodes for recovery, which not only increases energy consumptions, but also reduces the network lifetime.

Another recovery mechanism is reported in [6], where authors have simulated tracking for different motion features using current measurements or past records of object's motion. Prediction of target's next location is done by integrating its current information of location, velocity, and motion direction. In case of failure of prediction, the loss of target occurs. Now, the hierarchical recovery process in the network starts, considering the past record of motion of target. Authors in [17] have presented a foolproof recovery mechanism in target tracking using static clustered network. This protocol consists of four phases- Declaration of lost target, search, recovery and sleep. When target is missing, current CH declares loss of target, which initiates the recovery mechanism. The search phase is introduced to avoid false initiation of recovery mechanism, where the current CH waits for stipulated time to receive acknowledgement from downstream clusters. The actual recovery phase consists of N-level recovery. The first level recovery is initiated by waking up single hop clusters around the last location of target. If it is not captured here, all its two hop clusters are woken up, and this process is continued till the target is found. In the last phase, on successful recovery of the target, the CH and cluster members that are participating currently in tracking remain awake and rest go to sleep mode. However, in this algorithm when loss of target occurs, current CH wakes up all single or double hop cluster nodes as needed. Unnecessary waking up of large number of nodes wastes energy reducing the network lifetime. To improve this, we propose energy efficient and fast recovery algorithm using Kalman filter. The advantage of using Kalman filter is its prediction accuracy. Using this filter, the CH precisely selects a cluster to track the target.

3 Proposed Recovery Algorithm

The proposed DLTR algorithm has following assumptions at the time of deployment: a) sensor nodes know their location, b) sensing range and battery power is same for all sensor nodes, c)sensor nodes work in two modes, sleep and active (awake). d) boundary nodes are active all the time. For cluster formation the connectivity is considered. The CHs have information about its member nodes and neighbour CHs. The following subsection explains algorithm of lost target recovery during tracking.

3.1 Target Tracking

The first step in tracking is localizing the target. As soon as it enters into the network, and is detected by minimum three nodes, localization is performed using trilateration. Tracking is initiated with the current location of the target, using Kalman filter. Our recovery mechanism considers node failure issue for detecting lost target.

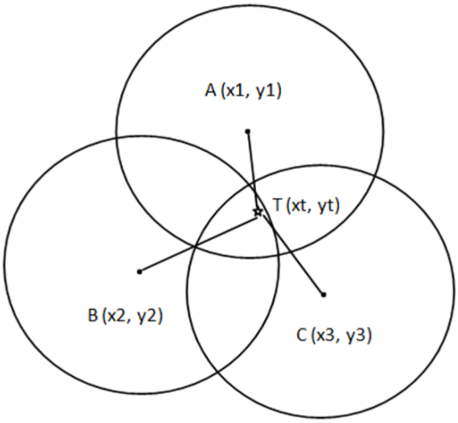


Fig. 1. Trilateration Method

Detection and Localization. Boundary nodes detect target and perform trilateration to find out location of target. In Figure 1 , let A, B and C are sensor nodes and T is target. Distances of target from A, B and C are d_1, d_2 and d_3 respectively. These distances can be obtained by received signal strength indicator (RSSI). Therefore, three equations can be obtained for these distances.

$$d_1^2 = (x_1 - x_t)^2 + (y_1 - y_t)^2 \quad (1)$$

$$d_2^2 = (x_2 - x_t)^2 + (y_2 - y_t)^2 \quad (2)$$

$$d_3^2 = (x_3 - x_t)^2 + (y_3 - y_t)^2 \quad (3)$$

By solving these equations, we get location of target (x_t,y_t). The boundary node which is nearest to the target sends an alert to the CH, which acts as current CH.

Prediction by CH. Current CH predicts next location of target using kalman filter [18]. The steps of kalman filter are illustrated in brief in this subsection. The Kalman filter works in two phases: Updation and prediction.

Let for each time-step (t), x_t be the state vector, F_t the state-transition model, H_t the observation model, Q_t the covariance of the process noise, R_t the covariance of the observation noise, and B_t the control-input model. The state and observation models can be expressed as-

$$x_t = (F_t x_{t-1}) + B_t u_t + w_t \tag{4}$$

$$z_t = (H_t x_{t-1}) + v_t \tag{5}$$

Prediction step: For the prediction phase, the state estimate from the previous step is used for producing an estimate of the current step state. The equations used in predict step are as under:

$$\hat{x}_{t|t-1} = F_t \hat{x}_{t-1|t-1} + B_t u_t + w_t \tag{6}$$

$$P_t = F_t P_{t-1|t-1} F_t^T + Q_t \tag{7}$$

Update step : In the update phase, state estimate is refined using the current predicted state estimate and current observation. Such an estimate is known as the a-posteriori state estimate. Predicted (a-priori) state estimate is expressed in equation 6 and Predicted (a priori) estimate covariance is as per equation7. The equations used in update step are as given in equation 8, 9, and 10.

$$K_t = P_{t|t-1} H_t^T S_t^{-1} \tag{8}$$

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t \tilde{y_k} \tag{9}$$

$$P_t = (I - K_t H_t) P_{t|t} \tag{10}$$

The update step incorporates the observation and predicts the next location.

Activation of CH. If predicted location lie inside the region of current cluster, then wake up message is sent to that member node who is nearest to the predicted location. Otherwise current CH finds next CH, and sends alert message for switching the cluster and handover tracking task. If selected next CH is not failed then it will send acknowledgement and start tracking. However, if it is failed then current CH will not receive any acknowledgement within stipulated time, then recovery mechanism is initiated.

3.2 Recovery Mechanism

As soon as the target is found to be missing, its presence is checked inside the present cluster. If it is not detected, then the recovery mechanism is required. The detection of lost target is performed in two steps, declaration of lost target, and recovery. On recovery of the target, all active nodes go to sleep mode, except current cluster. The three phases of recovery algorithm are explained as under:

Declaration. If acknowledgement is not received by current CH in stipulated time, it declares the loss of target, and the recovery process is initiated. It is same as discussed in [17].

Recovery. Declaration of loss of target requires time, meanwhile target is also moving. The current CH performs following steps

1. Predict next probable location as few steps ahead based on previous location and speed.
2. Wake up all the cluster members to search if target is present inside the cluster.
3. If target is found here, then continue tracking;
   Else
   a. Find all single hop clusters (not failed) surrounding the predicted location;
   b. Using the predicted location find the specific cluster, where target may be present and send wake up message to the corresponding CH. This CH now performs following tasks-
      - Send acknowledgement to the current CH;
      - Wake up cluster members to localize target;
      - If target is found;
        - Then become current CH;
        - Localize target with the help of nearest three cluster members;
        - Continue tracking;
      - Else Send message to current CH about target's unavailability;
4. Now wake up next two hop clusters surrounding the predicted location; If still not found then wake up three hop members;
5. When target is found, that corresponding cluster becomes current CH and continue tracking;

Sleep. Once target is recovered, only current tracking cluster remains in wake up mode and all other clusters go to sleep mode. This phase is same as in [17].

A Significant amount of energy and time is saved in recovery phase as least number of nodes are woken up for recovery. Thus, in proposed recovery algorithm, only one single hop cluster wakes up at a time which is nearest to predicted location. In this manner, less number of nodes is required to be in wake up state to recover the target. Algorithm is fast because there is no need to send wake up message to all single hop or multi-hop clusters. The amount of energy, and time required for recovery is mentioned in the next section.

4 Performance Results

In this section, the performance of the recovery mechanism is discussed with different trajectories. The results are compared with existing algorithm [17]. The performance metric used for comparison are recovery time, nodes awaken and energy consumption. We follow the energy model of [6]. Main parts of a wireless node are microcontroller unit(MCU), sensing unit, and transceiver. The typical values of power consumption by MCU for active and sleep mode are 360mW and 0.9 mW respectively. The sensing unit and transceiver's consumption are about 23mW and 720mW respectively. Here we assume that same amount of energy is consumed by transmitter and receiver unit[4].

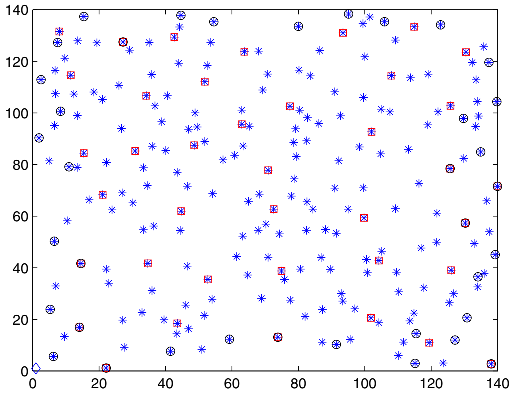


Fig. 2. Wireless Sensor Network for Tracking

Figure 2 shows the WSN used for tracking consisting of randomly deployed 225 nodes in [140mx140m] area with radio range of 15m. The Gaussian noise of 5 % is added to the radio range. The average node degree of the node in network is seven. These nodes are logically divided into three types - member nodes, boundary nodes and CH. The member nodes are represented with asterisk $*$, CHs are represented by $\boxast$, and $\circledast$ are boundary nodes.

Figure 3 shows the target tracking with existing approach and proposed approach. In this figure, target enters from location (1,1) and leaves at location (43,139). In this scenario, total seven numbers of nodes are failed in entire field. Thus,the target is lost and recovered twice due to four and three failed nodes respectively at different locations. The table 1 shows a comparison of existing and proposed approach in terms of nodes woken up and time required for recovery. The recovery time and total number of nodes awaken is less than the existing approach. The total energy consumed in tracking with recovery is 2.79J per node with existing approach whereas, it is about 1.5J per node with proposed approach. The loss of target is simulated with another trajectory as shown in figure 4. In this scenario, there are two regions depleted of nodes, with eight and six failed nodes. The target moves through first region twice, hence it is lost twice. With the second failed nodes region, loss of target occurs thrice. Here target enters the network from location (1,120) and leaves network at (140,27). The table 2 shows the comparison between existing and proposed recovery. The

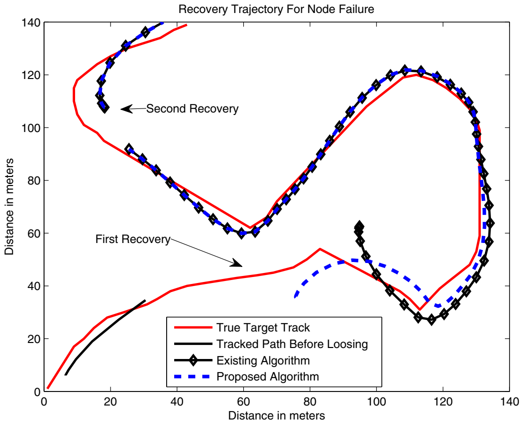


Fig. 3. Recovery of lost target with 7 failed nodes

Table 1. Comparison between Existing and Proposed Algorithm

Algorithm	1^{st} Recovery		2^{nd} Recovery	
	Nodes awaken	Time(s)	Nodes awaken	Time(s)
Existing Recovery	88	7.9397	38	3.6205
Proposed Recovery	23	5.2396	17	3.1994

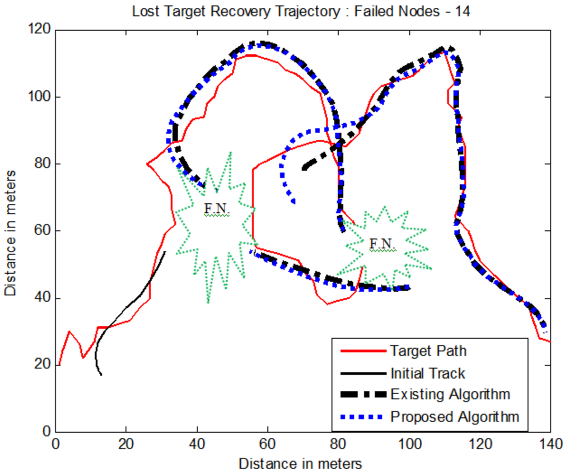


Fig. 4. Recovery of lost target with 14 failed nodes

Table 2. Comparison between Existing and Proposed Algorithm

Algorithm	1^{st} and 2^{nd}Recovery		3^{rd} Recovery	
	Nodes awaken	Time(s)	Nodes awaken	Time(s)
Existing Recovery	112	6.3095	71	6.1540
Proposed Recovery	21	3.1871	18	3.1708

total amount of energy spent per node for tracking with recovery(all three cases) is about 3.92J with existing and 2.13J with proposed approach. All the above tables and figures show that, the proposed approach outperforms the existing in terms of recovery time, energy consumed and total number of nodes awaken.

5 Conclusion

In this paper, we have proposed energy efficient and fast target recovery algorithm. In target tracking, our algorithm uses trilateration method for localization and kalman filter for prediction. For observing recovery of lost target during tracking, we have handled the issue of node failure. Different trajectories have been used for simulation. The proposed recovery approach has been compared with existing approach in terms of time and number of nodes required to be active during recovery of lost target. Similarly total energy consumed per node in the network for tracking with recovery is computed. Comparison of results show that proposed algorithm outperforms the existing algorithm. Further, we shall investigate the performance of algorithm with non-linear filters for other causes of losing target.

References

1. Yang, X., Ong, K.G., Dreschel, W.R., Zeng, K., Mungle, C., Grimes, C.A.: Design of a wireless sensor network for long-term, insitu monitoring of an aqueous environment. Sensors 2, 436–472 (2002)
2. Alan, M., David, C., Joseph, P., Robert, S., John, A.: Wireless sensor networks for habitat monitoring. In: Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications, pp. 88–97. ACM, New York (2002)
3. Bal, M., Xue, H., Shen, W., Ghenniwa, H.: A 3-d indoor location tracking and visualization system based on wireless sensor networks. In: IEEE International Conference on Systems Man and Cybernetics, pp. 1584–1590 (2010)
4. Rabiner, H.W., Anantha, C., Hari, B.: Energy-efficient communication protocol for wireless microsensor networks. In: Proceedings of the 33rd Hawaii International Conference on System Sciences, vol. 8, pp. 8020–8030. IEEE Computer Society, Washington, DC (2000)
5. Gupta, A., Patil, S., Zaveri, M.: Lost target recovery in wireless sensor network using tracking. In: International Conference on Communication Systems and Network Technologies, CSNT. IEEE (accpted, 2012)
6. Xu, Y., Winter, J., Lee, W.C.: Prediction-based strategies for energy saving in object tracking sensor networks. In: Proceedings of 2004 IEEE International Conference on Mobile Data Management, pp. 346–357 (2004)

7. Li, X., Shi, H., Shang, Y.: A sorted rssi quantization based algorithm for sensor network localization. In: Proceedings of 11th International Conference on Parallel and Distributed Systems, vol. 1, pp. 557–563 (2005)
8. Xiao, J., Ren, L., Tan, J.: Research of tdoa based self-localization approach in wireless sensor network. In: International Conference on Intelligent Robots and Systems, pp. 2035–2040 (2006)
9. Kuakowski, P., Vales-Alonso, J., Egea-Lpez, E., Ludwin, W., Garca-Haro, J.: Angle of arrival localization based on antenna arrays for wireless sensor networks. Computers and Electrical Engineering 36, 1181–1186 (2010)
10. Srinivasan, A., Wu, J.: Encyclopedia of Wireless and Mobile Communications, pp. 1–26. Taylor and Francis Group, CRC Press (2008)
11. Kung, H.T., Vlah, D.: Efficient location tracking using sensor networks. Wireless Communications and Networking 3, 1954–1961 (2003)
12. Wälchli, M., Skoczylas, P., Meer, M., Braun, T.: Distributed Event Localization and Tracking with Wireless Sensors. In: Boavida, F., Monteiro, E., Mascolo, S., Koucheryavy, Y. (eds.) WWIC 2007. LNCS, vol. 4517, pp. 247–258. Springer, Heidelberg (2007)
13. Yang, H., Sikdar, B.: A protocol for tracking mobile targets using sensor networks. In: IEEE International Workshop on Sensor Network Protocols and Applications, pp. 71–81 (2003)
14. Di, M., Joo, E.M., Beng, L.H.: A comprehensive study of kalman filter and extended kalman filter for target tracking in wireless sensor networks. In: IEEE International Conference on Systems, Man and Cybernetics, pp. 2792–2797 (2008)
15. Zhi-Jun, Y., Shao-Long, D., Jian-Ming, W., Tao, X., Hai-Tao, L.: Neural network aided unscented kalman filter for maneuvering target tracking in distributed acoustic sensor networks. In: Proceedings of the International Conference on Computing: Theory and Applications, ICCTA 2007, pp. 245–249. IEEE Computer Society, Washington, DC (2007)
16. Vemula, M., Miguez, J., Artes-Rodriguez, A.: A sequential monte carlo method for target tracking in an asynchronous wireless sensor network. In: 4th Workshop on Positioning, Navigation and Communication, pp. 49–54 (2007)
17. Khare, A., Sivalingam, K.M.: On recovery of lost targets in a cluster-based wireless sensor network. In: Ninth Annual IEEE International Conference on Pervasive Computing and Communications, Seattle, WA, USA, pp. 208–213 (2011)
18. Haykin, S.: Adaptive filter theory, 3rd edn. Prentice-Hall, Inc., Upper Saddle River (1996)

Generic Middleware Architecture Supporting Heterogeneous Sensors Management for Any Smart System

Soma Bandyopadhyay and Abhijan Bhattacharyya

Innovation Lab, Tata Consultancy Services,
Kolkata, India
{soma.bandyopadhyay,abhijan.bhattacharyya}@tcs.com

Abstract. The ever increasing demand to develop smart environments starting from small scale environment like smart home to more complex one like smart city, extensively require complex middleware to support interoperation among various diverse domains of applications and different heterogeneous sensors. The middleware is also responsible for providing abstractions to the application interfaces and device sensing. In this paper generic middleware architecture which supports modularity, sensor observations, and diverse sensor management and provides abstraction to sensors and applications is presented. Implementation of the proposed middleware architecture is described by smart irrigation and firming environment use case. The architecture presented here is solely based on object oriented concept and this can be further extended for any smart system. A future research scope of the proposed architecture is also discussed here.

Keywords: Ubiquitous Computing, Heterogeneous sensors, Middleware, Smart environment, Adaptation, Generic framework.

1 Introduction

A smart environment essentially comprises of sensing devices to capture the information of the environment and the actuators to function as per the obtained command. The services and applications require invoking the appropriate sensors to get the information on the environment, and to act as per the situation intelligently. The sensors and the applications/services are heterogeneous in nature and reside in a distributed environment. Therefore sensing and processing the raw sensed data and tunneling that data to the appropriate services and applications in appropriate formats are essential for building smart system. This is a fundamental requirement across all smart systems.

A middleware is required in order to achieve this requirement. It primarily acts as a bond across the heterogeneous sensors and heterogeneous applications/services. Its prime objective is to provide a homogeneous interface hiding diverse properties of heterogeneous sensors as well as applications/services.

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 395–404.
springerlink.com

This article focuses on design and development of a generic object oriented middleware architecture, supporting the following properties: 1) device discovery of certified devices, 2) device management of heterogeneous sensors, 3) resolution of semantics and syntaxes while interoperating among heterogeneous sensors, 4) extraction of context from the sensed data, 5) handling critical data based on the real-time requirements of the event to be processed and the state of the sensors, 6) creation of configuration schemas based on XML, 7) creation of device functionality schemas based on XML and, importantly, 8) providing an application adaptation interface. The main functional components are based on the functional components as proposed in [1] for IoT (Internet of Things) middleware.

The proposed architecture, as presented here, has a layered architecture with two abstraction layers. The architecture is modular based on an object oriented model supporting various generic modules like devices, certified- sensors, sensor_adaptor etc. with clearly defined inter-module relations. The abstraction layers are flexible enough to adapt new sensors or applications/services without requiring any modifications to the core components of the middleware. The new objects inherited from the existing ones are needed to be added to the framework and corresponding modifications are to be made in the XML schemas. Thus the proposed architecture can be used as a framework for developing a middleware capable to be extended to support any smart system as it addresses the generic requirement of these systems and does not require modifications inside the core components to support diverse configurations. Only the XML schemas specific to device configuration and functionality, its class path for dynamical execution are to be modified for this purpose.

The case study, as presented here, consists of a smart irrigation and agriculture system. Device interoperation is shown based on Zigbee and Ethernet based sensing devices. The class diagram depicts the relationship among the various objects related to this use case.

The remainder of this article is organized as below. First, the related work in middleware architecture and its review is presented followed by an overview of the proposed system. The architecture of the system along with the class diagram is then described. The next section describes a practical application of the middleware in a smart irrigation and farming system. The final section concludes this article with the future research scope on the proposed architecture.

2 Related Work

The middleware functionalities are widely studied to address the important aspects and to address various challenges of setting up smart environment in the domain of ubiquitous computing as well as IoT. The role of middleware in IoT is studied extensively and the various functional components of the middleware system are proposed in [2]. This survey paper concludes the open issues and challenges about the middleware. In [3], an interesting research project called ‘Munich’ (Mobile Users in a Non-intrusive Computing Hierarchy) on subjective sensing to meet the user needs intelligently and to support personalized services based on mobile phones with a layered architecture is presented.

In [4], an energy efficient mobile sensing system having a hierarchical sensing management scheme for setting up smart environment is presented. Here XML is used to define user state and user state transition rules, and these XML based state descriptors are used as inputs to turn sensors on and off.

In [5], a declarative model for Sensor Interface Descriptors (SID) based on OGC's (Open Geospatial Consortium)[9] SensorML[13] standard is presented to close the gap in standardization for the connection between the SWE (Sensor Web Enablement) and the underlying sensor layer with heterogeneous protocols. Here a generic SID interpreter capable of connecting sensors to Sensor Observation Services and Sensor Planning Services based on their SID has been developed. The system has two components, the SID interpreter and the SID interface. The SID interpreter runs on the data acquisition system and uses SID instances for the different sensors of the sensor network to translate between the sensor specific protocol and the SWE protocols. The interpreter is responsible to register a sensor at a SWE service and to upload sensor data to an SOS. Also, it is responsible for the opposite communication direction and forwards tasks received by an SPS to a sensor.

In [6], a middleware system 'LinkSmart' is proposed. This middleware combines semantic web services technology with SOA-based principles. It provides a mechanism for wrapping standard API interfaces of sensors and various physical devices with a defined web service extension, which is enhanced by a semantic description of provided or generated WSDL (Web Services Description Language) [12] files thereby connecting the devices and their local networks to the outside world through broadband and/or wireless networks.

Ref. [7] highlights the major issues in designing the middleware for WSN like heterogeneity, complex event processing, QoS support and access prioritization, etc. It divides the state-of-the-art middlewares into some distinguished categories like Interoperability Middleware, Web Service Enabler, Semantic Sensor Web Middleware and Information Processing Middleware based on the application usage. It proposes a middleware design to address the said challenges it has identified It proposes some generic layers as follow: connectivity layer, knowledge layer to form a knowledge-base for device management, semantic metadata, etc., information processing layer to handle the incoming queries and service provisioning layer to addresses the gap between traditional high-level application protocols and the underlying layers.

However, these research approaches do not make use of any generic middleware architecture which can be extendable to both fixed and mobile platforms to build any smart system. Importantly they do not specify about both the sensors as well as application abstractions and an object oriented based modular architecture which can be used as framework for building middleware for any smart system starting form interoperating among diverse sensors to post sensed data to application/services as per the user needs.

3 System Overview

The generic middleware for smart system proposed here consists of a key control layer shown as SMART Controller and two adaptation layers, sensor abstraction and application abstraction. The layered architecture is depicted in Fig. 1.

Fundamentally, the middleware is responsible to facilitate the exchange of information between the remote applications and the sensors embedded in the smart environment. The remote application connects to the middleware over the Internet. The middleware receives the information about the smart environment from the sensing devices and in turn provides this information to the remote applications based on the application requirements.

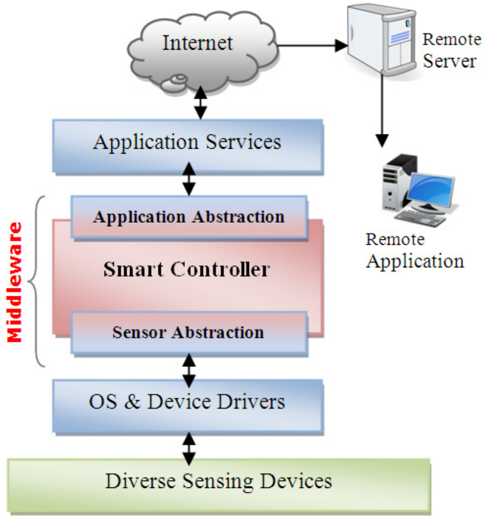


Fig. 1. System Overview

The middleware sits on top of the ‘OS (Operating System) & device drivers’ layer and below the ‘application services’ layer. The ‘device drivers’ layer takes care of the underlying protocols for communication between the middleware system and the sensing devices. The ‘application services’ layer runs the local services for the remote applications.

The heart of the middleware system, the ‘smart controller’, which handles all the core functionalities to communicate between the sensing devices and the remote applications as well as performs different device management activities, is responsible for all the operations like discovering the sensing devices, performing syntactic and semantic resolutions, make the raw data presentable to the web interface, performing device monitoring, posting the data to the application abstraction layer etc.

The smart controller interacts with two abstraction layers of the middleware. At the bottom is the ‘sensor abstraction layer’ and at the top is the ‘application abstraction layer’. These interfaces play a vital role to equip the system with generic interfaces.

The bottom sensor abstraction layer provides a general abstraction with respect to the diverse sensors so that the heterogeneous sensors can be interacted with a common format. Similarly the top level application abstraction layer provides a common interface to interact with the heterogeneous applications.

Thus these two abstraction layers make the middleware adaptive to any smart environment.

4 Architecture of Proposed System

The proposed architecture of the middleware for smart system is represented in the present section. A block diagram of the system architecture is depicted in Fig 2.

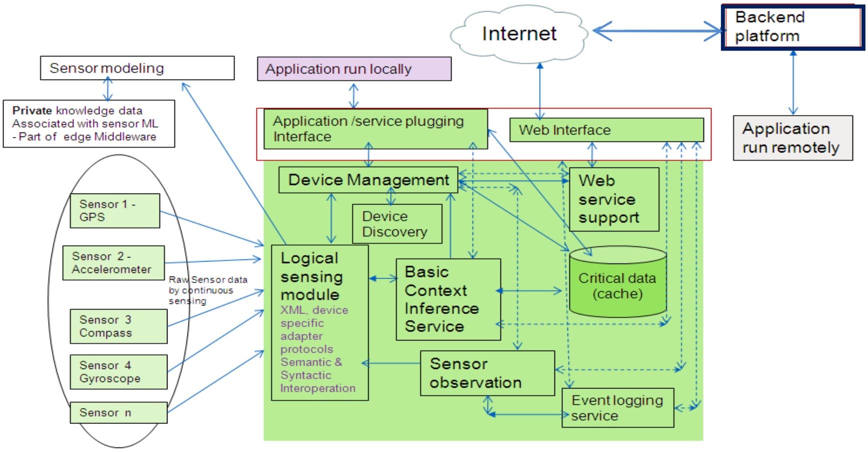


Fig. 2. System Architecture

The core part of sensor abstraction layer, the logical sensing module collects data from various diverse sensors like location, acceleration, angular velocity, temperature, humidity, light etc. using diverse communications and messaging protocols following different standards. This block is responsible to handle the interoperation issues, resolving the syntax and semantics of the messaging protocols by using the specific sensor adaptation protocols. It uses specific sensor descriptions/modeling standard like SensorML. This sub module fundamentally encapsulates all the details of the sensor interoperation and forms sensor abstraction layer.

The device management module resides in the smart controller layer. It is responsible for managing all active sensing devices by activating its different sub – modules and using associated handshaking messages with them. It controls posting of the sensor data to the application abstraction layer. It consists of sub components like device discovery which scans the environment through the available interfaces and communication protocols in regular intervals to detect the new sensing devices. It does a capability negotiation based on the various attributes like interface, protocol etc. of the sensor devices. Device discovery sub module executes SCAN command periodically. It broadcasts a handshake message to all the sensors, and accepts responses from the active sensing devices. It uses the device adapters and the associated protocols for those sensing devices made available during system configuration. Device management module maintains a device management table for the sensing devices having device-type; manufacturer identifier, Interface type (Zigbee, Bluetooth), protocol identifier and the associated service identifier. Some of the device specific information is taken from the configuration file like device type, interface type, and some of the information are generated dynamically like unique device/sensor identifier during device discovery and are maintained in the device table. It collects the inputs related to basic device properties from the XML based configuration schemas. A state flag is maintained by the device manager in the device management table depending on the response received from the sensing device during the periodic scanning. Fig. 3 illustrates the state diagram for the sensors. Respective events are triggered as per the application needs and a reduced set of sensors is generated in each scan depending on the sensor state idle or deep-sleep. The device management table also maintains a service identifier associated with the device. The other two important sub blocks of the smart controller layer are sensor-observation and context inference service which closely interact with device management module. The major activity of these two blocks is to collect data from sensors.

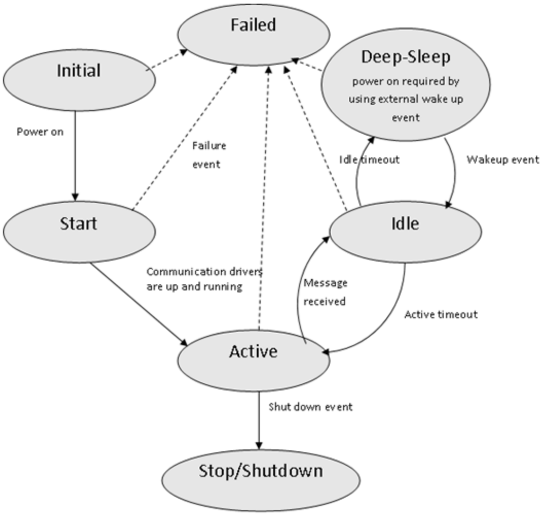
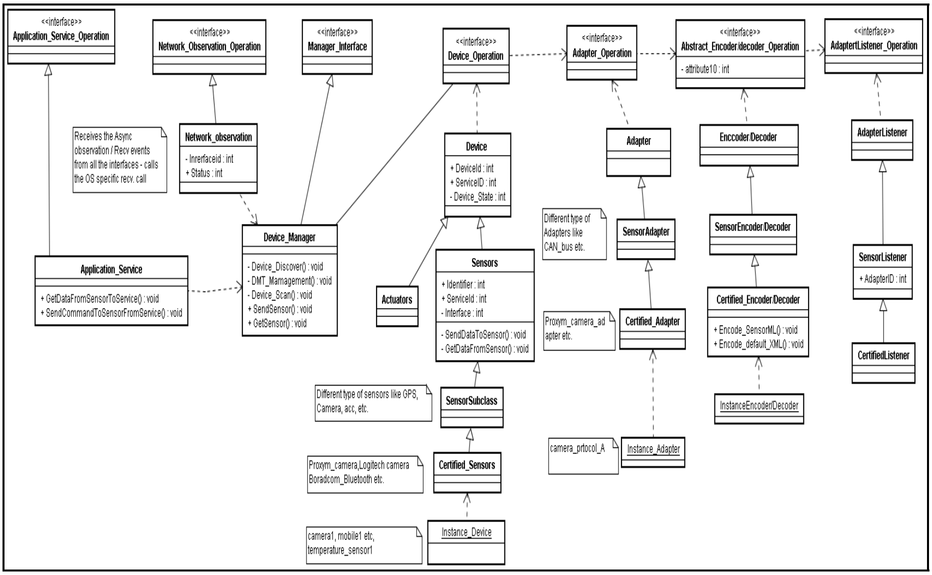


Fig. 3. State Diagram of Sensors

Sensor observation module interfaces with the logical sensing module to extract and post the data to the application. It gets an event from the device manager to start the posting of data. The context inference service block is part of smart controller layer. It retrieves contextual information both in synchronous and asynchronous mode after getting a triggering event from the device management module, which defines a structure for contextual data using the device properties defined in XML, and the demand of application/services. A tiny storage space is used as critical data cache to keep the critical data, mainly the contextual data, depending on the real-time requirement of the processing of that data as well as event identifier with applications/services – sensing-device mapping. Event logging module is interfaced with the sensor observation module. It keeps on posting of sensor observations as per the application's demand through web interface. It uses publish-subscribe model.

The applications abstraction layer consists of the web interface and application plug-in interface. The application plug-in interface provides mechanisms to register applications running locally on top of the middleware layer. Web interface supports the interfacing with the remote applications. Both these sub modules interact with the device management layer to post their demand as well as get the notifications related to the sensing devices from the smart controller layer via the device management module. Different OGC (Open Geo spatial Consortium) [9] based services can be used as a remote service and its counterparts as local applications.

Fig. 4 depicts the class diagram of the object oriented model used for designing the architecture. The various modules and their relationship are depicted in this figure. As an example it can be observed, device-operation, device-manager, devices, sensors, actuators, adapter, sensor-adapter are various modules/classes or interfaces where sensors and actuators are inherited from the device class, sensor-adapter is inherited from adapter class etc.



<<interface>>
Application_Service_Operation
<<interface>>
Network_Observation_Operation
<<interface>>
Manager_Interface
<<interface>>
Device_Operation
<<interface>>
Adapter_Operation
<<interface>>
Abstract_Encoder/decoder_Operation
- attribute10 : int
<<interface>>
AdaptertListener_Operation
Receives the Async observation / Recv events from all the interfaces - calls the OS specific recv. call
Network_observation
- Inrerfaceid : int
+ Status : int
Device
+ DeviceId : int
+ ServiceID : int
- Device_State : int
Adapter
Enccoder/Decoder
AdapterListener
Different type of Adapters like CAN_bus etc.
SensorAdapter
SensorEncoder/Decoder
SensorListener
+ AdapterID : int
Device_Manager
- Device_Discover() : void
- DMT_Mamagement() : void
- Device_Scan() : void
+ SendSensor() : void
+ GetSensor() : void
Application_Service
+ GetDataFromSensorToService() : void
+ SendCommandToSensorFromService() : void
Actuators
Sensors
+ Identifier : int
+ ServiceId : int
- Interface : int
- SendDataToSensor() : void
- GetDataFromSensor() : void
Proxym_camera_adapter etc.
Certified_Adapter
Certified_Encoder/Decoder
+ Encode_SensorML() : void
+ Encode_default_XML() : void
CertifiedListener
Different type of sensors like GPS, Camera, acc, etc.
SensorSubclass
InstanceEncoder/Decoder
camera_prtocol_A
Instance_Adapter
Proxym_camera,Logitech camera Boradcom_Bluetooth etc.
Certified_Sensors
camera1, mobile1 etc, temperature_sensor1
Instance_Device

Fig. 4. Class Diagram

5 Use Case: Smart Irrigation and Farming

This section describes the application of the proposed middleware in the context of a smart irrigation and farming system as a smart environment [10].

The smart irrigation and farming system 'senses' the different parameters from the environment using a composition of remote and in-situ sensors. The different heterogeneous sensors are used to measure temperature, rainfall, humidity, solar radiation, soil moisture, location, time, and also to detect presence of objects using remote camera.

Fig. 5 depicts the functional units of the use case showing the farmland equipped with the above mentioned heterogeneous sensors. The remote application service station is the unit where the relevant applications reside. This service station can be a mobile phone or a PDA (personal digital assistant) running suitable applications or it may be remote monitoring office monitoring the state of the environment and triggering necessary measures. The applications interacting with the middleware extracts the sensed data by using the sensor abstraction of the middleware and provide appropriate information to the farmers. An example of such information can be a message suggesting the farmer about the best suitable crop to choose for the prevalent environment. Another example application can be triggering security alarms in case unwanted cattles have intruded into the land. Another such useful application may be triggering the necessity to irrigate the land by measuring an inadequate soil moisture level and so on and so forth.

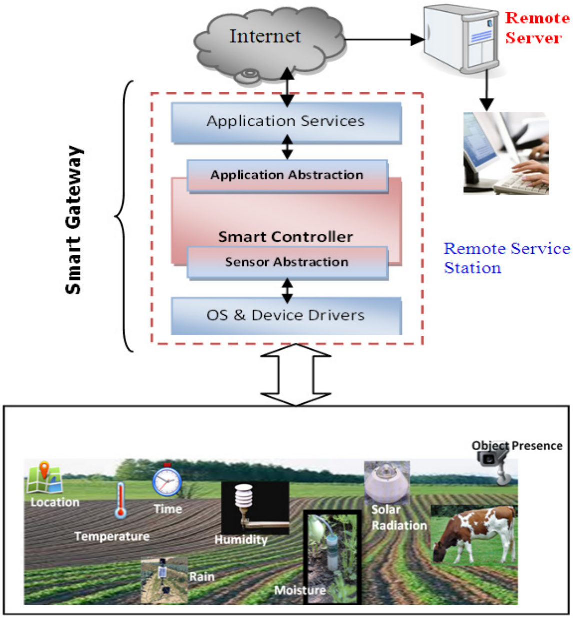
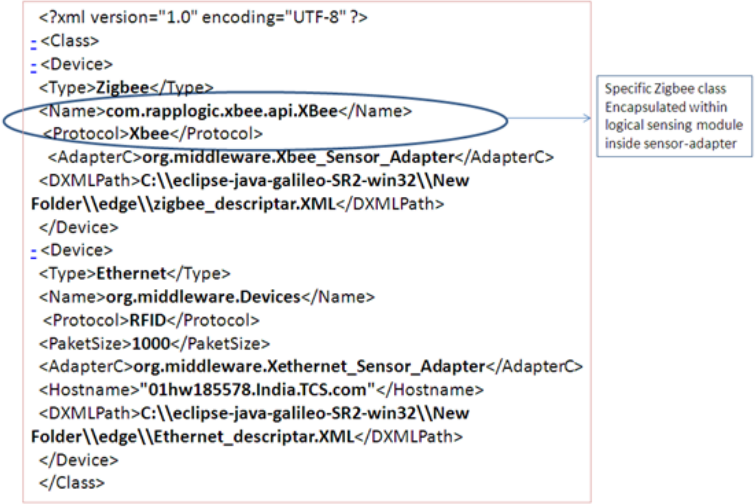


Fig. 5. Smart Irrigation and Farming Environment

Again, as it is evident that the list of sensors is not exhaustive, the system may need to cope up with new types of sensors. This enhancement can be easily done through the sensor abstraction layer by adding new sensors. Similarly new monitoring applications can be interfaced with middleware using its application abstraction.

Fig. 6 shows an excerpt from a typical device configuration schema written in XML. This configuration takes care of a Zigbee based temperature sensor and an Ethernet based remote camera for detecting presence of objects (e.g., cattle) in the farmland respectively. A close observation of Fig. 6 shows that the sensor class provides encapsulation for the device specific ‘Xbee’ [8] module.



```
 <?xml version="1.0" encoding="UTF-8" ?>
- <Class>
- <Device>
 <Type>Zigbee</Type>
 <Name>com.rapplogic.xbee.api.XBee</Name>
 <Protocol>Xbee</Protocol>
  <AdapterC>org.middleware.Xbee_Sensor_Adapter</AdapterC>
 <DXMLPath>C:\\eclipse-java-galileo-SR2-win32\\New
Folder\\edge\\zigbee_descriptar.XML</DXMLPath>
 </Device>
- <Device>
 <Type>Ethernet</Type>
 <Name>org.middleware.Devices</Name>
  <Protocol>RFID</Protocol>
 <PaketSize>1000</PaketSize>
 <AdapterC>org.middleware.Xethernet_Sensor_Adapter</AdapterC>
 <Hostname>"01hw185578.India.TCS.com"</Hostname>
 <DXMLPath>C:\\eclipse-java-galileo-SR2-win32\\New
Folder\\edge\\Ethernet_descriptar.XML</DXMLPath>
 </Device>
 </Class>
```

Specific Zigbee class Encapsulated within logical sensing module inside sensor-adapter

Fig. 6. Excerpt from the XML based sensor abstraction

6 Conclusion and Future Scope

In this article architecture of a middleware which can be used as a generic framework for middleware for interoperating with diverse sensing devices and diverse domain of applications is presented. The reference system architecture shows the interactions among every major block of the system. Managing the heterogeneous sensors based on the various states of the sensors is one of the challenges addressed here. The proposed architecture uses various configuration schemas which make it also extensible for different environments. Adding a new sensor requires the addition of its properties in the appropriate XML schemas. The proposed architecture follows ‘object-oriented’ concept and is adaptable for any smart environment. The presented UML based object oriented model gives a clear view of different class components of the system. The modular and object oriented architecture eases its development using Java and running it as a service of JVM (Java Virtual Machine) as well as OSGi(Open System Gateway Interface)[14] based platform which will be easy to port and implement.

The use case scenario on smart irrigation and agriculture system describes how the generic and adaptable characteristics of the middleware help to address the interoperation; mainly the semantic and syntactic and device management issues.

There are multiple scopes for future research works like, 1) to explore how to meet the real time requirements of interoperation considering the facts of energy constraints of the sensing devices, 2) to estimate what would be the optimized event handling mechanism to be adopted by the middleware's smart controller to actuate sensing/actuating device, 3) to define a unique device attributes list to generate device clustering algorithm by the device manager to achieve faster device interoperation, 4) to define the methods to achieve faster service collaboration providing sensing-device and service/application mapping. Currently we are conducting research on the 3rd and 4th points mentioned above in and developing the complete system on Android and Dalvik VM on top of OSGi [11].

References

1. Bandyopadhyay, S., Sengupta, M., Maiti, S., Dutta, S.: A Survey of Middleware for Internet of Things. In: CoNeCo 2011, Ankara, Turkey, June 26-28 (2011)
2. Bandyopadhyay, S., Sengupta, M., Maiti, S., Dutta, S.: Role of Middleware for Internet of Things: a Study. International Journal of Computer Science & Engineering Survey (IJCSES) 2(3) (August 2011), http://airccse.org/journal/ijcses/papers/0811cses07.pdf
3. http://research.microsoft.com/pubs/135844/SubjectiveSensing-statement.pdf
4. http://www.usc.edu/dept/ee/scip/assets/002/63910.pdf
5. Broering, A., Below, S., Foerster, T.: Declarative Sensor Interface Descriptors for the Sensor Web. In: Proceedings of WebMGS 2010: 1st International Workshop on Pervasive Web Mapping, Geoprocessing and Services, Como, Italy (August 2010)
6. Kostelník, P., Sarnovský, M., Furdík, K.: The Semantic Middleware for Networked Embedded Systems Applied in the Internet of Things and Services Domain. Scalable Computing: Practice and Experience (SCPE). Scientific International Journal for Parallel and Distributed Computing 12(3), 307–315 (2011)
7. Ganz, F.: Designing Smart Middleware for Wireless Sensor Networks. In: The 12th Annual Post Graduate Symposium on the Convergence of Telecommunications, Networking and Broadcasting, PGNET 2011. The School of Computing and Mathematical Sciences, Liverpool John Moores University (April 2011)
8. http://www.digi.com/products/wireless-wired-embedded-solutions/zigbee-rf-modules/point-multipoint-rfmodules/xbee-series1-module#overview
9. http://www.opengeospatial.org/standards/
10. Evaluation of Soil Moisture-Based on-demand Irrigation Controllers: Final Report, Prepared for and funded by: Southwest Florida Water Management District (August 2008)
11. http://www.knopflerfish.org/snapshots_trunk/current_trunk/docs/android_dalvik_tutorial.html
12. http://www.w3.org/TR/wsdl
13. http://www.opengeospatial.org/standards/sensorml
14. http://www.osgi.org/Main/HomePage

Improving TCP Performance in Hybrid Networks

O.S. GnanaPrakasi[1] and P. Varalakshmi[2]

[1] Computer Technology, [2] Information Technology, Anna University, MIT, Chennai

Abstract. It has been observed that TCP suffers from poor bandwidth utilization and extreme unfairness in wireless environment, and the utility of TCP in the multi-hop IEEE 802.11 network has been seriously questioned. TCP is the most dominant transport protocol that serves as a basis for many other protocols in wired and wireless networks. However, high transmission errors and varying latency inherent in wireless channel would have a seriously adverse effect on the performance of TCP. Thus, a novel and pragmatic cross-layer approach with joint congestion and contention window control scheme and a probabilistic approach for RTT measurement is proposed to improve the performance of TCP in multi-hop networks. The proposed design indeed provides a more efficient solution for frequent transmission loss by setting optimal congestion window to improve TCP performance in multi-hop network.

Keywords: TCP, Multi hop network, congestion window, contention window, RTT measurement.

1 Introduction

Wireless technologies eliminate the requirement of fixed cable infrastructures, thus enabling the cost-effective network deployment. In recent years, wireless communication networks have been extensively deployed and are generally specified in accordance with the IEEE 802.11 [6] standard. Multi-hop network comprises of number of wireless nodes to transfer packet from source to destination. Multi-hop networks has higher available bandwidth in a multi-rate 802.11 network and the power of transmission at the edges of the 802.11 nodes can be reduced resulting in lower interference.

1.1 TCP In Multi-hop Networks

Transmission Control Protocol (TCP) [6] is a reliable connection-oriented byte stream transport protocol for the Internet. TCP adjusts well in traditional networks comprising wired link and stationary hosts. TCP achieves congestion control or avoidance [12] by regulating the congestion window size in accordance with the estimated network congestion status in order to adjust the sending rate.

In multi-hop networks, TCP performance is degraded because of its two unique characteristics namely, location-dependent and spatial reuse. Packets may be dropped due to consistent link-layer contention, resulted from hidden/exposed terminal problem. Thus, packet loss in wireless network can be due to various factors.

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 405–415.
springerlink.com

1.2 Performance of IEEE 802.11 Protocols

In IEEE 802.11-based multi-hop network, the underlying MAC layer coordinates the access to the shared wireless channel and provides the link abstraction to upper layer such as TCP. The performance of IEEE 802.11 protocol will be degraded when bit error rate (BER) increases in the wireless channel as well. Unfortunately, wireless transmission links are noisy and highly unreliable. Path loss, channel noise, fading, and interference may result in significant bit errors.

Based when a packet is lost in a wired network because of congestion or is collided in a wireless network, the sender should slow down. When a packet is lost in a wireless network because of noise, the sender should try harder to reduce its TCP congestion window or exponentially increase its back off parameter value for retransmissions. However, if the sender is unable to identify causes of packet loss, it is difficult to make the correct decision. Characteristics of error-prone wireless channel make medium accesses in wireless LANs considerably more complex than in wired networks. However, even with the retransmission mechanism in the MAC layer, packets may still be lost without being handed over to the transport layer due to spurious interference or collisions [8-10]. Although TCP can successfully recover dropped packets, recovery routines inevitably degrade the TCP performance since they involve the end-to-end retransmission of the original packet. Hence, the cross layer interaction between the transport layer and the wireless MAC layer has a critical effect on the detection of erratic errors and on the control of congestion for multi-hop networks.

Thus in the proposed scheme by differentiating the packet loss due to congestion and transmission error, congestion window is not reset if the packet loss is due to transmission error. In case of congestion, instead of resetting the congestion window, an algorithm is proposed to find optimal round trip time (RTT) based on the probabilistic approach. The congestion window is set based on the RTT value calculated.

The paper is organised in such a way that Section 2 provides a literature review of various papers related to the proposed approach. Section 3 describes the proposed scheme. Section 4 describes the skeletal framework of this project. Section 5 illustrates the implementation of proposed work with detailed explanation. Section 6 describes the results generated by simulation. Section 7 concludes the paper

2 Related Work

The TCP performance improvement in wireless environment is discussed in various papers. In wireless networks, channel access contentions may occur between different flows passing through the same vicinity or between different packets within the same flow, which exacerbate the channel contention problem [17,5]. When multiple packets within the congestion window are lost in wireless links, conventional TCP schemes, such as Tahoe [17], Reno [17] and NewReno [11], etc., are only capable of recovering from single loss event per RTT time, and therefore result in high error recovery delays. In environments prone to significant loss, the performance of application is affected not only by the rate at which TCP restores its transmission, but also by its capability to recover from transmission error. Therefore, the congestion control mechanisms implemented in wired networks may not be entirely suitable for wireless environments. Accordingly, a requirement exists for a well-defined collaborative mechanism between TCP and the MAC protocol to reduce the effect of wireless interference on the TCP performance.

One method for dealing with erratic errors on a wireless link is to split the wireless portion of the network from the conventional TCP connection. The Indirect-TCP splits an end-to-end TCP flow into two separate TCP connections, i.e. wired network and a wireless TCP connection. Under this approach, any corrupted packets will be retransmitted directly through base stations on the wireless part of the path, and the wired connection is unaware of the wireless losses. However, a major drawback of this split-connection approach is that it fails to preserve the TCP end-to-end semantics because an ACK originating from the base station.

Alternate way for dealing with cross layer interaction between MAC and TCP is [4] in which a scheme to differentiate loss due to congestion and transmission error is proposed. In this method, the resetting of congestion window size due to collision will not be optimized one. So, the performance degradation will be there.

Another approach [13] is based on the decision of congestion window size based on RTT and channel bandwidth, which is can be measured periodically and change the maximum congestion window size to a new one. In [16], TCP- DAA they minimize unnecessary retransmissions by reducing number of duplicate ACKs for triggering a retransmission and the regular retransmission timeout interval is increased fivefold for compensating the maximum of four delayed ACKs.

3 System Architecture

The Proposed system architecture, in Figure 1, shows the data transfer of packets across a hybrid network from TCP source to TCP receiver with the wireless network being a multi-hop network. A snoop agent, inside the Base Station (BS) monitors very packet that passes through it in either direction. It maintains a cache of TCP packets sent from the TCP source that have not yet been acknowledged by the mobile host. Besides, the snoop agent also keeps track of the last acknowledgment sent from the mobile host

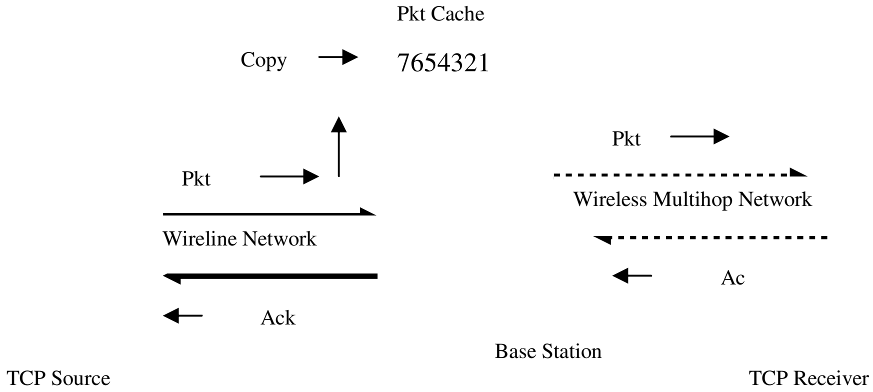


Fig. 1. System Architecture

4 Proposed Framework

4.1 Channel Status Estimation

In order to exploit the information about the actual channel status, we define the probability of transmission failure (*pf*), to be the probability that a frame transmitted by the station of interest fails. It is noted that busy period in the channel status includes collision, frame loss, and successful transmission.

Since the absence of an immediate positive acknowledgement following each data frame will be regarded as a failed transmission, *pf* can be obtained by counting the number of observed transmission failures, divided by the total number of transmission attempts on which the measurement is taken. Therefore, the probability of transmission failure can be defined as given in equation (4.1) [4] as,

$$pf = \frac{\text{Number of transmission failures}}{\text{Number of transmission attempts.}} \tag{4.1}$$

Now let us try to estimate the probability of transmission collision (pc), which is given as,

$$pc = 1 - \frac{1\text{-}pf}{1\text{-FER}} \tag{4.2}$$

Where pf defines Probability of Transmission failure and FER defines Frame Error Rate, which is set as 0.2 in our case.

4.2 Differentiating Loss due to Congestion and Noise

4.2.1 Packet Arrival from TCP Source

When a TCP data packet is received from wire-line network, the snoop agent will cache the TCP data packet in base station and monitor packet transfer at the base station. If the packet is lost in the wireless hop, the base station will automatically retransmit the TCP data packet. When the base station forwards a packet to a TCP receiver, the snoop agent will evaluate the probability of packet collision. If the packet is lost, the snoop agent will use the previous evaluation to adjust the contention window of the MAC layer. Also, a report stating the reason of packet lost will be sent to the TCP sender to help the TCP sender with resetting its congestion window size. Figure 2 shows the flowchart of how the proposed scheme processes the packet sent from TCP source.

When snoop agent receives TCP data packet, it determine if the packet has been received and cached in the base station. If new arrival, snoop agent will temporarily cache the packet into the buffer in base station. If this packet is in-order the snoop agent will take this situation as a general case and forward the packet to the TCP receiver (case 1: normal case). If the received TCP data packet is a newly arrived packet but is out-of-order indicates that the packet was lost in the wire-line TCP connection. The snoop agent will then forward this data packet to the TCP receiver. Afterwards, the TCP sender uses traditional congestion control mechanisms to retransmit the lost packet (case 2: packet lost in wired connection).

If the TCP data packet has been received by snoop agent but has not by the TCP receiver, snoop agent will evaluate the collision rate (*pc*) in the wireless channel. If *pc* is greater than the Frame Error rate, the loss is due to collisions, hence snoop agent compute Round Trip Time (RTT). Now reset Congestion window and contention window is calculated based on the optimal RTT. (Case 3: packet lost in wireless connection).

If the *pc* value is less than the frame error rate the congestion window size will be changed since the loss due to transmission error and hence retransmit the packet. (Case 4: packet lost in wireless connection due to transmission error).

If the packet has been received (i.e., is a retransmitted packet) and the last ACK number cached on the snoop agent is greater than or equal to the sequence number of the received TCP data packet. This represents that the TCP receiver has received the packet. Thus, snoop agent will directly drop the packet and send the last ACK number to the TCP source. (Case 5: ACK lost in wired TCP connection).

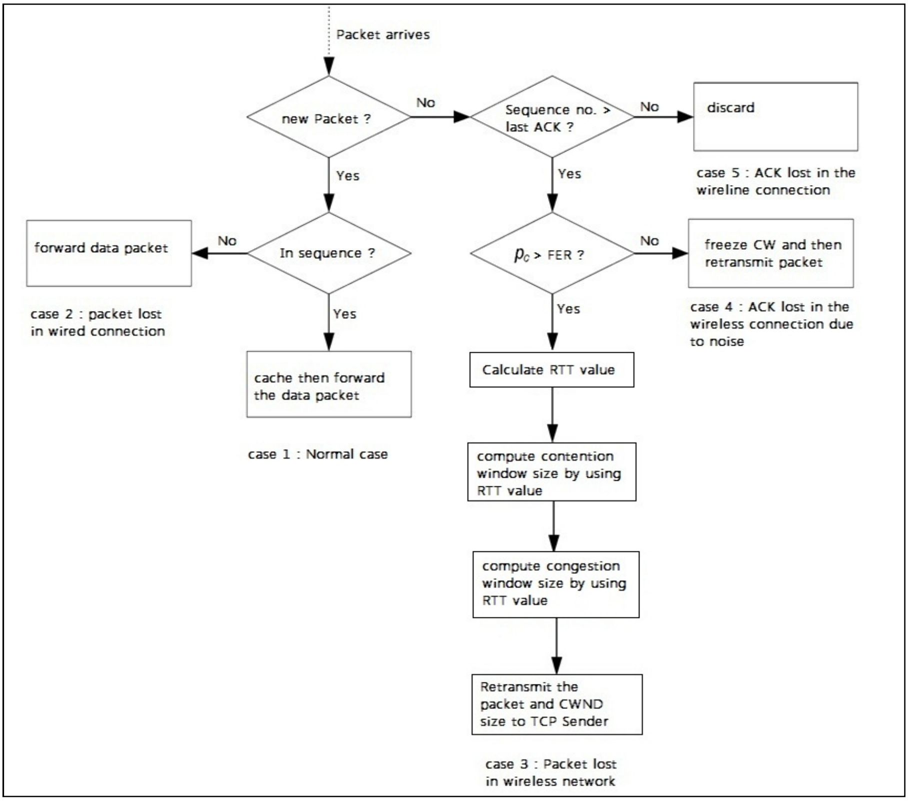

Packet arrives
new Packet ?
No
Sequence no. > last ACK ?
No
discard
case 5 : ACK lost in the wireline connection
Yes
Yes
forward data packet
No
In sequence ?
p_c > FER ?
No
freeze CW and then retransmit packet
case 2 : packet lost in wired connection
Yes
Yes
case 4 : ACK lost in the wireless connection due to noise
cache then forward the data packet
Calculate RTT value
case 1 : Normal case
compute contention window size by using RTT value
compute congestion window size by using RTT value
Retransmit the packet and CWND size to TCP Sender
case 3 : Packet lost in wireless network

Fig. 2. The operational procedure of snoop agent upon receiving TCP data

4.2.2 Ack Arrival from Wireless Station

When a TCP ACK is sent from the wireless station to the TCP source via the base station, snoop agent will follow the procedure shown in the flow diagram as given in figure 3

At first, snoop agent will inspect the arrived ACK. If the ACK is new (i.e., ACK number > last ACK number) the snoop agent will forward the ACK to the TCP source and clear the packet cached in the base station (case 1 and case 2). If the ACK is a duplicate, the snoop agent will investigate whether the packet was lost in the wire-line network or in wireless connection. If the packet was lost in the wire-line network, the ACK packet will be forwarded directly to the TCP source (case 3).

Conversely, if the packet was lost in the wireless connection, snoop agent will compute round trip time (RTT) based on probabilistic approach on sent packets and acknowledgements received. Congestion window and contention window is calculated based on the optimal RTT. A negative acknowledgement (NAK) option and congestion window size will be added in the ACK (associate with the same TCP connection) and the ACK will be forwarded to the TCP source at a later time (case 4). Hence, by inspecting the previous ACK sequence number and historical NAK information, the TCP sender clearly identifies corruption losses in the wireless links and therefore determines whether invoking congestion control mechanisms is necessary.

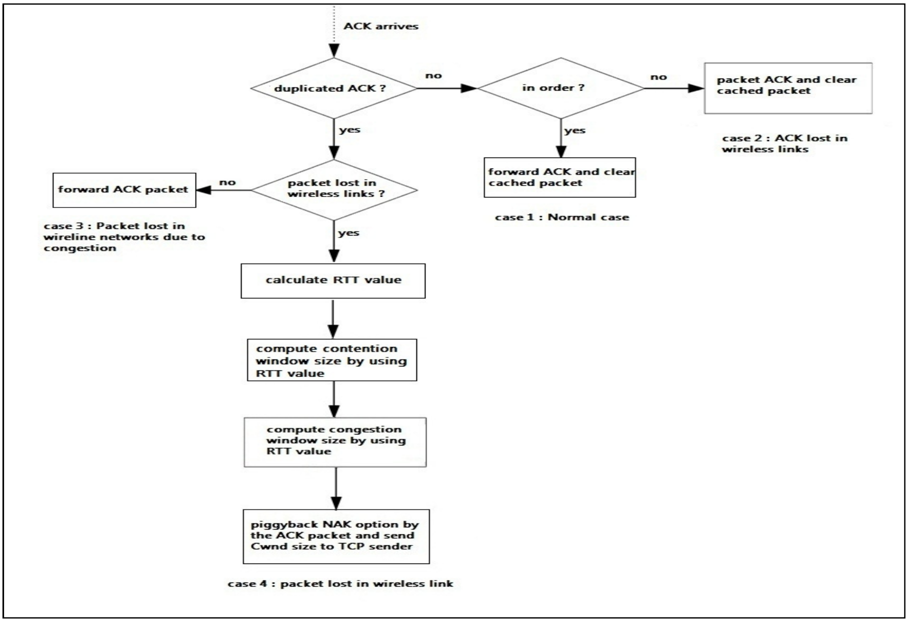

ACK arrives
duplicated ACK ?
no
in order ?
no
packet ACK and clear cached packet
yes
yes
case 2 : ACK lost in wireless links
forward ACK packet
no
packet lost in wireless links ?
forward ACK and clear cached packet
case 1 : Normal case
case 3 : Packet lost in wireline networks due to congestion
yes
calculate RTT value
compute contention window size by using RTT value
compute congestion window size by using RTT value
piggyback NAK option by the ACK packet and send Cwnd size to TCP sender
case 4 : packet lost in wireless link

Fig. 3. The operational procedure of snoop agent upon receiving TCP ACK

5 Implementation

The packet from TCP source (wired) to TCP receiver (wireless) transfers through snoop agent. At snoop agent, based on the packet sequence number and acknowledgement it received, it differentiates the loss due to congestion and transmission error.

5.1 Differentiate Congestion Loss and Transmission Error

This describes a mechanism to differentiate congestion loss and transmission error at the snoop agent.

5.1.1 Successful Transmission in MAC Layer

In the MAC layer of multi-hop network, check the channel status to see if it is idle. The numbers of attempts it had failed to provide successful transmissions are calculated in each node whenever it attempts to make a transmission as given equation (4.1).

5.1.2 Calculate Probability of Collision

Probability of collision (*pc*) is found by probability of failures and error bit rate as given in equation (4.2).

5.1.3 MAC Layer Transmission of Collision Probability to Snoop Agent

Each node whenever it sends acknowledgement for packet it received in MAC layer, it sends its collision probability value, which is compared with its own probability of collision and sends the maximum value to its next hop. The maximum collision probability that is found in the entire multi-hop link throughout its transmission from receiver to snoop agent is transferred to the snoop agent.

5.1.4 Decision Making

In the snoop agent, the value of collision probability is compared with (FER) which is set to 0.2. If collision probability is greater than FER, the loss is collision which invokes PA-RTT approach else it is due to noise and just retransmits the packet.

5.2 Calculate the Optimal RTT Value

Calculation of round trip time (RTT) is based on the history of sent packets and acknowledgements received at any instant of time. Given the two sequences S $=\{s_1, s_2, ., s_n\}$ and D = $\{a_1, a_2, ., a_m\}$,where $s_1, s_2, ., s_n$ are send packets and $a_1, a_2, ., a_m$ are ACK packets corresponding to the send packets in S. If a packet s_i in S is acknowledged by a packet a_j in D, we denote $s_i \alpha a_j$.

If all the packets are captured on a host in a connection chain at a period of time, the following conditions must be satisfied:

(1) Any send packet in S must be acknowledge by one or more packets in D; and any ACK packet in D must acknowledge one or more send packets in S;

(2) Packets both in S and E are stored in chronological order;

(3) For any two packets s_i, s_j in S and a_p, a_q in D,

if $s_i \alpha a_p$, $s_j \alpha a_q$,and $i < j$, then we have $p \leq q$.

Condition (1) indicates that the relationships between send and its corresponding ACK packets may be one-to-one, many to-one, or one-to-many. RTT of a send packet can be defined as the gap between the timestamp of a send packet and that of its corresponding ACK packet.

```
The algorithm for the optimal RTT calculation is
(1) Generate data sets $D_j$ ($1 \le j \le m$):
        $D_j = \{t(i, j) \mid t(i, j) = a_j \_ s_i,$
            $i = 1, \ldots, n \ \& \ t(i, j) > 0\}$.
(2) Combine data from sets $D_j$ to form clusters $C_u$:
    $C_u = \{t(i_j, j) \ \varepsilon \ D_j \mid \forall \ 1 \le j \le m \ \& \ i_1 < i_2 < \ldots < i_m\}$.
(3) For each cluster C:
    (a) if $x(i, j)$, $x(i, k) \ \varepsilon \ C$, $j < k$, then delete $x(i, j)$, and
    (b) if $x(i, j)$, $x(k, j) \ \varepsilon \ C$, $i < k$, then delete $x(k, j)$.
(4) Output $R = \{r_1, r_2, ., r_s\}$ ($1 \le s \le n$) which is the cluster C with smallest standard deviation among all $C_u$ ($1 \le u \le n^m$).
```

Conditions (2) and (3) guarantee that send packets must be replied sequentially. Each send packet must be acknowledged by one or more packets successfully at one time, and the value of a send packet RTT must be positive.

5.2.1 Formation of Data Set

To form data set compute the gaps between the timestamp of each ACK packet in D and that of all the send packets in S. Eliminate the negative values since RTT must be positive. Now group these differences in sets according to each ACK packet in D, forming data sets D_1, $D_{2,..,}$ D_m for ACK packets a1, a2, .,am, respectively.

$D_1 = \{s_1\, a_1, s_2\, a_1, ., s_n\, a_1\}$, $D_2 = \{s_1\, a_2, s_2\, a_2, ., s_n\, a_2\}$, ...

. . . .

. . . .

Dm = $\{s_1\, a_m, s_2\, a_m, ., s_n\, a_m\}$, where element $s_i\, a_j$ in D_j represents the gap $a_j – s_i$ between timestamp of the jth ACK packet in D and that of the ith send packet in S, where $1 \le i \le n$ and $1 \le j \le m$.

5.2.2 Formation of Clusters

Each send packet can be acknowledged by one or more packets successfully at one time, which indicates that in each data set D_j there is only one element to represent the real RTT of that send packet. To construct cluster C_u, take one element from each data set D_j ($1 \le j \le m$), and store them to C_u according to the chronological order of the ACK packets in D. Now find all possible combinations, there will be up to nm possible clusters, but only one of them represents the correct RTTs for all ACK packets.

$C_1=\{s_1a_1, s_2a_2, \ldots, s_na_m / (s_1 \ \varepsilon \ D_1) \infty (s_2 \ \varepsilon \ D_2) \ldots (s_n \ \varepsilon \ D_m) \infty (s_1 < s_2 \ldots < s_n)\}$

$C_2=\{s_2a_1, s_3a_2, \ldots, s_na_m / (s_2 \ \varepsilon \ D_1) \infty (s_3 \ \varepsilon \ D_2) \ldots (s_n \ \varepsilon \ D_m) \infty (s_1 < s_2 \ldots < s_n)\}$

....

....

$.D_n^m=\{s_na_1, s_{n+1}a_2, \ldots, s_{n+m}a_m / (s_n \ \varepsilon \ D_1) \infty (s_{n+1} \ \varepsilon \ D_2) \ldots (s_{n+m} \ \varepsilon \ D_m) \infty (s_1 < s_{2..} < s_n)\}$

Each cluster C_u ($1 \leq u \leq n^m$) has m elements while some of them may share a same ACK or send packet. Remove the send (ACK) packets which share the same ACK (send) packets but keep the one with smaller gap.

5.2.3 Finding Optimal RTT Based on Standard Deviation

Each cluster C_u is a candidate of the RTTs for the send packets in S while only one of them is the optimal one or close to the real one with high probability. We select the one with smallest standard deviation among all clusters C_u ($1 \leq u \leq n^m$).

From data set we calculate the RTT from all clusters.

$\text{Mean}_u=\sum t(i,j)_u/N$ for $(1\leq u\leq n^m)$, - - (5.1)

N is number of elements in $C_{u.}$

$\text{Standard\_deviation}_u = \sqrt{(x-\text{mean}_u)^2/N}$, where $C \varepsilon C_u$. -----(5.2)

Optimal RTT= mean_u with smallest $SD_{u.}$ - - (5.3)

5.3 Set Contention Window and Congestion Window

Contention window size is determined based on the RTT value calculated. Contention window size is increased or reduced relative to the RTT calculated and sent to the nodes in the wireless channel along the TCP packets, which is updated in the MAC layer for every node.

Retransmission timer is set based on the RTT calculated. When the timer fails, the congestion window is set as per Additive increase Multiplicative decrease mechanism.

6 Simulation Results

The simulation is done in NS2 simulator. The simulation is carried out in the multi-hop network with wired and wireless nodes which forms the multi-hop network. The simulation is carried out with a wired node being the TCP source and a wireless node as the TCP receiver. A wired node acts as a base station in which snoop agent is implemented.

The graph shows packet losses due to noise, congestion and total number of packets lost in our proposed scheme, which is compared with 2 existing techniques based on bandwidth measurement [13] and DAA approach. The graph for Cnu [13] depicts that the amount of packets lost is much higher than DAA and the proposed scheme. The proposed scheme implementation prevents the sudden increase and decrease in the congestion window size in the multi-hop network by setting the congestion window based on the RTT calculated in probabilistic manner. From the graph in fig 5 it is clear that the throughput is maintained even in the noisy environment in the proposed scheme. Compared with Cnu[13].

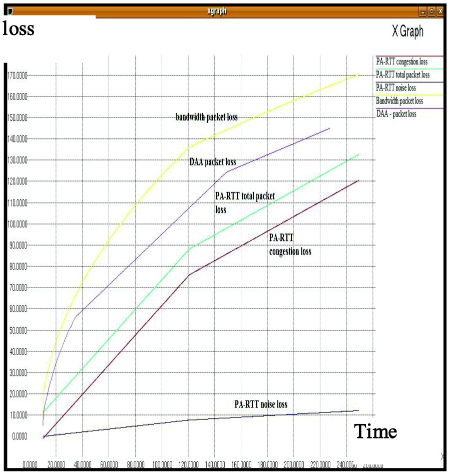


Fig. 4. Loss (Time Vs Loss)

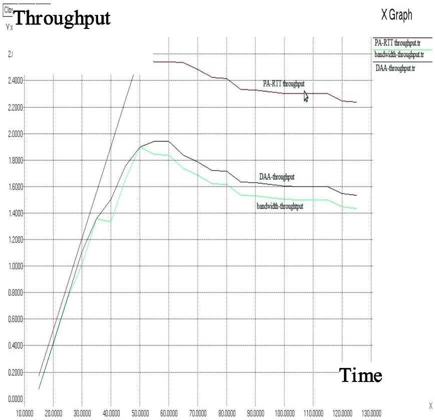


Fig. 5. Throughput (Time Vs Throughput)

7 Conclusion

The shared channel contention and erratic packet loss usually lead to a curbing of window size on the TCP, and thus limit the performance of TCP in multi-hop networks. This paper proposed a scheme to differentiate the loss due to congestion and transmission error and a probabilistic approach to find the optimal RTT to set congestion window size. The TCP performance improvement occurs in noisy scenarios which are prevalent by which it outsmarts other techniques. The future work includes extending this work in dynamic environment.

References

[1] Bakre, A.V., Badrinath, B.R.: Implementation and performance evaluation of indirect TCP. IEEE Transactions on Computers 46(3), 260–278 (1997)

[2] Balakrishnan, H., Seshan, S., Amir, E., Katz, R.H.: I”mproving TCP/IP performance over wireless networks. In: ACM MOBICOM, pp. 2–15 (1995)

[3] Balakrishnan, H., Padmanabhan, V.N., Seshan, S., Katz, R.H.: A comparison of mechanisms for improving TCP performance over wireless links. IEEE/ACM Transactions on Networking 5(6), 756–769 (1997)

[4] Cheng, R.S., Lin, H.T.: A cross-layer design for TCP end-to-end performance improvement in multi-hop wireless networks. Journal of Computer Communications 31, 3145–3152 (2008)

[5] Colandairaj, J., Irwin, G.W., Scanlon, W.G.: Wireless networked control systems with QoS-based sampling. IET Control Theory & Applications 1(1), 430–438 (2007)

[6] Deng, D.-J., Cheng, R.-S., Chang, H.-J., Lin, H.-T., Chang, R.-S.: A cross-layer congestion and contention window control scheme for TCP performance improvement in wireless LANs. In: Telecommunication Systems. Springer (2009)

[7] Dimitri, B., Robert, G.: Data networks, 2nd edn., pp. 1–556. Prentice Hall, New York (1992)

[8] Eckhardt, D.A., Steenkiste, P.: Measurement and analysis of the error characteristics of an in-building wireless network. ACM SIGCOMM, 243–254 (1996)

[9] Hamadani, E., Rakocevic, V.: Evaluating and Improving TCP Performance against Contention Losses in Multi-hop Ad Hoc Networks

[10] Ge, F., Tan, L., Zukerman, M.: Throughput of FAST TCP in asymmetric network. IEEE Commun. Lett. 12(2), 158–160

[11] Floyd, S., Henderson, T.: The NewReno modification to TCP's fast recovery algorithm. RFC, 2582, 1–12 (1999)

[12] Ge, F., Tan, L.: A partial super fast recovery algorithm for FAST TCP. In: Proc. 2nd International Conference on Wireless Broadband and Ultra Wideband Communications, pp. 57–61 (August 2007)

[13] Huh, I., Lee, J.Y., Kim, B.C.: Decision of Maximum Congestion Window size for TCP performance Improvement by Bandwidth and RTT measurement in Wireless Multi-hop Networks. International Journal of Information Processing Systems (March 2006)

[14] Leung, K.-C., Li, V.O.K.: Transmission control protocol (TCP) in wireless networks: issues, approaches and challenges. IEEE Commun. Surveys & Tutorials, 64–79 (4th Quarter 2006)

[15] Postel, J.B.: Transmission Control Protocol. RFC 793, 1–85 (1981)

[16] de Oliveira, R., Braun, T.: A Dynamic Adaptive Acknowledgment Strategy for TCP over Multihop Wireless Networks. In: 18th International Conference on Distributed Computing Systems (ICDCS)

[17] Yang, X., Nitin, H.V.: A wireless MAC protocol using implicit pipelining‘“““““. IEEE Transactions on Mobile Computing 5(3), 258–273 (2006)

Synchronization in Distributed Systems

Amritha Sampath and C. Tripti

Department of Computer Science
Rajagiri School of Engineering & Technology, Rajagiri valley, Cochin, India
amrithasampath@yahoo.com,
triptic@rajagiritech.ac.in

Abstract. In the present scenario, a demand for the highly reliable and synchronous systems is seen. As a result, there has been a gradual shift to distributed systems from the centralized systems. There are few disadvantages for this system too. The most important one is that in a distributed system, the different nodes maintain their own time using local clocks and their time values may not be same for the different nodes. I.e. there is no global clock within the system so that that the various activities in the distributed environment can be synchronized. The various clocks in the system even if set to a common time value at an instant, drift apart due to unavoidable reasons. Hence some kind of continuous mechanism for synchronization is needed so that they can coordinate and work together to achieve the objectives of the distributed system. Two types of synchronization are possible- external synchronization and internal synchronization. In a real time scenario, it is important for the system to be synchronous with each other and with a common external reference time. This is called external synchronization. But in certain systems, it is only necessary for the nodes in the system to be synchronized with each other. This is called internal synchronization. In many applications, the relative ordering of events is more important than actual physical time. Here event ordering is done without clock time values. Hence, depending on the area and type of application, clock synchronization techniques used differs.

In certain real time applications, the system requires to be both internally and externally synchronized. In such cases a centralized algorithm called the 'Cristian's' algorithm is used for synchronization. But this algorithm fails in situations where the time server fails. This paper suggests some methods to make the synchronization process distributed so that the disadvantages of the Cristian's algorithm can be nullified.

Keywords: Synchronization, Centralized Algorithms, Distributed Systems.

1 Introduction

The advantages of the distributed systems are so attractive that there is a gradual shift from the centralized system era to the distributed systems era. The distributed systems are faster and cheaper when compared to the centralized systems. In a distributed system, a set of processors, each with its own internally built-in hardware clock,

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 417–424.
springerlink.com

communicates by message transmission. But they do not have access to a central global clock. The hardware clock of each processor tends to drift apart even if all the processor clocks are set to a common time value. The drifting of these processors can be due to instabilities inherent in source oscillators and environmental conditions such as temperature, air circulation and mechanical stresses. For real time software applications and related processes, highly accurate and synchronized time is a necessity. Clock inaccuracy can cause a number of problems. Even if it is a difference of a minute or two, the outcomes may be unacceptable for the application.

A distributed system is designed to realize some synchronized behavior, especially in real-time processing in areas like factories, aircraft, space vehicles, and military. Synchronization of individual clocks becomes very important in case of certain hard and risky real time applications like, where predictable performance is of major concern, one need to preserve a total logical or temporal scheduling of the tasks in the system.

Clock inaccuracies occur due to certain instabilities inherent in source oscillators and environmental conditions such as temperature, air circulation and mechanical stresses. The clocks in the different nodes need to be synchronized to limit the inaccuracies and hence implement the objectives of distributed system in an efficient manner. Hence, clock skew and drifts[8] which forms the major source of clock inaccuracy needs to be monitored continuously. In certain applications it is not just enough to synchronize the various processes but also the various events that constitute them. This is called intraprocess synchronization. Intraprocess concurrency is captured by relation affects or causally affects. Bit matrix clocks and hierarchical clocks which evolved after the logical and vector clocks of Lamport[1] capture affects relation between events of process. However both have a major disadvantage in terms of increased storage and communication overhead. Difference clocks[2] captures intraprocess and interprocess concurrency, at the same time with reduced storage and communication overheads.

Section 2 describes about the various issues in clock synchronization. There are various methods of achieving clock synchronization depending on the requirements of the situation. In order to behave as a single, unified computing resource, distributed systems have need for a synchronization of clocks and several algorithms have been proposed on this topic. Section 3 describes the various clock synchronization algorithms and their advantages and disadvantages. Proposed solution for overriding the disadvantages of the clock synchronization algorithms is discussed in section 4. Section 5 gives the conclusion.

2 Issues in Clock Synchronization

Load balancing and resource sharing are the two main objectives of distributed systems. In order to achieve these objectives nodes should communicate with each other. In such an environment, it is necessary that the different nodes in the system should have a common time based on which they can order the events. The clocks of the communicating nodes should agree upon a common time value. If the system is working on real time applications like aviation traffic control and position reporting, radio and TV programming launch and monitoring, multimedia synchronization for real-time teleconferencing etc, then the clocks should match with Coordinated Universal Time, UTC[10].

Two factors that might cause errors are clock skew and drift rate. Two clocks are running at the exact same speed but with a constant difference called clock skew. Figure 1 demonstrates clock skew pictorially. Clock drift is another reason for mismatch in values between the various nodes. The clocks may run at different speeds and this difference in speed, even if small may get accumulated to a large value enough to cause errors in the intended working of the system. The difference in speed is mainly due to the type of quartz crystal being used in the clock. Other factors such as temperature and other mechanical effects on the crystal also causes change in frequency of oscillation and hence a drift in speed. Clock drifts are caused due to effects which cannot be removed permanently. Therefore, the clocks need frequent monitoring and adjustments in order to keep them synchronized.

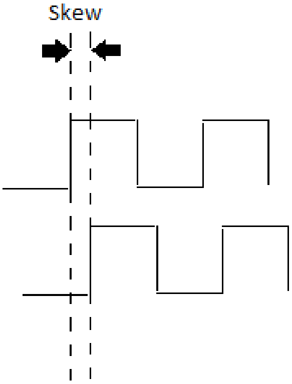


Fig. 1. Clock Skew

Thus, even after synchronization, clock values differs [3]. If C denotes the perfect clock's time, then ideally,(dC/dt) = 1. Perfect clock may be considered as that which provides value of UTC or some other external clock reference value. For a clock that is fast with respect to a perfect clock, (dC/dt) $>$ 1 and for a slow clock, (dC/dt)$<$ 1. The slow and the fast clocks drift in opposite ways when compared with the perfect clock. In both the cases, the clock values are incorrect after a particular time interval and needs to be resynchronized.

Clock synchronization requires that a node be able to read another node's time value. Errors occur mainly due to delay in the messages or the time values sent between the nodes. The minimum value of delay can be calculated by adding time taken to prepare, transmit and receive a message in absence of traffic and other errors in the system. But it is impossible to find the upper limit because it depends on the load in the communication system at the time of transmission[3].

Another issue in synchronization is that a computer clock usually can be adjusted forward but never backward[3]. The time of fast clock should be gradually corrected instead of setting it to the correct value at once. It is done using intelligent interrupt routine. An intelligent routine readjusts the amount of time to be added to the clock time for each interrupt.

3 Related Works

Various algorithms have been proposed for clock synchronization. Centralized algorithms maintain a node as time server which has a real-time receiver. Cristian's algorithm[10] is an passive time server centralized algorithm and Berkeley algorithm is example for active time server algorithm. These algorithms are not scalable and are subject to single point failure. Hence distributed algorithms are used. These can be global averaging or localized averaging distributed algorithms[3].

3.1 Cristian's algorithm

Cristian's algorithm[10] relies on the existence of a time server. In a centralized algorithm, time server is considered to be a perfect clock and whose value can be used as a reference for the other systems to set its time value, so that the entire system of nodes in the network remains synchronized with each other and with external reference time, i.e. it is both externally and internally synchronized. A client machine makes a procedure call to the time server and the server replies with the time value. The round trip time is used to calculate the propagation delay and is added with the server's time value to get the time value for the client clock.

- Client p sends request to time server S
- S inserts its time t immediately before reply is returned
- p measures how long it takes (TroundTrip=T1 - T0)
- p sets its local clock to t+TroundTrip/2

Cristian's algorithm takes several values of T1-T0 and those values which exceeds a particular threshold is considered unreliable and is discarded. Then average of the remaining values is calculated to get value of TroundTrip and half the value is added with 't' to get the value to which the client nodes must set its clock value. I.e. Precision of the passive time server centralized algorithm can be improved by taking several measurements and taking the smallest round trip or using an average after throwing out the large values.

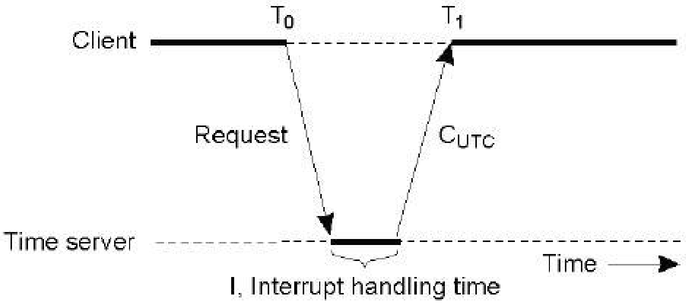


Fig. 2. Cristian's Algorithm[10]

3.2 Berkeley Algorithm

It is used in systems without UTC receiver. One computer is master, other are slaves[9].

- Server polls each client.
- Each client responds to the server with its local time.
- The server estimates the clients' local time (similar to Cristian's technique), and averages the time (including the server's own reading, but excluding those that may have drifted badly). It then tells each client their offset.

In case of failure of master, election is done to find a new master.

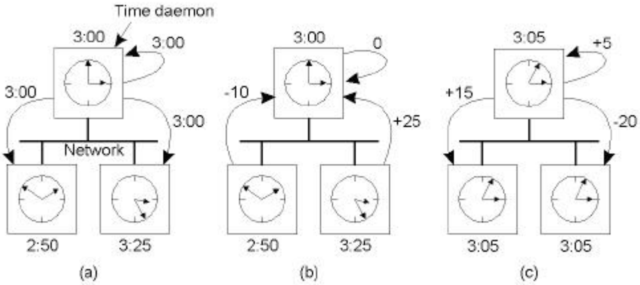


Fig. 3. Berkeley Algorithm[10]

3.3 Naimi-Trehel Algorithm- Token Based System

Several token based hierarchical algorithms have been proposed, most of which are tree based[4]. Naimi-Trehel's algorithm[5] is a token based algorithm for large hierarchical networks that maintains a logical dynamic tree structure such that the root of the tree is always the last node that will get the token among the current requesting nodes. Various extensions to this algorithm take into account the network topology, specially the latency gap between local and remote clusters of machines. It reduces the number of inter-cluster messages and gives a higher priority to local mutual exclusion requests.

In the first extension, on each cluster, excepting the one that initially holds the token, a dedicated process, called proxy is introduced. It is in charge of storing the last request to remote clusters. Before asking for a token which it believes belong to a node of a remote cluster, a node 'i' first a request to its corresponding proxy. If another node 'j' of the same cluster has recently asked for the token and the proxy is aware of it, the proxy redirects the request to 'j' avoiding transmission to the remote cluster.

The second extension aims at reducing the number of inter-cluster messages by aggregating remote requests. When a request has to be redirected to a probable owner, belonging to a remote cluster, the request is not sent to it but stored in a queue. This queue accumulates therefore requests for remote clusters. It is stored in the last node which will enter the critical section within the cluster.

Finally, a local preemption of the token is performed, giving a higher priority to requests originating from the local cluster in order to exploit cluster locality. A threshold that defines the degree of locality and avoids starvation is selected. When the number of local request is below this threshold, the requesting path is modified in order to serve local requests first.

3.4 Token Ring Based System

Paper by Latha CA and Dr. Shashidhara[7] proposes a token ring method that aims at distributing the job of the centralized time server to all the nodes within the system. Therefore, the effect of failure of the time server can be reduced to some extent.

Here the nodes in the system are arranged in the form of a ring and a token is circulated within. The node which possesses the token at an instant act as the time server for the system until the token is passed to its neighbor in the ring. It receives the time value from an external reference time source and broadcasts it to all the nodes in the system. The other nodes receive this value and set its time to the received value if it is within an expected range. Otherwise it requests for retransmission. Even after a particular number of retransmissions if system is unable to get an acceptable value, an error message is created. After a particular amount of time, the token is passed to the neighbor which then takes the role of the time server.

4 Proposed Solution

External synchronization techniques are usually implemented based on a centralized time server. In a centralized server, there is a time server which will be connected to external reference time or the UTC server and is responsible for synchronizing the other nodes within the system. But such a system has all disadvantages of a centralized approach, i.e. the burden of synchronizing all the nodes in the system lies with the dedicated time server. If the server fails, the entire system fails.

The proposed solution is based on the token ring based method discussed in [7] which reduces the centralized dependencies and aims at implementing a system which can realize both external and internal synchronization and hence can deal with the two issues in clock synchronization, that is, clock drift and clock skew.

Here, all nodes are arranged in form of a ring and token is passed from one node to another. Such a token based mechanism has a limit on the number of nodes that it can handle and distance between the nodes. Hence, this type of network organization has a disadvantage that it does not allow the system to be scalable. In order to make the network scalable, a hierarchical organization[6] of the nodes can be incorporated to the above mentioned scheme. Here, in the first layer of hierarchy, the external reference time can be broadcasted to nodes which may be a large distance apart and these nodes in turn act as the reference time source for local networks that work on the above mentioned token ring mechanism. I.e. nodes which are closely located will be arranged in the ring and work on token ring mechanism and these local groups get the external time value from a distant centralized server through broadcasting.

Hence this scheme implements a centralized approach in the first layer of hierarchy and a distributed approach in the second layer of hierarchy. The first layer helps to make the system more scalable while the second layer tries to reduce the centralized mechanism hence reducing the impact of node failures. In case of a node failure, only the nodes belonging to that particular ring will be affected. Trouble shooting also becomes easier since the entire system of nodes have been divided into smaller groups of nodes.

The time for which each node holds the token and hence work as the time server needs to be decided according to the requirements of the system. Also interval between each broadcast of the time value also needs to be decided based on the accuracy of the physical clocks in each node.

The advantages of the system are:

1. Synchronize the systems internally and externally, hence useful in real time systems
2. Makes the system more scalable.
3. Remove clock drift and clock skew
4. Nodes within the system will be internally synchronized even if connection to external reference time fails
5. Effect of failure of node possessing the token is minimal, since after a particular amount of time the next node in the ring takes up the job of the time server.
6. Easier troubleshooting.

Following are the disadvantages:

1. Requires broadcasting technique
2. Costly
3. Token loss, failure of node etc needs to be handled

5 Conclusion

In a distributed system, clocks of the individual nodes need to be synchronized with each other. If a system is synchronized with a universal reference time, the system is both internally and externally synchronized. Various centralized synchronization algorithms like Cristian's algorithm and Berkeley algorithm was discussed in section 3. They are capable of both internal and external synchronization but suffer from disadvantages of centralized systems. Hence, the proposed solution aims at making the system more distributed by introducing a hierarchical system together with a token ring based approach in the second layer of hierarchy. Here, the effect of failure of node that acts as time server is minimized. They can deal with clock skews and drifts. Above all it makes the system highly scalable.

References

1. Lamport, L.: Time. clocks and the ordering of events in a distributed system. Communications of the ACM 21(7), 558–564 (1978)
2. Vaidehi, S., Ram, D.J., Shukla, A.: Difference clocks-A new scheme for logical time in distributed systems. IEE Proc.-Comput. Digit. Tech. 143(6), 426–430 (1996)
3. Sinha, P.K.: Distributed Operating Systems: Concepts and Design, pp. 282–336. PHI Learning Private Limited (2009)
4. Housni, A., Trehel, M.: Distributed mutual exclusion token-permission based by prioritized groups. In: Proceedings of the ACS/IEEE International Conference on Computer Systems and Applications, pp. 253–259 (June 2001)
5. Bertier, M., Arantes, L., Sens, P.: Hierarchical token based mutual exclusion algorithms. In: IEEE International Symposium on Cluster Computing and the Grid, pp. 539–546 (2004)
6. Nishimura, T., Hayashibara, N., Enokido, T., Takizawa, M.: Causally Ordered Delivery with Global Clock in Hierarchical Group. In: Proceedings of the 2005 11th International Conference on Parallel and Distributed Systems, ICPADS 2005 (2005)
7. Latha, C.A., Shashidhara, H.L.: Clock Synchronization in Distributed Systems. In: 5th International Conference on Industrial and Information Systems, pp. 475–480 (July-August 2010)
8. Distributed Systems: Clock Synchronisation, UTC, Clock drift and skew (2010), http://www.krzyzanowski.org/rutgers/lectures/l-clocks.html
9. Distributed Systems: Principles and Paradigms, Physical and logical clocks (2010), http://net.pku.edu.cn/~course/cs501/2008/resource/steen_vrije/courses/ds-slides/2006/notes.06.pdf
10. Applied Computer Science Problems: Clock and State Synchronization, Clock Synchronisation algorithms (2010), http://www.sti-innsbruck.at/fileadmin/documents/teaching_archive/acsp0405/06_Ruff_Ausarbeitung.pdf

NAAS: Negotiation Automation Architecture with Buyer's Behavior Pattern Prediction Component

Debajyoti Mukhopadhyay[1], Sheetal Vij[2], and Suyog Tasare[2]

[1] Department of Information Technnology
[2] Department of Computer Engineering
Maharashtra Institute of Technology
Pune 411038, India
{debajyoti.mukhopadhyay,sheetal.sh,suyog59}@gmail.com

Abstract: In this era of "Services" everywhere, with the explosive growth of E-Commerce and B2B transactions, there is a pressing need for the development of intelligent negotiation systems which consists of feasible architecture, a reliable framework and flexible multi agent based protocols developed in specialized negotiation languages with complete semantics and support for message passing between the buyers and sellers. This is possible using web services on the internet. The key issue is negotiation and its automation. In this paper we review the classical negotiation methods and some of the existing architectures and frameworks. We are proposing here a new combinatory framework and architecture, NAAS. The key feature in this framework is a component for prediction or probabilistic behavior pattern recognition of a buyer, along with the other classical approaches of negotiation frameworks and architectures. Negotiation is practically very complex activity to automate without human intervention so in the future we also intend to develop a new protocol which will facilitate automation of all the types of negotiation strategies like bargaining, bidding, auctions, under our NAAS framework.

Keywords: NAAS, Negotiation architecture, Negotiation automation, Negotiation framework, Protocols, Agents, Web Services.

1 Introduction to Negotiation

Negotiation is the process between two or multiple entities where everybody comes to some useful consensus or agreement as a result. Like it or not everybody is a negotiator in some ways without even knowing it. We did it as kids for trading toys, cards and still we do it for the raise in salary, purchasing things in our personal lives. Many people try to avoid this blatant negotiation procedure consciously because they don't like it but end up in either negotiating or losing in the bargain. Now the question is how to automate this process using the latest technologies and advancements in computing. Here are some key concepts behind the theory of negotiation [6, 7, and 8].

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 425–434.
springerlink.com

1.1 Negotiation Types

1. Distributive or fixed pie negotiation. It involves people who have never had a previous interactive relationship neither they are likely to do so in near future. Everybody gets a fixed pie. 2. Integrative negotiation or everybody wins something or win-win scenario. This means to join several parts into whole. This needs some cooperation and higher degree of trust from every entity in negotiation. Ideally it is difficult to achieve and most difficult to automate because trust and forming relationships on that is human characteristic and difficult to implement in machines and computing [9].

1.2 Negotiation Tactics

The level of detail in the best negotiator actually understands the human mind and how to use this in different voice tones and expressions for the best possible negotiation outcome for all the parties, ideally. A few common tactics that are used in negotiation are outright refusal, conditioning, calling bluffs [10].

1.3 Negotiation Protocol Types

It is the set of rules that govern the interaction between participants. Depending upon the types the negotiation can be bidding, auction and bargaining. [11]

In this paper we are proposing a new architecture and a framework for automating the negotiation process. The key feature in this framework is a component for prediction or probabilistic behavior pattern recognition of a buyer, along with the other classical approaches of negotiation frameworks and architectures.

2 Related Work

Substantial research work has been done on various negotiation protocols, languages and set of parameters, frameworks on each of the negotiation types like auctioning, bidding, bargaining, simple request and response methods etc. We reviewed the following frameworks and architectures to come up with certain conclusions and assessed if any provision is given in those for observing the buyer's behavior.

Hudert, Guido and Ludwig [1] have defined a framework for augmenting WS-Agreement by Open grid Forum (OGF) standard which actually defines an XML based structural definition of Service Level Agreements [2]. This framework is proposed only in relation with WS-Agreement protocol where parties interested in negotiating an agreement first run the negotiation Meta protocol to establish which negotiation protocol is used. [1] Subsequently, the protocol is executed to determine the resulting, negotiated WS-Agreement document. Finally, winner is determined and acceptance, rejection is performed again according to the WS-Agreement protocol standard (Fig. 1).

Tung Bui and Gachet [3] found out that web services (using UDDI, WSDL, SOAP, and XML) can be used as a market broker, to help in discovering the supply/demand, arbitrate the pricing, find the most appropriate service for a given request, to modify the request and services and generate the contract. They have given the basic architecture as shown in Fig. 2.

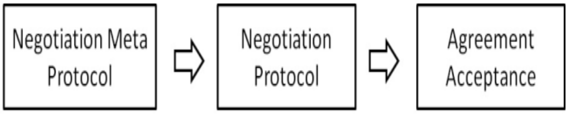


Fig. 1. Agreement creation process

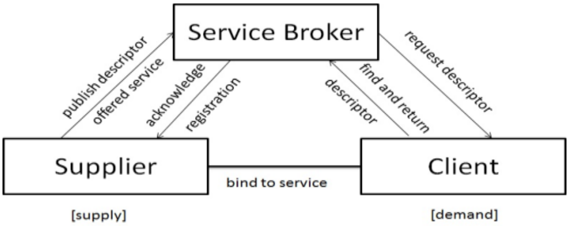


Fig. 2. Electronic market for web services

In this framework the authors have used BPEL4WS[4,5] and WSCI and proposed following support roles as information, Communication, Negotiation, Bargaining, Execution of transactions in pre-transaction, transaction and post transaction phases in electronic market. Authors have given detailed diagrams for the seven web services like topology of web services for negotiation and bargaining, service discovery, Adaptation and pricing, Service ranking, service bargaining, best price adaptation, contract composition, here the client is seen as a negotiation manager. This architecture has its limitations like assumption of trust and there is no competitive strategy to distort the cooperative spirit of e-market. There is no fixed pricing, brokerage and addressing of services, no technical requirements as security and logging mechanisms discussed.

Bin Wu and Chaozhen Guo [12] present a new web service negotiation mechanism and new web service composition coordinated negotiation architecture to solve the problems which mainly occur in such architectures of web service composition based on agent. The problems like unreasonable use of time and data link resources in the condition of multi-negotiation concurrency which leads to inefficiency of negotiation, lack of effective processing when confronting negotiation failure. Here the authors are applying asynchronous communication theory to the process of Web Service Negotiation and extend effective processing in case of negotiation failure. There are disadvantages as overhead of time spent in processing failure may be more than the benefit it brings, if the bad status of network (Fig. 3 and Fig. 4).

It divides NA into SPNA and SRNA. Here a pseudo algorithm of process handling negotiation failure is given. So the research focuses only on elaborating architecture (Improved Vs OWSCCNA) for proper processing in case of negotiation failure in web services.

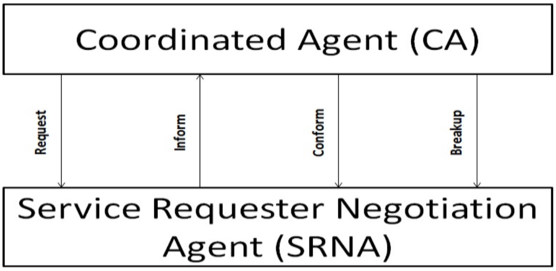


Fig. 3. Original web service composition coordinated negotiation architecture

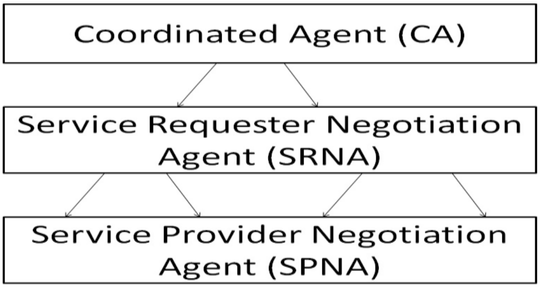

Coordinated Agent (CA)
Service Requester Negotiation Agent (SRNA)
Service Provider Negotiation Agent (SPNA)

Fig. 4. Improved web service composition coordinated negotiation architecture (IWSCCNA)

Jin and Segev [13] have proposed a framework for negotiation processes that provide a consistent model for supporting a comprehensive range of negotiations in dynamic next generation e-business environment. It has five components negotiation requirements, negotiation structure, negotiation process, negotiation protocol and strategy. The authors have described each of their framework components in stepwise details and claim that this model is the most flexible and practical where protocol and strategy are separated in the design, as shown in Fig. 5.

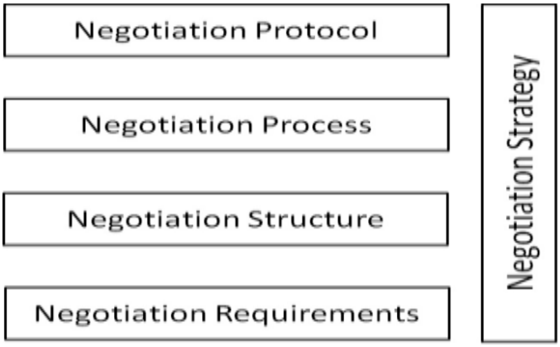


Fig. 5. Negotiation Framework

Chhetri, Lin, Goh, jun, Jian, Kowalczyk [14] have proposed an agent based coordinated negotiation architecture to ensure collective functionality, end to end QoS and coordination of complex services and they describe how this architecture can be used in different application domains and also how the negotiation system on the service provider's side can be implemented both as an agent based negotiation as well as a web service based negotiation system. This work is based on ASAPM, adaptive agreement and process management, aims at developing intelligent agent based techniques and tools to facilitate the adaptive service management and process management. The overall architecture of ASAPM includes following four components Fig. 6.

Coordinated negotiation architecture is proposed on the above as shown in Fig. 7.

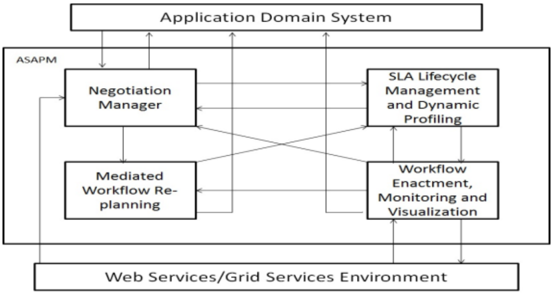

Application Domain System
ASAPM
Negotiation Manager
SLA Lifecycle Management and Dynamic Profiling
Mediated Workflow Re-planning
Workflow Enactment, Monitoring and Visualization
Web Services/Grid Services Environment

Fig. 6. ASAPM

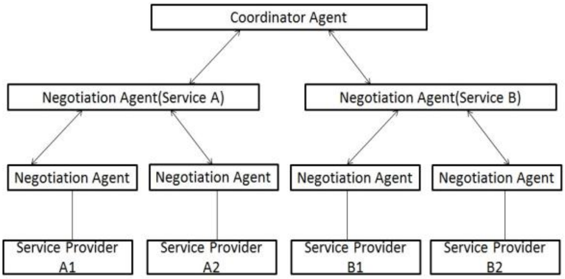


Fig. 7. Coordinated Negotiation Architecture

This two level approach enables the reuse of this architecture in any application domain. This paper does not contain the decision making strategies both at the negotiation level and at the coordination level.

All these architectures are mostly implemented in FIPA compliant JADE Agent Framework and WS2JADE toolkit which enables the integration of JADE agents and web services.

Stanley, Huang, Yihua, Haifei Li, Wang, Liu, Lee, Lam [15] have presented the design and implementation of a replicable, internet based negotiation server for conducting bargaining type negotiations between enterprises involved in e-commerce and e-business where enterprises can be buyers and sellers of product/services or participants of a complex supply chain involved in purchasing, planning and scheduling. The use of negotiation servers to conduct automated negotiation has been demonstrated by the authors. A content specification language for information registration, a negotiation protocol and its primitive operations, an automated negotiation process, a cost benefit decision model and the architecture of an implemented system have been described in this. This is based on object oriented, active database technology in contrast to the existing systems which are based mostly on distributed agent technology. Their negotiation server is analogous to web server which provide following negotiation services as a registration service, a negotiation proposal processing service, an event and rule management service, a cost-benefit analysis service. The system architecture consists of following components, external to WAN components as shown in Fig. 8.

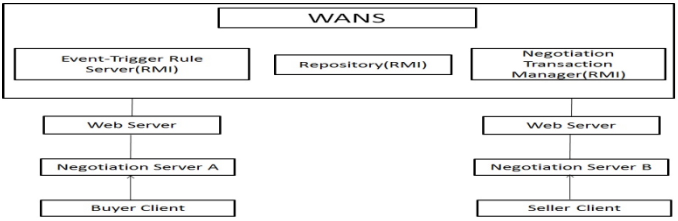


Fig. 8. System Architecture

Cao, Chi, Liu [17] have also described an automated negotiation architecture based on software multi agents using SOA (Service oriented architecture) and web services technology. Here is a reference to Negotiation as a Service where the authors observed that in businesses, negotiation process is better given as a service, not as software and more suitably the entities can be seen in a role of a service provider. Authors correlate it with SaaS (software as a service) where software providers deploy application software on their own servers and customers search, access software services via Internet, consume services as per demand and pay the software providers based on time and number of services consumed. So the users buy software service as an alternative to permanent license and SaaS subverts the traditional life cycle of software. User need not concern about the upgrade and maintenance of software. Here the authors have given a six layered architecture for the automated negotiation as in Fig. 9.

Though this prototype has been developed using SOA and UDDI, WSDL, XML technologies , it could not be a standard for practical implementation of negotiation automation due to many pervasive issues and conflicts in standards and overhead due to parsing of SOAP messages and XML documents on every exchange. Precisely in all the above architectures, even if they can be converted into a prototype of automated negotiation, they cannot be a fixed standard for the practical implementation of the same. Also there is no consideration of understanding and predicting the buyer's behavior in all the above architectures, which is if added it can be a very sophisticated and futuristic approach of negotiation automation.

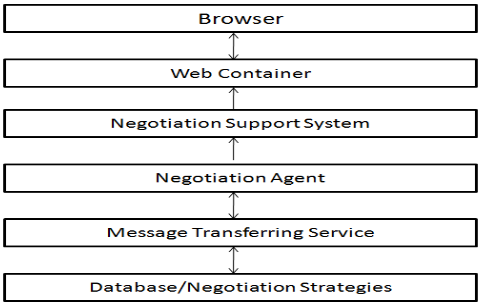

Browser
Web Container
Negotiation Support System
Negotiation Agent
Message Transferring Service
Database/Negotiation Strategies

Fig. 9. Architecture for Automated Negotiation

3 NAAS Automation Architecture

We are proposing an automated negotiation system in the form of web service calling it as NAAS (Negotiation as a Service). As we saw in survey, using SOA (Service oriented architecture) we can make the system running in any place as a service node that is integrated with third party e-commerce platforms so the system can play role of negotiation service provider in the real business environment. The benefits are , we can obtain stable visiting quantity, maintenance and upgrade of system can be completed on the server independently, saving human and material resources, automated negotiation system can make use of the existing basic facilities provided by e-commerce platform i.e. security, authentication, transaction management etc. ,saving costs of development.

With NAAS we are trying to overcome the issues we saw in the previous architectural styles. Specifically we are trying to add a module which can be developed on a strong algorithmic base i.e. using Association rules or Markov models to predict and then further analyzing the buyer's behavior for making the process of negotiation automation complete. The main reason for buyer's behavior prediction module in the basic design is to make the negotiation system intelligent, sophisticated and futuristic so that a better standard for negotiation automated can be created and existing designs can be enhanced.

NAAS Architecture Components: 1.Service registry (databases types, directory) 2.Negotiation support system (Negotiation support system provider, Negotiation service requestor) 3.Protocol on internet module, MTS and service discovery by using web services on UDDI, HTTP, SOAP 4.Advertisement publishing from provider/seller 5.Negotiation service requestor/ buyer 6.Buyer's behavior pattern prediction (proposed in auction, in bargaining, in bidding for further research by the authors) 7.Business logic module and agent management module (external to NAAS) 8.Strategies, decision modules (external to NAAS).

Working: The seller will publish its information about product and the prices etc. On the service registry via web services and web container to which a negotiation support system will interact on some negotiation protocol or all the existing protocols. A MTS is on the internet to transfer the requests from buyers as well as product information from the sellers. A database can be maintained for all these service related queries and information with different ontology, if this architecture is considered in details. In our buyer's behavior prediction component, some key features like age, gender, culture, type of product, feedback can be taken as input to the system for the decision making and predicting the buyer's choices, behavior according to the region, country and states. This will help us in tapping the particular market and applying further strategies in negotiation system according to the buyer's behavioral aspects for all the types in negotiation like bidding, bargaining and auctions for the desired product on sale (Fig. 10).

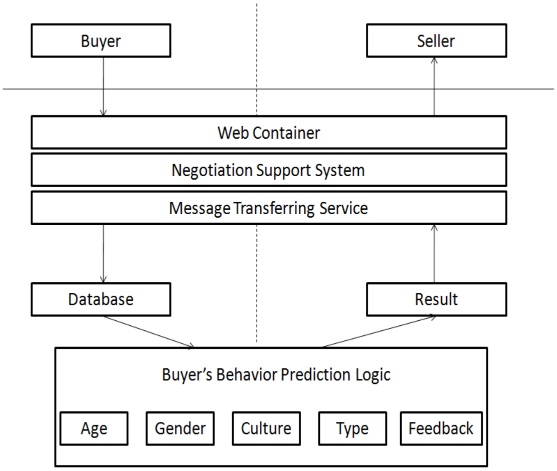

Buyer
Seller
Web Container
Negotiation Support System
Message Transferring Service
Database
Result
Buyer's Behavior Prediction Logic
Age
Gender
Culture
Type
Feedback

Fig. 10. NAAS Architecture

4 Discussion on Buyer's Behavior Prediction Component

Due to addition of the buyer's behavior pattern prediction, this architecture can be a flexible, reliable because if in advance of any negotiation process in a multiple buyers and sellers scenario on web based automated negotiation system, we are able to analyze how many potential buyers can be in this negotiation process and what kind of negotiation behavior the particular set of buyers go into. We are proposing a buyer's behavior prediction component as shown in Fig. 11.

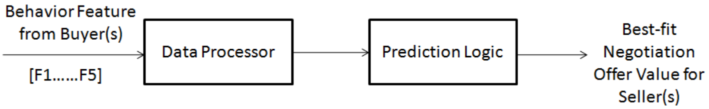

Behavior Feature from Buyer(s)
[F1……F5]
Data Processor
Prediction Logic
Best-fit Negotiation Offer Value for Seller(s)

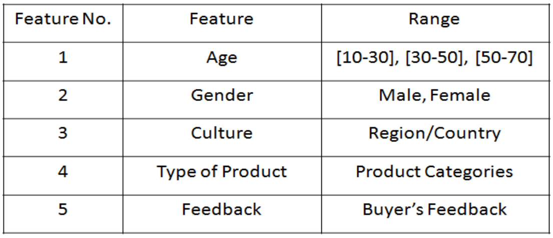
Fig. 11. Buyer's behavior prediction component

First, data processor processes the input and then forwards the required values to the prediction logic. Our proposed prediction logic computes the offer values according to the feature which is submitted by the buyer.

$$\text{Offer Value (OV)} \quad = \quad \frac{\text{Feature [F1.....F5]}}{\text{Count of Buyer(s)}} \quad * \; 100$$

According to the present proposed prediction logic, we can generate such five types of offer values according to the particular feature (e.g., age, gender, culture, type of product, buyer's demand or feedback)

For e.g. Feature (F1) = Age then we can divide this into 3 categories like Age[10 - 30] , [30 - 50], [50 – 70], we can allocate a weighted value for each of these age groups which will be divided by count (C) of buyer(s) to calculate the percentage. This percentage can be used by seller(s) to negotiate on the Offer Value (OV) of that particular feature (F1...F5). Please see Fig. 12.



Feature No.	Feature	Range
1	Age	[10-30], [30-50], [50-70]
2	Gender	Male, Female
3	Culture	Region/Country
4	Type of Product	Product Categories
5	Feedback	Buyer's Feedback

Fig. 12. Feature Categories

This leads to a very successful implementation of various negotiation strategies (auction, bidding, bargaining) etc. There is no existing architecture on negotiation which facilitates the intelligent buyer's behavior prediction in automated negotiation. We are working on the survey of appropriate existing algorithms for the prediction.

5 Conclusion and Future Work

The overall survey of existing architectures leads us to come up with shortfalls and requirements in negotiation automation. NAAS can provide the advantages and further applications related to this. The authors of this manuscript are trying to assess the work done in this area and come up with some conclusions about how to construct as well as deploy the buyer's behavior prediction component for the e-negotiation system. We intend to find out an appropriate algorithmic base to implement this scheme. We also intend to create a new negotiation protocol for our NAAS framework in the next phase of work.

References

1. Hudert, S., Ludwig, H., Wirtz, G.: A Negotiation Protocol Framework for WS-agreement. IEEE (2006)
2. Hudert, S., Eymann, T., Ludwig, H., Wirtz, G.: A negotiation Protocol Description Language or Automated service Level Agreement Negotiations. IEEE (2007)
3. Bui, T., Gachet, A.: Web Services for Negotiation and Bargaining in Electronic Markets: Design Requirements and Implementation Framework. IEEE (2005)
4. Introduction to BPEL4WS, `http://www.research.ibm.com/convsupport/papers/BPEL%20&%20Conversations.html`
5. Kim, J.B., Segev, A., Patankar, A., Cho, M.G.: Web Services and BPEL4WS for Dynamic eBusiness Negotiation Processes. IEEE (2004)
6. Lau, R.Y.K.: Towards a web services and intelligent agents-based negotiation system for B2B eCommerce. IEEE (2007)
7. Yao, Y., Ma, L.: Automated Negotiation for Web Services. IEEE (2010)
8. Negotiation Types, `http://www.negotiations.com/articles/negotiation-types/`
9. NegotiationTactics, `http://www.negotiationtactics.net/`
10. Negotiation Protocols, `http://www.w3.org/2001/03/WSWS-popa/paper19`
11. Wu, B., Guo, C.: The Reasearch and Improvement on the Coordinated- Negotiation Architecture of Web Service Composition based on Agent in the condition of Multi-Negotiation Concurrency. IEEE (2010)
12. Kim, J.B.: A Framework for Dynamic eBosiness Negotiation Processes. IEEE (2003)
13. Chhetri, M., Lin, J., Goh, S., Zhang, J., Kowalczyk, R., Yan, J.: A Coordinated Architecture for the agent–based Service Level Agreement Negotiation of Web Composition. IEEE (2006)
14. Su, S., Huang, C., Hammer, J., Huang, Y., Li, H., Wang, L., Liu, Y., Lam, C.L.: An internet based negotiation server for e-commerce. 2001 Springer Journal VLDB, 72–90 (2001)
15. Cao, M.-K., Chi, R., Liu, Y.: Developing Multi Agent automated negotiation service based on service oriented architecture. Service Science Journal 1 (2009)

Quantum DOT Sensor for Image Capturing and Routing Based on Temporal Power and Critical Factor

S. Indu Vadhani[1], G. Vithya[2], and B. Vinayagasundaram[3]

[1] Student, S.S.N College of Engineering, Kalavakkam, Chennai
induvadhanis@gmail.com
[2] Assistant Professor, St. Joseph's College Of Engineering, Chennai-119
vithyamtech@gmail.com
[3] Associate Professor, Computer Center M.I.T Campus, Anna University, Chennai
bvsundaram@annauniv.edu

Abstract. Wireless Multimedia Sensor Network(WMSN) is a system of interlinked wireless multimedia sensor nodes(WMSn) that are able to retrieve multimedia content such as video and audio streams, still images and scalar sensor data from the environment. The multimedia sensor nodes works with CMOS or CCD technology as the base. There is a serious need to overcome the drawbacks in CMOS technology such as low battery life ,high power requirements and inappropriate to be deployed in remote areas on account of frequent charging of the nodes and discrepancies in CCD technology such as delay in capturing due to row wise analysis of pixels. quantum dot sensors are introduced to overcome these serious drawbacks .The quantum film technology works by the attachment of very small quantum dots.the dots act as microelectronic units and captures images. The captured images with very high resolutions are transferred to the wireless multimedia sensor nodes(WMSn) which takes the data to multiple destinations. Jitter and delay is avoided because of the sharpness of the images and hence QOS is improved to a great extent. The proposed architecture ist called Quantum Dot sensors for image capturing and routing based on Temporal power (QDSTP) and Last Performance Time . Sensors that work based on quantum film technology achieve more selectivity, sensitivity, robustness compared to their classical counterparts. This also enables higher resolution and more low light performance.

Keywords: WMSN, WMSn, quantum film, quantum dots.

1 Introduction

In the near future ,there is a need to make extinct, the devices that use low battery power and replace them with high energy and long lasting power source. For managing real time data , batteries and efficient usage of sensors for collecting information there is a need of a source with enormous power. by using quantum techniques, it can be achieved.

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 435–445.
springerlink.com

Quantum films are introduced as a part of WMSn.the quantum films which are placed on the top layer of the CMOS chip increase the sensitivity by four times..A Photon ,an elementary particle is the basic unit of light. The energy of a photon is emitted in terms of quanta A quantum film is a collection of tiny semiconductors called quantum dots which absorb light and emit electrons within the polymer film. A quantum dot is a generic term applied to a particle which is a few nanometres in dimension. the film which uses embedded quantum dots instead of silver grains like photographic film, can produce images at a very high resolution.

One or two-dimensional molecular configuration that is suspended as a polymer backbone. by sharing the battery power ,the chances of draining of power in the WSMn is greatly reduced. The sensed images are transferred to the sink node with zero time delay which is a unique feature of QOS. High pixel quality and intensity, latency are supplemented with QOS.

2 Related Work

(Dowling JP 2011) in this paper, the basic computing and capturing of image has been given an overview. with this paper as a base the whole concept of quantum dot sensor was considered.

(Andrew Newell and Kemal Akkaya 2009) proposed a distributed camera actuation algorithm which turns on the least number of camera sensors during an event such that the amount of redundant multimedia data can be decreased while the adequate coverage can be achieved. At the same time camera sensors exchange their field of View (foV)s with their neighbours before they decide to be actuated. If the portion of the event area covered by a particular camera sensor has not already been covered by other camera sensors and the size of such area is significant enough, then the camera is actuated. The algorithm is completely distributed and requires only 1-hop communication for the nodes.

(GVithya and Dr B vinayagasundaram 2011)In the paper ,actuation sensor with adaptive routing with QOS aware checkpoint on wireless multimedia sensor network by gives a strong idea about the routing protocol used in wireless multimedia sensor network taking utmost care on power utilisation and lowering latency ,thus increasing the QOS to the end user.

(Kashif Zafar and Abdul Rauf Baig 2010) in the paper Optimisation of route planning and Exploration usinttg multi-agent system,optimal path for routing is chosen by considering factors such as energy ,time .This helps to arrive at a decision to choose the optimal path reducing the loss of energy and redundancy of data.

(Thang Hoang et al. 2011) in the paper"Single Photons emitted by Single Quantum Dots into waveguides:Photons guns on a Chip" helps in the analysing single photons emitted from quantum dots and application in circuitry.

(Sheng-Zhong et al.2005) in the paper "Quantum Communication For Wireless Wide Area Networks" have proposed the quantum routing mechanism between two nodes even if there is no platform sharing a common message.this has hepled in relating the quantum communication with different nodes.

(X.Liu et al.2011)In the paper, Thin film conductivity metrology using Photoluminescence of quantum Dots written by the extent to which conductivity of photons exist due to photoluminescence is proposed.this has helped in designing the smart node in QDSTP architecture . the concept has also paved way for changing the number of dots which are embedded in the film .

(Rahul Ratan et al 2006)in the paper "On Random routing and its Application to Quantum Interconnection Network " random path selected by quantum giving importance to qubits is proposed. The concept introduced in this paper has played an important role in the route path selection in QDSTP architecture.

(Ian F. Akyildiz 2006) , Tommaso Melodia and Kaushik R. Chowdhury discuss an algorithms ,protocols and hardware for the development of WMSN, and open research issues related to processing and compression of multimedia data for increased network lifetime and QOS provisioning which is required for multimedia data and issues at the application, transport, network, link and physical layers of the communication stack along with possible cross layer synergies and optimization.

(Jung-shian li et al 2011) paper on "quantum communication in distributed wireless sensor networks" has given a broad perspective of the communication in sensors.the concept introduced in this paper has served to be of great importance in the QDSTP architecture.

3 Overview of Proposed System

During the survey of energy management, it is observed that multimedia sensor networks are found to be badly in need of enormous energy to capture and transmit image. It is necessary to develop a technique for WMSN to flexibly perform the communication with minimum energy consumption, QOS and restricted execution time. Wireless Multimedia Sensor Network is deployed densely by WMS node called the Smart Quantum node(SQM).

By using packet scheduling algorithm, the temporal power of the cluster is taken into account and the event is maximized with least amount of redundancy. Once an event is detected with multimedia sensors, the sensors in the vicinity can be actuated to capture an image or video of the event until the event ends. There are 4 phases to support QDSTP Architecture.

3.1 Quantum Film Embedded with Dots Technique

A quantum film is prepared by sputtering process ,used for thin film deposition where atoms are ejected by bombardment of the target by many energetic particles. This technique evenly deposits fine particles on the surface. This technology also imparts a high temperature adhesive (HTA) technique and hence does not bubble or peel-away.the average heat rejection capacity of quantum films is 63%, which is quite high and hence it helps in maintaining the temperature without getting heated up.It covers and transfers hundred percent of each pixel capturing all the light that hits the top layer directly to the chip. This is opposed to the traditional CMOS battery which has to pass through many layers as a result a large number of photons are eliminated leading to reduction in quality of the images. The film can be included just as producing a photo resist on wafer.

All the properties of the quantum film composed of dots depends mainly on the size of the dot. Size changes the electrical and non-linear optical properties .smaller the dot, greater the band gap and hence the emitted light has higher energy and intensity. A single quantum dot has the capacity to function as a microelectronic unit such as a transistor. They have extremely low power requirements and has very high operating speed.

3.2 Smart Quantum Circuitry

For the deployment, different types of Quantum Wireless Multimedia Sensor Nodes(QWmsn) are taken into consideration. Optimised and a path with less loss to be chosen,two major kinds of nodes are put into use.most important is the Smart Quantum Multimedia Sensor node which comprises of major inclusions such as a 360 ^{o} rotational camera provided with a lens, quantum film, forming the sensing element.the processing part of the node consists of a memory element and a microprocessor in order to process the sensed image and the bottom most part plays an important role in communication between different nodes deployed in the area as it has transmitter and receiver as a part of it. Sensing layer having camera occupies top most position of the SQM node.it is capable of rotating in 360 degrees covering all possible information as it has a high memory backup provided along with it.Camera having quantum film consisting of quantum dots is used for better resolution as it has minimal layers and has high efficiency to capture light through quantum confinement. This replaces the use of photodetectors as in traditional CMOS battery which is used to sense each and every pixel emplying increase in cost and time .Quantum dots obey the pricipal of mechanical quantum confinement and the band gap of each dot determines the wavelength of the radiation and absorption, emission spectra.lesser the size of the dot higher is its energy band gap. A layer of capacitor also forms a part of the sensing layer below the quantum film in order to store the light energy absorbed through quantum confinement. The layer is succeded by two layers of metals such as CdSe and ZnS.This is efficiently processed by tranferring to the processing part followed by the part helping in communication. The transfer of images further continues by routing chosing the optimised path reducing time delay.

Fig. 1 gives a visualisation of how the SQM node will appear. The working and circuitry can be easily depicted ,divided into sensing layer, processing and communication layer with sub divisions for carrying out the assigned process respectively, as shown. The mounted high resolution camera starts rotating for every 60 seconds. The camera covers 6 degrees for every second approximately.

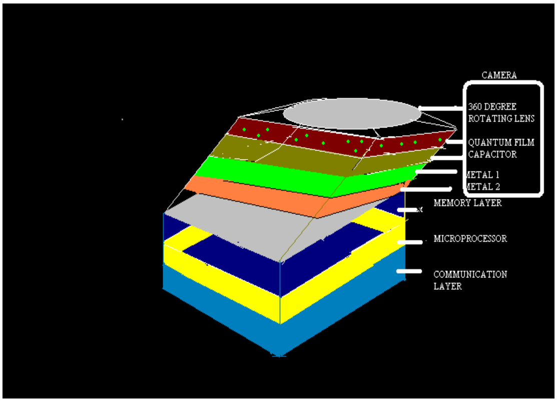


Fig. 1. Smart Quantum Multimedia node Circuitry(SQMn)

Scenario 1

A mishap or an intruder at an angle of 30° will be covered after the rotation has completed 5 seconds for the start. The video or static image is processed with the help of SQM node principally by calculating temporal power and the packets are routed by chosing optimal path to decrease the redundancy.

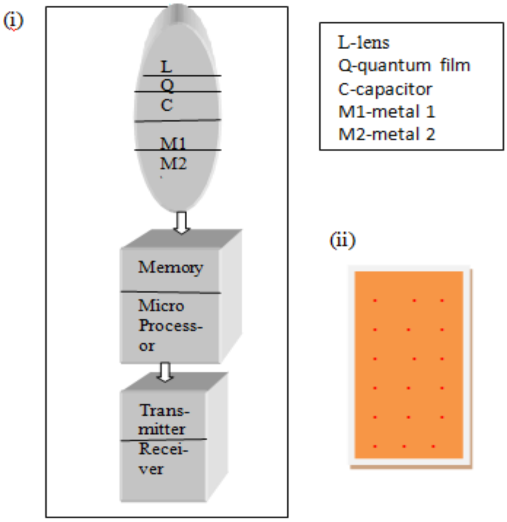


Fig. 2. (i) Flow of data in SQM node and (ii)cross-section of quantum film

3.3 Localisation of Smart Quantum Multimedia Node(SQMn)

Smart Quantum Multimedia node (SQMn) have appreciable funtionality when it is deployed by the principle of localisation.according to which, the number of high power SQM nodes to be installed depends on the coverage area which may vary to suit the situations. Each SQM node covers an area of 300 metres with its 360° rotation in the horizontal scale through all the directions. The important constraint is that the different SQM nodes are set up at different heights in order to get different views of the vicinity paving way for comparison of images and serving the purpose of analysis.

For instance,four different SQM nodes are deployes at a distance of 1 kilomtetre separation between them. Nodes are deployed at different heights such as 40 centimetre, 80 centimetre,120 centimetre and 160 centimetre above the ground surface. The images and videos captured by different SQM nodes are compared at the destination which is reached by tracing of optimal path through different sensor nodes which are deployed according to the location of SQM nodes.This serves to be of great importance in analysing the vicinity for nuances.

After the localisation of high power SQM nodes, other sensor nodes which are the main element in routing and packet scheduling are deployed as a cluster .The number of nodes to be distributed in a cluster is of random value ensuring that SE nodes are of greater number.depending on the number of sensor nodes and energy of each node, optimal path is chosen and the video frames reach the destination reducing redundancy and improving the quality of the end image.

3.4 Image Capture and Compression

The quantum film technology works by the attachment of very small quantum dots, which are minute semiconductors absorbing light and emitting electrons within a special polymer film. The film looks completely black and it is spun over a traditional CMOS wafer, this is used to capture the image and the technique is four times more efficient than the silver grains in photographic film.

Compression of images is done by intraframe compression technique. A video is the series of frames together. Each frame is compared with the next and only the difference between them is stored instead of utilising memory by storing the frame itself.this compression enables fast routing process , which inturn improves quality and speed of the outcome.

3.5 Multicast Multimedia Packet Scheduling

The ultimate usage of multicasting is to provide the end user with a better quality of service compared to other scheduling process and optimal utilisation of the network resources .Based on the energy and of the subsequent nodes,packet scheduling takes place by multicast process. The idea is to convey the same data to different destinations so that duplication of data and hiding of data is overcome.

The subsequent nodes that are set up in the vicinity area are allocated random locations and they are of two kinds based on energy Intermediate Energy nodes(IEn) and Super Energy nodes(SEn).low energy nodes are not deployed as it increases delay and causes redundancy of data. The capacity to transfer data or the Quality of service provided to the end user is a function of separating distance between the nodes, sum of energy of the nodes and the function of receiving time.

4 Critical Factor

Let E_i be sum of all the energies of the nodes chosen in the next hop. D_i is the sum of all distances of the nodes and T_k be the sum of all receiving time in each node chosen in the next hop. H_i is sum of energy factor and distance factor(E_i and D_j) The term C is denoted as the critical factor .Critical factor is the sum of the factors such as H and T_k respectively. The notations i*, i,j,k represent the number of nodes chosen in each cluster which are designed to be equal.

5 Energy Factor Based on Temporal Power

Temporal power of the node is based on the node history. It is the apparent power of the node or the capacity of the node to support signal transfer. A node has super high energy(SEn) in general when deployed. After it supports the transmission of signal once,it loses some of its energy depending on the strength of the signal. It charges itself for better performance for the support next time. During the charging process when it is approached by a signal,it does not reject but supports the signal acting as Intermediate Energy node(IEn).

The energy factor, sum of all energies of subsequent Intermediate Energy nodes(IEn) nodes covered by each SQM node towards all the destinations is given by

$$\mathbf{E_i} = \sum_{i=0}^{i=\max} \mathbf{E(i)} \qquad \mathbf{i \in 0 \leq \infty}$$

where i represents number of IE nodes chosen .

The energy of SE nodes is

$$\mathbf{E_{i*}} = \sum_{i*=0}^{i*=\max} \mathbf{E(i*)} \qquad \mathbf{i* \in 0 \leq \infty}$$

where i* represents the number of SE nodes chosen.

Distance of separation or the disatnce factor can be calculated from

$$\mathbf{D_j} = \sum_{j=0}^{j=\max} \mathbf{D(j)} \qquad \mathbf{j \in 0 \leq \infty}$$

The term H is the product of energy factor E_i and distance factor D_j

$$\mathbf{H = (E_i + E_{i*}) \times D_j}$$

the time factor is sum of receiving time of all the nodes in the cluster.

$$\mathbf{T_k} = \sum_{k=0}^{k=\max} \mathbf{T(k)} \qquad \mathbf{k \in 0 \leq \infty}$$

Critical factor is the sum of term H and time factor T_k

$$\mathbf{C = H + T_k}$$

5.1 Routing Path Discovery Based on Temporal Power and LPT

Each node group consisting of four subsequent nodes in the second hop if four SQM nodes are deployed has a particular critical factor based on which routing path is chosen. If another node group has approximately equal critical factor then the node group to be chosen depends on Last Performance Time (LPT).

Let c_i, c_j etc be the critical factors of different node groups. If two critical factors of the node groups are almost equal. The Last Performance Time denotes the seconds for which it has been idle after getting charged completely. The node with lower LPT value is given higher priority as it is in fully charged state with less amount of energy dissipated to the surrounding due to the external factors . a node with higher LPT value will have been idle for a comparitively longer time. Hence it would definitely be affected by surrounding at a greater rate than the node with lower LPT value.

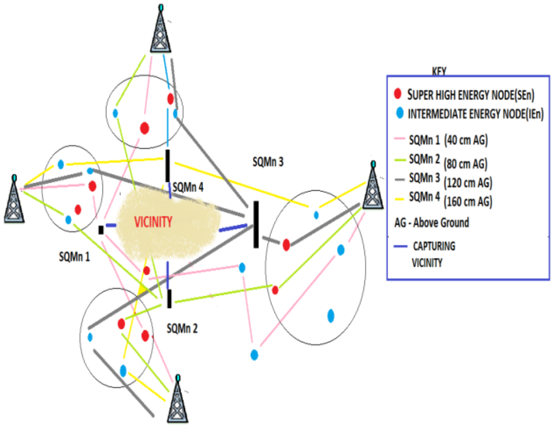


Fig. 3. Routing path discovery based on temporal power and LPT

In fig.3 Effective path for routing is judged from various factors such as critical factor based on temporal power and Last Performance Time (LPT) value.

Video packet scheduling is the next process that follows. Each video is divided into number of frames and each frame is further branched into several packets. Each video frame has its impact value upon which it can be characterised. repetition of video data in any form is omttied during the transmission. packet scheduling algorithm provided the path to be selected with optimum pattern so as to reduce the energy consumption and also loss.

6 Performance Evaluation

6.1 Simulation Setup

Assume 4 Smart Quantum nodes (SQM) are set with a relative distance of 32 metres between them forming an irregular polygonal shape. The smart nodes are setup with varying heights to provide a perspective view of the vicinity. The heights at which they are set up can be 40 cm, 80 cm, 120 cm and 180 cm above the ground. This setup facilitates capturing of images which is approximateely equal to an average human's height. To support each of the four SQM nodes, four SE nodes are set up at random locations so that they are suitable to reach the four different destinations. More than one destination is considered so that data is not duplicated and hiding of data is also avoided to a great extent.

6.2 SCENARIO

It is assumed that all the four SQM nodes start capturing images of the vicinity when there is a mishap or an intruder in the path of the site. Each node in different clusters are chosen so that each

SQM node situated at different heights is transmitted to all the four destinations. Signal from one SQM node reaches one node of each cluster.the signal from each of the node is transferred to the different destinations.critical factor is calculated based on energy factor, distance separating the nodes and receiving time.

The main task involved is to check the node to which data will be transferred next.once that is found out in each direction, and then the combining factor or the relating factor of all the nodes involved in transmission of data from one node is calculated. Based on this value, different set of nodes come into groups from different clusters. The selected nodes receive data from the main SQM node which is sending information currently and takes it to the next step closest to the destination or even to the destination straight if it is very close to the destination.

In the SQM node, each video is divided into number of frames and each frame is divided into number of packets. Each video frame has it own characteristic impact value based on which it is characterised. Repeated video packets are omitted during the transmission. By the usage of packet scheduling algotrithm, the correlation among different reference frames is selected and this selects the path with minimal amount of distortion which is automatically decided by the critical factor. This process is neither power nor time consuming because of the involvement of quantum dots which are a repository of energy.average life time of a node is the sum of the energy E(node)available and its capacity ,CS(node) to serve a signal of particular strength.

Average life time=E (node)+CS(node)

In the Fig 4, the average life time of quantum dot sensor node which is the Smart Quantum node (SQM) is compared with other image sensors such as CMOS and CCD. CCD is the acronym of Charge Coupled Device. It records the images based on the charge in each pixel.this requires continous supply of energy and hence results in highest consumption of energy. Each row of pixels are read and coupled to an amplifier which further amplifies the signal. And the point to note is that the pixels are read row by row which increases the time and also energy consumption.CMOS is the Complimentary Meta Oxide Semiconductor.here, tiny circuitry is etched on to the chip. Each pixel requires the energy of a transistor connected to it. It shows fair results compared to CCD but the results are not better compared to Quantum film image sensor.

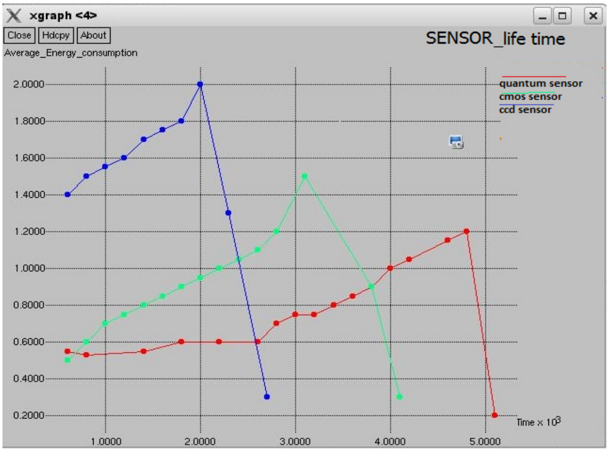


Fig. 4. Average life time of quantum sensor node

7 Conclusions

The paper presenting the architecture of Quantum Dot Sensor for Image capturing and Routing based on Temporal Power(QDSTP), a scheme for efficient and high resolution video capturing in 360^o and communication over a set of nodes to wide ranges of destinations comprises of the main sensor node, known as the Smart Quantum Multimedia node, packet scheduling algorithm which ensures that only the nodes with the highest critical factor which is decided by the sum of energy , path separation and receiving time . they are calculated based on the temporal power and Last Performance Time(LPT) is chosen for routing. Moreover, the proposed scheme utilizes an intelligent video packet scheduling algorithm which selectively drops non significant packets prior to their transmission hence it improves the video transmission rate.

The end quality of image is very high due to the usage of quantum films embedded with dots technique. Quantum films enhance the resolution and since the sensed data reaches many destinations, duplication of data and hiding is avoided providing the user with high quality images.

References

[1] Dowling, J.P.: Quantum sensors, computing, metrology and imaging (2011)

[2] Newell, A., Akkaya, K.: Department of Computer Science Southern Illinois. In: Self-actuation of Camera Sensors for Redundant Data Elimination in Wireless Multimedia Sensor Networks Communications (2009)

[3] Vithya, G., Vinayagasundaram, B.: Actuation sensors with Adaptive Routing and QOS aware checkpoint arrangement on multimedia sensor network. In: ICRTIT 2011 (2011)

[4] Zafar, K., Baig, A.R.: Optimisation of route planning and Exploration using multi-agent system (August 2010)

[5] Hoang, T., Beetz, J., Lermer, M., Kamp, M., Hofling, S., Balet, L., Chauvin, N., Li, L., Fiore, A.: Single Photons emitted by Single Quantum Dots into waveguides: Photons Guns on a Chip (2011)

[6] Cheng, S.-Z., Wang, C.-Y., Tao, M.-H.: Quantum Communication For Wireless Wide Area Networks (July 2005)

[7] Liu, X., Wu, X.M., Ren, T.L.: Thin film conductivity metrology using Photoluminescence of quantum Dots (June 2011)

[8] Ratan, R., Shukla, M.K., Oruc, A.Y.: On Random routing and its Application to Quantum Interconnection Network (2006)

[9] Akyildiz, I.F., Melodia, T., Chowdhury, K.R.: A Survey on Wireless Multimedia Sensor Network (November 2006)

[10] Li, J.-S., Yang, C-F(co author): Quantum communication in distributed wireless sensor networks. Inst. of Comput. & Commun. Eng., Nat. Cheng Kung Univ., Tainan, Taiwan

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

Checkpointing and Recovery Using Node Mobility among Clusters in Mobile Ad Hoc Network

Suparna Biswas[1] and Sarmistha Neogy[2]

Department of Computer Science & Engineering
[1] West Bengal University of Technology, India
mailtosuparna@gmail.com
[2] Jadavpur University, India
sarmisthaneogy@gmail.com

Abstract. In this paper we propose a new mobility aware checkpointing and failure recovery algorithm for cluster based mobile ad hoc network (MANET). Here we introduce a parameter ‘cluster-change-count’, maintained by each member node to count number of clusters a mobile node traverses through during a single checkpoint interval. A mobile node increments ‘cluster-change-count’ by 1 each time the mobile node leaves a cluster and joins another. Each mobile node saves a checkpoint independently if its ‘cluster-change-count’ exceeds a predefined threshold. This measure is important because each mobile node saves logs, important data required for its recovery at different cluster heads it traversed through. If the node fails, these data are to be searched and retrieved for recovery along with last saved checkpoint. This search and retrieval cost of dispersed data of a failed node increases with increasing ‘cluster-change-count’ and gets added to total recovery cost of a failed mobile node. To limit this cost, a threshold value of ‘cluster-change-count’ is set. In MANET no node has stable storage. This makes checkpointing in MANET more challenging. Checkpoint placement needs to be done efficiently to ensure minimum recovery cost and improved recovery probability. We have analyzed performance of the proposed algorithm.

Keywords: checkpoint, clustering, mobile adhoc network, recovery.

1 Introduction

Rapid development of communication technology from wired to wireless network led almost all service oriented systems to “all time everywhere” service from “anywhere anytime” service. This has been possible due to stabilization of portable devices e.g. laptop, smart phones, mobile phones and advancement of wired communication to wireless infrastructured as well as wireless infrastructureless communication. Todays portable devices are equipped with sufficient resources hence can do computing as well as communication while on the move. In this kind of systems,

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 447–456.
springerlink.com

computing devices as well as communication links are failure prone. Hence applications running on the mobile nodes must be fault tolerant. Checkpointing is an established technique to provide fault tolerance in wireless cellular network [2],[3],[4],[5],[6]. As mobile adhoc network has some limitations and unique characteristics [1], traditional checkpointing algorithms [2],[3],[4],[5],[6] suitable for distributed and mobile systems need to be modified to implement in MANET. Here checkpoint placement is an important issue for successful recovery of failed applications with minimum cost. MANET applications has been extended to disaster management, war zone, collaborative researchers working in remote areas e.g. coal mines etc. Research in checkpointing in MANET has been a subject of interest of many researchers. Some of the existing literatures in checkpointing in MANET is being discussed here.

In [7], P.K.Jaggi et.al.presented Staggered Checkpointing and Recovery in Cluster Based Mobile Ad Hoc Networks. In their work, concurrent checkpoint initiation can cause stable storage access contention over limited bandwidth in MANET. This has been eliminated by staggering checkpoints. In [8], Juang and Liu presented an efficient asynchronous recovery algorithm in cluster based MANET. In this work they have studied the fundamental problem of crash recovery in mobile distributed environment. Yi, Heo, CHo and Hong presented an adaptive mobile checkpointing facility for wireless sensor networks in [9]. As a node cannot save all the checkpoints of its neighbor nodes due to limited storage space, each node can decide adaptively to save checkpoint or not depending on a probability as mentioned in this work.

Studying related works, we find following problems that still exists:

- Most of the works do not address the issue of saving checkpoints in absence of stable storage in MANET except in [9].
- Above mentioned related works do not consider random mobility of cluster members and cluster heads as a factor of saving checkpoints.

Our Contribution: a) A mobile host saves back up copy of a checkpoint in one of its neighbor node that has sufficient resources as no node has stable storage.

b) Each mobile host saves checkpoint based on its movement from one cluster to another. To ensure minimum clusterhead change, the clusterhead selection algorithm in [10] is used. c). Backup copy of checkpoint of a cluster member is saved in its neighbor node that have sufficient resources, logs and other data required for recovery are saved in cluster head, log record of received computation messages are saved in gateway node, coordination at the time of recovery of a failed node is done by cluster head. This reduces cluster head's overhead.

1.1 Data Structures and Notations

CH_j= Cluster head, j=1……m, **G_k**= Gateway nodes, k= 1……p, **mh_i** , i = 1.......n, **n** = number of Mobile Hosts, **c3**=cluster_change_count, **c3_th** = threshold of cluster change count, **intv** = checkpoint interval, **seq** = sequence number of computation message, **mh_dep [ ]** = during current interval, list of mhs from which computation messages received, **CH_mh_list[ ]** = list of mobile hosts connected to a clusterhead currently, **CH_mh_list[i]** = 1, mh$_i$connected, **CH_mh_list[i]** = 0, mh$_i$disconnected, **CH_mh_list[i]** = -1, mh$_i$failed, **CH_traversed_list[ ]** = Array of cluster heads through which a mobile host traverses during a checkpoint interval, **mhi_s** = mobile host that sends computation message, **mhi_r** = mobile host that receives computation message, **m$_c$**= computation message, **m$_{co}$** = coordination message, **log** = log$_{s,\ r,intv,seq}$ = **s**=sender mh, **r** = receiver mh, **intv**. = interval, **seq** = sequence number of computation message, e.g. **log $_{(1,*,\ 1,\ *)}$: s = mh$_1$, intv = 1, r = seq = 'any'**, **UID** = UID $_{s,\ r,intv,seq}$= s=sender mh, r = receiver mh, intv. = interval, seq = sequence number of computation message, e.g. UID $_{(*,\ 1,\ 1,*)}$: r = mh$_1$, intv = 1, s = seq = 'any'

Checkpoint_ file:	MH_id, CH_id Process status Data, intv	**Log_file:**	log$_{s,\ r,intv,seq,}$ copy of m$_c$

1.2 Assumptions

1. Each mobile host saves last checkpoint, logs and log records of current checkpoint interval at any instant of time.
2. The mobile host that saves checkpoint and the mobile host that carries the copy will not fail simultaneously
3. Application process runs on nodes, application process failure is generally termed as node failure for simplicity.

2 Proposed Checkpointing and Recovery

During a checkpoint interval mobile node computes, sends and receives computation messages to and from other mobile nodes. A mobile node saves log of sent computation message and forwards to current cluster head. The mobile node that receives computation message adds sender mobile node in its dependence list. Dependent list is saved per mobile host, per checkpoint interval basis. A cluster head saves logs of its member nodes, coordinates recovery of its failed member nodes besides own computations. A mobile node increments 'cluster-change-count' by 1 each time the mobile node leaves a cluster and joins another. Each mobile node saves a checkpoint independently if its 'cluster-change-count' exceeds a predefined threshold. If a mobile node fails before saving a checkpoint during current checkpoint interval, the mobile node rolls back to last checkpoint, replays logs, sends request message to the mobile nodes saved in its dependent list to replay received computation messages before failure. The recovered node restarts execution.

In figure 1 we consider a small cluster based adhoc network consisting of three clusters, each have a single cluster head and multiple cluster member nodes. A member node can communicate with another member node of different cluster through cluster head. Cluster heads communicate with each other through gateways.

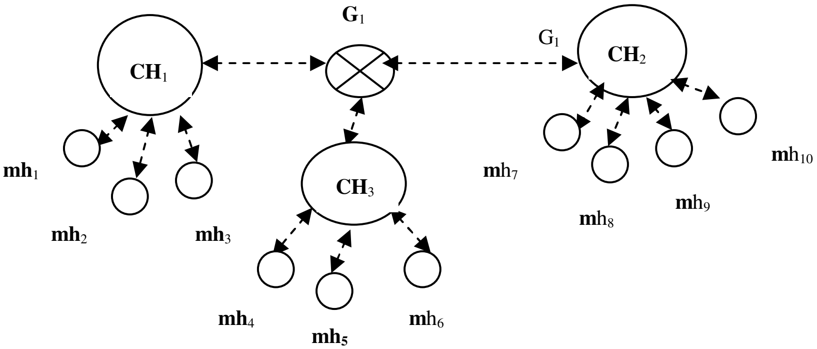


Fig. 1. Cluster based adhoc network consisting of 3 clusters, 3 cluster heads, 1 gateway and 10 cluster members

Here for $mh_1$and CH_1specific data structures it saves initially are described here, these are true for all other mobile hosts and cluster heads.

mh_1: dep_list [ ] =

1	2	3	4	....	n
1	**0**	**0**	**0**		**0**

mh_1 : CH_traversed_list [ ] =

1	2	....	c3_th
CH_1^1			

mh_1: $C^3 = 0$, log = empty, UID = NIL, chk_intv = 1

CH_1: CH_mh_list[ ] =

mh_1^1		

We consider following scenariso of different non-deterministic events : mh_1 is connected to CH_1 . mh_1 sends computation message to mh_3, saves $\log_{(1,\ 3,\ 1,\ 1)}$ and sends to CH_1, mh_3 receives computation message from mh_1, updates dependence list, saves $UID_{(1,3,1,1)}$ and sends to G_1, Mh_1 sends 'leave message' to CH_1 along withcopy of dependence list and CH_traversed_ list then leaves, mh_1 joins CH_2, updates CH_traversed_ list, mh_1increments cluster_change_count by 1, checks if cluster_change_count exceeds threshold (NO), mh_1 receives computation message from mh_4, updates dependence list, saves $UID_{(4,1,1,1)}$ and sends to G_1, mh_4 saves $\log_{(4,\ 1,\ 1,\ 1)}$, sends to CH_3, mh_1 sends 'leave message' to CH_2 along withcopy of dependence list and CH_traversed_ list then leaves, increments cluster_change_count by 1, mh_1 joins CH_3, updates CH_traversed_ list, mh_1 increments cluster_change_count by 1, checks if cluster_change_count exceeds threshold, (YES), mh_1 saves checkpoint, saves a copy of it in its neighbor suitable as per criteria described in section 2, mh_1sends ID of mh_1_neighbor to its current checkpoint : CH_3, mh_1sends delete_ log $_{(1,*,\ 1,\ *)}$, delete_UID$_{(*,\ 1,\ 1,*)}$to CH_3which broadcasts to other cluster heads through G_1. G_1 separates log $_{(1,*,\ 1,\ *)}$ from delete_ log $_{(1,*,\ 1,\ *)}$, compares with saved UIDs in it, finds a match with $UID_{(1,3,1,1)}$, G_1does not forward delete_ log $_{(1,*,\ 1,\ *)}$message, sends a message to CH$_{1,}$ first element in CH_traversed_list of mh$_1$to forward saved log that matches with log $_{(1,*,\ 1,\ *)}$ to G$_{1,}$ CH$_1$ forwards log$_{(1,\ 3,\ 1,\ 1)}$ to G$_{1,}$ G$_1$saves log$_{(1,\ 3,\ 1,\ 1)}$because the UID$_{(1,3,1,1)}$ is still not deleted i.e. mh$_3$has not saved checkpoint of the receive event of the computation message, G1 deletes all the UIDs that matches with UID$_{(*,\ 1,\ 1,*)}$, mh$_1$enters next checkpoint interval, refreshes CH_traversed_list, refreshes dep_list, C^3 = 0.

2.1 Algorithm

1. mobile hosts compute, communicate with each other through message passing.
2. Each mobile host sends a beacon message by which its authenticity and connectivity is checked.
3. If mh$_{i_s}$wants to send a m$_c$ to mh$_{i\_r}$, mh$_{i_s}$saves a log, forwards m$_c$ and log to current cluster head. c_ch saves log in its memory, checks own ch_mh [ ], if finds ‘1’ in corresponding array field then mh$_{i\_r}$ is in cluster 1, if finds ‘0’ then mhi_r is in other clusters, if finds ‘-1’, then mh$_{i\_r}$ failed.
   case 1: mh$_{i\_s}$ and mh$_{i\_r}$ in same cluster:
   c_ch forwards m$_c$ to mh$_{i_r}$andmh$_{i_r}$sends ack to mh$_{i\_s}$ throughc_ch.
   case 2: mh$_{i\_s}$ and mh$_{i\_r}$ in different cluster:
   c_ch of mhi_s broadcasts a look up (mh$_{i\_r}$) message to all cluster heads through gateway nodes. All cluster heads searches their corresponding ch_mh [ ]. The CH that founds 1 replies to c_chof mh$_{i\_s}$, c_ch forwards m$_c$ and UID of m$_c$ gateway node which saves UID of m$_c$ and forwards m$_c$ to reply ch, ch forwards to mh$_{i\_r}$.
   case3: mh$_{i\_r}$ fails:
   mh$_{i_s}$stops sending m$_c$ temporarily, will try to send later.
4. mh$_{i\_r}$ receives m$_c$,updates dep_list, sends UID of m$_c$ to gateway node.
5. mh$_{i_s}$moves randomly, leaves current cluster and joins another cluster .
6. In cluster based adhoc network each mobile host increments cluster_change_counter (c3) each time a mobile host changes cluster.
7. The mobile host saves ID of cluster heads being traversed in CH_traversed_list[ ].
8. Before the mobile host leaves a cluster head, sends different data structures e.g. dep_list, CH_traversed_list etc. saved being connected to current cluster head along with a leave message to current cluster head.
9. The mobile host saves ID of previous cluster head in CH_traversed_list[ ], e.g. CH_traversed [0] = CH$_1$, if mh leaves CH$_1$.
10. The mobile host next joins another cluster head say CH$_2$
11. repeat steps 2 to 6 .
12. The mobile host saves ID of previous cluster head in CH_traversed_list[ ], e.g. CH_traversed [1] = CH2, if mh leaves CH2.
13. Repeat steps 2 to 6.
14. The mobile host saves ID of previous cluster head in CH_traversed_list[ ], e.g. CH_traversed [2] = CH3, if mh leaves CH3.
15. If c3 > c3_th, mobile host invokes checkpoint procedure ( ).
    15.1.The mh takes a snapshot of current state of computation of running application process.

15.2 Checkpointing mh saves a back up copy of checkpoint in one of its neighbor with highest remaining memory and the mh sends check point_backup_node = ID of$mh_{neighbour}$ to CH_2.

16. The mobile host sends delete_logof prev_interval message to CH_2 which forwards this message tothe cluster heads that are saved in CH_traversed_list of the mobile host through gateway nodes.

16.1 If any gateway node finds a match between log_ID and UID, sends do_not_delete message to the cluster head.

16.2 The cluster head forwards log to the gateway node.

16.3 Gateway node saves log.

17. The mobile host enters into next checkpoint interval.
18. The mobile host saves current cluster head in its CH_traversed_list[ ]
19. repeat steps 2 to 6 till $c3 \leq c3\_th$.
20. The mobile host fails, last checkpoint can be recovered from the host itself or backup node

20.1 **CASE1:** Last checkpoint can be recovered from the failed host itself, the host rollsback upto last checkpoint.

20.1.1 The failed host is mhi_s, Replay_ log (UID) message to c_ch, c_ch checks if the log is saved in its memory, if yes replays log to mhi_r, else broadcasts replay_log (UID) message to the chs saved in failed host'sch_traversed_list.

20.1.2 The failed host is mhi_r, Receive_m_c (UID) message to c_ch, c_ch checks if the log is saved in its, memory, mhi_r receives mc from, c_ch, else c_ch broadcasts receive_$m_{c,}$ (UID) message to the chs saved in mhi_s' ch_traversed_list.

20.1.3 The failed host is both mhi_s and mhi_r, Do tasks in case 1 and case2

20.1 **CASE 2:** Last checkpoint needs to be recovered from backup node, Current cluster head sends recovery_message of the mobile host to the cluster head saved in first field of ch_traversed_list, The cluster head finds ID of $mh_{neighbor}$ of failed node, retrieves copy of last checkpoint of failed node and forwards to current cluster head of failed node. The cluster head forwards copy of checkpoint to failed node.

21. The failed host rollsback upto last checkpoint and recovers following any of the above three cases .

2.2 Correctness Proof

Theorem 1: The algorithm ensures consistent recovery

Proof: With the help of following two lemmas the above statement can be proved.

Lemma 1: There is no orphan message.

Proof: save log of each sent $m_{c,}$ TRUE for $\forall$ mh
if (mh fails without saving checkpoint)
sent m_c is retrieved from saved logs, TRUE for $\forall$ sent m_c

Lemma 2: There is no lost message.

Proof: save UID of each received $m_{c,}$ TRUE for $\forall$ mh
if (mh fails without saving checkpoint)
received m_c is retrieved according to saved UIDs, TRUE for $\forall$ mh

Theorem 2: Proposed algorithm causes minimum checkpoint and log overhead per mobile host per checkpoint interval

Proof: Each mh saves log, $\forall$ sent computation message, interval=current
Each mh saves UID, $\forall$ received computation message, interval=current
Each mh saves checkpoint, C3>C3_th, delete logs and UIDs, interval=next
Each mh saves log, $\forall$ sent computation message, interval=current
Each mh saves UID, $\forall$ received computation message, interval=current
Each mh saves checkpoint, C3>C3_th, delete previous checkpoint, delete logs and UIDs, interval=next
So, per interval each mh saves checkpoint, logs and UIDs of single checkpoint interval.

2.3 Performance Analysis

We have done simulation by c programming. Adhoc mobile environment is simulated by creating non-deterministic events randomly. Simulation parameters are:

parameter	value
Checkpoint Size	2000 B
Log Size	50 B
computation message size	50 B
coordination message size	2.5 B
UID size	2.5 B
time to transfer checkpoint per hop through wireless channel	0.08s
time to transfer log or computation message per hop through wireless channel	0.002s
time to transfer coordination message or UID of a log.	0 .0001s

- cluster change count(c3) vs. failed node's recovery time (assuming mobile host fails when c3 = c3_th):

As mh moves from one cluster to another cluster, c3 increases by 1and distance between old and new CH also increases by 1 hop. Here we assume that a node fails when c3 = (c3_th). At this point recovery information, logs have to be retrieved from (c3_th) number of cluster heads and checkpoint will be transferred from the cluster head which is at a distance of (c3_th) number of hops from recovery CH. Thus log transfer cost and checkpoint transfer cost both will be maximum in a single interval as per our assumption. So it is justified that if checkpoint is saved based on threshold of c3 then maximum recovery cost can be calculated beforehand. This will help to set the value of (c3_th) as per application.

Recovery time = cost to transmit recovery request message to (c3_th) no. of CH + cost to transfer logs saved in c3_th no. of CH + cost to transfer last checkpoint from the CH which is at a distance of c3_th hops

$$= (\text{Cm}_{\text{co}} * (\text{c3\_th}) + \text{Clog\_transfer} * (\text{c3\_th}) + \text{Ccheckpoint\_transfer}*(\text{c3\_th}) \text{ hops }) \text{ unit}$$

$$= ((\text{Cm}_{\text{co}} + \text{Clog\_transfer} + \text{Ccheckpoint\_transfer}) * (\text{c3\_th}))\text{unit}$$

$$= ((0.0001 + 0.002 + 0.08) * (\text{c3\_th})) \text{ unit}$$

$$=(0.0821 * (\text{c3\_th})) \text{ unit} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{(i)}$$

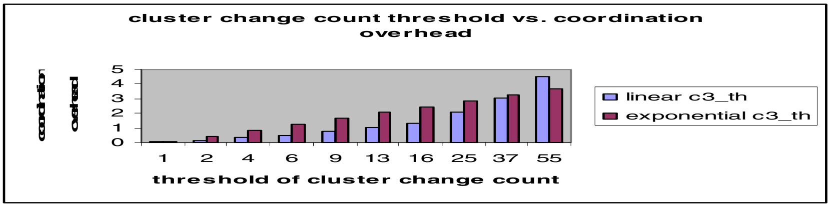


Fig. 2. Maximum coordination overhead at the time of recovery varies with cluster change count threshold

A mobile node fails when c3= (c3_th), just before saving checkpoint. Hence number of clusterheads where the node's recovery information are scattered is c3_th. For simplication we consider that distance between last clusterhead where the ID of the clustermember node that saves backup copy of last checkpoint of failed node is saved and the clusterhead where the failed node recovers is (c3_th) hops. Here checkpoint transfer cost is bounded by (c3_th).

- Number of cluster member nodes vs. coordination overhead :

Coordination message for Checkpoint recovery

= recovery message$_{\text{mh-CCH}}$ + transfer_checkpoint_message$_{\text{CCH-CH}}$ + transfer_checkpoint_message$_{\text{CH-mh}}$

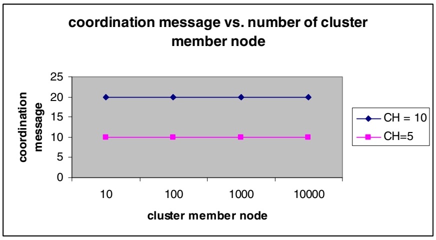
$= \text{m}_{\text{co}} + \text{CH}* \text{m}_{\text{co}} + \text{m}_{\text{co}} = \text{m}_{\text{co}}(1+\text{CH}+1) \approx \text{CH}$ (for high value)

coordination message for Log recovery

=log recovery message$_{\text{mh-CH}}$+ transfer log message$_{\text{CH-CH}}$

$\approx (1+\text{CH})*\text{m}_{\text{co}} \approx \text{CH}$, hence total coordination message = CH + CH = 2*CH

Hence total coordination message ∞ CH ....(ii)



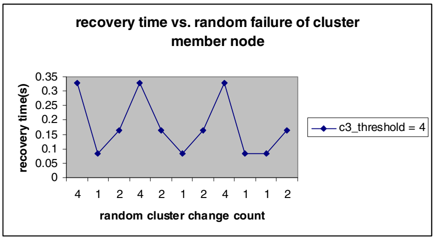


Fig. 3. Recovery coordination message varies with cluster head irrespective of number of cluster member nodes

Fig. 4. Recovery time of a cluster member node that fails randomly

In figure 3 it is shown that coordination among different nodes at the time of recovery of a failed node is restricted to cluster head level.

Recovery time vs. random failure of a cluster member node:
Here cluster change count threshold is set to 4. Cluster change count is set initially as 0. From equation (i) recovery cost of a failed node for any cluster change count (c3) is recovery cost = (0.0821 * c3) unit

If a node fails when c3 =0, log and last checkpoint remains in same cluster hence no cost to transfer log, transfer checkpoint. This is justified by above equation. Here we consider that a node fails randomly and value of cluster change count at failure time may be any value between 1 and 4. Accordingly recovery cost will vary and this is shown in figure 4.

3 Conclusion

In this work we propose a mobility based checkpoint and rollback recovery algorithm combined with message logging for failure prone cluster member nodes in a cluster based mobile adhoc network. What makes more challenging the task of checkpointing in MANET than in wireless cellular network is lack of stable storage. In MANET checkpoint and log placement is also a very trivial issue besides saving them. Checkpoint, logs, log record of received computation message etc. are saved in different nodes at different levels to reduce storage overhead of clusterhead and for efficient recovery . Cluster change count threshold can be set depending on mobility rate, failure rate, send and receive computation message frequency etc. This gives flexibility in implementation of this algorithm in different types of MANET with different applications. Failure prone components are attack prone also. Moreover insecurity leads to distrust and vice versa. Security of checkpoints in MANET is important if the application domain is financial transactions, military communication or war field or any other application that demands confidentiality of application data along with fault tolerance. Present work will be extended to make checkpointing trusted in cluster based mobile adhoc network.

References

1. Forman, G.H., Zahorjan, J.: The challenges of Mobile Computing. Journal Computer 27(4) (April 1994)
2. George, S.E., Chen, I.R., Jin, Y.: Movement-Based Checkpointing and Logging for Recovery in Mobile Computing Systems. In: MobiDE, pp. 51–58 (2006)
3. Park, T., Woo, N., Yeom, H.Y.: An Efficient recovery scheme for fault-tolerant mobile computing systems. Future Generation Computer System 19(1), 37–53 (2003)
4. Men, C., Xu, Z., Wang, D.: An Efficient Handoff Strategy for Mobile Computing Checkpoint System. In: Kuo, T.-W., Sha, E., Guo, M., Yang, L.T., Shao, Z. (eds.) EUC 2007. LNCS, vol. 4808, pp. 410–421. Springer, Heidelberg (2007)
5. Quaglia, F., Ciciani, B., Baldoni, R.: Checkpointing Protocols in Distri-buted Systems with Mobile Hosts: a Performance analysis. In: Workshop on Fault-Tolerant Parallel and Distributed Systems, pp. 742–755 (2006)

6. Prakash, R., Singhal, M.: Low Cost Checkpointing and Failure Recovery in Mobile Computing Systems. IEEE Transactions on Parallel and Distributed Systems 7 (October 1996)
7. Jaggi, P.K., Singh, A.K.: Staggered Checkpointing and Recovery in Cluster Based Mobile Ad Hoc Networks. In: Nagamalai, D. (ed.) PDCTA 2011. CCIS, vol. 203, pp. 122–134. Springer, Heidelberg (2011)
8. Ying, T., Juang, T., Liu, M.C.: An efficient asynchronous recovery algorithm in wireless mobile adhoc networks. Journal of Internet Technology Special Issue on Wireless Internet: Applications and Systems! 3(2), 147–155 (2002)
9. Yi, S., Heo, J., Cho, Y., Hong, J.: Adaptive Mobile Checkpointing Facility for Wireless Sensor Networks. In: Gavrilova, M.L., Gervasi, O., Kumar, V., Tan, C.J.K., Taniar, D., Laganá, A., Mun, Y., Choo, H. (eds.) ICCSA 2006. LNCS, vol. 3981, pp. 701–709. Springer, Heidelberg (2006)
10. Basu, P., Khan, N., Little, T.D.C.: A Mobility Based Metric for Clustering in Mobile Ad Hoc Netwo

Design of Broadband Optical Sources for OCDMA/WDMA Applications

Lakshmi Priya[1], M. Meenakshi[2], and G. Geetha[3]

Department of Electronics and Communication, Anna University,
Chennai, India
priya_isanaka@yahoo.com,
meena68@annauniv.edu,
geetha@annauniv.edu

Abstract. Motivation behind this work is the increasing need for high capacity communication systems. Broadband optical sources are an integral part of multichannel high speed fiber optical communication networks based on all-optical WDM, CDM and WDM. FWM (Four-Wave Mixing) effects and SC(Supercontinuum) phenomenon in fibers are used in the design of broadband sources. The spectral slicing of the broadband spectra has been proposed in literature as a simple technique to create multi-wavelength optical sources for wavelength division multiplexing applications. The objective of this work is to develop an accurate model for simulating FWM and SC based broadband optical spectra and compare their performances. The modeling work is carried out using SIMULINK in MATLAB 7.10.0(R2010a).

Keywords: Four Wave Mixing, Supercontinuum, Non Linearity, Schrodinger Equation, SMF, DSF, PCF.

1 Introduction

FWM is a dominant non-linear effect present in wavelength division multiplexed (WDM) networks. The fiber non-linear effects in a single-mode fiber impose a fundamental limitation on the capacity of multi-channel optical communication systems. But due to its ability of generating new frequency components, FWM being a spectral broadening phenomenon and can be used in designing a source. When two or more wavelengths interact in a nonlinear medium, they give rise to a fourth wavelength which is formed by the scattering of the incident photons, producing the fourth photon. This phenomenon is known as four wave mixing.[8]

Given inputs fi, fj, and fk, the nonlinear system will produce

$$\pm fi \pm fj \ \pm fk \tag{1}$$

Supercontinnum generation is defined as the process of generating new frequency components by focusing intense light in a highly non linear medium. Due to this ability SC effect proves to be advantageous to use in the design of broadband sources.

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 457–466.
springerlink.com

Supercontinuum generation is a process where laser light is converted to light with a very broad spectral bandwidth (i.e., low temporal coherence), whereas the spatial coherence usually remains high. The spectral broadening is usually accomplished by propagating optical pulses through a strongly nonlinear device, such as an optical fiber. Of special interest are photonic crystal fibers, mainly due to their unusual chromatic dispersion characteristics, which can allow a strong nonlinear interaction over a significant length of fiber.[8]

2 Fiber Model

In general, analytical solutions to the full Maxwell wave equation for a nonlinear optical system do not exist. Even numerical solutions to the wave equation are extremely difficult to implement due to the dimensionality of the problem. The vector form of the wave equation is a four-dimensional (three spatial, one temporal), second-order partial differential equation. Thus, approximations based on propagation conditions and experimental results are needed in order to solve an approximate scalar form of the wave equation, i.e. the nonlinear Schrödinger equation.[2]

$$\frac{\partial A}{\partial z} + \beta_1 \frac{\partial A}{\partial t} + \frac{j}{2}\beta_2 \frac{\partial^2 A}{\partial t^2} - \frac{1}{6}\beta_3 \frac{\partial^3 A}{\partial t^3} = j\gamma |A|^2 A - \frac{\alpha}{2} A \tag{2}$$

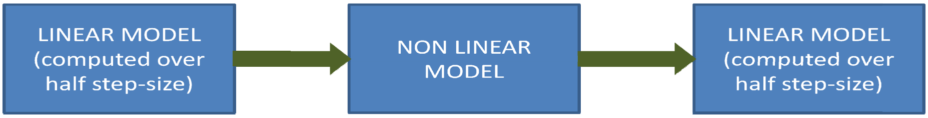

LINEAR MODEL (computed over half step-size)
NON LINEAR MODEL
LINEAR MODEL (computed over half step-size)

Fig. 1. Block Diagram of Fiber

The SSFM is the technique of choice for solving the NLSE due to its easy implementation and speed compared to other methods, notably time-domain finite difference methods. The finite difference method solves the Maxwell's wave equation explicitly in the time-domain under the assumption of the paraxial approximation. The SSFM falls under the category of pseudo spectral methods, which typically are faster by an order of magnitude compared to finite difference methods.[2]

The major difference between time-domain techniques and the SSFM is that the former deals with all electromagnetic components without eliminating the carrier frequency. As shown in the previous chapter, the carrier frequency is dropped from the derivation of the NLSE. Thus, finite difference methods can account for forward and backward propagating waves, while the NLSE derived for the SSFM cannot.

Since the carrier frequency is not dropped in the form of the electric field, finite-difference methods can accurately describe pulse propagation of nearly single-cycle pulses. While the finite difference method may be more accurate than the SSFM, it is only at the cost of more computation time.[2-3]

The mathematical terms due dispersion and nonlinearity are separate and de-coupled in the NLSE. It is this fact that allows the use of the SSFM for solving the NLSE. By looking at NLSE, the operators D and N can be written to correspond to the dispersive (and absorptive) and nonlinear terms respectively

$$\hat{D}= -\frac{\alpha}{2} - \sum_{m=2} \frac{i^{m-1}}{2^{m-1}} \beta_m \frac{\partial^m}{\partial t^m} \tag{3}$$

and

$$\hat{N} = i\gamma\left(\left|E(z,t)\right|^2 + \frac{2i}{\omega_0 E(z,t)} \frac{\partial}{\partial t}\left(\left|E(z,t)\right|^2 E(z,t) \right) \right) \tag{4}$$

Note: N operator multiplies the field solution and is a function of the solution E(z,t). The D operator is a differential operator expressed in terms of time derivatives that operate on E(z,t).

The fiber model designed incorporates the effect of varying β values around central wavelength. Most of the existing models are not inclusive of this effect. Hence our model tends to provide highly precise results which are comparable to the results that one obtains in a real time scenario.

3 FWM Based Broadband Source

3.1 System Model

The block diagram given below depicts the overall source model.

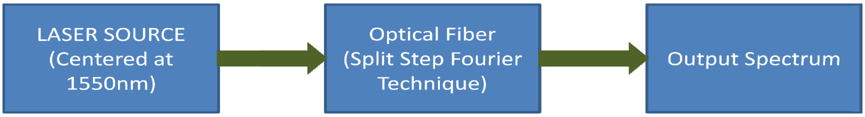
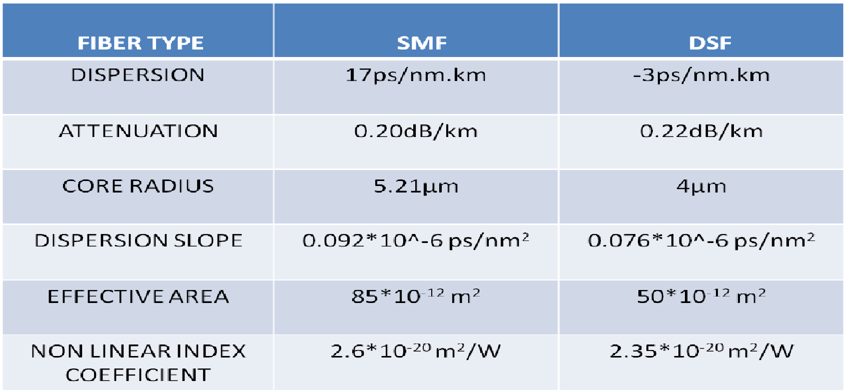


Fig. 2. Block Diagram of FWM based Broadband Source

For the simulation, three laser sources were used at 193.99THz (8.5THz), 193.415THz (8.7THz) and 194.111THz (8.9THz) centered at 1550nm respectively.

3.2 System Parameters

Table 1. Parameters used for fiber design

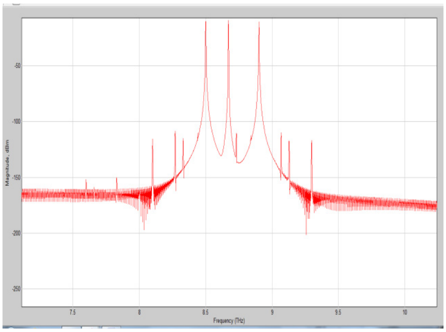


FIBER TYPE	SMF	DSF
DISPERSION	17ps/nm.km	-3ps/nm.km
ATTENUATION	0.20dB/km	0.22dB/km
CORE RADIUS	5.21μm	4μm
DISPERSION SLOPE	0.092*10^-6 ps/nm^2	0.076*10^-6 ps/nm^2
EFFECTIVE AREA	$85*10^{-12}$ m^2	$50*10^{-12}$ m^2
NON LINEAR INDEX COEFFICIENT	$2.6*10^{-20}$ m^2/W	$2.35*10^{-20}$ m^2/W

3.3 Simulation Results

CASE 1: Varying Input Power

Overall length of the fiber considered for simulation is 10m with step-size taken as 1m.



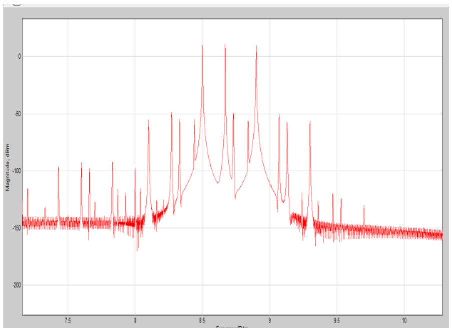


Fig. 3. Spectrum for Pin=100mW

Fig. 4. Spectrum for Pin=1W

From the above simulation results it was observed that for very low values of input power, the effect of FWM could not be observed. But with increase in input power additional components were obtained.

CASE 2: Varying overall Length if the Fiber

The input power is maintained at 100mW, and the following results were observed at the end of the single mode fiber.[4]

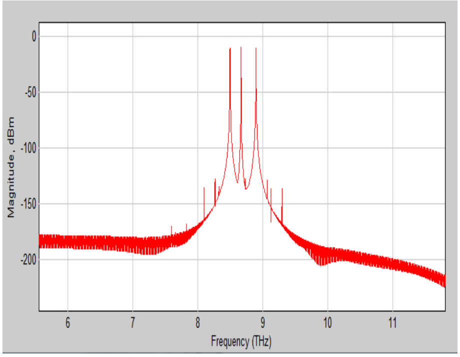


Fig. 5. Spectrum for h=0.1m, L=1m

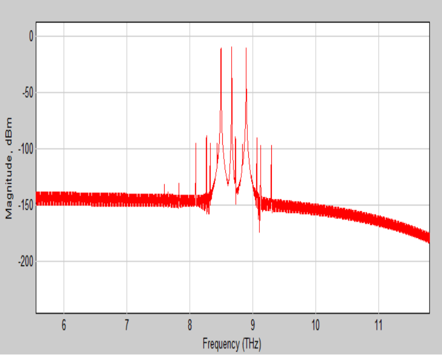


Fig. 6. Spectrum for h=10m, L=100m

From the above simulation results it was observed that for shorter lengths of the fiber, the effect of FWM could not be observed. But with increase in overall length of the fiber along with appropriate increase in step-size additional components could be observed.

CASE 3: Different types of fiber.
The following results were observed with input power as 100mW and length of the fiber was taken 10m with step-size as 1m.

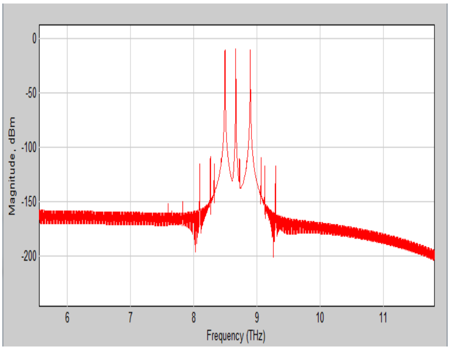


Fig. 7. Spectrum of SMF

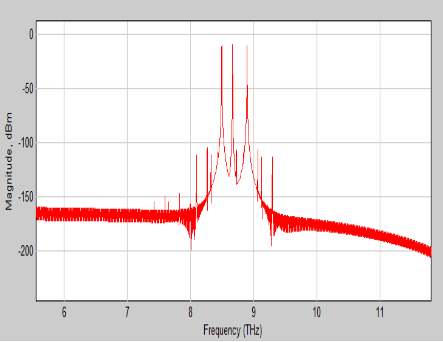


Fig. 8. Spectrum of DSF

From the above simulation results it was observed that for different types of fiber, the effect of FWM differed. The non linear effect of FWM was observed to be more dominant in Dispersion Shifted fiber than in Single Mode fibers.

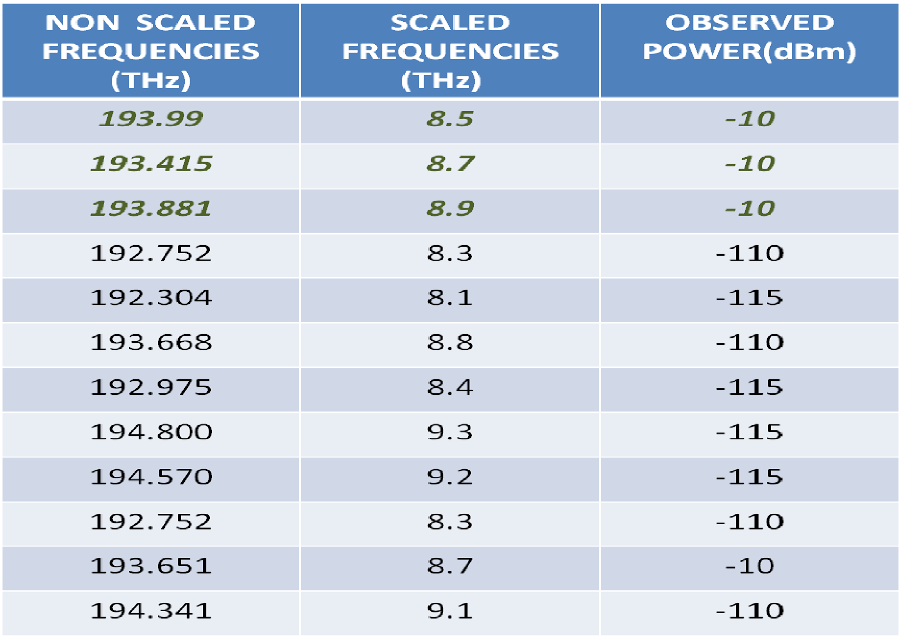
3.4 Theoritical Analysis

For the purpose of theoritical analysis, we consider the case were input power is taken to be 100mW for a length of 10m with step-size of 1m in a single mode fiber.

The figure given below depicts the input spectrum with the components at 8.5Thz, 8.7THz and 8.9THz (scaled frequencies)respectively.

The theoretical power of the FWM components was calculated and compared with the observed values from figure 7 and the results were tabulated as follows.[1]

Table 2. Observed power values of FWM components



NON SCALED FREQUENCIES (THz)	SCALED FREQUENCIES (THz)	OBSERVED POWER(dBm)
193.99	*8.5*	*-10*
193.415	*8.7*	*-10*
193.881	*8.9*	*-10*
192.752	8.3	-110
192.304	8.1	-115
193.668	8.8	-110
192.975	8.4	-115
194.800	9.3	-115
194.570	9.2	-115
192.752	8.3	-110
193.651	8.7	-10
194.341	9.1	-110

The computed theoretical values for the FWM components is around 87dBm.The trend to be observed is that the values for all the components(degenerate and non-degenerate) have approximately similar power values ie. 110-115dBm.

4 Supercontinuum Based Broadband Source

4.1 System Model

The block diagram given below depicts the overall source model.

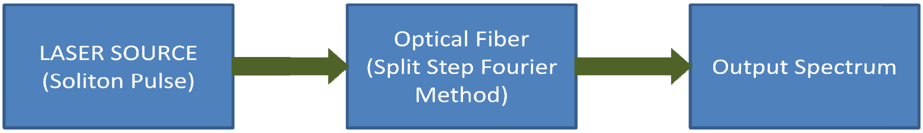


Fig. 9. Block Diagram of Supercontinuum Source

The fiber used here is a Photonic Crystal Fiber with central wavelength as 850nm.[5-7]

4.2 System Parameters

Table 3. Parameters used for PCF design

FIBER TYPE	PCF
DISPERSION	2.5ps/nm.km
ATTENUATION	0.1dB/km
CORE RADIUS	1µm
DISPERSION SLOPE	1*10^-6 ps/nm^2
EFFECTIVE AREA	3.14*10^{-12} m^2
NON LINEAR INDEX COEFFICIENT	3*10^{-20} m^2/W

4.3 Simulation Results

The input pulse is of width 70ps with a hyperbolic secant profile and the overall length of the fiber considered is 25cm and the peak power used for simulation is 100W.[8-9]

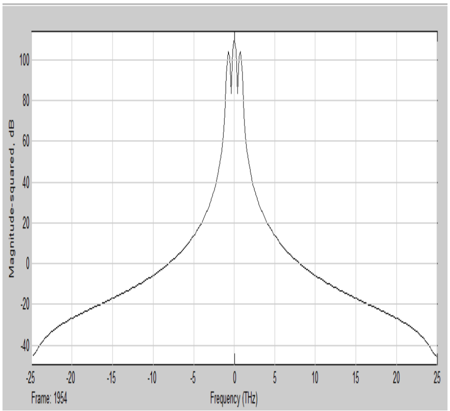


Fig. 10. Spectrum of Input Signal

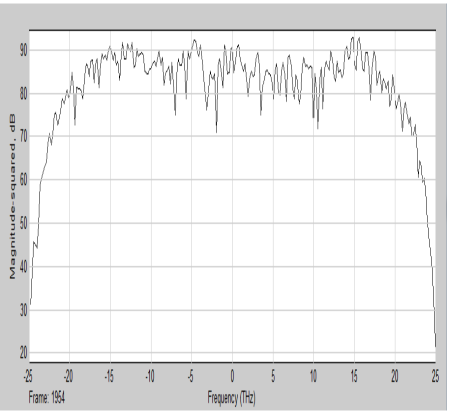


Fig. 11. Output Spectrum

The output spectrum is observed to be broadened widely. This is a combined effect of the spectral broadening phenomena that are prevalent in a optical fiber like Self Phase Modulation (SPM), Cross Phase Modulation(XPM) , Stimulated Raman Scattering(SRS) and Four Wave Mixing (FWM) respectively.

5 Performance Analysis

The two sources designed, namely the FWM based source and SC based source were incorporated in a Hybrid OCDMA/WDMA network and their performance was compared with a arrayed laser source.

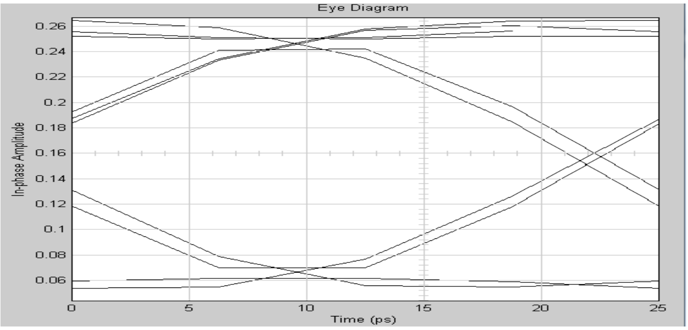


Fig. 12. Eye Diagram of Arrayed Laser Source

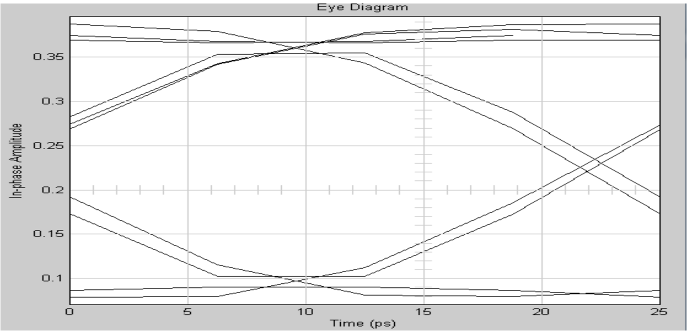


Fig. 13. Eye Diagram of FWM based Source

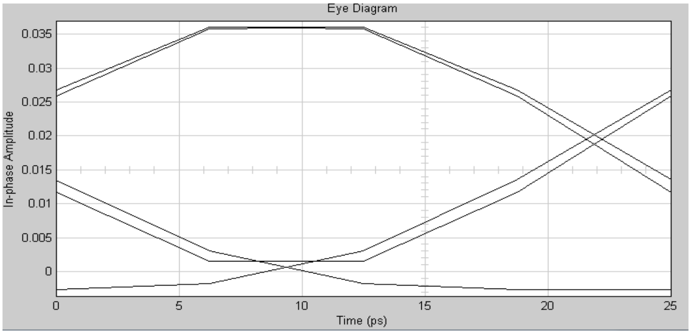


Fig. 14. Eye Diagram of SC based Source

Based on the eye opening width, it can be concluded that FWM based source has wider eye opening in comparison with that of the arrayed laser source and SC based source.

6 Conclusion

From the results obtained through simulation it was observed that the strength of four wave mixing components increases with the increase in input power and also increases with increase in the length of the optical fiber used. Also with increased length we can observe increase in number of components at higher lengths of the fiber due to the possibility of cascaded four wave mixing. The effect of four wave mixing is observed to be more prominent in dispersion shifted fiber than single mode fiber.

In the case of the broadband source designed based on the supercontinuum phenomenon, the output spectrum is observed to be depended on the input power of the hyperbolic secant pulse and the length of the photonic crystal fiber.

The sources designed can be each suited for a different situation. The FWM based source shows superior performance with respect to interference, whereas the SC based source provided a highly broadened spectra and can be ideal for the case where the optical fiber capacity needs to be utilized to the maximum.

7 Future Work

The analysis need to be extended to the case of asynchronous transmission and also spectral efficiency of the sources is to be studied.

References

1. Ram Prasad, A.V., Meenakshi, M.: A study of statistical Behavior of Receiving Currents in the presence of Four Wave Mixing Effect in DWDM Optical systems. Information Technology Journal (2006)
2. Binh, L.N., Gan, I., Tan, W.: SIMULINK Model for Optically Amplified Transmission Systems: Part V: Linear and Nonlinear Fiber Propagation Models. Department of Electrical and Computer Systems Engineering. Monash University (2005)
3. Bandelow, U., Demrican, A., Kesting, M.: Simulation of Pulse Propa-gation in Nonlinear Optical Fibers. Physics and Astronomy Classification (2003)
4. Arismar, S., Marconi, J.D., Fragnito, H.E.: Efficient Generation of Cascaded Four-Wave Mixing in very Short Optical Fibers. In: Proceedings of PIERS (2009)
5. Coen, S., Chau, A.H.L., Leonhardt, R., Harvey, J.D., Knight, J.C., Wadsworth, W.J., Russell, P.J.: White-light supercontinuum with 60 ps pump pulses in a photonic crystal fibre. Optics Letters 26, 1356–1358 (2000)
6. Cherif, R., Zghal, M.: Nonlinear phenomena of ultra-wide-band radia-tion in a photonic crystal fibre. Optics Letters (2000)

7. Gaeta, A.L.: Nonlinear propagation and continuum generation in microstructured optical fibers. Optics Letters 27(11), 924–926 (2002)
8. Agrawal, G.P.: Fibre-Optics Communications Systems. John Wiley & Sons, Inc. (1997)
9. Agrawal, G.P.: Nonlinear Fibre Optics. Academic Press, San Diego (2001)

Localization in Wireless Sensor Network: A Distributed Approach

Shailaja Patil and Mukesh Zaveri

Department of Computer Engineering,
Sardar Vallabhbhai National Institute of Technology, Surat, India
{p.shailaja,mazaveri}@coed.svnit.ac.in

Abstract. With the recent advances in radio technology and MEMs, Wireless Sensor Network has got a wide gamut of applications. Many applications like monitoring, surveillance and tracking need location of the object to carry out specified task.In this paper, we propose a distributed approach for localization, namely, Multidimensional Scaling with refinement using trilateration (MDS-DRT). The algorithm has been analyzed for varying number of node densities, number of anchors, and radio ranges. The simulation results show that the proposed algorithm performs better than the existing algorithms in terms of accuracy with reduced computational complexity.

Keywords: Localization, Wireless Sensor Network, distributed, Multidimensional Scaling.

1 Introduction

Wireless Sensor Network (WSN) has a very wide plethora of applications for the real time environment. WSN is composed of tiny, spatially distributed, battery powered nodes to observe, sense and monitor surrounding environment, event or area of interest like fire spread in forest, chemical flume, military surveillance etc. These nodes may be deployed in hostile environment, where they need to operate with minimum attention. The sensed data is collected from each node and sent to a central base station to process it locally or remotely. To identify the place of occurrence of event, localization is required. The process of estimating the position or spatial coordinates of sensor nodes is known as localization.

WSNs have to deal with various challenges such as scarcity of resource, communication failure, dynamic network topology, heterogeneity of nodes, and scalability to large scale of deployment. Due to these issues, the localization becomes a very challenging and important area of research. In WSN, this problem is being studied from many years for various applications like battlefield monitoring and tracking [1], building health structure monitoring [2], location aware routing [3], indoor people tracking [4], security [5] etc. As the sensor node is resource constrained i.e. it has limited battery power, processing speed and memory therefore, the localization algorithm needs to be energy efficient, robust to noise and

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 467–476.
springerlink.com

node failures. Also, the environment in which these nodes are deployed needs the algorithm to maintain the network self-organized.

Earlier, we have proposed a centralized localization algorithm, namely, MDS using refinement with trilateration (MDS-CRT) in [6]. The disadvantage of centralized algorithm is that it is prone to a single point of failure. To overcome this limitation, in this paper we propose a distributed localization algorithm, namely, MDS-DRT, where D stands for distributed localization and RT is refinement with trilateration. For the development of the algorithm it is assumed that the complete sensor network is divided into clusters (sub-network) and the localization is performed on individual clusters in the network.

The rest of the paper is organized as follows. Section 2 deals with literature survey and related work. Section 3 describes the proposed approach. The performance analysis is discussed in section 4, and conclusion is presented in section 5.

2 Related Work

The communication in WSN depends on the radio range of the node. Nodes in the vicinity of in-range-nodes can communicate with each other. The internode distances are required for localization. Depending on distance measurement techniques, the localization algorithms are broadly classified as range based and range free. The former one uses techniques such as received signal strength (RSS) [7], time of flight (TOF)/time difference of arrival (TDOA) [8], angle of arrival (AOA) [9], and latter one uses only the connectivity information of who is in vicinity of whom [10]. Range based techniques except RSS require specific hardware mounted on the node. The use of specific hardware with node not only increases the cost but also increases the power consumption. Due to this, most of the algorithms obtain the range information using RSS as a distance measure. However, the radio signal is sensitive to channel noise, interference, and reflections [11]. This makes a significant impact on signal amplitude, and uncertainty in the process of estimating locations, which makes localization more challenging.

Depending on the type of computation performed, the localization algorithms are classified as centralized or distributed. Thus, centralized and distributed algorithms may be range free or range based. Range-free distributed localization algorithms have been proposed in [12,13,14], whereas the range based distributed algorithms are reported in [15,16,17]. Multidimensional scaling (MDS)[18] based methods are also being exploited in localization techniques, for example the work reported in [19,20,21,22,23] have different localization techniques with its variants. MDS based algorithms are further classified as metric or classical MDS [19] and non-metric [23] MDS methods. It is possible to use classsical MDS (CMDS) algorithm for both range-aware and range-free localization. For range based distributed localization a metric MDS based algorithm, namely, MDS-MAP(P,R), has been proposed in [19]. In [21], a metric MDS based algorithm has been proposed for range based localization. However, these algorithms work based on local convergence property and hence, it implies poor localization performance if a good initialization is not available. A non-anchor based localization algorithm is introduced in [17], namely, simple hybrid absolute relative positioning (SHARP). This hybrid algorithm uses MDS and adhoc positioning system (APS) for localization. The process of localization consists of three phases. In the first phase, randomly a set of reference nodes are selected. In the second phase, CMDS is applied on these reference nodes. In third phase, an absolute localization is performed using APS. However, SHARP gives poor performance for anisotropic networks. Another hybrid approach is presented towards reducing complexity of MDS based algorithms in [24]. Here authors use MDS and proximity-distance map (PDM), in a phased manner. The phased approach yields comparable complexity to PDM which is less than MDS. However, performance of the algorithm degrades with reduction in anchors.

In all the MDS based algorithms, it is observed that, MDS-MAP(P,R) is best as it has the ability to localize isotropic, and anisotropic networks. MDS-MAP(P,R) algorithm works in three stages. First, the local maps are formed from distance matrix using CMDS. These maps are patched together one by one and affine transformation is applied on this stitched map in second stage. Thus, in this step local map is converted to global map to perform localization. The third stage refines these locations of the map using optimization. The initial maps are patched to form a global map and refined centrally. Though, the relative locations of local maps are estimated in distributed manner, these are required to be patched together to apply affine transformation with anchor nodes which requires centralized computation. The main disadvantage of centralized algorithms is that the node which performs the centralized computations is heavily loaded and these algorithms are prone to a single point of failure. When the energy of controlling node gets exhausted, complete network fails. In this paper we propose a distributed algorithm (MDS-DRT) and compare the performance of this algorithm with MDS-MAP(P,R) [19].

3 Proposed MDS-DRT Algorithm

The proposed algorithm is divided into two phases. The first phase consists of clustering and computing local maps. Clustering is used for dividing the whole topology into small networks. Nodes are clustered according to degree of connectivity, and a node with highest connectivity becomes the cluster head (CH). The CMDS is applied on these small clusters and their first estimate i.e. relative map of locations is obtained by CH. With the help of anchor nodes this relative map is converted to global map which yields locations of non anchor nodes in the network. In the second phase, these locations are further refined using the optimization technique, trilateration by adjustment (TBA). Figure 1, shows randomly spread anchor and non-anchor nodes. The nodes with IDs 1, 3 and 6 are anchors. Let the proximity between two points (i, j) is be given by Euclidean distance as expressed in equation 1.

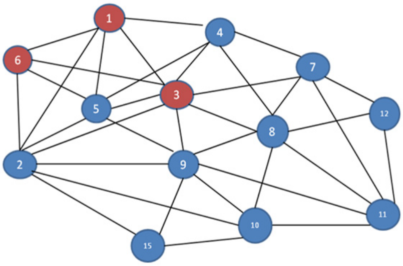


Fig. 1. The network of randomly spread nodes

$$d_{ij} = \sqrt{(X_j - X_i)^2 + (Y_j - Y_i)^2} \tag{1}$$

Thus, all the nodes gather the single hop neighbour distances. Each node maintains an information table as shown in table 1 for node ID [1]. Communication is started by first anchor node by sending hello packet and its own location. Each single hop neighbour node will respond to the hello packet with its ID. After receiving all responses, the cluster head (CH) updates distance information in the information table. The anchor node having greater degree of connectivity becomes the first CH and is responsible for final location estimation of its cluster members which is described in the following section.

The cluster localization consists of two phases. During the formation of initial cluster it may be the case where it is not possible to have three anchors, and in such cluster it is not possible to run first phase of localization algorithm. Here it is assumed that at least three anchor nodes are in range of each other. Hence, the cluster having only three anchor nodes run the proposed algorithm and estimate the locations for the cluster members, which are called as initial locations in the table 1. It is important to note that, the estimated locations are not final one and needs to be refined using optimization technique which is described in second phase. After refinement performed by CH, these nodes will receive the final coordinates. Now, these cluster members will be able to act as anchor nodes and we call them as pseudo-anchor nodes as these are other than original anchor nodes. Using these pseudo anchor nodes new clusters are formed, this will ensure that cluster now may have three anchor nodes and hence, first phase can be repeated for the cluster members whose locations are not yet estimated.

Table 1. Information Table of a node with ID [1]

Node-ID	Distance	Initial-location	Final-Location
(1,6)	2.1596	[5.99,8.00]	[5.93,7.91]
(1,8)	1.9124	[5.47,8.36]	[5.34,8.31]

The distributed localization algorithm is described as below. The following three steps are executed by CH:

1. Form single hop neighbours cluster including minimum three anchors.
   - **Phase I:**
     - Obtain distance matrix for all cluster members. Complete the squared distance matrix D^2.
     - Estimate the double centered matrix $B =$ C D^2 C; where C is the centering matrix expressed as $C = I - n^{-1}11'$ Here I is Identity matrix of nxn size, for n number of nodes and 1 is vector of n ones,
     - Compute singular value decomposition (SVD) of B as shown in equation 2

       $$SVD(B) = Q\Lambda Q' \tag{2}$$

       where Λ is $diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$, the diagonal matrix of Eigen values and Q is the matrix of corresponding eigenvectors.
     - Modify SVD output of matrix B according to dimensions (here two dimension). A reformed matrix is obtained for B, expressed as in equation 3.

       $$SVD(B_+) = Q_+\Lambda_+Q'_+ \tag{3}$$

     - Obtain coordinates from matrix B using following equation.

       $$X = Q_+\sqrt{\Lambda_+} \tag{4}$$

     - Transform relative map to absolute map using anchor nodes.
   - **Phase II:** In this phase, the estimate of locations obtained in phase I will be refined using optimization technique. In our proposed algorithm we use trilateration by adjustment (TBA) method for optimization. TBA uses the estimate of locations obtained in phase I as an initial estimate and are further refined during each iteration. For trilateration, three anchors (A, B, C) and one non-anchor node (D) is required. The objective function for this technique can be written as-

     $$f(d_{ij}) = d_{ij} - \sqrt{(X_j - X_i)^2 + (Y_j - Y_i)^2} = 0 \tag{5}$$

     Where d is the distance between i^{th} and j^{th} node.
     - Estimate weights from uncertainty of each link.

       $$W_j = \frac{\sigma_0^2}{\sigma_j^2} \tag{6}$$

       Where σ_0 is the variance of unit weight and σ_j is the variance of j^{th} link

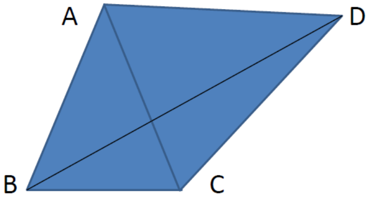


Fig. 2. A cluster of four nodes

- Evaluate values of objective functions F_1, F_2, F_3 at X_D and Y_D using equation 5 and estimate the following vector.

$$f = -\begin{bmatrix} F_1(X_D, Y_D) \\ F_2(X_D, Y_D) \\ F_3(X_D, Y_D) \end{bmatrix}$$

- Estimate correction in distances (d_{ij}), coordinates of unknown node and the precision of every node after every adjustment. Repeat above step until estimation of position error satisfies the predefined precision requirement or reaches the final count of iterations

2. Update the location information received from CH and if >3 locations are received, respond with candidature hello packet and wait for the response of current CH to become next CH.
3. Now, the newly formed CH node sends another hello packet to its single hop neighbour to collect the distance information. Steps enumerated in 1 are followed to estimate locations.
4. Steps 1-3 are repeated till all the nodes in the network are localized.

In this algorithm, initially clusters are formed and are localized with CMDS. The complexity of this step is $O(n^3)$ for n number of nodes in a cluster. The refinement step of trilateration on local map takes $O(m^2)$, where m are non anchor nodes. For k number of local maps which are stitched together, it becomes kO(n^3).

4 Performance Results

The performance of proposed algorithm has been examined for various radio ranges, anchors and node densities. To introduce error in radio link, the radio range is blurred with Gaussian noise. The performance measure used for evaluation is root mean square error (RMSE). We have performed simulations for 30 to 100 nodes with number of anchors varying from 3 to 5. Here we are presenting results of 50 and 100 nodes for 3,4 numbers of anchors; with and without noise. We have compared performance of our proposed algorithm with distribute localization algorithm MDS-MAP(P,R).

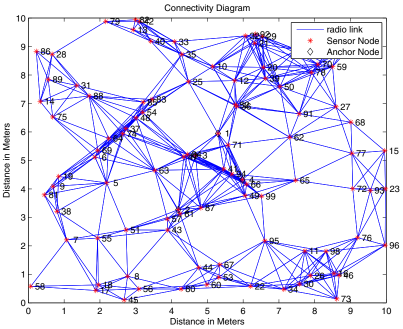
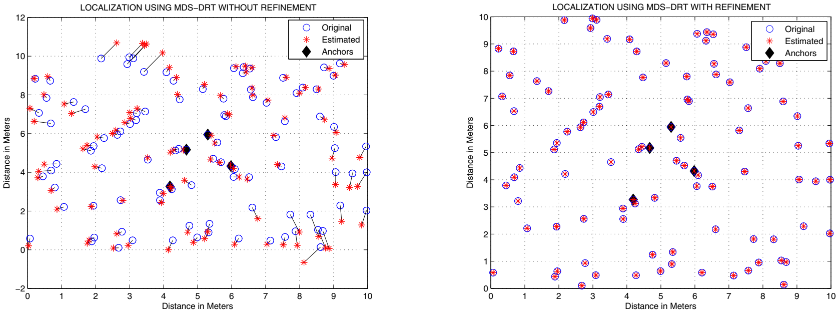


Fig. 3. Connectivity of nodes at radio range of 2m in Isotropic Network



(a) Localization without refinement

(b) Localization with refinement

Fig. 4. Localization with MDS-DRT at 5% range error

Figure 3 shows an isotropic network of 100 nodes placed randomly in $10mx10m$ square unit area. The lines joining nodes are radio links. This configuration is obtained at radio range of 2m with range error of 5% and connectivity of 13. Here, connectivity represents the average number of nodes connected with each other in the network.

Figure 4 shows scatter plot of nodes with original and estimated locations. The symbol ◦ shows original, * is estimated, and diamonds are anchor locations. This configuration is obtained at radio range of 2 with 5% range error in presence of 4 anchor nodes. Figure 44(a), and 44(b) shows localization without and with TBA refinement. An average RMSE with MDS-DRT in this case is 2.15, and 0.0537 respectively. The simulation has been performed for node density of 50, with range error of 0% and 5% at the connectivity level of 12.

The table 2 shows performance comparison of existing and proposed approach. The localization is performed with 50 number of nodes using 3 and 4 anchors. The RMSE is obtained with and without adding the range error of 5%. The

Table 2. Performance comparison of algorithms for density of 50 nodes

Anchors	Density	Connectivity Level	MDS-MAP (PR) [RMSE %]	MDS-DRT [RMSE %]
3	50 (RE-0%)	12	35.65	29.32
	50 (RE-5%)	12	37.34	31.75
4	50 (RE-0%)	12	30.54	27.59
	50 (RE-5%)	12	32.01	28.54

Table 3. Performance comparison of algorithms for density of 100 nodes

Anchors	Density	Connectivity Level	MDS-MAP (PR) [RMSE %]	MDS-DRT [RMSE %]
3	100 (RE-0%)	11	13.5	12.08
	100 (RE-5%)	11	16.43	14.65
4	100 (RE-0%)	11	13.00	11.73
	100 (RE-5%)	11	16.02	13.87

normalized values of RMSE are shown in the table 2. Table 3 shows the performance of algorithm for node density of 100 and connectivity level of 11 with 0% and 5% range error. The results of Table 2, and Table 3 very clearly show that MDS-DRT outperforms the MDS-MAP(P,R).

5 Conclusion

MDS-MAP is one of the pioneering algorithms in localization based on CMDS. However, the high computational complexity of algorithm increases the energy consumption of power constrained sensor node, reducing the network lifetime. We have proposed a distributed localization algorithm namely, MDS-DRT, with reduced complexity. Simulations have been performed with 50 to 100 numbers of nodes by varying number of anchors and connectivity. The algorithm performance is analysed in presence of noise. From the results it is evident that the algorithm is not only computationally less expensive but also performs better than existing algorithms. In future we shall perform energy analysis of the algorithm for varying node density, number of anchors, and range errors.

References

1. Alsharabi, N., Fa, L.R., Zing, F., Ghurab, M.: Wireless sensor networks of battlefields hotspot: Challenges and solutions. In: Proceedings of 6th International ICST Symposium on Modeling and Optimization, pp. 192–196. IEEE (2008)
2. Sukun, K., Shamim, P., David, C., James, D., Gregory, F., Steven, G., Martin, T.: Health monitoring of civil infrastructures using wireless sensor networks. In: Proceedings of the 6th International Conference on Information Processing in Sensor Networks, pp. 254–263. ACM, New York (2007)

3. Seung-Chul, M.W., Suresh, S.: Scalable routing protocol for ad hoc networks. Wirelss Network 7, 513–529 (2001)
4. Maria, C.J., la Torre Fernando, D., Aritz, S.: Arizaga: Indoor people tracking based on dynamic weighted multidimensional scaling. In: Proceedings of the 10th ACM Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems. MSWiM 2007, pp. 328–335. ACM, New York (2007)
5. Adrian, P., John, S., David, W.: Security in wireless sensor networks. Communications of the ACM 47, 53–57 (2004)
6. Patil, S., Zaveri, M.: Mds and trilateration based localization in wireless sensor network. Wireless Sensor Network 3, 198–208 (2011)
7. Li, X., Shi, H., Shang, Y.: A sorted rssi quantization based algorithm for sensor network localization. In: Proceedings of 11th International Conference on Parallel and Distributed Systems, vol. 1, pp. 557–563 (2005)
8. Xiao, J., Ren, L., Tan, J.: Research of tdoa based self-localization approach in wireless sensor network. In: International Conference on Intelligent Robots and Systems, pp. 2035–2040 (2006)
9. Kuakowski, P., Vales-Alonso, J., Egea-Lpez, E., Ludwin, W., Garca-Haro, J.: Angle of arrival localization based on antenna arrays for wireless sensor networks. Computers and Electrical Engineering 36, 1181–1186 (2010)
10. John, N.B., Heidemann, J., Estrin, D.: Gps-less low cost outdoor localization for very small devices. IEEE Personal Communications Magazine 7, 28–34 (2000)
11. Whitehouse, K., Karlof, C., Woo, A., Jiang, F., Culler, D.: The effects of ranging noise on multihop localization: an empirical study. In: Fourth International Symposium on Information Processing in Sensor Networks, IPSN, pp. 73–80 (2005)
12. Niculescu, D., Nath, B.: Ad hoc positioning system (aps). In: Proceedings of GLOBECOM, pp. 2926–2931 (2001)
13. He, T., Chengdu, H., Blum Brian, M., Stankovic John, A., Tarek, A.: Range-free localization schemes for large scale sensor networks. In: Proceedings of the 9th Annual International Conference on Mobile Computing and Networking. MobiCom 2003, pp. 81–95. ACM, New York (2003)
14. Savarese, C., Rabaey, J.M., Langendoen, K.: Robust positioning algorithms for distributed ad-hoc wireless sensor networks. In: Proceedings of the General Track of the Annual Conference on USENIX Annual Technical Conference, ATEC 2002, pp. 317–327. USENIX Association, Berkeley (2002)
15. Shi, Q., He, C., Chen, H., Jiang, L.: Distributed wireless sensor network localization via sequential greedy optimization algorithm. IEEE Transactions on Signal Processing 58, 3328–3340 (2010)
16. Priyantha, N.B., Balakrishnan, H., Demaine, E.D., Teller, S.J.: Anchor-free distributed localization in sensor networks. In: Proceedings of the 1st International Conference on Embedded Networked Sensor Systems, SenSys, pp. 340–341. ACM, Los Angeles (2003)
17. Ahmed, A., Shi, H., Shang, Y.: Sharp: a new approach to relative localization in wireless sensor networks. In: 25th IEEE International Conference on Distributed Computing Systems Workshops 2005, pp. 892–898 (2005)
18. Cox, T.F., Cox, M.A.: Multidimensional Scaling. Chapman Hall, London (1994)
19. Shang, Y., Ruml, W.: Improved mds-based localization. In: Proceedings of Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies, vol. 4, pp. 2640–2651 (2004)
20. Yi, S., Wheeler, R., Ying, Z., Markus, F.: Localization from connectivity in sensor networks. IEEE Transactions on Parallel Distributed Systems 15, 961–974 (2004)

21. Ji, X., Zha, H.: Sensor positioning in wireless ad-hoc sensor networks using multidimensional scaling. In: Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM, vol. 4, pp. 2652–2661 (2004)
22. Jose, A.C., Neal, P., Hero Alfred III, O.: Distributed weighted multidimensional scaling for node localization in sensor networks. ACM Transactions on Sensor Network 2, 39–64 (2006)
23. Nhat, V.D.M., Vo, D., Challa, S., Lee, S.: Nonmetric mds for sensor localization. In: 3rd International Symposium on Wireless Pervasive Computing, ISWPC 2008, pp. 396–400 (2008)
24. Cheng, K.Y., Lui, K.S., Tam, V.: Localization in sensor networks with limited number of anchors and clustered placement. In: Wireless Communications and Networking Conference, WCNC 2007, pp. 4425–4429. IEEE (2007)
25. Bal, M., Liu, M., Shen, W., Ghenniwa, H.: Localization in cooperative wireless sensor networks: A review. In: 13th International Conference on Computer Supported Cooperative Work in Design, pp. 438–443 (2009)

Web Mining and Security in E-commerce

Shaikh Mohammed Atiq[1], Dayanand Ingle[1], and B.B. Meshram[2]

[1] Bharati Vidyapeeth College of Engineering, Navi Mumbai, India
cooldudes2786@gmail, dringleus@yahoo.com
[2] Veer Jijamata Technological Institute, Mumbai India
bbmeshram@vjti.org.in

Abstract. This paper is based on e-commerce web sites how to use web mining technology for providing security on e-commerce web sites. The connection between web mining ,security and e-commerce analyzed based on user behavior on web. Different web mining algorithms and security algorithm are used to provided security on e-commerce web sites. Based on customer behavior different web mining algorithms like page rank algorithm and trust rank algorithm is used for developing web mining framework in e-commerce web sites. We have developed false hit database algorithm and nearest neighbor algorithm to provide security on e-commerce web site. In existing web mining framework is based on only web content mining, We have proposed Web mining framework system which is based on web structure mining analysis, Web Content Mining analysis, decision analysis and security analysis.

Keywords: Web mining, Security, E- commerce, Web usage mining.

1 Introduction

Web mining is a rapid growing research area. It consists of Web usage mining, Web structure mining, and Web content mining. Web usage mining refers to the discovery of user access patterns from Web usage logs. Web structure mining tries to discover useful knowledge from the structure of hyperlinks. Web content mining aims to extract/mine useful information or knowledge from web page contents. Web content mining is related to data mining because many data mining techniques can be applied in Web content mining. It is also quite different from data mining because Web data are mainly semi-structured and/or unstructured, while data mining deals primarily with structured data. In the past few years, there was a rapid expansion of activities in the Web content mining area. However, due to the heterogeneity and the lack of structure of Web data, automated discovery of targeted or unexpected knowledge information still present many challenging research problems [1].

The rest of the paper is organized as follows. Section 2 deals with Web Mining Framework System Section 3 deals with Web structure mining analysis, Section 4 deals with Web content mining, Section 5 gives Decision analysis, section 6 gives Security analysis and Section 7 conclude the paper.

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 477–487.
springerlink.com

2 Web Mining Framework System

Web mining is the use of data mining techniques to automatically discover and extract knowledge from web documents.web mining is the information service centre for news, e-commerce, and advertisement, government, education, financial management, education, etc[2]. We have developed Web mining framework for evaluating ecommerce web sites .In general web mining task can be classified into web content mining, web structure mining and web usage mining. Some of the well-known classification techniques for web mining such as like, page rank algorithm and trust rank algorithm is used in this paper. Our proposed web mining framework consists of four phase's web structure mining analysis, Web Content Mining analysis, decision analysis and security analysis[3].

3 Web Structure Mining Analysis

This phase analyses a web site by using both page rank algorithm and trust rank algorithm[4]. The ranking of a page is determined by its link structure instead of its content. The trust rank algorithm is procedure to rate the quality of web sites. The output is quality based score which correspond to trust assessment level of the web site. The initial step is collects information from web sites and stores those web pages into web repository[5].

A. Page Rank Algorithm

Page Rank algorithm used by search engine .We have computed page rank of web sites by parse web pages for links, iteratively compute the page rank and sort the documents by page rank engine[6] .Page Rank algorithm is in fact calculated as follows

PAR(A)=(1-d)+d(PAR(T1)/OG(T1)+….+PAR(TN)/OG(TN) Where PAR(A)

is the PageRank of page A

0G(T1) is the number of outgoing links from page T1

d is a damping factor in the range $0<d<1$,usually set to 0.85

The PageRank of web page is calculated as sum of the PageRank of all pages linking to its divided by the number of links on each of those pages its outgoing links.

B. Trust Rank Algorithm

The trust rank algorithm is procedure to rate the quality of web sites. Taking the linking structure to generate a measure for quality of a page.

Steps of Trust Rank algorithm.

1-The starting point of the algorithm is the selection of trusted web pages.

2-Trust can be transferred to other page by linking to them.

3-Trust is propagating in the same was as Page Rank

4-The negative measure is propagating backwards and is a measure of bad pages

5-For the ranking algorithm both measures can be taken into account.

Trust Rank algorithm is in fact calculated as follows

Trust Rank=M*x

Where the matrix m is given by

M=1-dt

With Tij=1/cj(if page j is linking to page i) Tij=0 otherwise

d is damping factor and x is the source vector of the trust

The invrse PageRank is given by Minv*xinv

With Minv=1-dinvTinv

The inverse transition matrix Tinv is definied by

Tij=1/nj(if page i is linking to page

Tij =0 otherwise

d is damping factor and xinv is the source vector of the bad pages and n is number of incoming links on page j.Minv is nether the transparent nor the inverse matrix of M. From this we can say that pages are bad which are linking to bad pages. While pages are good which are linking good pages.

4 Web Content Mining Analysis

Web content mining is defined as searching of new information from web data. Data is retrieved for desired topic by user[7]. In Web content mining analysis we have taken example job categories and the associated skills needs prevalent in the computing professions. We performed a cluster analysis on the ads in two phases. Hierarchical agglomerative clustering is the first step to identify unique skill set clusters. The classification of ads is validated into clusters by performing k-means cluster analysis[8]. Module1: User Identification, Module2: Job Definition, Module3: Data Collection, Module4: Data Analysis.

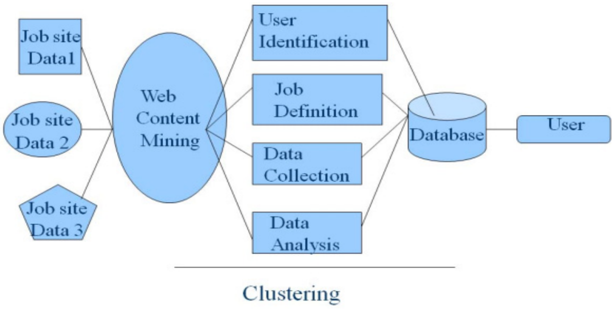


Fig. 1. Module Diagram

A. Hierarchical Agglomerative Clustering

Agglomerative hierarchical clustering is a bottom-up clustering method where clusters have sub-clusters, which in turn have sub-clusters, etc. The classic example of this is species taxonomy. Gene expression data might also exhibit this hierarchical quality (e.g. neurotransmitter gene families). Agglomerative hierarchical clustering starts with every single object (gene or sample) in a single cluster. Then, in each successive iteration, it agglomerates (merges) the closest pair of clusters by satisfying some similarity criteria, until all of the data is in one cluster[9]. The hierarchy within the final cluster has the following properties: Clusters generated in early stages are nested in those generated in later stages. Clusters with different sizes in the tree can be valuable for discovery. A Matrix Tree Plot visually demonstrates the hierarchy within the final cluster, where each merger is represented by a binary tree. **Process**: Assign each object to a separate cluster. Evaluate all pair-wise distances between clusters[10]. Construct a distance matrix using the distance values. Look for the pair of clusters with the shortest distance. Remove the pair from the matrix and merge them. Evaluate all distances from this new cluster to all other clusters, and update the matrix. Repeat until the distance matrix is reduced to a single element. **Advantages:** It can produce an ordering of the objects, which may be informative for data display. Smaller clusters are generated, which may be helpful for discovery [11].

B. K-Means Cluster Analysis

In statistics and machine learning, k-means clustering is a method of cluster analysis which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean[12]. It is similar to the expectation-maximization algorithm for mixtures of Gaussians in that they both attempt to find the centers of natural clusters in the data as well as in the iterative refinement approach employed by both algorithms. **Process:** The dataset is partitioned into K clusters and the data points are randomly assigned to the clusters resulting in clusters that have roughly the same number of data points. For each data point: Calculate the distance from the data point to each cluster. If the data point is closest to its own cluster, leave it where it is. If the data point is not closest to its own cluster, move it into the closest cluster. Repeat the above step until a complete pass through all the data points results in no data point moving from one cluster to another. At this point the clusters are stable and the clustering process ends. The choice of initial partition can greatly affect the final clusters that result, in terms of inter-cluster and intra cluster distances and cohesion. **Advantages:** With a large number of variables, K-Means may be computationally faster than hierarchical clustering (if K is small).K-Means may produce tighter clusters than hierarchical clustering, especially if the clusters are globular.

C. Module1: User Identification

Users are of different categories. New Users will get registered in the system. Existing users can logon to their account. Administrator has the highest priority. Generate user profiles based on their access patterns. Cluster users based on frequently accessed URLs. Use classifier to generate a profile for each cluster. We have developed the web site using dot net as front end and sql as back end.

D. Module2: Job Definition

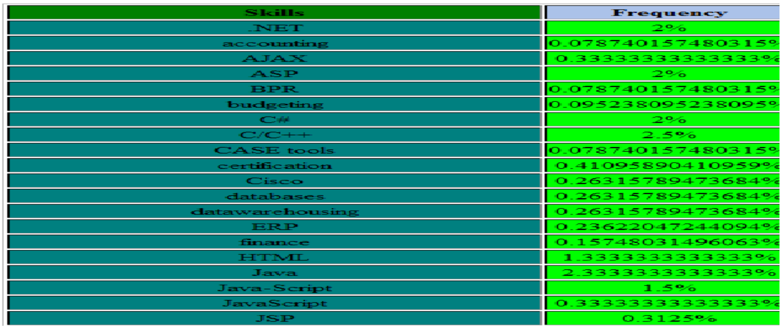
In this module we have authenticated users can proceed with the features provided by the web based learning site[13]. The materials based on few subjects are given in the site and the users can utilize it.

E. Module3: Data Collection

Collecting the job definitions based on grouping .Collecting the values of job title, job description and the skills required by the company of the candidate. The job definitions is being clustered based on the job title[14].

F. Module4:Data Analysis Module

The data are analyzed based on the Data Collection Module. Dagger (†), Asterisk (*).The data is mined based on the previous modules.

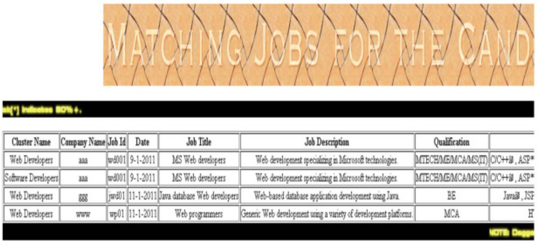


Skills	Frequency
.NET	2%
accounting	0.078740157480315%
AJAX	0.333333333333333%
ASP	2%
BPR	0.078740157480315%
budgeting	0.095238095238095%
C#	2%
C/C++	2.5%
CASE tools	0.078740157480315%
certification	0.410958904109589%
Cisco	0.263157894736842%
databases	0.263157894736842%
datawarehousing	0.263157894736842%
ERP	0.236220472440944%
finance	0.157480314960631%
HTML	1.33333333333333%
Java	2.33333333333333%
Java-Script	1.5%
JavaScript	0.333333333333333%
JSP	0.3125%

Fig. 2. Skills Frequency

G. Results

Using a Web content data mining application, few unique IT job descriptions from various job search engines are extracted and distilled each to its required skill sets. We examined these, revealing few clusters of similar skill sets that map to specific job definitions. It makes job search faster and gives the results according to user preference.



MATCHING JOBS FOR THE CAND

sk(*) Indicates 80%+.

Cluster Name	Company Name	Job Id	Date	Job Title	Job Description	Qualification	
Web Developers	aaa	wd001	9-1-2011	MS Web developers	Web development specializing in Microsoft technologies.	MTECH/ME/MCA/MS(IT)	C/C++â , ASP*
Software Developers	aaa	wd001	9-1-2011	MS Web developers	Web development specializing in Microsoft technologies.	MTECH/ME/MCA/MS(IT)	C/C++â , ASP*
Web Developers	ggg	jwd01	11-1-2011	Java database Web developers	Web-based database application development using Java.	BE	Javaâ , JSP
Web Developers	www	wp01	11-1-2011	Web programmers	Generic Web development using a variety of development platforms.	MCA	H

NOTE: Dagge

Fig. 3. Matching job

H. Performance analysis

We propose a system in which all the Information about the system can be logged for future reference. Analysis of the student's and the fresher's performance measurements could be done. Graduates can get an exact job and the fresher's can measure their gap in the industry and learn accordingly.

5 Decision Analysis

This phase uses the total trust of web page generated from Web structure mining analysis phase. Two processes are performed (a)Trust calculation of web site and (b) Application of suitable statistical techniques to analyses the result of the evaluation We consider three trust levels

A. Trust calculation of web site

The value of the trust value variable has converted into degrees of membership fuction defined on variable Let as consider Trust Value :0.11

Un trust Web sites:0.78

Moderate Trust Web sites:0.22

High Trust Web sites:0.00

We can say that

If trust value is un trust web sites then trust level is none. If trust value is Moderate Trust web sites then trust level is limited. If trust value is High Trust web sites then trust level is full. From the above method we can calculate trust of the web site

B. Application of suitable statistical techniques to analyses the result of the evaluation

Analyzing information from website is important. Using statistics, we can able to evaluate your website. Descriptive Statistics is mainly used to describe populations using random samples of Web data collected from web sites. It provides a statistical summary of the web data with a view to understand the population that sample represent. Central Tendency and dispersion measures are used in descriptive statistics. Measures of central tendencies describe the central values of a collected sample of web data. For an ungrouped set of web data measures are mean, median and mode. Pareto principal -The first 50% of un trusted web site is banned, next 25% of the un trusted web site is banned, next 12.5% takes same effort and so on..By application of suitable statistical techniques we can evaluate results[15].

6 Security Analysis

We perform complete security analysis in this phase. 89% of web development companies has not follow industry standard in developing and hosting the websites they make[16]. The customers who use the web sites do know the difference between a secure website and insecure website. We have developed trust path intermediaries building algorithm, false hit database algorithm and nearest neighbor algorithm to provide security on e-commerce web site[17]. Multi-step processing is used for nearest neighbor and similarity search in application involving web data and/or costly distance computations. C AMNC-to reduce the size of False Hit database. The query is authenticated. A server maintains dataset database signed by trusted authority False hit Database to reduce hang or lag in the server[18]. Provides accurate data as well as NN result-set. We have developed following modules for providing security on e-commerce web sites. Module 1: Authentication, Module 2: Query processing, Module 3: Similarity search and Module 4: False hit reduction [19].

A. Module 1: Authentication

In authentication module Member or user access the search facility and admin check false hits and updates the database.

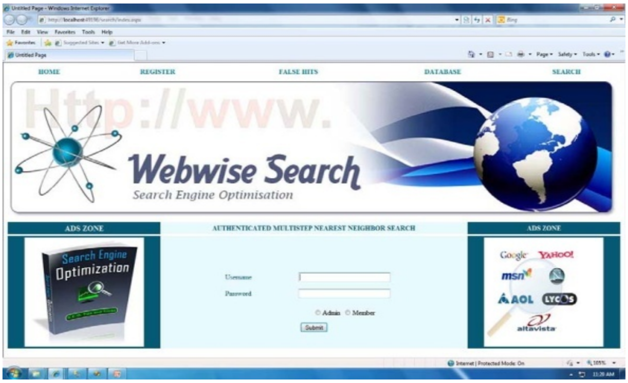


Fig. 4. Webwise search web site

B. Module 2: Query processing:

This module describes the server and user communication process where the client posts the query and the server delivers the result based on the criteria.

C. Module 3: Similarity search

The similarity search proposes the criteria of retrieving the relevant information from the database based on similar keyword.

D. Module 4: False hit reduction:

Admin checks the false hits recorded. He then posts the necessary responses to the search database for future verification.

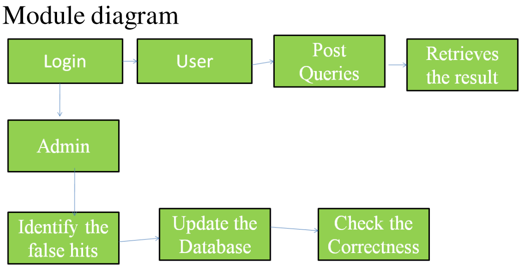


Fig. 5. Module Diagram

Working (Default)

User enters the search keyword in the web site. Admin checks the false hits recorded. He then posts the necessary responses to the search database for future verification.

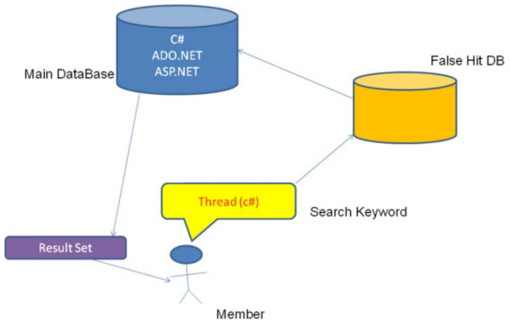


Fig. 6. Working

Case 1

If the search keyword is not present ,search keyword is updated in the false hit database.

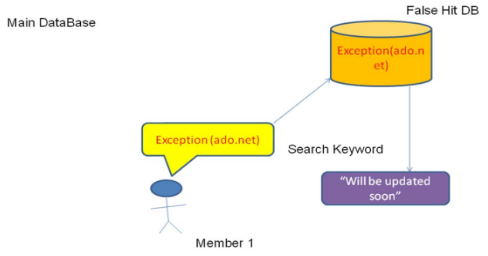


Fig. 7. Keyword is not present false hit is updated

Case 2:

If the search keyword is present then posts the necessary responses to the search database for future verification.

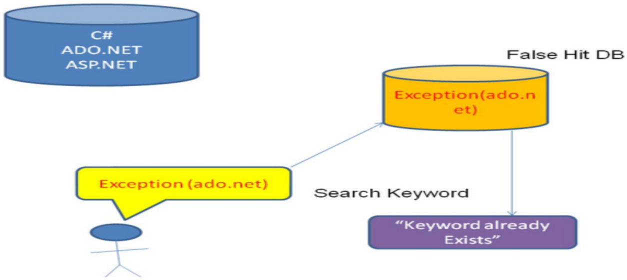


Fig. 8. Keyword is present post the response

Admin Updates the Database

User access the search facility and admin check false hits and updates the database.

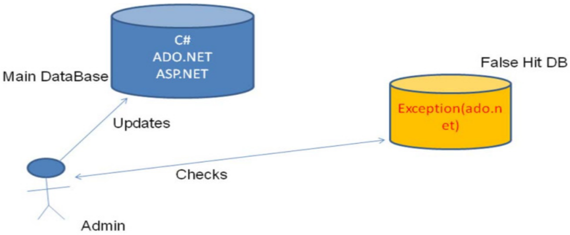


Fig. 9. Admin Updates the Database

The importance of authenticated query processing increases with the amount of information available at source that are untrustworthy, unreliable, or simply unfamiliar. This is the first work addressing authenticated Similarity retrieval from such sources using the multistep NN framework. We show that a direct integration of optimal NN search with an authenticated data structure incurs excessive communication overhead. From security module we provide security to web site.

7 Conclusion

In this paper we have proposed web mining framework for e-commerce web sites. In web mining framework we have developed four phases' web structure mining analysis, Web Content Mining analysis, decision analysis and security analysis. In web structure mining analysis we have used page rank algorithm and trust rank algorithm.

In Web Content Mining analysis we have used Hierarchical agglomerative clustering and k-means cluster analysis. In decision analysis we have used trust calculation of web site and statistical techniques to analyses the result of the evaluation. In security analysis we developed trust path intermediaries building algorithm, false hit database algorithm and nearest neighbor algorithm to provide security on e-commerce web site.

References

[1] Litecky, C., Aken, A., Ahmad, A., Nelson, H.J.: Southern Illinois University, Carbondale, Mining computing jobs. IEEE Software (January/February 2010)

[2] Tao, Y., Yi, K., Sheng, C., Kalnis, P.: Quality and Efficiency in High Dimensional Nearest Neighbor Search. In: SIGMOD (2009)

[3] Nasraoui, O., Member, IEEE, Soliman, M., Member, IEEE, Saka, E., Member, IEEE, Badia, A., Member, IEEE, Germain, R.: A Web Usage Mining Framework for Mining Evolving User Profiles in Dynamic Web Sites. IEEE Transactions on Knowledge and Data Engineering 20(2) (February 2008)

[4] Liu, B., Grossman, R., Zhai, Y.: University of Illinois at Chicago, Mining Web Pages for Data Records. Published by the IEEE Computer Society (December 2004)

[5] Pal, S.K., Fellow, IEEE, Talwar, V., Student Member, IEEE, Mitra, P., Student Member, IEEE: Web Mining in Soft Computing Framework: Relevance, State of the Art and Future Directions. IEEE Transactions on Neural Networks 13(5) (September 2002)

[6] Atif, Y.: United Arab Emirates University, Building Trust in E-Commerce. IEEE Internet Computing (January-February 2002)

[7] Korn, F., Sidiropoulos, N., Faloutsos, C., Siegel, E., Protopapas, Z.: Fast Nearest Neighbor Search in Medical Image Databases. In: VLDB (1996)

[8] Zhang, Q., Segall, R.S.: Web mining: a survey of current research, Techniques and software. The International Journal of Information Technology & Decision Making 7(4), 683–720 (2008)

[9] Yang, Q., Wu, X.: 10 challenging problems in data mining research. Int. J. Inform. Technol. Decision Making 5(4), 597–604 (2006)

[10] Etzioni, O.: The world wield web: Quagmire or Gold Mining. Communicate of the ACM 11(39), 65–68 (1996)

[11] Han, J., Chang, C.: Data mining for web intelligence. Computer, pp. 54–60 (November 2002), http://cs.uiuc.edu/hanj/pdf/computer02.pdf

[12] Barsagade, N.: Web usage mining and pattern discovery: A survey paper, Computer Science and Engineering Dept., CSE Tech Report 8331. Southern Method ist University, Dallas (2003)

[13] Chau, R., Yeh, C.-H., Smith, K.A.: Personalized Multilingual Web Content Mining. In: Negoita, M.G., Howlett, R.J., Jain, L.C. (eds.) KES 2004. LNCS (LNAI), vol. 3213, pp. 155–163. Springer, Heidelberg (2004)

[14] Kolari, P., Joshi, A.: Web mining: Research and practice. Comput. Sci. Eng., 42–53 (July/August 2004)

[15] Liu, B., Chang, K.: Editorial: Special issue on web content mining. SIGKDD Explorations 6(2), 1–4 (2004)

[16] Semantic Web Mining: State of the art and future directions Web Semantics: Science, Services and Agents on the World Wide Web 4(2), 124–143 (June 2006)

[17] Long, F., Zhang, H., Feng, D.D.: Fundamentals of content \ based imagereal, http://www.cse.iitd.ernet.in/~pkalra/siv864/Projects/h01_Long_v40

[18] Zhang, H., Chen, Z., Li, M., Su, Z.: Relevance feedback and learning in content-based image search. World Wide Web 6(2), 131–155 (2003)

[19] Chen, L., Lian, W., Chue, W.: Using web structure and summarization techniques for web content mining. Inform. Process. Management: Int. J. 41(5), 1225–1242 (2005)

Performance Improvement in MIMO Systems Using Rotating Codebooks

J. Julia[1] and M. Meenakshi[2]

[1] P.G. Scholar
Anna University, Chennai, India
juliajoseph08@yahoo.com
[2] Associate Professor
Anna University, Chennai, India
meena68@annauniv.edu

Abstract. MIMO systems provide with the advantages of diversity and capacity. If the information regarding the channel is known to the transmitter, then the transmitter can accordingly weigh and transmit the data to achieve better performance. Receiver can estimate the channel conditions and the information can be sent to the transmitter through the control channel. But the bandwidth of the control channel is limited. The main aim of this paper is to improve the performance of the MIMO systems while employing lesser number of feedback bits. In this paper, the rotating mixed codebook technique is presented which helps in attaining the goal. Simulation results are also presented to show the gain achieved.

Keywords: MIMO precoding, Limited feedback precoding.

1 Introduction

Multiple input multiple output systems provide performance benefits as against the conventional systems in terms of diversity and capacity. These performance benefits can be still more improved or enhanced if the transmitter has knowledge about the channel. The receiver obviously should know about the channel statistics in the MIMO systems. And this information can be fedback to the transmitter through the control channel. But the bandwidth of the control channel through which this information is sent is limited. Also the overhead which increases linearly with the product of number of antennas, the frequency selectivity and the feedback frequency can be very large [6].

The information regarding the channel may be instantaneous CSI. However the partial CSI, that is certain statistics like the channel mean or channel covariance[7] can also be sent. But these methods do not perform well like the instantaneous feedback because they do not track the rapid fluctuations in the channels. This method of sending the instantaneous information regarding the channel to the transmitter is called as limited feedback precoding. Many works are available in literature[8],[9] regarding the limited feedback precoding.

The basic idea of limited feedback precoding can be explained as: Codebooks are generated and are made available at both the transmitter and the receiver. Receiver with the channel knowledge will choose particular precoder using a selection criteria.

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 489–496.
springerlink.com

The index of the chosen precoder is transmitted back to the transmitter using the control channel.The transmitter will choose the precoder matrix corresponding to the index, multiply it with the data to be sent and transmits the data. Thus using this limited feedback scheme, the number of feedback bits is reduced considerably when compared to the direct quantisation of the channel.

However, it could be found that the precoding performance increases with the number of feedback bits. Practically this is limited to 4 bits/subband. But this is very less when compared to the MIMO systems with higher order of transmit and receive antennas. In this paper, the rotating mixed codebook scheme is presented which helps in improving the precoding performance with lesser number of feedback bits.

The paper is organized as follows. The limited feedback preliminaries is provided in section 2. The rotating codebook precoding technique is presented in section 3. Different types of codebooks are studied and the rotating mixed codebook scheme is presented in section 4 and simulation results are presented in section 5. This is followed by conclusion in section 6.

Notation. We use $(.)^H$ to denote the transposition of a matrix , $\|.\|^F$ to denote the Frobenius norm of a matrix and I_m denotes the identity matrix of order m X m.

2 Limited Feedback Precoding

Let us consider a linear precoding MIMO system [4] where $S = [S_1, S_2, . . . , S_M]^T$ contains a vector of M modulation symbols . The symbol vector S is then multiplied by an $M_T \times M$ precoding matrix F, producing a vector X = FS of length M_T. The received signal can be given by

$$Y = H x + w \quad (1)$$

where H is an $M_R \times M_T$ channel matrix ,where $M_R \geq M$ is the number of receive antennas, and N is a length-M_R noise vector. In closed-loop precoding, the receiver chooses an $M_T \times M$ precoding matrix F. The transmitter precodes that is, multiplies the symbol vector S with F and then sends the data . The equivalent channel is HF.

The above steps are depicted in Fig.1:

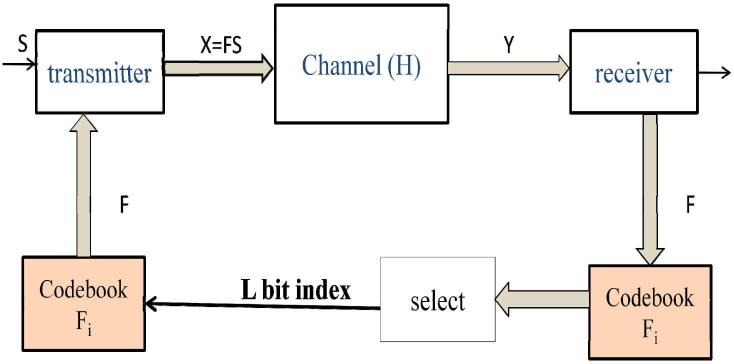


Fig. 1. Block diagram of limited feedback precoding

It could be observed that two factors influence the precoding performance.

(a) codebook generation method
(b) Precoder selection function

In this paper , the precoder matrix which maximizes the capacity is chosen. And the capacity is given by,

$$C(\mathbf{F})=\log\det(I+\gamma\,(\mathbf{HF})^H(\mathbf{HF}))=\log\det(I+\ \gamma\,\mathbf{F}^H\,\mathbf{H}\,^H\mathbf{HF}\,) \qquad (2)$$

Where γ is the ratio of the transmit signal power to the receiver noise power . The best precoder that maximizes the capacity is given by

$$\hat{}\mathbf{F} = \mathrm{argmax}\ C(\mathbf{F}) \qquad (3)$$

ˆF is further confined to a limited number of choices, i.e., ˆF ∈ F_L, where F_L is a finite set of unitary matrices of size $|F_L| = 2^L$

$$F_L = \{F_0, F_1, \ldots, F_i, \cdots F_{2^L-1}\} \qquad (4)$$

called a codebook of precoding matrices. The size of the codebook is limited by the number of bits per feedback L. The 2^L precoding matrices are independent of the current channel state condition .They are known to both the transmitter and receiver. Based on the estimated channel knowledge the receiver selects a particular F which satisfies (3) and conveys the L bit index to the transmitter via the control channel.

3 Rotating Codebook Scheme

The performance of the limited-feedback precoding system depends on the size of the codebook, which is determined by the number of feedback bits. Increasing the number of feedback bits no doubt improves the performance. However, the increase in the number of feedback bits, in turn, requires a significant increase in control channel bandwidth. The goal of the rotating codebook scheme [4] shown in Fig. 2, is to improve closed-loop precoding gain without increasing the number of feedback bits.

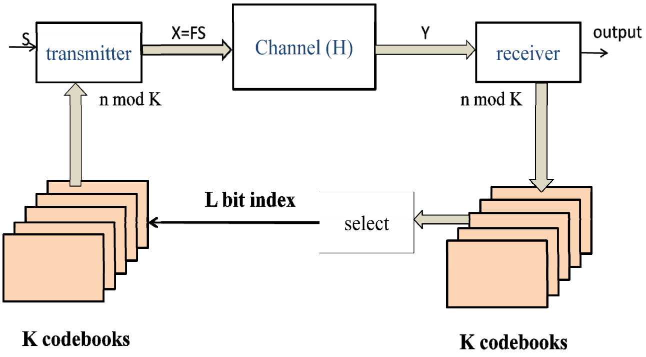


Fig. 2. Block diagram of rotating codebook scheme

K different but equivalent codebooks of size 2^L are generated. One out of the K codebooks should be selected at a time for use. And this is accomplished by sequence , s = n mod K; 0≤ s ≤K-1, at time t= nT. This scheme actually gives a chance to use more precoder matrices when compared to the single codebook scheme. But the number of feedback bits used is the same.

Using this rotating codebook scheme, virtually the size of the codebook is increased. But the number of feedback bits is the same. To make this scheme more effective, the default precoding matrix scheme is used.The default precoding matrix scheme reserves an index $\tilde{i} \in \{0, . . . , 2^L - 1\}$ to serve as an indication to the transmitter that the default precoding matrix (i.e., the precoding matrix used in the previous transmission, which is a matrix optimized through the previous precoding matrix selection process) should be used for current transmission. This allows the optimization results from the previous codebooks to be used in the optimization process within the current codebook.

Simulations were performed for a 2 × 1 alamouti scheme. The symbols were QAM modulated. A slow, non frequency selective channel was considered. It is assumed that the channel conditions are perfectly known at the receiver. The codebooks were designed using Grassmannian packing criteria. For simulation, the Sloane packings [5] are considered and 16 codebooks are taken. The results obtained is tabulated in Table1. It could be observed that the rotating codebook scheme with 16 codebooks performs well when compared to the conventional scheme, while both the schemes use only 3 feedback bits.

Table 1. SNR Values To Achieve Ber Of 10^{-3} (averaged over 20 iterations)

Single codebook scheme	Rotating codebook scheme, 16 sloane codebooks
10.2 dB	8.3 dB

4 Rotating Mixed Codebook Scheme

The rotating codebook scheme performs well in MIMO systems, and this has motivated to study the effect of using different combinations of codebooks for the rotating scheme. For the rotating mixed codebook scheme, the following types of codebooks are considered:

- Codebook based on Grassmannian packings

  (1) Sloane packings[5]
  (2) Ideal Grassmannian packings[5]
  (3) Nearly Grassmannian packings [2]

- DFT codebook [10]
- Kerdock codebooks [1] based on Sylvester Hadamard matrix

The block diagram of the rotating mixed codebook scheme is given in fig.3

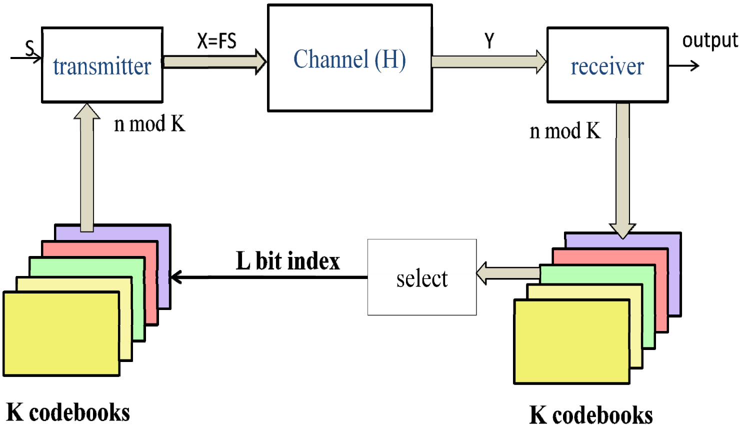


Fig. 3. Rotating mixed codebook scheme

4.1 Codebook Based on Grassmannian Packings

Codebook design problem can be formulated as Grassmannian subspace packing problem. The performance measure in Grassmannian subspace packing [10] is the chordal distance, which is defined as

$$d(W_k,W_l) = \frac{1}{\sqrt{2}} \| W_k W_k^H - W_l W_l^H \|. \quad (5)$$

The optimum codebook is designed to maximize the minimum chordal distance ∂min = min k≠l;1≤ k;l ≤ L d(W$_k$,W$_l$), where L is the number of feedback bits. Fig. 3 shows the Precoding matrix and chordal distance.

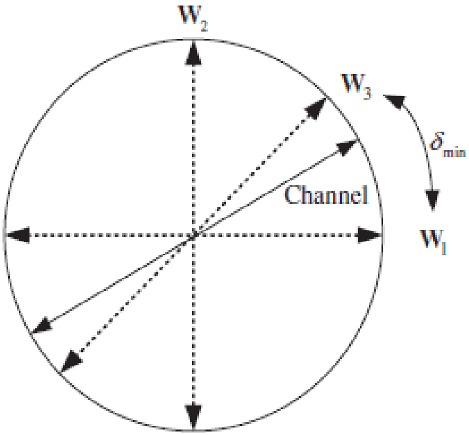


Fig. 4. Precoding matrix and chordal distance

The packings provided by Sloane and Grassmann are considered as codebooks for the rotating mixed codebook scheme. A numerical method is described in [2] for generating nearly Grassmannian matrices.

4.2 DFT Codebook

Solving the Grassmannian packing problem for arbitrary NT, codeword length M, and codebook size L is quite time-consuming and not straightforward. Instead , a suboptimal yet practical design method is considered. One particular design method is to use DFT matrices [10] given as

$$F=\{ W_{DFT}, \Theta W_{DFT},\ldots\ldots \Theta^{L-1} W_{DFT} \} \tag{6}$$

The first codeword W_{DFT} is obtained by selecting M columns of N_T X N_T DFT matrix, of which the $(k,l)^{th}$ entry is given as $e^{j2\pi(k-1)(l-1)/N_T}/\sqrt{N_T}$, k,l =1,2,….N_T. Θ is the diagonal matrix given as

$$\Theta= \mathrm{diag}([\ e^{j2\pi u_1/N_T}\ e^{j2\pi u_2/N_T}\ldots\ldots e^{j2\pi u_{NT}/N_T}\ ]\ ) \tag{7}$$

The free variables u_i , i=1,2,….N_T should be determined such that the minimum chordal distance between the codewords must be maximised.

4.3 Kerdock Codebooks

The procedure [1] for generating the Kerdock codebooks is as follows:

Let $\hat{}H_2$ denote the Hadamard matrix and is given by,

$$\hat{}H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

Let $\hat{}H_{N_t}$ denote the size N_t X N_t Sylvester-Hadamard matrix . And is given by

$$\hat{}H_{N_t} = \hat{}H_2 \otimes \hat{}H_2\ ; \tag{8}$$

for B times, where B is the number of feedback bits.

Construct the diagonal matrices D_n for n=0, ……N_t-1.These are called the generator matrices.

Each base is constructed by

$$S_n = (1/\sqrt{N_t})D_n \hat{}\ H_{N_t}) \tag{9}$$

The codebook F is given by

$$F=[\ S_0\ \ S_1\ \ S_{N_t} - 1]. \tag{10}$$

5 Simulation Results

Simulations were performed to study the performance of individual codebooks. Table 2 shows the SNR (in dB) values required for BER of 10^{-3} for various types of codebooks when a single codebook was used (not rotating case).And table 3 compares the SNR values required to achieve BER of 10^{-3} in the single and rotating codebook scheme. For not rotating scheme, a single sloane codebook is considered.For rotating scheme, 16 sloane codebooks were considered.

Table 2. SNR values required to achieve a particular BER for different codebooks (single codebook case)

Codebooks	10^{-3}
DFT	11.189dB
Sloane	10.1535dB
Ideal grassmannian	9.73965 dB
Nearly grassmannian	10.60 dB
Grass_4_2_4	10.47 dB
Grass_6_3_6	12.1553 dB
Kerdock	9.8147 dB

Table 3. Comparison of rotating and not rotating (i.e) single codebook scheme

Not rotating scheme	Rotating Scheme
10.2 dB	8.36 dB

Table 4 gives the values of SNR required to achieve a BER of 10^{-3} for rotating and rotating mixed schemes. For rotating scheme, the sloane codebook is considered.However for mixed codebook scheme, codebooks discussed in section 4 were used. It is found that the rotating mixed codebook scheme performs well when compared to the rotating scheme.But the number of feedback bits is same for both the cases.

Table 4. Comparison of rotating and rotating mixed codebook schemes(averaged over 10 iterations)

Rotating codebook scheme	Rotating mixed codebook scheme
9.3604 dB	7.0933 dB

6 Conclusion

Feedback of CSI to the transmitter can enable the transmitter to better exploit channel conditions to improve MIMO performance. However, the amount of channel information fed back to the transmitter is limited by the feedback control channel bandwidth gain. By using rotating codebook precoding approach the effective size of the codebook is expanded without the need to increase the number of feedback bits. By using the rotating mixed codebook scheme proposed in this work, the performance is shown to further improve without the need to increase the storage capacity at the transceivers and also without the need to increase the number of feedback bits.

References

1. Innoue, T., Heath, R.W.: Kerdock codes for limited feedback precoded MIMO systems. IEEE Transactions on Signal Processing (November 2007)
2. Dhillion, S., Heath, R.W., Strohmer, T., Tropp, J.A.: Constructing Packings in Grassmannian Manifold Via alternating Projections. IEEE Transactions on Experimental Mathematics 17 (2008)
3. Shor, P.W., Sloane, N.J.A.: Family of optimal packings in Grassmannian manifold (1996)
4. Jiang, C., Yang, C., Shu, F., Wang, J., Mao, M., Sheng, W.W., Chen, Q.: MIMO precoding using rotating codebooks. IEEE Transactions on Vehicular Technology 60(3), 1222–1227 (2011)
5. Love, D.J.: Tables of Complex Grassmannian packings (2004), http://www.ece.purdue.edu/~djlove/grass.html
6. Raghavan, V., Heath Jr., R.W., Sayeed, A.M.: Systematic Codebook Designs for Quantized Beamforming in Correlated MIMO Channels
7. Visotsky, E., Madhow, U.: Space-time transmit precoding with imperfect feedback. IEEE Trans. Inf. Theory 47(6), 2632–2639 (2001)
8. Love, D., Heath Jr., R.: Limited feedback unitary precoding for spatial multiplexing systems. IEEE Trans. Inf. Theory 51(8), 2967–2976 (2005)
9. Love, D., Heath Jr., R.: Limited feedback unitary precoding for orthogonal space-time block codes. IEEE Trans. Signal Process. 53(1), 64–73 (2005)
10. Cho, Y.S., Kim, J., Yang, W.Y., Kang, C.G.: MIMO-OFDM Wireless Communications with MATLAB

Efficient Techniques for the Implementation of AES SubByte and MixColumn Transformations

K. Rahimunnisa[1], M. Priya Zach[1], S. Suresh Kumar[2], and J. Jayakumar[3]

[1] Department of Electronics and Communication Engineering, Karunya University, Coimbatore, India
[2] Department of Electronics and Communication Engineering, Dr. N.G.P Institute of Technology, Coimbatore, India
[3] Department of Electrical and Electronics Engineering, Karunya University, Coimbatore, India

Abstract. The Advanced Encryption Standard, AES, is commonly used to provide data confidentiality and authentication in several security systems. Designing efficient hardware architecture with small hardware resource usage is a challenge. In this paper, a new technique for the FPGA implementation of the Subbyte and MixColumn transformations, an important part of AES, is introduced. Sub-byte transformation in AES is operated using S-box for each byte. The hardware complexity in AES is dominated by AES substitution box (S-box). S-box is considered as one of the most complicated and costly part of the system due to its non-linear structure. It has high power consumption and high design complexity. In this paper, S-box is optimized by using multiplexer logic design. It is compared to the typical ROM based lookup table and the combinational logic designs. The MixColumn is also optimised by shifting the bytes and reusing the resources. This is also done using the multiplexer logic.

Keywords: AES, S-box, LUT, MixColumn Transformations.

1 Introduction

The Advanced Encryption Standard (AES) can be used to provide security services such as data confidentiality or authentication. Data confidentiality provides protection of data from being disclosed to unauthorized parties. Data authentication is the assurance that the received data has not been replayed or affected by modification, insertion, or deletion, and also the sender is authenticated. AES was standardized by the National Institute of Standards and Technology (NIST) in 2001[1]. NIST selected Rijndael as the proposed AES algorithm. Rijndael has many advantages. Hardware implementation has high speed. AES has a wide range of applications in realtime, which requires confidentiality of data transferred. AES is a symmetric block cipher which uses same key for encryption and decryption. The specification of the AES block cipher, defines two functions: encryption that generates ciphertext and decryption that produces plaintext. The AES has a block length of 128 bits and key length of 128,192 or 256 bits. The basic unit of processing in the AES algorithm is a byte. The AES operates on a 4x4 array of bytes which is called a state. The state undergoes 4

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 497–506.
springerlink.com

transformations, namely the AddRoundKey, SubByte, ShiftRow and MixColumn transformation.

In AES, the two expensive transformations in terms of computational resources are MixColumns and SubBytes transformations. This paper discuss few techniques to optimize these transformations and hence the algorithm.

2 An Introduction to AES

The AES algorithm operates on 128 bits of data and generates 128 bits of output. The length of the key used to encrypt this input data can be 128, 192 or 256 bits. In this paper, 128 bits of data and key are used. N_b which defines the number of columns of 32 bits is, N_b =128/32=4. Similarly N_k which defines the number of columns of 32 bits of key is, N_k =128/32= 4. The number of rounds N_r =10 when N_k= 4.

The AES algorithm basically consists of four byte oriented transformation and a key expansion block. The blocks are repeated for 10 rounds by applying the inputs to produce cipher text block. For the first nine rounds all four blocks are repeated but for the final round the MixColumns block is excluded. The basic building block of AES, SubBytes, ShiftRows, MixColumns and AddRoundKey are shown in Fig. 1.

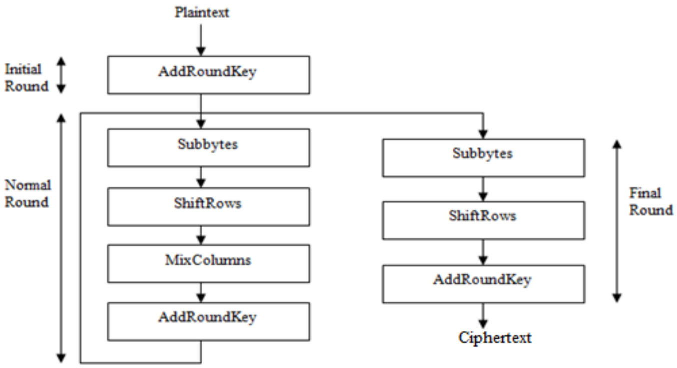


Fig. 1. Basic building block of AES

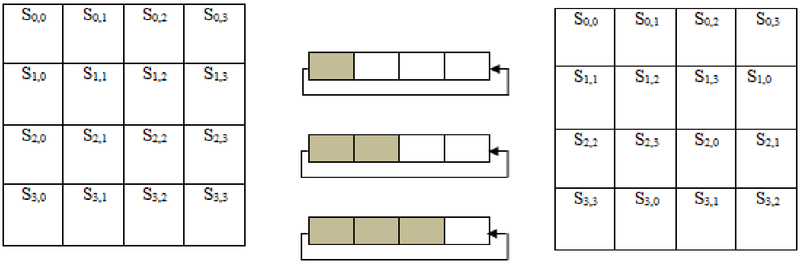
3 Individual Block Description

3.1 SubBytes

AES defines a 16×16 matrix of byte values, called an S-box [1]. SubByte transformation is a nonlinear substitution that operates on individual bytes using a substitution table (S-Box), which contains a permutation of all 256 possible 8-bit values. S-Box is also defined as the multiplicative inverse in the finite field GF(2^8) with the irreducible polynomial m(x)=$x^8+x^4+x^3+x+1$ followed by an affine transformation.

3.2 ShiftRows

ShiftRows essentially consists of shifting the bytes in the row. It is a transposition step on the row of the state where each row of the state is shifted cyclically by certain number of steps. The first row (row 0) is unaltered. The second row (row 1) is shifted by one byte, the third row is shifted by two bytes and final row is shifted three bytes. It also ensures that each byte in each row does not interact solely with their corresponding bytes. The transformation is shown in Fig.2.



$S_{0,0}$	$S_{0,1}$	$S_{0,2}$	$S_{0,3}$
$S_{1,0}$	$S_{1,1}$	$S_{1,2}$	$S_{1,3}$
$S_{2,0}$	$S_{2,1}$	$S_{2,2}$	$S_{2,3}$
$S_{3,0}$	$S_{3,1}$	$S_{3,2}$	$S_{3,3}$

$S_{0,0}$	$S_{0,1}$	$S_{0,2}$	$S_{0,3}$
$S_{1,1}$	$S_{1,2}$	$S_{1,3}$	$S_{1,0}$
$S_{2,2}$	$S_{2,3}$	$S_{2,0}$	$S_{2,1}$
$S_{3,3}$	$S_{3,0}$	$S_{3,1}$	$S_{3,2}$

Fig. 2. Row transformation

3.3 MixColumns

The MixColumns function takes four bytes as input and outputs four bytes, where each input byte affects all four output bytes. Together with ShiftRows, MixColumns provides diffusion in the cipher. Each column is treated as a polynomial over GF (2^8) and is then multiplied modulo $x^4 + 1$ with a fixed polynomial $c(x) = \{03\}x^3 + \{01\} x^2 + \{01\} x + \{02\}$.

3.4 AddRoundKey

In the AddRoundKey transformation, a Round Key is added to the State by a simple bit wise XOR operation. Each Round Key consists of words from the key schedule. Those words are added into the columns of the state.

3.5 KeyExpansion

KeyExpansion generates a total of $N_b(N_r + 1)$ words. The algorithm requires an initial set of N_b words, and each of the N_r rounds requires N_b words of key data. The resulting key schedule consists of a linear array of 4-byte words, denoted $[w_i]$, with i in the range $0 \leq i < N_b(N_r+1)$.

The KeyExpansion has three steps: Byte Substitution subword( ), Rotation rotword ( ) and XOR with RCON (round constant). The subword ( ) function takes a four byte input and applies the byte substitution operation and produces an output word. The rotword ( ) takes a word $[w_0,w_1,w_2,w_3]$ as input and performs a cyclic permutation to produce $[w_1,w_2,w_3,w_0]$ as output word. SubWord(RotWord(temp)) is XORed with Rcon[j] – the round constant. The round constant is a word in which the three rightmost bytes are zero. It is different for each round and defined as: Rcon[j] = (RC[j], 0,0,0), where RC[1] = 1, RC[j] = 2 * RC[j-1]. Multiplication is defined over GF (2^8). The structure of key expansion is shown in Fig.3 [2].

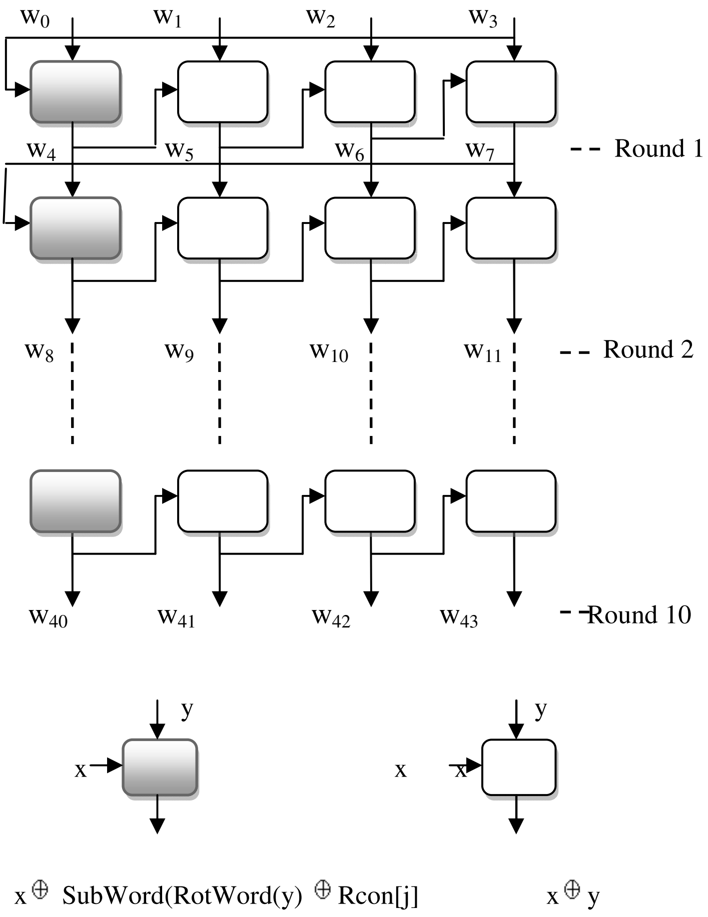


Fig. 3. Data path for key expansion

The first N_k words of the expanded key are filled with the input key. With the help of these initial words rest the words are generated iteratively. Each round key has 128 bits, and is formed by concatenating four words.

4 Related Work

S-box is implemented by different ways. Traditionally, it was implemented by look up tables (LUT) [3] which store all 256 it predefined values of S-box in a ROM. The advantage of LUT is it offers a shorter critical path. But, it requires larger area to implement. Another way is to design the S-box circuit using combinational logic [4]. It will be explained in section 5. It has less number of transistors and smaller area than ROM based implementation. It also has shorter critical path and is more flexible for speed optimization.

There are many ways to implement the MixColumns also. An originally proposed one in the AES takes the form of xtime. Another method was using the counter for shifting the bytes. These methods will be explained in section 6.

5 Implementation of S-Box

The major factors that influence the implementation techniques are speed and area cost. The efficiency of AES hardware implementation in terms of size, speed, security and power consumption depends mainly on the AES architecture. As S-Box is considered as a full complexity design and causes high power dissipation in AES, this paper is focused on the way to implement it efficiently. Out of four transformation S-box consumes more power.

The S-Box design uses combinational logic [4] to solve the unbreakable delay in look-up table. The S-box has 8 bit input and 8 bit output. It has 256 bit vectors. The logic function to realize each byte is derived from the Boolean expression using k-map. Each input byte is represented as a,b,c,d,e,f,g and h. The expressions are derived based on the 8-bits in previous work [5]. The 4 bit data input of least significant bit (LSB) will be the input of the sixteen module logic function (M1, M2, M3… M16) derived using Boolean simplification based on Karnaugh map. Another 4 bit data of most significant bit (MSB) will be the selection input of 16 to 1 multiplexer that will derive the output for S-box. Based on the MSB bits each module is selected. Each module in the architecture implies the rows in the sbox. The Boolean equation is derived for each row by taking the 4bit LSB as variables. This architecture can be used for SubByte Transformation. The S-Box architecture is shown in Fig.4.

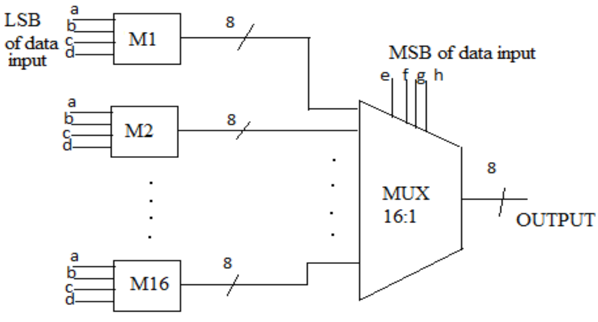


Fig. 4. Sbox architecture using combinational logic

In the proposed technique 16×16 S-box is divided into four blocks. Each block consists of 8×8 S-box with 64 values in each blocks. The blocks are selected using a 4:1 multiplexer. In S-box the 4-bit MSB is represented as row and 4-bit LSB is represented as column. To select each block, the first MSB bit and first LSB bit is taken. These bits are given as select line input to the multiplexer. If the select line is 00, the left half of the upper part of the S-box is selected. The right half is selected when the select line is 01. Similarly, the lower left and right part is selected when the select line is 10 and 11 respectively. The structure for this method is shown in Fig.5. The number of input bits to the LUT is be minimized by using a demultiplexer. So that the area is minimized compared to the basic implementation.

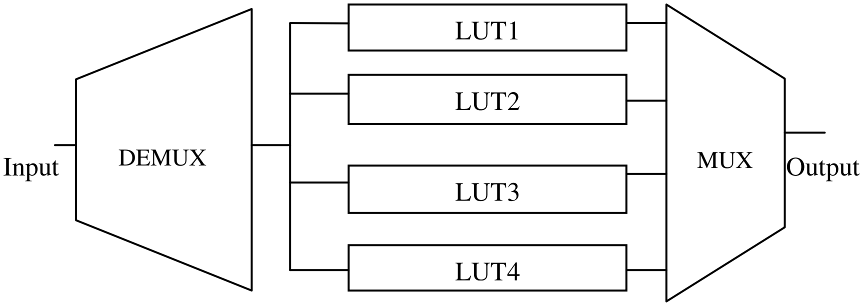


Fig. 5. Structure of S-box using MUX

6 Implementation of MixColumns

In the AES algorithm, the MixColumns are hardware demanding operations. Various architectures have been proposed for the implementation of the MixColumns transformation. By analyzing the basic operations employed in MixColumns, it is found that the modular multiplier is the vital calculation module.

The matrix multiplication of MixColumn could be represented as shown in Equation.1 [6]. The function *xtime* is used to represent the multiplication with ‘02’, modulo the irreducible polynomial $m(x)= x^8 + x^4 + x^3 + x + 1$. Implementation of function *xtime*() includes shifting and conditional XOR with ‘11B’.

$$\begin{aligned}
b_0 &= \mathit{xtime}\,(a_0 \oplus a_1) \oplus (a_0 \oplus a_1 \oplus a_2 \oplus a_3) \oplus a_0 \\
b_1 &= \mathit{xtime}\,(a_1 \oplus a_2) \oplus (a_0 \oplus a_1 \oplus a_2 \oplus a_3) \oplus a_1 \\
b_2 &= \mathit{xtime}\,(a_2 \oplus a_3) \oplus (a_0 \oplus a_1 \oplus a_2 \oplus a_3) \oplus a_2 \\
b_3 &= \mathit{xtime}\,(a_3 \oplus a_0) \oplus (a_0 \oplus a_1 \oplus a_2 \oplus a_3) \oplus a_3
\end{aligned} \tag{1}$$

From above representations, the MixColumn could be designed easily using just one basic module which imposes one *xtime* block, two or three byte-XOR logics and additional data path selector. This idea is depicted in Fig.6 [7]. The basic module of MixColumn is represented by the dashed line box in Fig.6.

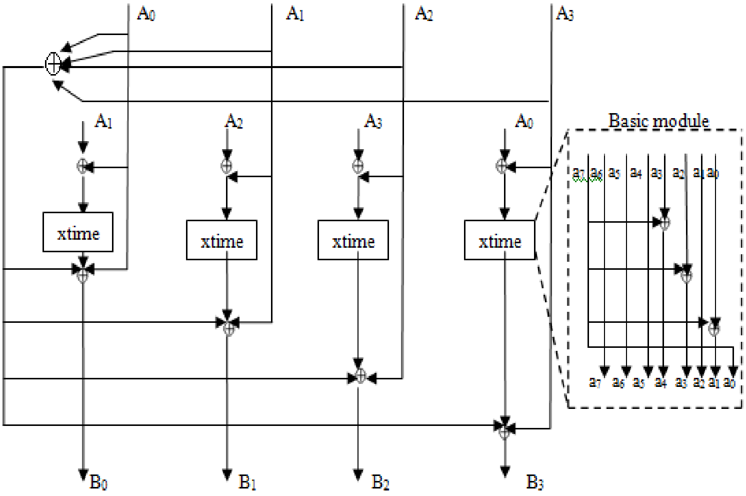


Fig. 6. MixColumns and its basic module

Another method used for implementing MixColumn is by counter for shift operation. By using the counter the bytes of each column are shifted in each clock cycle. The structure is shown in Fig.7 [8].

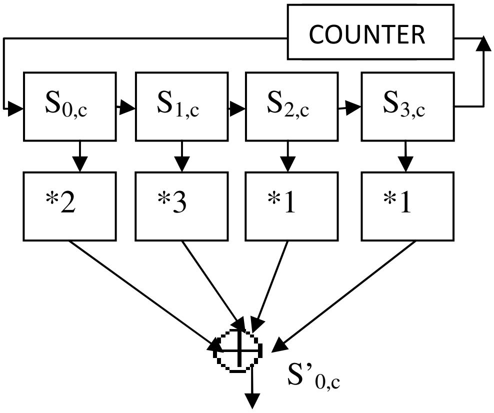


Fig. 7. Structure of MixColumns using counter

$S_{i,c}$ are byte-format and assumed to be loaded into the multiplication register either in parallel or serial manner before the computation starts. Data path is 8 bits wide. The computation of each transformed component takes one clock cycle. The next component can be computed with same set of data and multipliers, after the cyclic shift of $S_{i,c}$, i.e., $S_{i+1,c.}$ As a result, the computation is a word-serial scheme. One column transform takes 4 clock cycles. The next data set will implement $S_{i,c+1}$.

In the proposed method, the MixColumn is implemented using multiplexer based on the matrix form. The structure is used for four columns operation for 128 bit input. Five 4:1 multiplexer is used in this structure. This is used to shift the bytes in each column. By shifting the bytes in each column the multiplication with values 2 and 3 in the matrix can be reused. So that the area can be optimised. The structure using multiplexer for MixColumn is shown in Fig.8. In the MixColumn transformation each byte from the multiplexer will be multiplied with 2,3 or 1 respectively based on the matrix given in Equation.2.

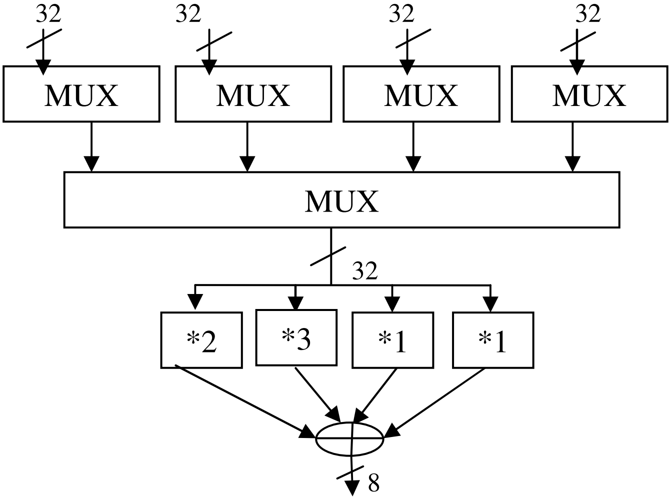


Fig. 8. Structure using MUX for MixColumn

$$\begin{pmatrix} S'_{0,c} \\ S'_{1,c} \\ S'_{2,c} \\ S'_{3,c} \end{pmatrix} = \begin{pmatrix} S_{0,c} & S_{1,c} & S_{2,c} & S_{3,c} \\ S_{1,c} & S_{2,c} & S_{3,c} & S_{0,c} \\ S_{2,c} & S_{3,c} & S_{0,c} & S_{1,c} \\ S_{3,c} & S_{0,c} & S_{1,c} & S_{2,c} \end{pmatrix} \begin{pmatrix} 2 \\ 3 \\ 1 \\ 1 \end{pmatrix} \tag{2}$$

Based on the select lines the multiplexer select each word. When the select line is {00}, the MUX selects {$S_{0,c}$, $S_{1,c}$, $S_{2,c}$, $S_{3,c}$} and it will multiplied by the corresponding coefficient. When the select line {S0,S1}is {01} the word will be shifted by one byte, {$S_{1,c}$, $S_{2,c}$, $S_{3,c}$, $S_{0,c}$}. The operation thus continues for other select lines. By shifting the bytes in each column the multiplication with values 2 and 3 in the matrix can be reused. Because the state of AES algorithm consist byte of arrays, the most operations could be processed by unit of byte. In the MixColumn transformation each byte from the multiplexer will be multiplied with 2, 3 or 1 respectively based on the matrix.

7 Results and Discussion

The design of S-box and MixColumn implementation is done in Verilog code. The power and area are found using Synopsys tool. Reducing the usage of hardware resources results in a smaller device. For instance, using a smaller device can be an important factor in reducing the hardware cost. In application, such as space application, that have low power requirements, a smaller device size decrease the overall power consumption.

7.1 Results for S-Box Implementation

The power and area can be found using Synopsys tool. The results are shown in Table.1. In the table, S-box designed LUT, combinational logic and multiplexer are compared. The area can be reduced by combinational method compared to the other technique. It is clear that the multiplexer logic has less power consumption and combinational logic requires less area but has very high power consumption.

Table 1. Results for Subbyte

Type	Area (μm^2)	Power (μW)
LUT	2146	57.21
Combinational logic	728	121.8
MUX	848	54.81

7.2 Results for MixColumns Implementation

The Synopsys results are given in Table.2. The proposed design is compared with other methods.

Table 2. Results for MixColumn on Synopsys

Type	Area (μm^2)	Power (μW)
Matrix	1296	225.5
Equation	1399	315.43
Counter	1168	123.6
MUX	879	32.29

In the multiplexer method of MixColumn, an FPGA design is proposed which uses less hardware compared to previous work. The MixColumn with multiplexer is having less area and low power consumption.

8 Conclusions

In this paper, an FPGA implementation of S-box and MixColumns transformation was proposed with less hardware utilization. Due to more efficient resource sharing, the proposed design for S-box and MixColumns transformation provide the smallest hardware usage on an FPGA. Thus a design with minimum area and low power requirement was designed. The S-box implementation with the multiplexer and MixColumns implementation with the multiplexer method is found to be optimized. With these optimised techniques a compact AES architecture can be developed. Thus the overall AES encryption implementation with the proposed architecture reduces usage of hardware resources. The proposed design can be used to provide security services such as confidentiality or authentication.

References

[1] National Institute of Standards and Technology (U.S.), Advanced Encryption Standard, `http://csrc.nist.gov/publications/fips/fips197/fips-197.pdf`

[2] Reddy, S.K., Sakthivel, R., Praneeth, P.: VLSI Implementation of AES Crypto Processor for High Throughput. International Journal of Advanced Engineering Sciences and Technologies 6(1), 022–026 (2011)

[3] Xinmiao, Z., Parhi, K.K.: High-speed VLSI architectures for the AES algorithm. IEEE Trans. on VLSI Systems 12, 957–967 (2004)

[4] Ahmad, N., Hasan, R., Jubadi, W.M.: Design of AES S-Box using combinational logic optimization. In: IEEE Symposium on Industrial Electronics and Applications (ISIEA 2010), Penang, Malaysia, October 3-5 (2010)

[5] Rachh, R.R., Ananda Mohan, P.V.: Implementation of AES S-Boxes using combinational logic. In: IEEE International Symposium on Circuits and Systems, pp. 3294–3297 (2008)

[6] Kim, M., Kim, J., Choi, Y.: Low Power Architecture of AES Crypto Module for Wireless Sensor Network. World Academy of Science, Engineering and Technology (2005)

[7] Ahmad, E.G., Shaaban, E., Hashem, M.: Lightweight MixColumns Implementation for AES. In: Proceedings of the 9th WSEAS International Conference on Applied Informatics and Communications, pp. 253–258

[8] Noo-intara, P., Chantarawong, S., Choomchuay, S.: Architectures for Mix-Column Transform for the AES. In: ICEP (2004)

Feature Selection for Detection of Ad Hoc Flooding Attacks

Sevil Sen and Zeynep Dogmus

[1] Department of Computer Engineering, Hacettepe University
ssen@cs.hacettepe.edu.tr
[2] Faculty of Engineering and Natural Sciences, Sabanci University
zeynepdogmus@sabanciuniv.edu

Abstract. In recent years ad hoc networks have become very attractive for many applications such as tactical and disaster recovery operations. However they are vulnerable to many attacks. The vulnerabilities of wired networks such as denial of service (DoS), eavesdropping, spoofing and the like, becomes more acute in these networks. Especially it is hard to differentiate DoS attacks in these highly dynamic systems. In this research, we design an intrusion detection model using Support Vector Machines (SVM) in order to detect a popular DoS attacks on these networks, namely ad hoc flooding attacks. We evaluate its performance on simulated networks with varying traffic and mobility patterns. Furthermore we investigate to choose the relevant features using Genetic Algorithms (GA) in order to increase SVM performance on detection of these attacks.

1 Introduction

Mobile ad hoc networks are one of the fastest growing areas of research. This new type of self-organizing network combines wireless communication with a high degree node mobility. They do not have any fixed infrastructure such as base stations or centralized management points as in conventional networks. The nodes cooperate with each other to provide basic functionality such as routing in a network, independent of any fixed infrastructure or centralized management. This flexibility makes them attractive for many applications such as military applications, disaster recovery operations, and virtual conferences.

These networks have different properties than conventional networks such as lack of central points, dynamic topology, resource-constraints and the like. This new networking is by its very nature more vulnerable to attacks than wired networks. Furthermore their specific features present a challenge for security solutions such as intrusion detection systems. Especially the impact of mobility on intrusion detection is a complex matter. It has both positive and negative impacts on security. On one hand, mobility helps security that victims could receive (direct or promiscuously) only parts of the falsified packets due to link breakages caused by mobility. So, the attacker might partially achieve his goal. On the other hand, the attacker hides himself from the detection system under high mobility which makes it difficult to differentiate normal behaviour of the network from anomaly behaviour in this environment. We particularly focus on ad hoc flooding attacks in the second group. These highly mobile, complex systems should

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 507–513.
springerlink.com

be modelled in order to detect attacks against them. In this research we employ SVM algorithms to achieve that. We evaluate our model created by SVM on networks with varying traffic and mobility patterns. Moreover, we investigate the selection of relevant features by using GA in order to model these complex systems better. We know of no other proposed approach in the literature on selecting features for intrusion detection in these networks.

2 Related Work

The specific features of ad hoc network make application of existing intrusion detection approach to this environment problematic. Therefore researchers have been working on new approaches or adaptation of existing approaches to ad hoc networks.

One of the most commonly proposed techniques on these networks is specification-based intrusion detection. This technique has been applied to a variety of routing protocols on ad hoc networks [1][2]. A few signature-based IDSs have also been proposed. In [3], an approach is proposed based on a stateful misuse detection technique which defines state transition machines for detecting known attacks on AODV [4]. In [5], an IDS is proposed which uses a specification-based technique for attacks that violate the specifications of AODV directly and an anomaly-based technique for other kinds of attacks such as DoS. Since wireless nodes can overhear traffic in their communication range, promiscuous monitoring can also be used to detect attacks such as dropping and modification. Mobile agents have been suggested as another way to provide communication between IDS agents.

Few artificial intelligence based intrusion detection systems have been proposed to explore the complex behavioural space of ad hoc networks. In the first proposed intrusion detection systems for these networks [6], statistical anomaly-based detection is chosen over misuse-based detection, so unknown attacks could be detected. The SVM Light and RIPPER classifiers are employed and compared in that research. In [7] a Markov-chain based local anomaly detection model is proposed for a Zone-Based IDS architecture. The network is partitioned into zones based on geographic location. Another approach which constructs an anomaly-detection model automatically by extracting the correlations among monitored features is proposed in [8]. Furthermore, they introduce simple rules to determine attack types and sometimes attackers after detecting an attack using cross-feature analysis. In [9], an approach which takes into account limited power issues of nodes by using multi-objective evolutionary computation techniques is proposed. In [10], four classification algorithms are applied for intrusion detection in ad hoc networks and compared.

SVM has been applied to ad hoc networks in order to detect known attacks [6][10]. In [6], while SVM shows a good performance on the routing protocol DSR, its performance on distance-vector routing protocols such as DSDV and AODV is poorer. In [10], SVM is shown to be the best algorithm among four classification methods on detection of some known attacks against AODV (average detection rate: %77, average false positive rate: %0.97). In this research we aim to investigate if we could achieve a good performance with SVM by using an expanded feature set.

3 Intrusion Detection by Using Support Vector Machines

Support Vector Machines (SVM) is a supervised learning algorithm used mainly for classification and regression analysis. It is one of the popular techniques for intrusion detection in conventional networks due to their good performance both in unbalanced and balanced datasets. In this research, we employ this promising technique to ad hoc networks environment.

3.1 Support Vector Machines (SVM) Model

SVM algorithm basically constructs a hyperplane or a set of hyperplanes. In order to have a good separation, the hyperplane's distance to the nearest training point of any class should be maximized. The larger the distance the better classification is achieved. In our experiments, we use libSVM library [11].

In this research the networks are simulated by ns-2 [12]. Mobility patterns of the nodes are simulated by the Random Waypoint model which is created using BonnMotion [13]. Different network scenarios are created with different mobility levels and traffic loads. 50 nodes are placed in a topology of 1000 m by 500 m. TCP traffic is used for communication. The maximum number of connections is set to 20 and 30 to simulate different traffic loads. The maximum speed of nodes is set to 20 m/s and the pause time between movements is set to 40, 20, and 5 s to simulate low, medium, and high mobility respectively. AODV periodic hello messages are used for local link connectivity. The simulations run 5000 s for training and 2000 s for testing.

Table 1 shows the features maintained at each node. This feature set is wider than other approaches which use SVM for intrusion detection in ad hoc networks [6][10]. The features can be categorised into two main groups: mobility-related and packet-related features. Mobility-related features help reflect the mobility model of a node or the network. Some of the mobility features give information about mobility directly such as changes in the number of neighbours. Others can be the results of mobility such as changes in the routing table (*e.g.* number of new routes, number of invalidated routes) in a time interval. Packet-related features include information about the frequency of the routing protocol control packets (RREQ, RREP, RERR) sent, received, or forwarded in a time interval. All features are local to a node, so no communication with other nodes is needed to gather them.

In order to evaluate our model, we focus on two metrics: detection rate and false positive rate. The detection rate (DR) shows the ratio of correctly detected intrusions to the total intrusions on the network. The false positive rate (FPR) shows the ratio of normal activities that are incorrectly marked as intrusions to the total normal activities on the network. An acceptable low rate of false alarms is as important as a high detection rate.

For training we use a dataset obtained from a network under medium mobility and traffic. It is an unbalanced dataset where normal cases are much more than abnormal cases. That is the reason we use weight parameter during the construction of our model. We try different weight parameters empirically and obtain different trade-offs between detection rate and false positive rate. Based on that, we choose the following weight parameters: 0.005 for normal cases, 0.1 for abnormal cases. We use C-SVC algorithm

Table 1. Features

	Features (of a node)
1	no. of neighbours
2	no. of added neighbours
3	no. of removed neighbours
4	no. of active routes
5	no. of routes under repair
6	no. of invalidated routes
7	no. of added routes by route discovery mechanism
8	no. of added routes by overhearing
9	no. of updated routes (modifying hop count, sequence number)
10	no. of added routes under repair
11	no. of invalidated routes due to expiry
12	no. of invalidated routes due to other reasons
13	no. of received route request packets destined to this node
14	no. of received route request packets to be forwarded by this node
15	no. of broadcasted route request packets from this node
16	no. of forwarded route request packets from this node
17	no. of received route reply packets destined to this node
18	no. of received route reply packets to be forwarded by this node
19	no. of initiated route reply packets from this node
20	no. of forwarded route reply packets from this node
21	no. of received broadcast route error packets (to be forwarded or not)
22	no. of broadcasted route error packets from this node
23	no. of received total routing protocol packets
24	no. of received total routing protocol packets to be forwarded
25	no. of initiated total routing protocol packets from this node
26	no. of forwarded total routing protocol packets by this node

which is the default algorithm in libSVM. The cost parameter between 5 and 50 is evaluated, and is chose to be 5. The model trained with these parameters which happens to be the optimal model considering trade-offs between detection rate and false positive rate is chose empirically.

3.2 Experimental Results

We evaluate our model on six networks under varying mobility and traffic levels and the results are demonstrated in Table 2. SVM shows a good performance on detection of ad hoc flooding attacks. As it is shown clearly, detection rate decreases and false positive rate increases under high traffic. It is the traffic level which affects the performance of the model much more than mobility.

Table 2. SVM Performance on Detection of Ad Hoc Flooding Attacks

Simulations	Detection Rate	False Positive Rate
low mobility, medium traffic	97.03%	0.83%
low mobility, high traffic	94.17%	2.10%
medium mobility, medium traffic	97.86%	0.76%
medium mobility, high traffic	96.20%	1.70%
high mobility, medium traffic	97.40%	0.95%
high mobility, high traffic	90.02%	1.83%

As it is stated before, SVM is one of the most promising techniques used for intrusion detection in wired networks. That is the reason we aim to increase its performance for ad hoc networks successfully in this research. Since the right choice of features is much important for any machine learning method, we mainly focus on reducing the features given in Table 1 for a better performance by using genetic algorithms.

4 Selection of Features by Using Genetic Algorithms

The choice of which network characteristics can be used for machine learning is very important. They must contain sufficient information to allow the fundamentals to be developed. However irrelevant and too many features could degrade the performance of the learning algorithms. In this research, we investigate if we could increase the performance of our model with the selection of right features for training.

Since ad hoc networks have complex properties, we use all possible features which could represent its behaviour at the routing level as given in Table 1. However some of these features could not be representative for detecting ad hoc floodding attacks. Genetic algorithms have been used for selection and reduction of features in many areas successfully [14,15,16,17], that's why we use this technique to increase the performance of our model by using the relevant features.

4.1 Genetic Algorithms (GA)

Genetic Algorithms is an evolutionary computation technique inspired from biological evolution. The algorithm starts with creating individuals (generally randomly) which are the candidates solutions for the problem. Traditionally, individuals are represented in binary as strings of 0s and 1s. Each individual is assigned a *fitness* value which shows how the individual solves or comes close to the solution. Some genetic operators (selection, crossover, mutation, reproduction and etc.) are applied on individuals based on their fitness values until the termination criteria is satisfied. The aim is to provide better individuals in the new population.

4.2 Feature Selection

At first, random individuals are created for the solution. These individuals represent which features are used for the SVM model, and which not. SVM algorithm is run for each individual and a fitness value is assigned to each individual based on the formula below. The GA algorithm continues until a defined generation is reached.

$$fitness = detection\,rate - false\,positive\,rate \tag{1}$$

We use ecj 20 toolkit [18] for the GA implementation in our experiments. The GA parameters are selected as follows: 100 for population size, 100 for generation size, 0.9 for crossover probability, 0.1 for reproduction probability. Other parameters used are the default parameters of the toolkit. At each generation 100 individual is evaluated by creating a SVM model for each individual. Since our training dataset is huge, fitness values might not be obtained in a reasonable time. That's the reason we use a balanced small subset of our training data and do not employ any weight parameter consequently.

4.3 Experimental Results

GA algorithm is run ten times and the feature set with the highest fitness value is selected. A SVM model is run with this feature set ($\{1, 2, 9, 10, 16, 19, 26\}$ in Table 1) and all training dataset, and evaluated on different network simulations again. The results are demonstrated in Tablo 3. Both an increase in detection rate and a decrease in false positive rate is seen in the results. The false positive rate is below 1% in most of the cases. A noticeable performance increase is achieved by the feature reduction.

Table 3. SVM Performance after Feature Selection

Simulations	Detection Rate	False Positive Rate
low mobility, medium traffic	97.69%	0.23%
low mobility, high traffic	96.49%	1.29%
medium mobility, medium traffic	99.67%	0.19%
medium mobility, high traffic	97.57%	1.08%
high mobility, medium traffic	97.44%	0.39%
high mobility, high traffic	93.25%	1.00%

5 Conclusion

In this research, we aim to detect ad hoc flooding attacks. We investigate the use of SVM in this environment and evaluate its performance on networks with varying traffic and mobility patterns. In order to increase its performance, we explore the selection of relevant features by using GA. It is shown that the performance of SVM has increased with the reduced feature set obtained by GA, both an increase in detection rate and a decrease in false positive rate is observed. As far as we know this is the first attempt on selecting relevant features by using artificial intelligence based techniques for intrusion detection in these networks. In this research, the parameters of SVM is chose empirically. In the future, the effects of these parameters could be investigated by using GA as well.

References

1. Tseng, C.-Y., Balasubramayan, P., Ko, C., Limprasittiprn, R., Rowe, J., Lewitt, K.: A Specification-Based Intrusion Detection System for AODV. In: Proceedings of the ACM Workshop on Security in Ad Hoc and Sensor Networks, pp. 125–134 (2003)
2. Tseng, C.H., Wang, S.-H., Ko, C., Levitt, K.N.: DEMEM: Distributed Evidence-Driven Message Exchange Intrusion Detection Model for MANET. In: Zamboni, D., Kruegel, C. (eds.) RAID 2006. LNCS, vol. 4219, pp. 249–271. Springer, Heidelberg (2006)
3. Vigna, G., Srinivasan, K., Belding-Royer, E.M., Kemmerer, R.A.: An Intrusion Detection Tool for AODV-Based Ad hoc Wireless Networks. In: Proceedings of the 20th Annaual Computer Security Applications Conference, pp. 16–27. IEEE Computer Society (2004)
4. Perkins, C.E., Royer, E.M.: Ad-hoc on demand distance vector routing. In: Proceedings of IEEE Workshop on Mobile Computer Systems, pp. 90–100 (1999)
5. Huang, Y.-A., Lee, W.: Attack Analysis and Detection for Ad Hoc Routing Protocols. In: Jonsson, E., Valdes, A., Almgren, M. (eds.) RAID 2004. LNCS, vol. 3224, pp. 125–145. Springer, Heidelberg (2004)
6. Zhang, Y., Lee, W., Huang, Y.: Intrusion detection techniques for mobile wireless networks. Wirel. Netw. J. 9(5), 545–556 (2003), doi:10.1023/A:1024600519144
7. Sun, B., Wu, K., Pooch, U.: Zone-based intrusion detection for mobile ad hoc networks. Int. J. of Ad Hoc and Sens. Wirel. Netw. 2 (2003)
8. Huang, Y., Fan, W., Lee, W., Yu, P.S.: Cross-feature Analysis for Detection Ad-hoc Routing Anomalies. In: Proceedings of the 23rd International Conference on Distributed Computing Systems, pp. 478–487. IEEE (2003)
9. Sen, S., Clark, J.A.: Evolutionary Computation Techniques for Intrusion Detection in Mobile Ad Hoc Networks. Comput. Netw. 55(15), 3441–3457 (2011)
10. Mitrokotsa, A., Tsagkaris, M., Douligeris, C.: Intrusion Detection in Mobile Ad Hoc Networks Using Classification Algorithms. In: Proceedings of the Seventh Annual Mediterranean Ad Hoc Networking Workshop-Advances in Ad Hoc Networking, pp. 133–144. Springer (2008)
11. LibSVM: A Library for Support Vector Machines, http://www.csie.ntu.edu.tw/~cjlin/libsvm/
12. The network simulator, http://www.isi.edu/nsnam/ns/ (cited February 15, 2012)
13. BonnMotion: A mobility scenario generatin and analysis tool, http://net.cs.uni-bonn.de/wg/cs/applications/bonnmotion/ (cited February 15, 2012)
14. Wroblewski, J.: Finding Minimal Reducts Using Genetic Algorithms. In: Proceedings of the Second Annual Joint Conference on Information Sciences, pp. 186–189 (1995)
15. Lanzi, P.L.: Fast Feature Selection with Genetic Algorithms: A Filter Approach. In: Proceedings of IEEE Conference on Evolutionary Computation, pp. 537–540 (1997)
16. Yang, J., Honavar, V.: Feature Subset Selection Using A Genetic Algorithm. IEEE Intell. Syst. 12(2), 44–49 (1998)
17. Huang, C.-L., Wang, C.-J.: A GA-Based Feature Selection and Parameters Optimization for Support Vector Machines. Expert Syst. Appl. 31, 231–240 (2006)
18. ecj20: A Java-based Evolutionary Computation Research System, http://cs.gmu.edu/~eclab/projects/ecj/ (cited February 15, 2012)

Performance Analysis of SCTP Compared to TCP and UDP

Nagesha[1] and S.S. Manvi[2]

[1] Instrumentation Technology Department,
JSS Academy of Technical Education, Bengaluru, India
[2] ECE Department, Reva Institute of Technology and Management, Bengaluru, India
nageshashiva@gmail.com, sunil.manvi@revainstitution.org

Abstract. The SCTP (Stream Control Transmission Protocol) is relatively newer transport layer protocol which incorporates the core features of TCP (Transmission Control Protocol) and UDP (User Datagram Protocol) and also has many unique features like multihoming and multistreaming. Thus it is necessary to analyze the performance of SCTP and its advantages over TCP and UDP. This paper presents the performance analysis of SCTP compared to TCP and UDP by using two topologies with NS2 simulator. The two topologies are: dumb-bell topology with single home and dual home SCTP. Measured performance parameters are throughput, delay, jitter and packet loss. The results indicate that the throughput of SCTP is better than the throughput of TCP and UDP. The jitter problem is less in SCTP compared to TCP. Transmission delay of SCTP is more compared to TCP and UDP transmission delay. Packet loss is zero for all three protocols. The single home SCTP and dual home SCTP gave similar performance.

Keywords: SCTP, throughput, jitter, delay, packet loss, multihoming, single home SCTP, dual home SCTP.

1 Introduction

The number of mobile devices (e.g., mobile phone) equipped with multiple network interfaces such as IEEE802.11, IEEE802.16, IEEE802.15, etc. are increasing and also demanding many multimedia streaming applications. In this context, gaining capability to support multimedia communication in the next generation heterogeneous Internet is important. The Internet Engineering Task Force (IETF) has taken initiative and proposed a new transport layer protocol named stream control transmission protocol (SCTP) [1][2] to facilitate multimedia communication with multiple IP addresses attached to ends.

The SCTP is relatively newer protocol which incorporates the core features of TCP (connection oriented, reliable data transfer) and UDP (preservation of message boundaries) and also has many unique features like multihoming and multistreaming

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 515–524.
springerlink.com

[1][2]. Multihoming allows to attach multiple IP addresses to an end, and multistreaming is used to avoid head of line blocking and dynamic address reconfiguration (for mobile communication).

Multihoming allows an association (connection) between two endpoints with multiple IP addresses. An example of SCTP multihoming is shown in figure 1 where both endpoints A and B have two interfaces bound to an SCTP association. One of the address is designated as a primary, while the other can be used as a backup (secondary) in case of failure of the primary address. Retransmission of lost packets can also be done over the secondary address.

Table 1 Shows the comparison of SCTP, TCP and UDP features [3][4].

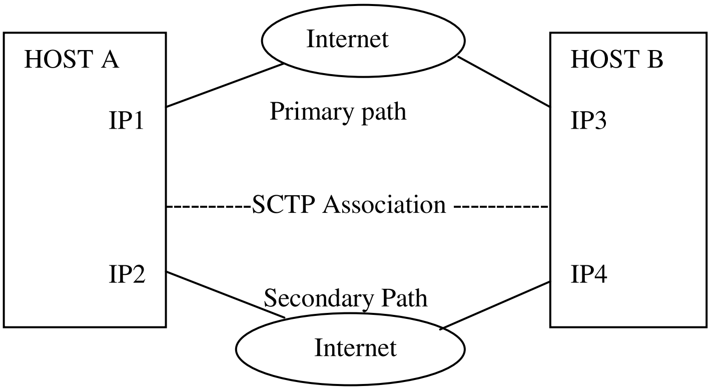


Fig. 1. Principle of multihoming

Some related works are as follows. An analytic model which approximates the steady state throughput of an SCTP session is given in [7]. The model which establishes a relationship between throughput and congestion control mechanism is given in [8]. The comparison of SCTP and TCP in the viewpoint of the throughput performance over the Linux platform is given in [9]. The work given in [10] reports throughput of SCTP and TCP streams when they co-exist together in the same channel under self-similar traffic environment. The Single Direction Delay Difference (S3D) method presented in [11] can bring better throughput when the send and receive direction delay are quite different.

In this paper, the performance parameters throughput, jitter, delay and packet loss of SCTP, TCP and UDP are measured by using dumb-bell topologies with single home and dual home SCTP. The results indicate that the throughput of SCTP is better than TCP and UDP. The SCTP is competent enough with TCP and UDP when compared to other performance parameters.

Table 1. Comparison of SCTP, TCP and UDP

Features	**SCTP**	**TCP**	**UDP**
Full duplex data transmission	yes	yes	Yes
Connection oriented	Yes	Yes	No
Reliable data transfer	Yes	Yes	No
Partially reliable data transfer	Yes	No	No
Preservation of message boundaries	Yes	No	Yes
Congestion and flow control	Yes	yes	No
Ordered data delivery	Yes	Yes	No
Unordered data delivery	Yes	No	Yes
Protection against SYN flood attacks	Yes	No	NA
Multihoming	Yes	No	No
Multistreaming	Yes	No	No
Dynamic address configuration	Yes	No	No
Selective acks	Yes	optional	No

The rest of the paper is organized as follows. Section 2 deals with the simulation setup. Section 3 explains the performance parameters. Results and discussions are presented in section 4. Finally, conclusions are given in section 5.

2 Simulation Setup

The NS2 [5] based simulation is used to plot graphs throughput Vs traffic, jitter Vs traffic, delay Vs traffic and packet loss Vs traffic. The SCTP, TCP and UDP performance parameters are then compared.

To conduct simulation experiments NS2.35 and its SCTP patch (version 3.8, Delaware University) is used. We ran simulation from 0.5s to 7.5s in NS2 (i.e., for total of 7s). The topologies used for simulation and the performance parameters measured are given in this section.

We have used two topologies as given in figures 2 and 3: Single home and dual home SCTP. Link bandwidth and delays used in the simulation are given in figures. Other parameters used in the simulation are as follows: drop tail queues, IEEE 802.3

MAC with MTU = 1500 bytes, duplex links, delayed acks, one stream, CBR traffic for UDP and TCP with ftp of packet size of 1000 bytes. We measure the performance by considering three types of protocols applied separately in the simulation: TCP, UDP, and SCTP.

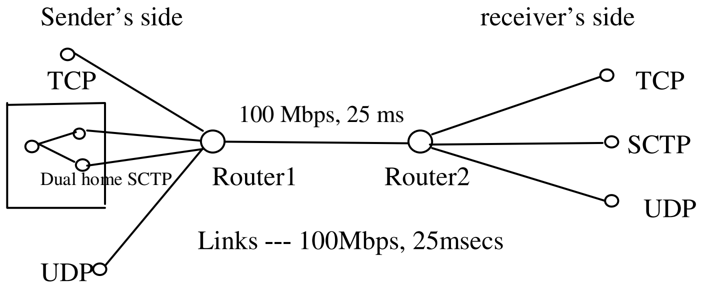


Fig. 2. Topology 1- Dual home

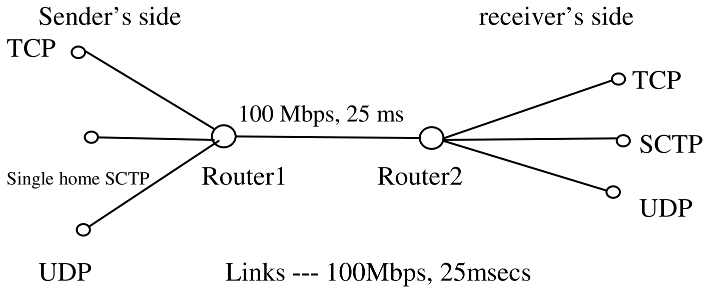


Fig. 3. Topology 2 – Single home

3 Performance Parameters

Throughput is measured by using the following: (N*S*8bits) / (Duration*1000000) Mbps, where N – Numbers of packets received, S – Size of packets in bytes, Duration=last packet received time in second minus the first packet sent time in seconds.

The delay is measured using difference of packet received time at the destination and packet sent time from the source. Jitter is about the different packets of data experiencing different delays in the network. The jitter is computed using the following: Arrival interval[i] = arrival time[i+1] – arrival time[i], where i=packet number.

The packet loss is measured by using difference of the number of packets transmitted and number of packets received.

4 Results and Discussion

This section presents the results in terms of throughput, delay, jitter and packe t loss. The dual home topology gave zero packet loss for SCTP, TCP and UDP data transmission.

4.1 Throughput

The figure 4 shows that the throughput of SCTP is far better than UDP and TCP. The throughput of SCTP rises to 3.19714 Mbps and then increases and decreases between 3.19714Mbps 5.21404Mbps. The TCP throughput increases to 1.00588Mbps (observe that TCP is sitting on 1.0000Mbps mark line) and stays constant. The UDP throughput increases to 0.914035Mbps and then stays constant.

We have also observed that the number of packets received in dual home topology by SCTP receiver, TCP receiver and UDP receiver is 2434, 865 and 866, respectively (even though all links are of same bandwidth (100Mbps) and same delay (25ms)).

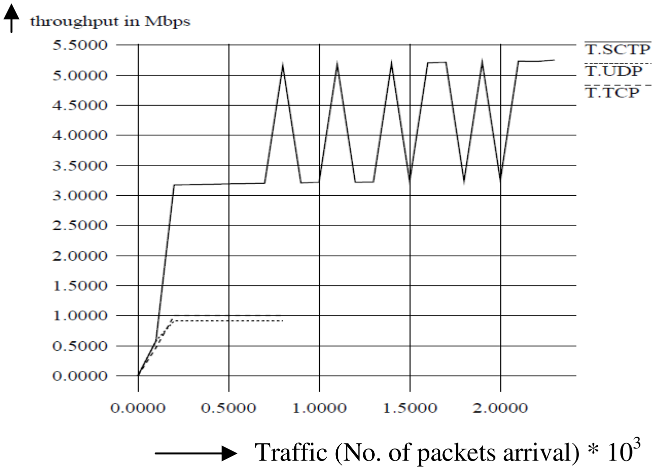


Fig. 4. Throughput Vs Traffic

4.2 Delay

We analyze the delay for dual home topology. The figure 5 shows the SCTP and UDP delay. For SCTP, delay is different for first 136 packets compared to the rest of the packets. The UDP delay is (shown in figure 5 along with SCTP) 0.240 ms. The figure 6 shows TCP delay which is also different for first 150 packets compared to the rest of the packets.

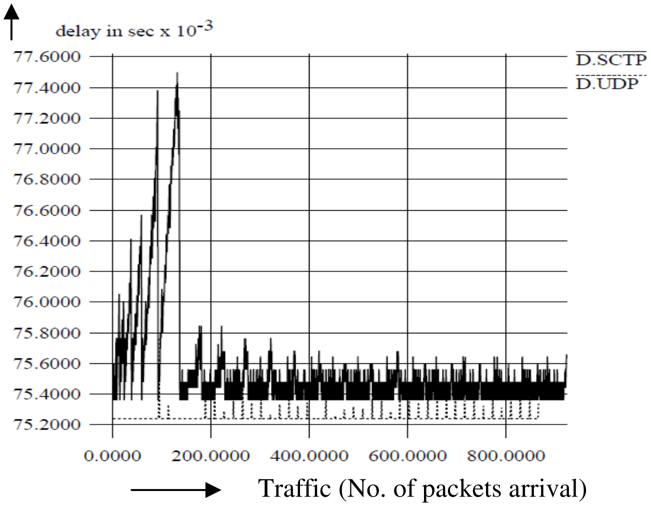


Fig. 5. Delay Vs traffic

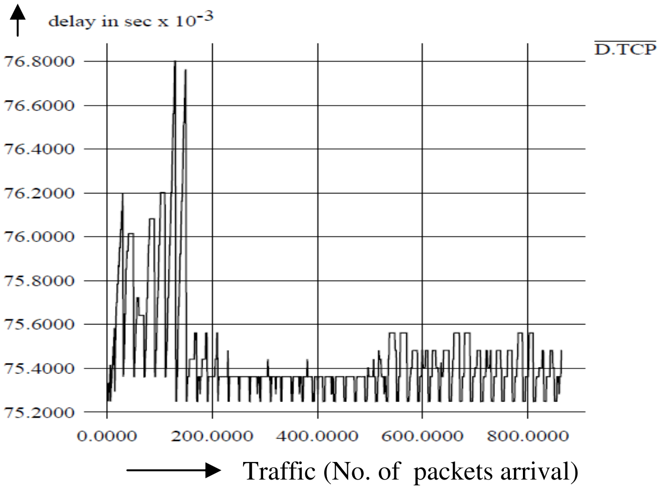


Fig. 6. TCP delay Vs Traffic

The delay characteristic of SCTP and TCP which is different for first 150 packets is shown in figure 7. Figure 8 shows the comparison of SCTP, TCP, and UDP delay. The delay experienced by TCP packets is less than (for almost all packets) the delay experienced by SCTP packets. The delay experienced by UDP is less compared to TCP and SCTP delay.

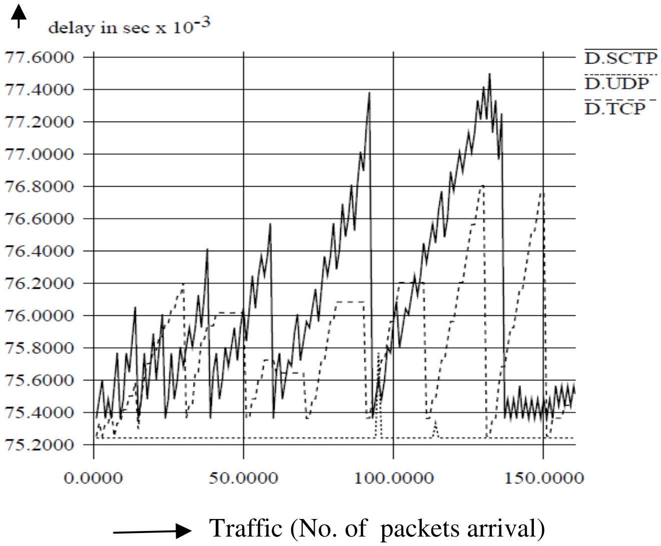


Fig. 7. First 150 packets delay characteristic

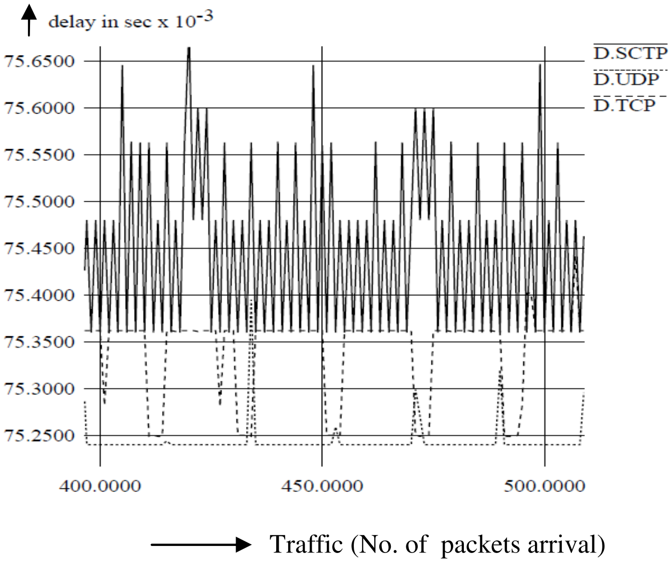


Fig. 8. Comparing SCTP, TCP & UDP delay

4.3 Jitter

The jitter of SCTP, TCP and UDP are shown in the figure 9 for dual home topology. The time gap between successive packets (for most of the packets) of SCTP is 0.000125s, TCP is 0.000323s and UDP is 0.008s. But we can see the spikes (rise in time gap between arrival of successive packets) in case of SCTP and TCP for some packets. But we can also observe that more number of spikes in TCP transmission compared to SCTP data transmission. UDP is not experiencing jitter problem but time gap between successive packets is more (8ms) compared to SCTP and TCP.

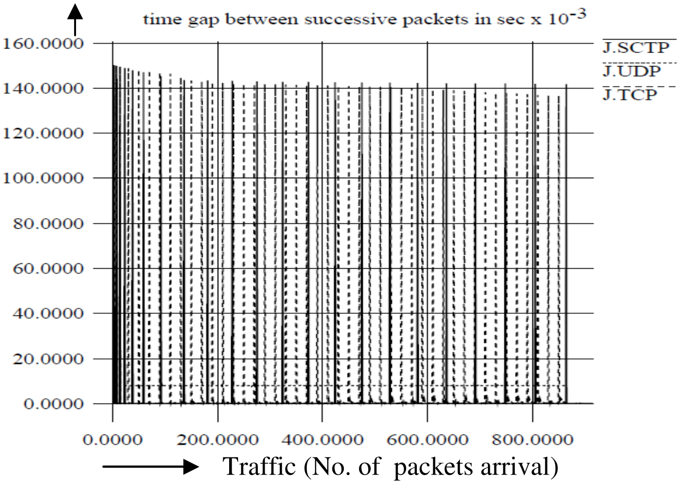


Fig. 9. Jitter of SCTP, TCP & UDP

4.4 SCTP Single Home and Dual Home SCTP Comparison

Fig. 10 shows the throughput of SCTP, TCP and UDP when single home SCTP (Fig. 3) is used. Comparison of fig.10 with fig.4 (dual home SCTP throughput) reveals that the single home SCTP throughput is exactly same as (even numeric values match) dual home SCTP throughput. The TCP and UDP throughput also remains same. Fig.11 shows the delay of SCTP, TCP and UDP when single home SCTP (fig.3) is used. Comparison of fig.11 with fig.5 (dual home SCTP delay) reveals that the single home SCTP delay is exactly same as dual home SCTP delay. The TCP and UDP delay also remains same. Similarly, jitter also follows the similar pattern of results.

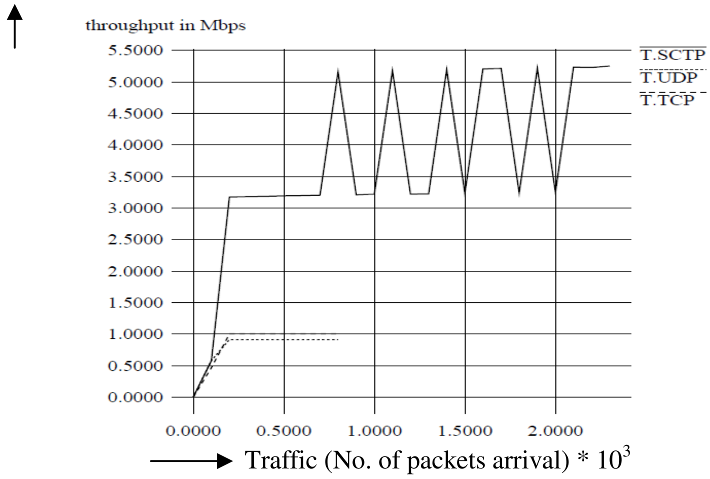


Fig. 10. Single home SCTP: Throughput Vs Traffic

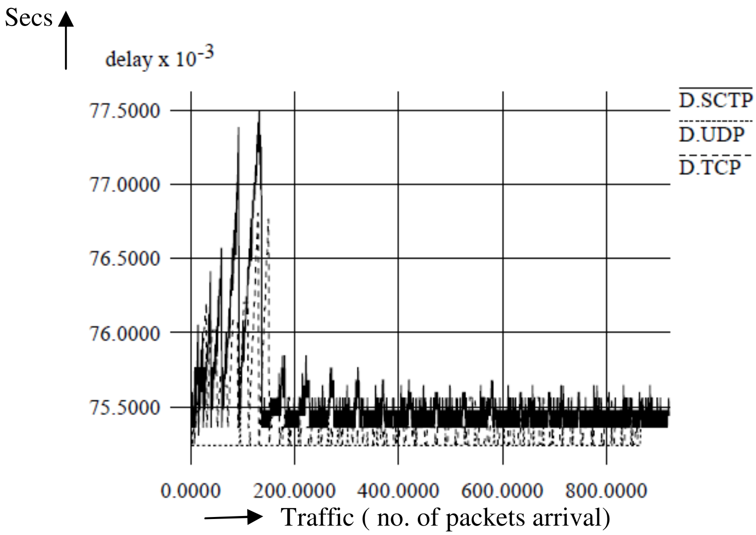


Fig. 11. Single home SCTP: Delay Vs traffic

Thus single home SCTP performance results are same as dual home SCTP performance results. In other words, the multihoming feature of SCTP is not degrading its performance. Also in the network it is not affecting the performance of either TCP or UDP.

5 Conclusions

The performance parameters throughput, jitter, delay and packet loss are measured for SCTP, UDP and TCP data transmission and then the parameters are compared. The SCTP is far better than UDP and TCP in throughput. The SCTP is also better in jitter when compared to TCP. The UDP is not experiencing any jitter. The delay is more in both cases of SCTP and TCP initially and later settles down to little above total link delays in the network. The delay is more in SCTP compared to TCP and UDP. The UDP delay is constant from first packet to last packet which is just above total link dealys. The packet loss is zero for all three protocols. The performance of single home SCTP and dual home SCTP is also compared. They gave exactly the same performance results.

References

1. Stewart, R.R., et al.: Stream control transmission protocol. IETF RFC 4960 (September 2007)
2. Dreibholz, T., Rathgeb, E.P., Rüngeler, I., Seggelmann, R., Tüxen, M., Stewart, R.R.: Stream control transmission protocol: Past, current and future standardization activities. IEEE Communications Magazine 49(4), 82–88 (2011)
3. Caro Jr., A.L., Iyengar, J.R., Amer, P.D., Ladha, S., Heinz II, G.J., Shah, K.C.: SCTP: A proposed standard for robust internet data transport. IEEE Computer Magazine 36(11), 56–63 (2003)
4. Natarajan, P., Baker, F., Amer, P.D., Leighton, J.T.: SCTP: What, why and How. IEEE Internet Computing 31(5), 81–85 (2009)
5. NS-2, The ns Manual (formally known as NS Documentation), `http://www.isi.edu/nsnam/ns/doc`
6. Lachlan, A., Cesar, M., Sally, F., Lawrence, D., Romaric, G., Wang, G., Lars, E., Ha, S., Rhee, I.: Towards a common TCP Evaluation suite. In: Proceedings of International Workshop on Protocols for Fast Long Distance Networks, Manchester, UK (2008)
7. Wallace, T.D., Shami, A.: An Analytic Model for the Stream Control Transmission Protocol. In: Proceedings of Global Telecommunications Conference (GLOBECOM), Florida, USA, Deccmber 6-10, pp. 1–5 (2010)
8. Cao, Y., Liu, C.: An Extended Throughput Analysis Model for SCTP with Scalable Congestion Control. In: International Forum on Information Technology and Applications (IFITA), Guangzhou, China, July 16-18, vol. 1, pp. 365–368 (2010)
9. Ha, J.-S., Kim, S.-T., Koh, S.J.: Performance Comparison of SCTP and TCP over Linux Platform. In: Huang, D.-S., Zhang, X.-P., Huang, G.-B. (eds.) ICIC 2005. LNCS, vol. 3645, pp. 396–404. Springer, Heidelberg (2005)
10. Charoenwatana, L., Rattanabung, S.: Coexistence of SCTP and TCP variants under self-similar network. In: Proceedings of Eighth International Joint Conference on Computer Science and Software Engineering (JCSSE), Nakhon Pathom, Thailand, May 11-13, pp. 17–22 (2011)
11. Yan, H., Gao, D., Song, F., Zhang, L.: Path Selection Based on Single Direction Delay Difference. In: Proceedings of 7th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM), Wuhan, China, September 23-25, pp. 1–4 (2011)

MSRCC – Mitigation of Security Risks in Cloud Computing

D. Chandramohan, T. Vengattaraman, M.S.S. Basha, and P. Dhavachelvan

Department of Computer Science, Pondicherry University,
Pondicherry, India
{pdchandramohan,vengat.mailbox,
smartsaleem1979,dhavachelvan}@gmail.com

Abstract. Cloud computing represents the latest phase in the evolution of Internet-based computing. In spite of the attractive properties it provides, security and privacy issues loom large for cloud computing. In this paper, we concentrate on the security threats involved in cloud by providing an Authenticated Key Exchange Protocol (AKExP). In this protocol, we make use of the identity management scheme and symmetric key encryption algorithm so that unauthorized users cannot get access to the sensitive data stored in the cloud. This proposed approach minimizes the computation cost and communication cost when compared with another authentication protocol, Identity-based Authentication Protocol. This paper provides an efficient and secure way to authenticate the user in the cloud both by the owner and cloud service provider.

Keywords: Cloud Request Dispatcher, Symmetric Encryption, Key Exchange, Cloud Security.

1 Introduction

Cloud computing is a class of next generation highly scalable distributed computing platform in which computing resources are offered 'as a service' leveraging virtualization and Internet technologies. Cloud computing represents a significant opportunity for service providers and enterprises. Cloud-based services include Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), and Software-as-a-Service (SaaS). Relying on cloud computing, enterprises can achieve cost savings, flexibility, and choice for computing resources. Because of these advanced features users are looking up to cloud computing to expand their on-premise infrastructure, by adding capacity on demand. Now, recession-hit companies are increasingly realizing that simply by tapping into the cloud they can gain fast access to best-of-breed business applications or drastically boost their infrastructure resources, all at negligible cost. Amazon's Elastic Compute Cloud (EC2) [3] and IBM’s Blue Cloud [4] are examples of cloud computing services. These cloud service providers allow users to instantiate cloud services on demand and thus purchase precisely the capacity they require when they require based on pay-per-use or subscription-based model. Although cloud computing provides a number of advantages that include economies

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 525–532.
springerlink.com

of scale, dynamic provisioning, increased flexibility and low capital expenditures, it also introduces a range of new security risks [5]. As more and more in sequence on those issues and companies info placed in the cloud, concerns are commencement to grow about just how safe a cloud environment is. We are able to identify the following threats in cloud environment

i. Maltreatment and despicable Use of Cloud Computing
ii. lacking confidence Application Programming Interfaces
iii. Malevolent Insiders
iv. Shared Technology Vulnerabilities
v. Data Loss/Leakage
vi. Account, Service & Traffic Hijacking

In cloud computing, the security is the main issue that catches the attention of many researchers. So that new models are developed in order to reduce the vulnerabilities in cloud. Cloud contains large amount of sensitive information which the malicious users and hackers wants their hands on it. Our proposed model provides a mechanism which allows only the authenticated user can gain access to the sensitive information stored in the cloud. We make use of the identity management schemes to provide unique identity to the users and owners. The owner authenticates the user before granting permission to access the data. The computation which is done on the owner part is negligible because it is a onetime process. Hence our model drastically minimizes the computational cost and the communication cost. The rest of the paper is organized as follows: Section 2 discusses the related work. Section 3 presents the proposed model. In Section 4 we will discuss the merit of our approach and in Section 5 we will conclude the paper.

2 Related Work

Identity Based Authentication for cloud computing [1] proposes an identity based authentication model for cloud computing (IBHMCC) and present a new identity based authentication protocol for cloud that provides authentication for the user by making use of encryption and signature schemes. Identity based Cryptography for Cloud Security the same applies to the work of Crampton et al. [3] who examine how identity-based cryptography can be used to secure Web services in general. Lishan Kang et al [6] improved the authentication method by placing the trust authentication component on the cloud end. This system will also produce hindrance if the service provider is not trusted. Qin Liu et al [7] provide secure means of sharing the data in cloud by making use of the hierarchical identity based encryption algorithm. Using their scheme, user needs to encrypt the file only once and store the corresponding cipher text in a cloud. Our paper overcomes the drawbacks that are present in the prior research work and provide security with minimal computation cost.

3 Proposed Model and Its Approach

3.1 Cloud Request Dispatch Manager

The sole purpose of this approach is to provide authentication to the authorized user, so that malicious user do not gain access to the sensitive information stored in the cloud. Our proposed model is composed of three entities: cloud service provider (CSP), the owner and the user as shown in

```
  RScmp = (RS1 + RS2 … + RSn)/Sn

  RScmp          →Resource computing from global
source

  {RS1, RS2, RSn} →  Retrieved  Resources  after
computing
```

Fig. 1. The user is the one who request authentication for accessing data in the cloud. The cloud service provider stores all the information which the user is trying to get access to. The owner processes the request to provide authentication for the data the user wants to gain access. Our model comprises of two phases:

1. Authenticated Key Exchange Phase.
2. Key Verification Phase

Preliminaries

C_{pu} – Cloud Service Provider's Public Key.
C_{pr} – Cloud Service Provider's Private Key.
U_{ID} – User Unique Identity.
O_{ID} – Owner Unique Identity.
H_{1} – User's Hash Function.

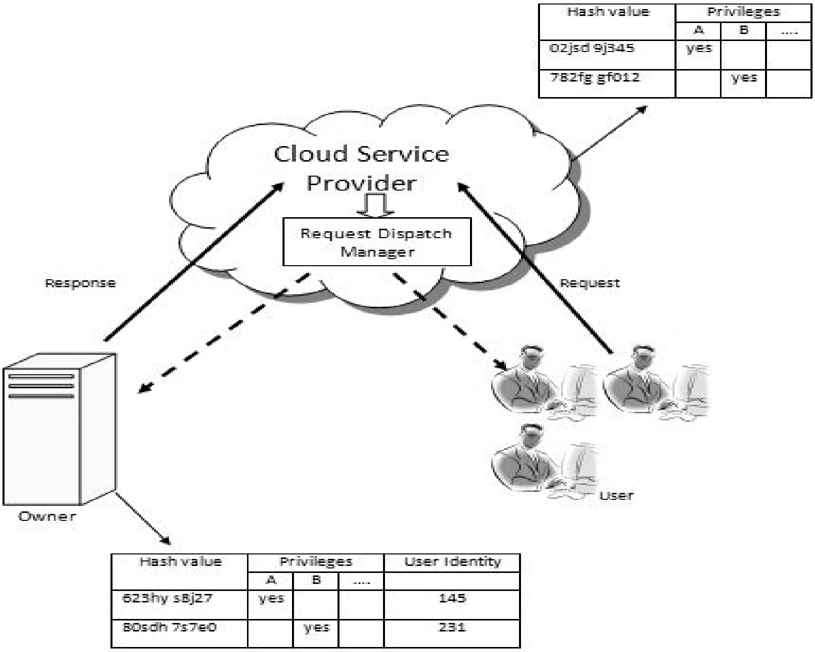
H_{2} – Owner's Hash Function.
$A_{Key.ID}$ – Access Key of the User.
P_{ID} – Access Privilege of the User.
T_{ID} – Identity Table maintained by the Owner.
T_A – Access Table maintained by the User.

KGA – Key Generation Algorithm.
S_k – Symmetric Key Encryption.
D_K – Decryption Key of the user.
Req – Requesting permission to access the data.

3.2 Authenticated Key Exchange Phase

3.2.1 Key Request Phase

In this phase, the user sends request the CSP to access the data stored in it. The request message consist of the hashed unique identity $H_1(U_{ID})$ and the corresponding owner identity $O_{ID.}$ The U_{ID} and the O_{ID} are the key entities which uniquely identifies each user and the corresponding owners in the cloud. The H_1 (U_{ID}) and O_{ID} are encrypted by using the public key of the accessing CSP. The CSP receives the encrypted request message and decrypts it using the private key. The CSP checks the owner's identity O_{ID} and instructs the request dispatch manager to send the hashed unique identity H_1 (U_{ID}) of the user to the corresponding owner.



Hash value	Privileges		
	A	B	....
02jsd 9j345	yes		
782fg gf012		yes	

Hash value	Privileges			User Identity
	A	B	....	
623hy s8j27	yes			145
80sdh 7s7e0		yes		231

Fig. 1. Authenticated Key Exchange Protocol for Cloud

3.3 Request Message

```
  Cpub((H1(UID),OID))                  (1)  encrypted
message

  Cpr(Cpub((H1(UID),OID)))             (2)

  H1 (UID), OID)                       (3)  decrypted
message

  OID=H1 (UID)                         (4)
```

3.4 Key Generation Phase

The owner after receiving the hashed value maps this value with the identity table which consists of all the hashed value of its users and their corresponding access privileges. If the hashed value of the user matches with any of the hashed value present in the table, then the owner runs a key generation algorithm to generate the access key for the user $A_{Key.ID.}$ The owner performs two operations with the generated access key. First, the owner encrypts the access key $A_{Key.ID}$ along with the privilege assigned to the user using the symmetric encryption algorithm AES and then performs a hash operation on the access key H_2 ($A_{Key.ID}$). The owner encrypts the hashed access key H_2 ($A_{Key.ID}$) and the privilege assigned to that user P_{ID} by using the CSP's public key.

3.5 Generation of Access Key

```
If H (TID) == H1 (UID)              (5)

Then Identify ID

AKey.ID=KGA                         (6)
```

3.6 Key Acknowledgement Phase

The owner then sends the encrypted message C_{Pub} (H_2 ($A_{Key.ID}$), P_{ID}) and the S_K ($A_{key.ID}$, P_{ID}) to the CSP. The CSP decrypts the message using its private key and stores the hash value and the corresponding privilege of the user on the access table. The CSP sends the encrypted access key and the user access privileges to the requested user by the request dispatch manager. The user U_{ID} is used for decrypting the encrypted message send by the request dispatch manager. Acknowledgement:

```
Cpu(H2(AKey.ID), PID, Sk(AKey.ID, PID))   (10)

Cpr(H2(AKey.ID), PID, Sk(AKey.ID, PID))   (11)

H2 (AKey.ID), PID, Sk(AKey.ID, PID)       (12)
```

User

$\boldsymbol{D_K (S_K(A_{Key.ID}, P_{ID}))}$

$A_{Key.ID}$, P_{ID} – is user access key and access privileges.

3.7 Key Verification Phase

This is the final phase, the user receives the access key and the privilege required to access the data in the cloud. The key exchange operation performed by the owner comes to a completion, once all the authorized users are provided with their respective access key and access privileges. The users who were all declared as authenticated can directly access the cloud by using their access keys and access privileges. The user receives the corresponding access key and privileges through which the access to the data is granted by the service provider. The message received from the CSP is decrypted by making use of the unique identity of the user. In order to access the data the user sends a request message to the service provider consisting of the hashed value of the access key and the access privileges. The CSP maintains an access table consisting of the hashed value of the access key of the users and their corresponding access privileges pair. The access privileges are used to efficiently map the users with their corresponding hashed value. The CSP receives the access request from the user to decrypt the message using its private key C_{pr} (H_1 ($A_{Key.ID}$), PID) and maps the received hashed access key and access privileges with the access table. If a match is found then the CSP start to process the request by retrieving the data from the cloud and sends the data back to the user.

```
Verification
Cpu(Req(H1(AKey.ID),PID))              (7)

Cpr(Req(H1(AKey.ID),PID))              (8)

H1(AKey.ID) ==H2(AKey.ID)              (9)

and PID == P'ID then Access Granted.
```

Consider an example where Bob is the user who is eager to access the information in cloud. For Bob to gain access, he must be authenticated has an authorized user. Bob request the owner say Alice for an access key via CSP. Suppose Bob's secret identity is U_{Bob}, he request the CSP by sending a request message $C_{pu}(H_1(U_{Bob}),O_{Alice})$. The CSP decrypts the message using C_{pr} and instruct the request dispatch manager to redirect the request to Alice. Alice maintains an identity table T_{ID} which consist of H_2 (U_1… U_n), U_1…U_n and P_1,……,P_n. Alice maps $H_1(U_{Bob})$ with T_{ID}. If match is found, Alice generates a random access key A_{Bob}. Alice runs the symmetric encryption algorithm for A_{Bob} and P_{Bob} using U_{Bob}. Alice sends an acknowledgement message to the CSP which is given as$C_{pu}(H_2(A_{Bob}),P_{Bob1},S_K(A_{Bob},P_{Bob}))$.The CSP decrypts the message and stores $H_2(A_{Bob})$,P_{Bob} in its access table T_A, then $S_K(A_{Bob},P_{Bob})$ is send to Bob. Bob decrypts it using U_{Bob} and obtains the access key A_{Bob} and P_{Bob}. Now Bob is proved as an authorized user. The CSP provides access to Bob only if $H_2(A_{Bob})$= $H_1(A_{Bob})$ and P_{Bob1}= P_{Bob}. As this methodology uses symmetric key encryption, the computation time reduces to a considerable level. The user uses hashed user identity which makes it more difficult to decrypt the user identity which adds to an additional level of security. Even though the owner participates in authenticating the user by providing access key which increases the computation time, this is a onetime process which is negligible.

4 Advantages of the Proposed Approach

This paper proposes an approach which is an alternate to the trusted third party's that are used for key generation for the authentication of the user and it reduces the risks that are associated with the use of trusted third parties. The main advantage of the proposed approach is:

1. Authentication without disclosing unencrypted data. This prevents unnecessary data disclosures.
2. Protection of identity data from untrusted hosts. If the data reach an unintended destination, the attacker cannot retrace the identity of the user because the hashing function which is used is one way.
3. Access privileges are used so that CSP can efficiently map the authorized user. So that access is granted with minimum time period.

Table 1. AKExP- Authenticated Key Exchange Protocol encryption and decryption stages

Request Message Phase-RMP		Key Generation Phase-KGP		Key Acknowledge Phase-KAP		Key Verification Phase-KVP	
RScmp = (RS1 + RS2 … + RSn)/Sn							
Encrypted	Decrypted	Encrypted	Decrypted	Encrypted	Decrypted	Encrypted	Decrypted
Cpub((H1(UID),OID))	Cpu(Cpub((H1(UID), OID)))	OID=H1 (UID)	AKey.ID= KGA	Cpur(H2(AKey.ID), PID, Sk(AKey.ID, PID))	Cpu(H2(AKey.ID), PID, Sk(AKey.ID, PID))	Cpur(Req(H1(AKey.ID), PID))	H1(AKey.ID) ==H2(AKey.ID)
H1 (UID), OID)	H(ODI)	H (TID) == H1 (UID)	Akey= AKGP(ID, H1, Hn)	D_K (S_K($A_{Key.ID}$, P_{ID}))	H2 (AKey.ID), PID, Sk(AKey.ID, PID)	Cpur(Req(H1(AKey.ID) ,PID))	PID == P'ID

A model of encryption and decryption format acquire in normal key exchange while accessing cloud information are described in Table.1 based on AKExP-Authenticated Key Exchange Protocol encryption and decryption technique and key exchanges as per cloud users and requestors input request phenomenon.

5 Conclusion

With the immense growth in the popularity of cloud computing, the privacy and security issues are the major concern for both the public and private sectors. In this paper, we propose an Authenticated Key Exchange Protocol (AKExP) which provides authenticated access to authorized user. The computation cost and the communication cost are drastically reduced in our proposed model. Our future work will be extending the proposed protocol for the hybrid cloud.

References

1. Li, H., Dai, Y., Tian, L., Yang, H.: Identity-Based Authentication for Cloud Computing. In: Jaatun, M.G., Zhao, G., Rong, C. (eds.) Cloud Computing. LNCS, vol. 5931, pp. 157–166. Springer, Heidelberg (2009)
2. Identiy Based Cryptography for Cloud Security, `http://eprint.iacr.org/11/169`
3. Crampton, J., Lim, H.W., Paterson, K.G.: What can Identity-Based Cryptography offer to Web Services? In: SWS 2007: Proc. of the 2007 ACM Workshop on Secure Web Services, pp. 26–36. ACM (2007)
4. Amazon Elastic Compute Cloud, `http://aws.amazon.com/ec2` (access on October 2009)
5. IBM Blue Cloud project, `http://www-03.ibm.com/press/us/en/pressrelease/22613.wss/` (access on October 2009)
6. Abawajy, J.: Determining Service Trustworthiness in InterCloud Computing Environments. In: 10th International Symposium on Pervasive Systems, Algorithms and Networks (I-SPAN 2009), pp. 784–788 (2009)
7. Kang, L., Zhang, X.: Identity-Based Authentication in Cloud Storage Sharing. In: IEEE International Conference on Multimedia Information Networking and Security, pp. 851–855 (2010)
8. Nanda Kishore, M.S., Jayakumar, S.K.V., Satya Reddy, G., Dhavachelvan, P., Chandramohan, D., Soumya Reddy, N.P.: Web Service Suitability Assessment for Cloud Computing. In: Wyld, D.C., Wozniak, M., Chaki, N., Meghanathan, N., Nagamalai, D. (eds.) NeCoM 2011, WeST 2011, WiMoN 2011. CCIS, vol. 197, pp. 622–632. Springer, Heidelberg (2011)

Topology Construction of 3D Wireless Sensor Network

Sarbani Roy and Nandini Mukherjee

Department of Computer Science and Engineering
Jadavpur University, Kolkata-32
India
{sarbani.roy,nmukherjee}@cse.jdvu.ac.in

Abstract. The design of 2D wireless sensor network (WSN) has been mostly considered in recent research approaches, while the case of 3D sensor network has not been so much explored. In reality, most wireless sensor networks operate in three-dimensions. Moreover, the 3D scenarios of WSN represent more accurately the network design for real world applications. One of the main design challenges in WSNs is energy efficiency to prolong the network operable lifetime. An efficient planning of WSN can control the energy consumption of the whole network. In this paper, by using computational geometry theoretic, we propose a framework for dynamic topology construction of 3D WSN for a given 3D space monitoring application.

Keywords: 3D WSN, Voronoi diagram, Delaunay triangulation, coverage, node scheduling.

1 Introduction

Generally, it is assumed that all nodes of a network reside on a plane in 2D WSN design of terrestrial networks. This assumption is not always valid if a network is deployed in space or ocean, where nodes of a network are distributed over a 3D space. For example, 2D design of underwater sensor network is not appropriate it requires 3D design. Ocean column monitoring requires the nodes to be placed at different depths of the water, thus creating a three-dimensional network. The design of WSNs for application like environment monitoring also requires 3D design rather than 2D design, where WSNs deployed on the trees of different heights in a forest. The 3D WSN design is most suitable for monitoring cave as compared to 2D WSN design. Similarly, 3D WSN design is the obvious choice for monitoring railway tunnels, underground tunnels in the mine. Monitoring a large area with WSNs requires a very large number of sensor nodes which entails more energy consumption. One of the main design challenges in the WSNs is energy efficiency to prolong the network operable lifetime.

Deployment of sensor nodes is one of the major issues in WSN environment. A sensor network deployment can usually be categorized as either a dense deployment or a sparse deployment [1]. A region of interest must be covered with

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 533–542.
springerlink.com

the deployment of sensor nodes. Thus, while planning a wireless sensor network, coverage and connectivity are two important issues which need to be dealt with. The coverage problem in WSN can be classified into three classes: area coverage, point coverage and barrier coverage [2][3]. Deployment and configuration of sensor networks to ensure the desired level of connectivity and coverage is fundamentally more challenging in 3D rather than 2D [4]. Energy conservation and network performance are two other issues and probably the most critical issues in wireless sensor networks. To increase the lifetime of the network, one solution that has been proposed in the literature [5] is to switch set of nodes between active and inactive states.

This paper intends to deal with the above mentioned issues while focusing on design and planning of 3D WSNs. The specific applications which are chosen for this work are monitoring applications for 3D space e.g., forest, cave, underground mine tunnel, railway tunnel etc. The paper presents an algorithmic framework to plan an energy efficient 3D WSN for the monitoring applications. The methodology proposed within the framework centers around dynamic activation of a subset of sensor nodes in a densely deployed environment, thereby reducing the initial topology and helping to conserve energy for the entire WSN. The entire framework is built up on the basis of computational geometry algorithms.

The remainder of this paper is organized as follows. Section 2 reviews related work. The proposed framework and methodology for planning of wireless sensor networks are discussed in Section 3. In Section 4, we evaluate our proposal. Concluding remarks and future research directions are given in Section 5.

2 Related Work

Several research works for minimizing energy consumption and prolonging the network lifetime have been proposed in the literature. In this section we mention some of the prominent proposals reported in the literature and discuss their properties in relation to our work.

Ravelomanana [6] investigated several basic characteristics of randomly deployed WSNs regarding sensing and transmission ranges for connectivity and coverage in 3D WSNs. In [7], Huang et al. addressed the coverage problem in 3D WSNs by reducing the geometric problem from a 3D space to a 2D space. Xing et al. [8] discussed the dependencies between coverage and connectivity on the basis of the relationships between the radius of the communication range and the radius of the sensing range of the sensors. Alam and Haas proposed a solution for the coverage and connectivity problem in a 3D underwater sensor network in [9]. In [10], Chen et al. proposed a probability based K-coverage approach for WSNs in the monitoring of 3D space. In [11], Ammari et al. proposed an integrated concentric sphere model to address coverage and connectivity in 3D WSNs using the concepts of continuum percolation.

In [4], Poduri et al. emphasize some of the challenges in designing algorithms for 3D and discussed possible extensions of existing 2D designs for the deployment and configuration to 3D design. Andersen et al. [12] proposed a technique to optimize sensor deployment in the presence of constraints such as restrictions on sensor locations and non-uniform sensing regions for the 3D WSNs.

A sponsored area algorithm is proposed in [5], which aims at providing complete coverage by its off-duty eligibility rules. In [13], the authors concentrate on the issue defined as *density control* which arises in high density network. Zhang et al also prove a fundamental result, i.e., if the transmission range is at least twice the sensing range, a complete coverage of a convex area implies connectivity of the working nodes [13].

In our present work, the objective is to propose an algorithmic framework for designing and planning of an energy efficient WSN for a given 3D space monitoring application. Tian et al. in [5] pointed out that nodes can have different sensing ranges due to initial set up or due to effect of changes during their lifetime. In [14], the problem of selecting a minimum energy-cost connected sensor cover, when each sensor node can vary its sensing and transmission radius is discussed. Although, our work has been inspired from these research works, unlike the previous works, in our proposed methodology, each sensor node is deployed in 3D space and can dynamically change its sensing range and transmission range. Computational geometry based techniques are used to compute sensing and transmission ranges of sensor nodes. Initially sensor nodes are randomly deployed in the 3D space. Thus, the deployment is dense in nature. Sensing and transmission ranges are adjusted dynamically and then a subset of sensors is selected to be active so that the given region is covered and the selected set of sensors form a connected communication graph. The initial topology is thus reduced to control the energy consumption of the whole network subject to maintaining the network performance. This is in contrast with the works of other researchers [15] where each sensor node is statically assumed to be either in active (powered on) or in inactive state.

Thus, starting from an initial topology, we propose to reduce the topology by separating the nodes into an active and an inactive sets with a goal that the region of interest is fully covered while keeping the energy consumption as low as possible. The energy costs due to sensing and transmission activities of all the active sensor nodes in the network are considered here. The proposed framework also identifies a region within the area of interest, where more sensors need to be deployed. To deal with above mentioned issues, computational geometry techniques are applied. Thus, in this paper, we will focus on the sensing and transmission ranges of sensor nodes when discussing deployment issues in 3D WSN.

3 Proposed Framework

Let us consider a set of sensors, $S = \{s_1, s_2, ..., s_n\}$ which are randomly distributed in a 3-dimensional space to monitor a desired area. The wireless sensor network is modeled as a graph $G(S, E)$, where S represents a set of sensor nodes (vertices) and E is a set of links (edges) between sensor nodes. The sensor density should be high enough so that any point can be monitored by at least one sensor. The sensing and transmission areas of the sensors are modeled as sphere and radii of such sphere are different (non unit sphere) for different sensors. It has been already mentioned that the objective of the framework proposed in this paper is to dynamically control the sensing range and transmission range in order to reduce the energy consumption of the entire network. The sensing range of a sensor s_i is modeled by a sphere of radius sr_i. The transmission range of sensor s_i is modeled by a sphere of radius tr_i. The transmission range can be adjusted by changing the transmission power of the radio transceiver. The proposed framework consists of two main processes namely topology construction and topology control.

Topology construction process aims at building a reduced topology in order to save energy while preserving network connectivity and area coverage. Once the initial reduced topology is created, the network starts performing the tasks for which it has been designed. In the proposed approach, each sensor node autonomously and periodically can change its sensing range and transmission range. In such a dynamic environment Voronoi diagram [16] is used to determine the sensing range of sensor nodes and the concept of Delaunay triangulation [16] is used to determine the transmission range of sensor nodes. K-coverage algorithm is used to identify extra nodes in the initial random deployment. The initial topology is reduced and the selected sensor nodes are scheduled according to the proposed node scheduling algorithm. Appropriate routing algorithm needs to be designed in the next phase.

Topology control process, shares most, if not all, of the above steps that are used for topology construction. In the topology control phase two important activities take place - (1) Monitoring and (2) Maintenance. Monitoring energy consumption of sensor networks is of great significance to prolong the network lifetime and to maintain the network connectivity. The proposed framework mainly considers three metrics for monitoring purpose: (1) Energy (2) Link quality (3) Usage. Both available and consumed energy of sensor nodes need to be monitored. Lifetime of a sensor node may be measured using these data. The LQI (Link Quality Indicator) measurement is a characterization of the strength or quality of a received packet by a sensor node. RSSI (Received Signal Strength Indicator) measures the strength (power) of the signal for the packet. Topology can be maintained in reactive or proactive manner.

This paper mainly deals with the topology construction process. Therefore, topology control has been kept out of its scope. Nevertheless, within this framework, topology control process reconstructs the WSN topology with the help of similar algorithms that are used in the topology construction process and thereby optimizes the performance of the network. Following subsections describe each of the phases of topology construction process of WSN planning in detail.

3.1 Determining Sensing Range

As we have already discussed, the 3D Voronoi diagram of S i.e., $V(S)$ decomposes the space into regions around each point $s_i \in S$, such that all the points in the region around s_i are closer to s_i than any other point in S. Thus one of the vertices of the convex polyhedron forming the Voronoi cell is the farthest point from the point (sensor node) inside it. Therefore 3D Voronoi diagram can be used in determining WSN coverage in 3D networks. We assume that srm_i is the maximum possible sensing range of the sensor. In this algorithm srv_i is computed which is the distance between s_i and the farthest point fs_i in its region i.e., Voronoi cell $\nu(s_i)$. So, the sensing range of s_i could not be less than srv_i, as s_i is responsible for monitoring any event occurrence in its region $(\nu(s_i))$. Here, srv_i and srm_i are two extreme points, where srv_i indicates the minimum possible range and srm_i indicates the maximum possible range. Each sensor node s_i is set with a sensing range sr_i, which is the average of these two extreme values.

3.2 Determining Transmission Range

In [13], connectivity is trivialized by assuming that the transmission range is at least twice of the sensing range. We focus on a more generic connectivity condition that can be used even when the transmission range is less than twice the sensing range. In this algorithm, 3D Delaunay triangulation of S i.e., $D(S)$ is computed. Each edge $e_{ij} \in D(S)$ is assigned a weight, which is the Euclidian distance between s_i and s_j. The 3D Delaunay triangulation, $D(S)$ is then converted to a weighted 3D Delaunay triangulation. The transmission range of sensor s_i denoted by tr_i is then defined as the maximum value of all the weights those have been assigned to the edges incident to s_i. Here, we assume that trm_i is the maximum possible transmission range of the sensor node s_i and $tr_i < trm_i$, as large number of sensor nodes are deployed in the field.

3.3 Coverage

Coverage problem in sensor networks is one of the fundamental issues, which reveals how well a defined area is monitored by sensors. Since sensors are randomly distributed, the sensing regions of some sensors overlap and holes may also be created in some areas. The sensing ranges of sensors can be unit spheres or non-unit spheres. In this paper, the sensing and transmission range of sensor nodes are modeled by non unit spheres as discussed earlier. As output of the previous steps, each sensor $s_i, i = 1, ..., n$, has a sensing range of sr_i, i.e., it can monitor any event that occurs within a sphere of radius sr_i and centered at s_i. The area monitored by sensor node s_i is denoted by A_i and is said to be *k-covered* if it is within the sensing ranges of at least k sensors. According to our assumption, the WSN may be overly covered by too many sensors in certain areas. In such cases, the area A_i monitored by s_i can be entirely covered by its k immediate neighbor nodes. The objective of this step is to find the value of k for each sensor node.

3.4 Node Scheduling

As suggested in [5], if there are more sensors than necessary, we may turn off some redundant nodes to save energy. These sensors may be turned on later when other sensors run out of energy. Tian and Georganas [5] proposes a node-scheduling scheme to guarantee that the level of coverage of the network area after turning off some redundant sensors remains the same. This paper proposes a node scheduling algorithm on the basis of variable sensing and transmission range as discussed earlier. Let us define the *active set* of sensor nodes AS ($AS \subseteq S$) which contains the sensor nodes that are turned on. Similarly, *inactive set* of sensor nodes contains nodes which are turned off ($IS \subseteq S$). After the initial node scheduling the WSN is reduced and this reduced WSN should be revisited to verify the connectivity. Connectivity can be assured by controlling the transmission range of the sensor nodes from the active set and if required selecting some nodes from the inactive set of sensor nodes and making them partially active for transmission purpose only. Sensor nodes in *partially active set* ($PAS \subseteq S$) are used for transmission purpose only. These sensor nodes are not used for sensing activity. The transmission range of each sensor node from the active set is revised. If the revised transmission range of any sensor node s_i exceeds its maximum possible value (trm_i), then a sensor node $s_k \in IS$ is selected as a partially active sensor node (for transmission purpose only).

4 Result

The topology construction phase has been tested in a simulation environment in order to demonstrate its functioning and its effectiveness. The simulation environment is generated by using MATLAB version 7.11.0. Topology construction process of WSN for a given application and the performance evaluation of the framework are presented in the following subsections.

4.1 Topology Construction

The demonstration of the design and planning of WSN using our proposed framework for an application is presented in this section. Here, we simulate the environment of a sensor network with 30 sensor nodes. Consider our deployment area as a $10m$ cubic area. In our model, we initially deploy sensors randomly in the area to be monitored.

Voronoi diagram is used for decomposing the area into 30 polyhedras i.e., Voronoi cells, such that each cell corresponds to one of the sites (sensor nodes).

Transmission range of a sensor node is calculated from the weighted $3D$ Delaunay triangulation. Here, weight indicates the distance between the sensors. Maximum weight among the edges incident to a sensor node is assigned as the required transmission range of that particular sensor node. For example, in the current scenario, sensor node 5 is the farthest node from the sensor node 29, while sensor node 17 is the farthest from the sensor node 5. So, the transmission range of the sensor node 29 is the distance between sensor nodes 29 and 5, while the transmission range of the sensor node 5 is the distance between sensor nodes 5 and 17. Transmission range of each sensor node is depicted by dotted lines in the Figure 1 (a). Figure 1(b) shows the topology with active sensors.

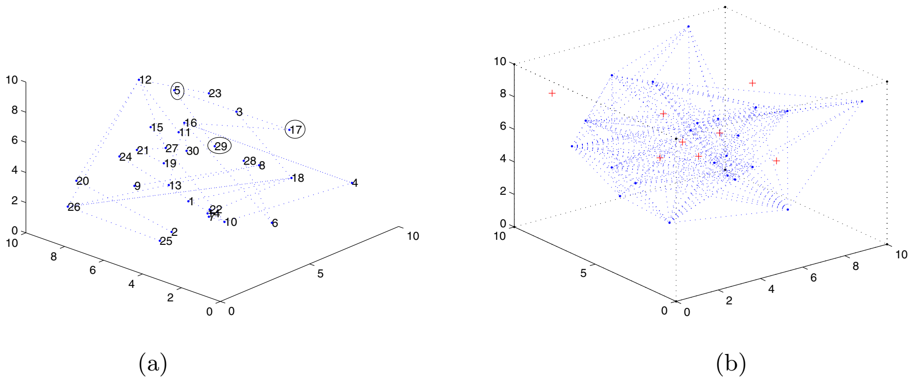


Fig. 1. (a) Transmission range using 3D Delaunay Trainagulation (b) Topology with active sensor nodes

4.2 Performance Evaluation

In this section, we examine the performance of the proposed solution through extensive simulations. The performance of our proposed framework in terms of energy conservation was investigated. We consider an energy consumption model for analysis of sensor networks as proposed in [17]. We have carried out a set of experiments with various topologies. The key metric for evaluating the proposed algorithms was the energy consumption used for data transmission. Simulations are conducted with different parameter settings as described earlier to evaluate the proposed framework. The sensor networks with 10 to 50 nodes simulating the sensor nodes with randomly generated positions in a $10m$ to $50m$ cubic region were considered in our experiments.

The performance of the proposed algorithm depends on various parameters: the number of sensor nodes in a target area and the transmission range of sensor nodes. In the simulation, we consider three ways to set transmission range of sensor nodes.

1. As in [13], we set transmission range as twice of sensing range.
2. Transmission range is set to maximum possible range.
3. Transmission range is set using our proposed algorithm.

Performances of the following two cases are compared:

case a: All sensor nodes are active in the WSN.
case b: Only some sensor nodes are activated using our algorithm.

Energy consumption in both cases with transmission range according to our proposed algorithm is depicted in Figure 2. These results clearly illustrate that the sensor network with active sensor nodes consume much less energy than the sensor network with all nodes in the network. Figure 3 shows the energy

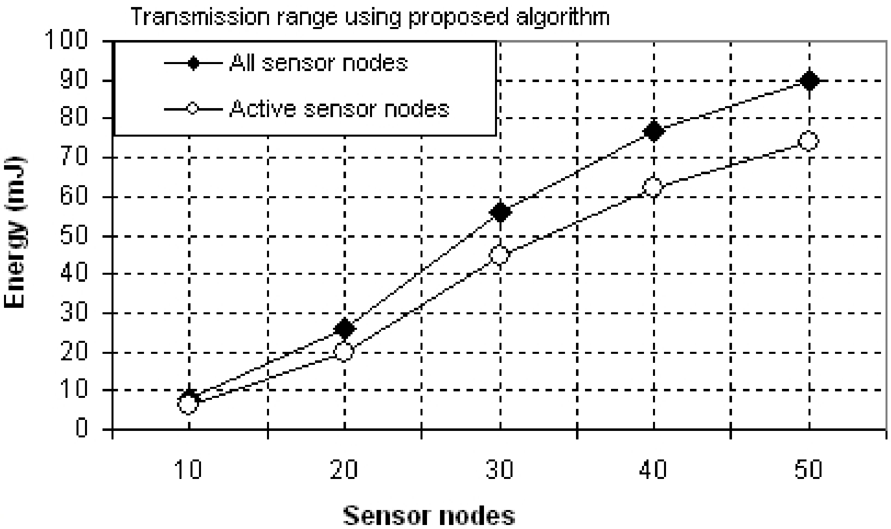


Fig. 2. Comparison of energy consumption in case (a) and case (b) with transmission range using proposed algorithm

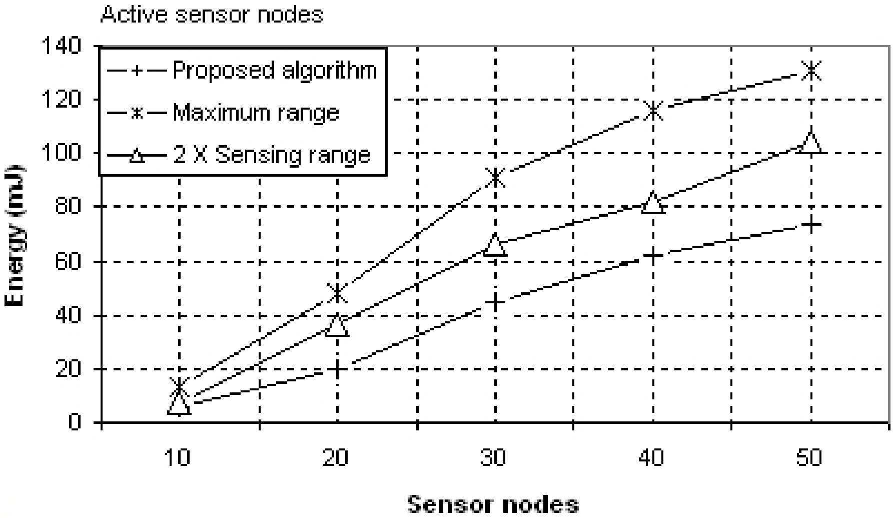


Fig. 3. Energy consumption with active sensor nodes

consumption in sensor network with active sensor nodes. Generally, energy consumption depends on the power settings used for transmission, reception and idle listening (partially active mode). In the simulation, we assume that the energy consumed during inactive mode is negligible. According to our framework, sensing ranges and transmission ranges of sensor nodes are dynamically restructured with changing topology.

Simulation results indicate that transmission ranges have great impact on energy consumption and network lifetime. As a consequence, network lifetime decreases with an increase of transmission range. These results clearly illustrate that the sensor network with active sensor nodes (case b) consume much less energy than the sensor network with all nodes (case a) in the network. The amount of energy saved in case b is due to significant number of inactive nodes. Note that the energy savings achieved by the proposed algorithm is always higher than those in the other situations. From the simulation results we have observed that how well the proposed algorithm reduces energy consumption.

5 Conclusion

In this paper, a new algorithmic framework is proposed to address the need for an energy efficient planning mechanism for 3D wireless sensor network for an environmental monitoring application. Topology construction and topology control are two main processes in the framework. This paper mainly focuses on the topology construction process. The proposed framework utilizes the properties of 3D Voronoi diagram and 3D Delaunay triangulation to limit the sensing range and transmission range of sensor nodes. Robustness of the algorithm has been demonstrated by simulation. Future work will aim at developing efficient topology monitoring and maintenance schemes as a part of the topology control process of the proposed framework. There are still many areas to be explored within this research topic. This initial set of experiments serves to demonstrate the fruitfulness of the concept. Future work also includes extending the framework to different application scenarios and investigating the impact of different network parameters and performance metrics in the design of 3D WSNs.

References

1. Mulligan, R., Ammari, H.M.: Coverage in Wireless Sensor Networks: A Survey. Journal of Network Protocols and Algorithms 2(2), 27–53 (2010)
2. Commuri, S., Watfa, M.K.: Coverage Strategies in Wireless Sensor Networks. International Journal of Distributed Sensor Networks 2(4), 333–353 (2006)
3. Cardei, M., Wu, J.: Coverage in Wireless Sensor Networks. In: Ilyas, M., Mahgoub, I. (eds.) Handbook of Sensor Networks. CRC Press, Boca Raton (2004)
4. Poduri, S., Pattem, S., Krishnamachari, B., Sukhatmeet, G.S.: Sensor network configuration and the curse of dimensionality. In: Proceedings of The Third IEEE Workshop on Embedded Networked Sensors (EmNets), May 30-31, Harward University (2006)
5. Tian, D., Georganas, N.D.: A Coverage-preserving Node Scheduling Scheme for Large Wireless Sensor Networks. In: Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications (WSNA 2002), pp. 32–41. ACM press, New York (2002)

6. Ravelomanana, V.: Extremal Properties of Three-Dimensional Sensor Networks with Applications. IEEE Transactions on Mobile Computing 3(3), 246–257 (2004)
7. Huang, C., Tseng, Y., Lo, L.: The Coverage Problem in Three- Dimensional Wireless Sensor Networks. In: Proceedings of Global Telecommunications Conference (GLOBECOM 2004), November 29-December 3, pp. 3182–3186. IEEE (2004)
8. Xing, G., Wang, X., Zhang, Y., Lu, C., Pless, R., Gill, C.: Integrated Coverage and Connectivity Configuration for Energy Conservation in Sensor Networks. ACM Transactions on Sensor Networks (TOSN) 1(1), 36–72 (2005)
9. Alam, S.M.N., Haas, Z.J.: Coverage and connectivity in three-dimensional networks. In: Proceedings of 12th Annual International Conference on Mobile Computing and Networking (MobiCom 2006), September 23-29, pp. 346–357. ACM, New York (2006)
10. Chen, F., Jiang, P., Xue, A.: Probability-Based Coverage Algorithm for 3D Wireless Sensor Networks. In: Proceedings of 4th International Conference on Intelligent Computing (ICIC), Shanghai, China, September 15-18, pp. 364–371 (2008)
11. Ammari, H.M., Das, S.K.: Critical Density for Coverage and Connectivity in Three-Dimensional Wireless Sensor Networks Using Continuum Percolation. IEEE Transactions on Parallel Distributed Systems 20(6), 872–885 (2009)
12. Andersen, T., Tirthapura, S.: Wireless sensor deployment for 3D coverage with constraints. In: Proceedings of 6th International Conference on Networked Sensing Systems (INSS 2009), pp. 78–81. IEEE Press, Piscataway (2009)
13. Zhang, H., Hou, J.C.: Maintaining Sensing Coverage and Connectivity in Large Sensor Networks. In: NSF International Workshop on Theoretical and Algorithmic Aspects of Sensor, Ad hoc Wireless and Peer-to-Peer Networks, pp. 251–262 (February 2004); also a technical report with reference number UIUCDCS-R-2003-2351 in the department of computer science. University of Illinois at Urbana-Champaign
14. Zhou, Z., Das, S., Gupta, H.: Connected k-coverage Problem in Sensor Networks. In: Proceedings of the 13th International Conference of Computer Communications and Networks (ICCCN), October 11-13, pp. 373–378 (2004)
15. Zhou, Z., Das, S., Gupta, H.: Variable Radii Connected Sensor Cover in Sensor Networks. ACM Transactions on Sensor Networks (TOSN) 5(1), Article 8 (February 2009)
16. de Berg, M., van Kreveld, M., Overmars, M., Schwarzkopf, O.: Computational Geometry: Algorithms and Applications, 3rd edn. Springer, Heidelberg (2008)
17. Heinzelman, W., Chandrakasan, A., Balakrishnan, H.: An Application-Specific Protocol Architecture for Wireless Microsensor Networks. IEEE Transactions on Wireless Communications 1(4), 660–670 (2002)

Design and Analysis of Dual Capacitively Loaded C-PIFA

Kirti Dhwaj, Rachit Garg, Gaurav Mishra, and Neetesh Purohit

Indian Institute of Information Technology Allahabad, India
kdhwaj@gmail.com, rachitgarg91@gmail.com,
mishragaurav27@gmail.com, neetesh15@yahoo.com

Abstract. A dual capacitively loaded planar inverted-F antenna (C-PIFA) is proposed and studied. The capacitive loading of the proposed structure reduces the resonant length from λ/4 to less than λ/6. While inheriting the attractive features of PIFAs, such as easy fabrication and low cost, the proposed design exhibits desired GSM bandwidth with high gain. This antenna could be used in mobile devices requiring less weight and high mechanical strength. The proposed structure can be implemented with very thin metal sheet, which accounts for its less weight.

Index Terms: Finite ground plane, MoM methods, patch antenna, GSM, PIFA.

1 Introduction

The need for smaller antennas with low profiles is increasing day-by-day with the reduction in size of the wireless sensors and communication devices. The room for antenna design has been marginalized. Various designs for PIFA have been discussed which address these demands [1]-[3]. PIFA is an electrically small antenna with resonant length approximately equal to λ/4. The PIFA has an omnidirectional pattern in the plane of the patch. The PIFA are reported to have their resonant wavelengths less than λ/8 [4].

The PIFA has been studied in detail in recent years. A lot of effort has been made in recent past to study the effects of various parameters on performance parameters. Effect of varying width of feed plate and shorting post on bandwidth has been studied [5] It was shown that the height [6], shorting plate width [7], and meandered shorting strip [8] can be used to increase the bandwidth of the Inverted F Antenna. The centrally fed planer inverted F-antenna (CPIFA) is derived from the PIFA. It is fed at the centre-line giving the antenna its name [9].

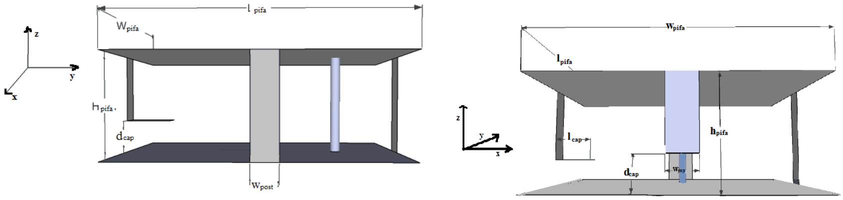
A λ/4 x λ/4 microstrip antenna is formed by placing two shorting walls on the adjacent edges of a square microstrip antenna. The shape produces an antenna with one-fourth the area of a square patch antenna [10]. This provides a good option where volume is restricted. The antenna design proposed in this paper is a λ/4 x λ/4 variant of the C-PIFA. The design reduces the resonant length of the antenna to less than λ/6 preserving the square shape for efficient volume utilization.

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 543–550.
springerlink.com

Section 2 introduces the structure of the proposed antenna. The computational results describing the effects of various parameters are discussed in Section 3. In Section 4, a compact antenna design for DCS communications is presented. The conclusions are presented in Section 5.

2 Structure

The structure of the proposed antenna is shown in Fig. 1. The top plate dimensions of the PIFA are *(l_{pifa}, w_{pifa})* = (45 mm, 45 mm). The height of the PIFA is h_{pifa} = 10 mm. Air is taken to be the substrate. The shorting posts are centered along the edges. The antenna is fed along the centerline at (0, 36mm, 0) giving the antenna it's name centrally fed planer inverted-F antenna. The feeding medium is taken to be coaxial cable with a reference impedance of 50 Ώ. The adjacent edges of the antenna are loaded with capacitances to maintain the symmetry of the structure. The capacitive loads are formed by folding the open ends of the PIFA towards the ground plane and adding plates parallel to ground. The length of the top plates of the capacitors l_{cap} is 2mm. The ground plane is taken to be square with the patch lying on its centre i.e (0, 0, 0).This structure serves as basis for computations in next section.



(a) (b)

Fig. 1. Side views of the capacitively loaded centrally fed planar inverted F antenna. The dimensions of the patch used for simulation purposes in Section 3 are (l_{pifa}, w_{pifa}, h_{pifa}) = (45 mm, 45 mm, 10mm).

Table 1. Design Parameters

Notation	Quantity	Notation	Quantity
w_{post}	width of the shorting post	h_{pifa}	height of the substrate i.e. air
l_{pifa}	length of the rectangular patch	l_{cap}	length of the capacitance
w_{pifa}	width of the rectangular patch	w_{cap}	width of the capacitance
		d_{cap}	depth of the capacitance

3 Computational Results

The performance parameters i.e. impedance bandwidth , gain, field patterns and input impedance are determined using the Method of Moments (software FEKO) [11]. Theoretically, the resonant frequency is 1.66 GHz if the antenna is assumed to be a λ/4 patch.

A. Shorting Post

The ground plane dimensions are taken to be 50mm x 50mm. The capacitor depth and width are (d_{cap}, w_{cap}) = (2mm, 4mm). It can be seen from Fig. 2 that the reduction of the width of the shorting posts reduces the resonant frequency of the antenna. The bandwidth is considerably decreased . The quality of matching is improved. The maximum gain remains almost the same in the XY plane (Fig . 3). The reduction in the frequency is attributed to the increase in the inductance of shorting posts on decreasing its width [9].

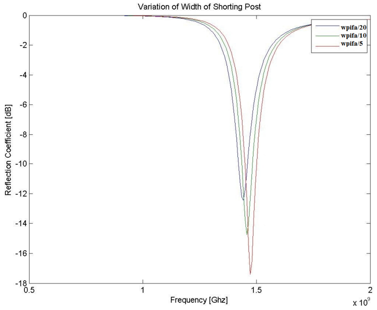


Fig. 2. The reflection coefficient against frequency curves for different values of the shorting post width. w_{post} is varied as a fraction of the width of the PIFA, w_{pifa}.

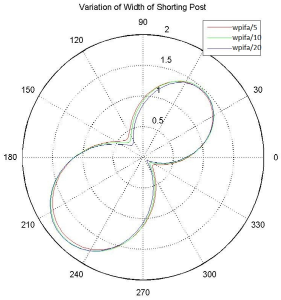


Fig. 3. The far field pattern in XY plane with the variation of the width of the shorting post w_{post}. w_{pifa}/10 signifies the width of the capacitor plate to be one-tenth of the patch width .

B. Capacitive Loads

The inductance introduced by the shorting posts has to be balanced out by an equivalent capacitance. The introduction of load capacitances at the edges opposite to the shorting post further reduces the operating frequency of the PIFA.

Here, we present the effect of varying the load capacitance on antenna characteristics. As the capacitance, depends upon W_{cap} and d_{cap} essentially the variation of the two parameters is discussed in this section. The variations are done for both the capacitances simultaneously.

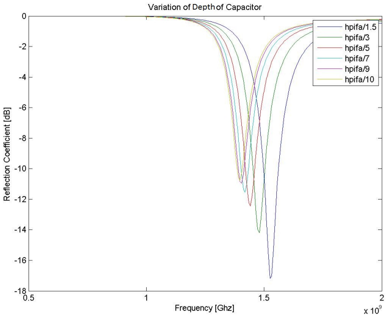


Fig. 4. The reflection coefficient versus frequency curves are plotted by varying the depth of the capacitor. The depth is varied as a function of the height of the substrate i.e. $h_{pifa}/3$ signifies depth of the capacitor, d_{cap} to be one-third of the height of substrate.

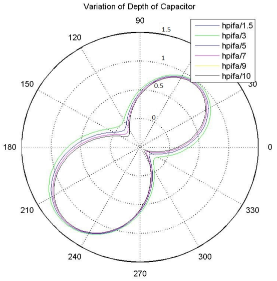


Fig. 5. The far field radiation patterns of the antenna are plotted by varying the depth of the capacitor. The depth is varied as a function of the height of the substrate i.e. $h_{pifa}/3$ signifies depth of the capacitor, d_{cap} to be one-third of the height of substrate.

For variation of d_{cap} , the antenna introduced in Section 2 is used with w_{post} = 9mm and w_{cap} = 4mm. It can be seen from Fig.4 that the reduction in the depth of the capacitors reduces the resonant frequency of the antenna. There is a considerable penalty in bandwidth and matching. Maximum gain is also reduced marginally with reduction in capacitor depths (Fig. 5).

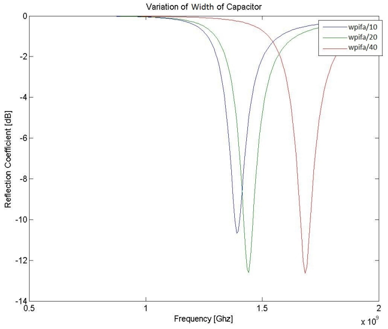


Fig. 6. The reflection coefficient against the frequency for variation of the width of the capacitor w_{pifa}. $w_{pifa}/10$ signifies the width of the capacitor plate to be one-tenth of the patch width

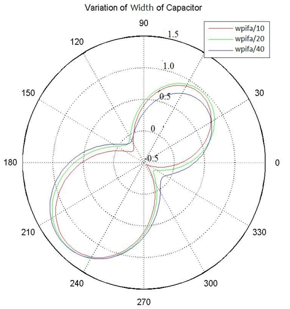


Fig. 7. The far field patterns with the variation of the width of the capacitor w_{cap}. $w_{pifa}/10$ signifies the width of the capacitor plate to be one-tenth of the patch width

Similarly, for variation of w_{cap}, the antenna in section 1 is again used with w_{post} = 9mm and the capacitor depth d_{cap} = 1.11 mm. It can be seen from Fig. 6 that decreasing the width of capacitor results in continuous increase in resonant frequency but reflection coefficient attains a constant value after a threshold width (which can be exploited as optimum width of capacitor).

Correspondingly, the bandwidth is incremented with w_{cap} *(*as a function of w_{pifa} *).* The X-Y plane field patterns for different capacitor widths are shown in Fig. 7. The above results validate (1).

C. Finite Ground Plane

The effects of finite ground plane on a PIFA have been studied extensively [10]-[12]. The antenna used in the simulations present in this section has the structure presented in Section 1. The width of the shorting post is fixed at 9mm and the capacitance parameters are (d_{cap}, w_{cap}*)* = (1.11 mm, 4.5 mm).

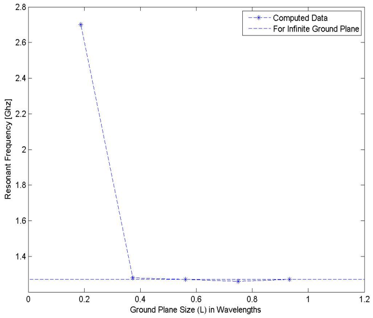


Fig. 8. The plot of resonant frequencies with the variation in size of the ground plane. The ground plane size i.e the length and the width is shown as a factor of the resonant wavelength obtained with the ground plane size taken to be infinite.

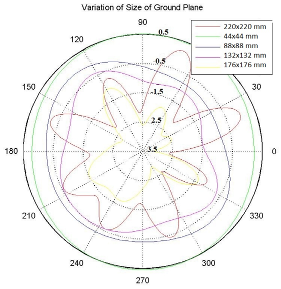


Fig. 9. The far field radiation of the antenna in XY plane with the variation of the size of the ground plane

The variation of operating frequency with the change in size of the ground plane is shown in Fig. 8. It can be seen that for ground plane size more than that of the patch, the resonant frequencies remain almost constant .The corresponding X-Y field patterns are shown in Fig. 9. Similarly the bandwidths are plotted in Fig. 10. The antenna is first simulated with an infinite ground plane. The resonant wavelength comes out to be 23.6 mm. Fig. 8 and Fig. 10 depict the variation of the ground plane size as a factor of the above mentioned wavelength (i.e 23.6 mm) on their X axes. Fig. 8 concludes that there is no fixed pattern of operating frequency variation with the ground size plane until the ground plane size reduces to a certain size, in this case being 0.2λ. To sum up, it can be said that the ground plane size has a significant effect on the radiation pattern of the antenna.

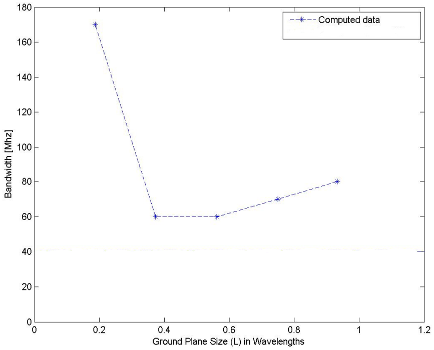
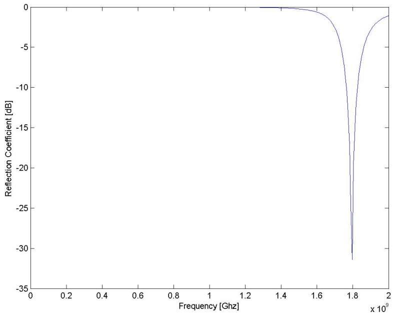


Fig. 10. The plot of bandwidth with the variation in size of the ground plane. The ground plane size (i.e. the length and the width) is shown as a factor of the resonant wavelength obtained with the ground plane size taken to be infinite.

4 Proposed Antenna Design

Using the computational results, we design a minimum volume antenna to function at the 1.8 GHz band .The dimensions of the antenna are scaled to make the operating frequency as 1.795 GHz (Fig. 11).



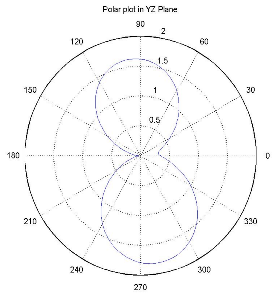


Fig. 11. The reflection coefficient versus frequency curve for the antenna operating in 1.8 GHz band. The figure shows the good quality matching the antenna enjoys with a 50 Ω coaxial cable

Fig. 12. The far field pattern in YZ plane for the proposed antenna.

The bandwidth is 70 Mhz. The specifications are shown in Table 2. The Y-Z, X-Z and X-Y plane radiation patterns are shown in Fig. 12, Fig 13 and Fig. 14 respectively. The efficiency of the capacitively loaded patch antenna is lower than the corresponding PIFA as the current distribution on the top plate will be more uniform and larger in magnitude in a capacitively loaded PIFA [8].

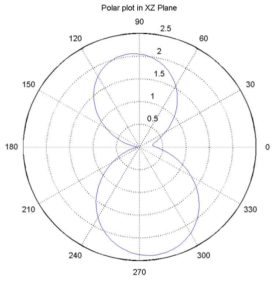


Fig. 13. The far field pattern in XZ plane for the proposed antenna

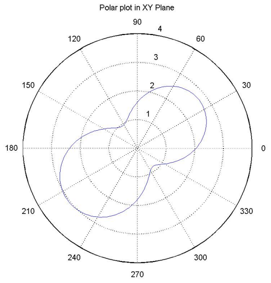


Fig. 14. The far field pattern in XY plane for the proposed antenna

This design approach offers a low volume antenna solution for the GSM 1800 band and can be easily integrated in mobile communication handsets.

Table 2. Design Parameters for 1.8 GHz Antenna

Quantity	Value
Length of the patch	32.5 mm
Width of the patch	32.5 mm
Height of the substrate	10 mm
Feed probe coordinates	(0, 13.5 mm, 0)
Width of the shorting post	9 mm
Length of the capacitance	2 mm
Width of the capacitance	4.5 mm
Depth of the capacitance	1.11 mm
Length of the ground plane	40 mm
Width of the ground plane	40 mm

5 Conclusion

We have shown that the addition of two capacitances in a PIFA significantly reduces its resonant wavelength. The loss of matching can be overcome by changing the size of the ground plane and the position of feeding probe [9]. The effect of the size of shorting post, capacitance values and the ground plane on the antenna characteristics have been presented in section 3.

A design to operate at 1.8 GHz band is presented which can be used where volume is a consideration. Proper scaling of the antenna dimensions can be done to make it operate at other frequencies.

References

[1] Pedersen, G., Andersen, J.: Integrated antennas for hand-held telephones with low absorption. In: IEEE 44th Veh. Technol. Conf., Stockholm, Sweden, vol. 3, pp. 1537–1541 (June 1994)

[2] Taga, T.: Analysis of planar inverted-F antennas and antenna design for portable radio equipment. In: Hirasawa, K., Haneishi, M. (eds.) Analysis, Design and Measurement of Small and Low Profile Antennas, ch. 5. Artech, Boston (1992)

[3] Taga, T., Tsunekawa, K.: Performance analysis of a built-in planar inverted-F antenna for 800 MHz band portable radio units. IEEE J. Select Areas Commun. SAC-35, 921–929 (1987)

[4] Rowell, C.R., Murch, R.D.: A capacitively loaded PIFA for compact mobile telephone handsets. IEEE Transactions on Antennas and Propagation 45(5), 837–842 (1997)

[5] Chattha, H.T., Huang, Y., Lu, Y.: PIFA Bandwidth Enhancement by Changing the Widths of Feed and Shorting Plates. In: IEEE Antennas And Wireless Propagation Letters, vol. 8, p. 637 (2009)

[6] Liu, D., Gaucher, B.: The Inverted-F Antenna Height Effects on Bandwidth. IBM T. J. Watson Research Center. IEEE, Yorktown Heights, NY (2005)

[7] Hall, P.S., Song, C.T.P., Lin, H.H., Chen, H.M., Lin, Y.F., Cheng, P.S.: Parametric study on the characteristics of planar inverted-F antenna. Proc. Microw., Antennas Propag. 152(6), 534–538 (2005)

[8] Chan, P.W., Wong, H., Yung, E.K.N.: Wideband planar inverted-F antenna with meandering shorting

[9] Bancroft, R.: Rectangular Microstrip Antennas. In: Microstrip and Printed Antenna Design, 2nd edn., ch. 2, sec 2.5, pp. 36–38. SciTech Publishing, Raleigh (2009)

[10] Bancroft, R.: Unpublished Witnessed/Notarized Engineering Notebook, October 23 (1998)

[11] Davidson, D.B., Theron, I.P., Jakobus, U., Landstorfer, F.M., Meyer, F.J.C., Mostert, J., van Tonder, J.J.: Recent progress on the antenna simulation program FEKO. In: Proc. South African Symp. on Communications and Signal Processing, COMSIG, pp. 427–430 (1998)

[12] Liang, M.C., Lai, K.L., Lin, S.T.: The effect of finite ground on a rectangular C-patch antenna. In: Antennas and Propagation Society International Symposium, June 22-27, vol. 2, pp. 732–735. IEEE (2003)

[13] Huynh, M.-C., Stutzman, W.: Ground plane effects on planar inverted-F antenna (PIFA) performance. IEE Proceedings of Microwaves, Antennas and Propagation 150(4), 209–213 (2003)

Jong Nang 3-Input NOR Channel[*,**]

Moon Ho Lee[1], Md. Hashem Ali Khan[1], and Daechul Park[2]

[1] Institute of Information and Communication,
Chonbuk National University , Jeonju, 561-756, Korea
[2] Hannam University, DaedeokGu, Daejeon 306-791, Korea
{moonho,hashem05ali}@jbnu.ac.kr, fia4joy@yahoo.co.kr

Abstract. In this paper, we introduce the root of digital human binary coded Jong Nang communications as the wooden gate in Korea Jeju Island custom, has been using after about 1234 years. We compare the digital logic with modern traffic signal codes and analysis of both cases. Shannon channel capacity of the Jong Nang are 3 input NOR and binary erasure multiple access channel. Jong Nang communications is normal 3 rafters placed on two vertical stones with three holes to convey the family's whereabouts that is deterministic signal, nowadays it is applied to backhaul in mobile base station.

Keywords: Human binary coded communications, Jong Nang, NOR channel, Erasure channel, Shannon capacity.

1 Introduction

Thirty of years ago, in western society, F. G. Heath described the development of the binary code from Francis Bacon's "two-letter alphabet" which was conceived at the beginning of the seventeenth century, in "Origins of the Binary Code" [1]. Subsequently, Jacquardt's punch card operated loom (1805) and Boole's logical algebra (1854) led to the introduction of a binary telegraphic alphabet by Baudot (1875). Volker Aschoff, a Germany professor, reported the early history of the binary code in [2]. However, in oriental society, especially in Korea, concerned with 760 years old Jeju Island custom, Jeju *Jong Nang* Code is considered as one of the earliest Human Binary Coded Communication (HBCC) in the world with a definite "1" or "0" system.

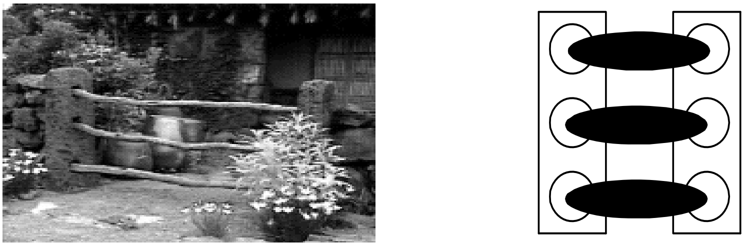
Jong Nang, the wooden gate in Korea Jeju Island dialect, had three wooden rafters placed on Jong-Ju-Mok (two large vertical stones with three holes) to convey the family's whereabouts[*]. A product of the wisdom of Jeju Island people in Korea, the *Jong Nang* was a unique custom of local culture. As there was no gate at the house in Jeju Island, timbers were used to prevent cattle or horses from entering and having the

[*] This work was supported by WCU R32-2012-000-20014-0, FR2010-0020942, MED 2012-002521, NRF, Korea and Hannam University while taking a sabbatical leave in SRM University, India.

[**] Available on line: This work was partial introduced to KBS TV Broadcasting in Korea 27 Oct. 2011, also link to China network:
http://v.youku.com/v_show/id_XMzE3NjMyNTky. html

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 551–562.
springerlink.com

barley and millet that were spread out in the yard. Later, *Jong Nang* was developed into the means of informing visitors whether the residents were at home or not [3-8]. The *Jong Nang* used the binary system similar to digital communications and computers today. Three timbers were exactly like three binary digits. The *Jong Nang* system could convey eight different messages. One of three Jong-Nang placed between the Jong-Ju-Mok, or "100" indicated there was no one at home, but the family would soon return form a neighboring area. Two *Jong Nangs*, or "101" meant the family was visiting a neighboring town and it would be a while before they returned. All these *Jong Nangs*, or "111" announced the family was out of town for a long time, as shown in Fig 1. When none of the *Jong Nang* was placed, or "000", this meant the family was at home, as shown in Fig 1. This system derived from the life of the Jeju Island people.



(a) House with Jong Nang

(b) The Jong Nang Code with 3 bits

Fig. 1. Korea Jeju Island Jong Nang

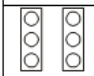
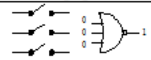
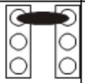
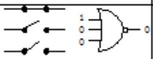
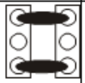
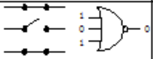
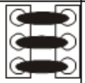
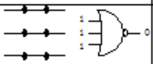
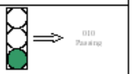
Although it is not exactly known when the *Jong Nang* first appeared, is not clear, but it is considered to be about 760 years ago, during the Koryo dynasty. It seems to have started after the construction of stone fences. Stone fences were first erected to show the boundary of land ownership. During King Kojong's era, government officials and judges were sent to Jeju Island. In A.D.1234, Ku Kim was appointed as a judge and he ordered residents to build stone fences along their boundaries.

The dispute over boundaries and the damage of agricultural products caused by grazing cattle and horses subsequently disappeared. The stone fences also acted as windbreaks. In addition, the collection of stones used to construct the fences contributed to easier cultivation and the quality of Island life. The Jong-Nang did not appear simultaneously with the erection of stone fences. To pasture cattle and horses in the Jung-sangan (=mountain) village on Jeju Island, the people built fence-like enclosures to prevent the animals from intruding onto farms. During the Chosun dynasty (A.D. 1392-1910), the practice became widely used across the Island.

The installation of 'Salchaegi' or 'Sombi' at the entrance became a starting point of the *Jong-Nang*. 'Salchaegi' or 'Sombi' originated as a field door set up vertically to pasture cattle and horses at the foot of Mt. Halla (1950m). It consisted of four to five thin logs or branches. At first, 'Sombi' was supposed to set up so that a horse could get in and eat crops located in the yard. When two *Jong Nang*s were used, large cattle or horses could not get in, but calves or ponies could. When one more *Jong Nang* was added, the calves and ponies could not get into the house at all.

A general communication system conveys information from its source to a destination some distance away using common symbols. The *Jong Nang* HBCC of Jeju Island, however, differs from pre-set wireless communication in that the people wanting to communicate had to visit to confirm the message.

Table 1. Jong Nang messages

Jong Nang Pattern	Jong Nang Comm.	Jong Nang Digital code	Jong Nang Switching NOR Channel/ Logic gate [8]	Traffic Signal Codes
	Staying at home	000		111 No Crossing
	Visiting next door for a while	100		No signal
	Visiting a neighboring village	101		010 Stand by
	Out of town for a long time	111		

The digital logic analysis of *Jong Nang* is presented in Table 1.

The mapping from Boolean algebra into digital switching was first noted in the paper “A symbolic analysis of relay and switching circuits”. Computer logic and digital system have been developed from this circuit-switching concept. But the *Jong Nang* system based on NOR circuit has represented the output as “0” and “1” easily from about 760 years before. Table 2 which was proposed by author shows Jong Nang Switching Channel and Modern Traffic.

Table 2. Relationships between Traffic Light and Jong Nang Code Signals

Code/Light	Traffic Signal Code	Jong Nang Code	Relationship
Red	111	000	1’s Complementary
Yellow	010	101	1’s Complementary
Green	000	111	1’s Complementary

In Table 1, we present the relationship between the *Jong Nang* switching channel/logic gate and modern traffic light as follows. We can make sense that modern traffic signal codes are 1’s complementary of *Jong Nang* codes in Table 2.

2 The Jong Nang Digital Logic Analysis

Definition 1: *Jong Nang* system is composed of 3 bits as shown Table 1 and each bit conveys different meanings. MSB (Most Significant Bit) as an existing bit indicates whether people are at home or not. When the existing bit is “0”, it means someone is at home. Second bit is a spatial bit, so if this bit marks “0”, it carries the information that the landlord stays near outside from his house. LSB (Least Significant Bit) is a temporal bit. “0” represents that no one is in house but returns in a short time. For each bit, “1” means the opposite case, as shown in Fig 2. This custom is similar fashion as space-time coding scheme.

Definition 2: The information priority of 3 bit *Jong Nang* system is MSB, Medium bit, and LSB in a sequence order, i.e. MSB>Medium>LSB.

Counter case 1) “001”: Existing MSB “0” shows someone is at home, spatial second bit “0” means nobody at home for visiting near village and temporal LSB “1” informs long time outgoing. Therefore, it is contrary to Definition 2, because the information priority does not follow the ordering.

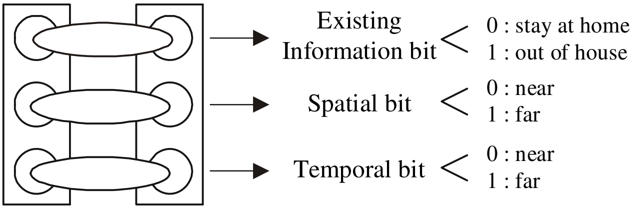


Fig. 2. Jong Nang information code protocol

Counter case 2) “110”: Existing MSB “1” indicates house is empty, spatial second bit “1” denotes long distance visiting and temporal LSB “0” shows short time outgoing. But long distance visiting generally requires long time to come back home. This case also is contrary to Definition 2 by the same reason of the counter case 1.

The above cases are not appropriate to the Definition 2, so they are not permitted in the Jong-Nang logic as shown in Table 3 (shaded entries).

Definition 3: The larger Hamming weight of *Jong Nang* message, the longer the outgoing time. For example, “111” represents that it takes longer outgoing time than “100” or “101”. Table 3 shows the different *Jong Nang* messages.

In view of Jong Nang HBCC development, it’s evident that they used to place the upper *Jong Nang* first, then lower, and middle one last. The introduction and Table 2 also explain its history, which is based on the oriental philosophy. But, today it is often used “001” (place the lower *Jong Nang* only) and “011” (place the middle and lower *Jong Nang*) patterns respective form earlier *Jong Nang* Philosophy.

Table 3. The comparison of decimal and binary number in Jong Nang

Decimal	Binary	Comparison
0	000	Staying at home
1	001	Not permitted
2	010	Not permitted
3	011	Not permitted
4	100	Visiting next door for a while
5	101	Visiting neighboring village
6	110	Not permitted
7	111	24 hours out of home

3 Capacity of the Jong Nang Channel as 3-Input NOR Channel

When the Jong Nang channel is viewed as a coordinated 3-input noiseless multiple access NOR channel with the following transition probabilities: $p(y=1|x=100)=p(y=0|x=100)=p(y=0|x=101)=p(y=0|x=111)=1$, the capacity of the channel can be calculated [4, 9].

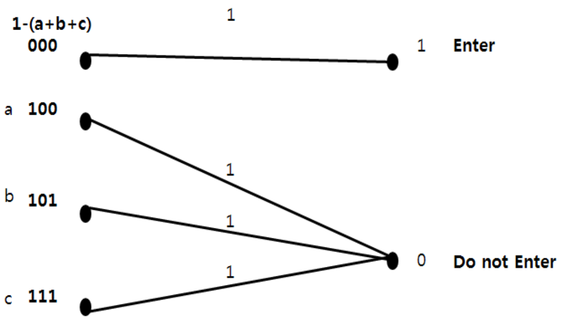


Fig. 3. The Jong Nang Channel

Theorem. The capacity of the Jong Nang Channel modelled as 3-input noiseless NOR channel is one.

Proof. Let $p(x=100)=a$, $p(x=101)=b$, $p(x=111)=c$, and $p(x=000)=1-(a+b+c)$. The output of the channel then has the following distribution:

$$p(y=0)=a+b+c, \quad p(y=1)=1-(a+b+c) \tag{1}$$

Furthermore, we can calculate the joint distributions:

$$p(y=1,x=000)=p(y=1|x=000)\,p(x=000)=1-(a+b+c)$$
$$p(y=0,x=100)=p(y=0|x=100)\,p(x=100)=a$$

$$p(y=0,x=101)=p(y=0\,|\,x=101)\,p(x=101)=b$$
$$p(y=0,x=111)=p(y=0\,|\,x=111)\,p(x=111)=c \qquad (2)$$

Now, to compute the mutual information between the input and the output of the channel $I(X;Y)=H(Y)-H(Y\,|\,X)$,we evaluate $H(Y)$ and $H(Y\,|\,X)$. Let $q=a+b+c$, we have

$$H(Y)=-(1-q)\log(1-q)-q\log(q)=h(q) \qquad (3)$$

$$H(Y\,|\,X)=-(1-q)\log(1)-a\log(1)-b\log(1)-c\log(1)=0 \qquad (4)$$

Hence,

$$I(X;Y)=H(Y)-H(Y\,|\,X)=h(q) \qquad (5)$$

Maximizing the mutual information with respect to q ,

$$\ln(2)\,dI(X;Y)/dq=1+\ln(1-q)-1-\ln(q)=\ln((1-q)/q)=0\,, \qquad (6)$$

which implies that $q=1/2$.

Therefore, the capacity of the channel is $h(1/2)=1$. **QED**.

That the capacity of the channel is 1 and can be achieved when $q=1/2$ implies that the family can communicate at most one bit of information per three timbers using this channel, with the requirement that the family is home half of the time.

4 The Jong Nang Binary Erasure Multiple Access Channel

We can see that the Jong Nang code does not have any error. In the case that the channel has noise, which may result from a timber falling down from its place by natural or manmade sort of event, an error detecting code maybe desired. As shown Fig 3~ 4:

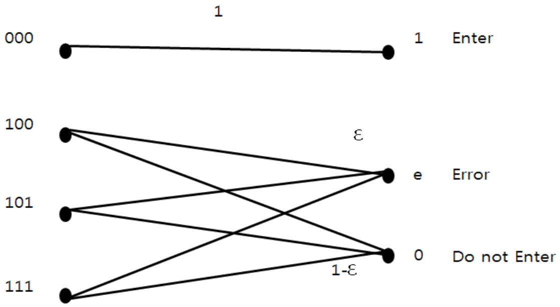


Fig. 4. The Jong Nang Binary Erasure Channel

As we mentioned, the channel does not take care the error number. It only takes whether the error is occurred or not. So all errors can be detected by using the codes, but the number of errors cannot be detected. In the following analysis, we assume that errors can occur when timbers fall down from its original place.

As shown Fig 5, the occured error includes all case of the errors; first, we transmit '111' which has the message 'out of town' and noise also occured, and we receive '101' which has the message 'visiting neighboring Village', this channel is transmitting the wrong message by ocuuring the error. Second, we transmit '111' which has the defined message and noise also occured. We receive '110' which has the undefined message.

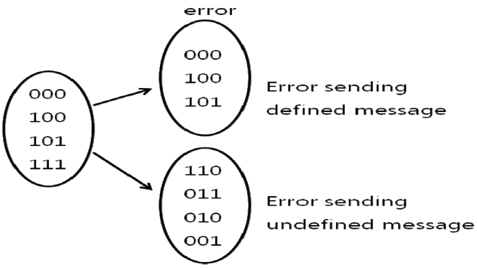


Fig. 5. The event diagram of the occured error

Consequently, the only allowed words are 000, 011, 101, 110, 001, 010, 100. Furthermore, to reflect the situation in which the error is detect, the "erased" channel output is created, and let probabilities that an erasure occurs given channel inputs 000, 100, 101 and 111 are the same and equal to ε. That is when a visitor sees 110, 001, 010 or 001, channel output is set to "erased." Under these conditions, the channel transition probabilities are given,

$$p(y=1 \mid x=000)=1$$

$$p(y=0 \mid x=011)=p(y=0 \mid x=101)=p(y=0 \mid x=110)=1-\varepsilon$$

$$p(y=e \mid x=011)=p(y=e \mid x=101)=p(y=e \mid x=110)=\varepsilon \tag{7}$$

Theorem. The capacity of the Noisy Jong Nang Channel as described above is 1.

Proof. Let $p(x=000)=a,\ p(x=100)=b,\ p(x=101)=c,\ p(x=111)=1-(a+b+c)$, and $q=a+b+c$. The output of the channel then has the following distribution:

$$p(y=1)=1-(a+b+c)=1-q,\ p(y=e)=\varepsilon(a+b+c)=\varepsilon q$$

$$p(y=0)=(1-\varepsilon)(a+b+c)=(1-\varepsilon)q \tag{8}$$

Furthermore, we can calculate the joint distributions:

$$p(y=1,x=000)=p(y=1 \mid x=000)\,p(\ x=000)=1-(a+b+c)$$

$$p(y=0,x=100)=p(y=0 \mid x=100)\,p(\ x=100)=a(1-\varepsilon)$$

$$p(y=0,x=101)=p(y=0\,|\,x=101)\,p(\,x=101)=b(1-\varepsilon)$$
$$p(y=0,x=111)=p(y=0\,|\,x=111)\,p(\,x=111)=c(1-\varepsilon)$$
$$p(y=e,x=100)=p(y=e\,|\,x=100)\,p(\,x=100)=a\varepsilon$$
$$p(y=e,x=101)=p(y=e\,|\,x=101)\,p(\,x=101)=b\varepsilon$$
$$p(y=e,x=111)=p(y=e\,|\,x=111)\,p(\,x=111)=c\varepsilon \tag{9}$$

Now, to compute the mutual information between the input and the output of the channel $I(X;Y)=H(Y)-H(Y\,|\,X)$, we evaluate $H(Y)$ and $H(Y\,|\,X)$.

The capacity of the channel (C) is maximum of the mutual information ; that is, when $H(Y\,|\,X)=0$, we find the capacity of the channel.

Let $q=a+b+c$, we have

$$H(Y)=-(1-q)\log(1-q)-\varepsilon q\log\varepsilon q-(1-\varepsilon)q\log(1-\varepsilon)q \tag{10}$$
$$H(Y\,|\,X)=-(1-q)\log(1)-\varepsilon q\log\varepsilon-(1-\varepsilon)q\log(1-\varepsilon) \tag{11}$$

Hence,

$$I(X;Y)=H(Y)-H(Y\,|\,X)=-(1-q)\log(1-q)-\varepsilon q\log\varepsilon q-(1-\varepsilon)q\log(1-\varepsilon)q$$
$$+(1-q)\log(1)+\varepsilon q\log\varepsilon+(1-\varepsilon)q\log(1-\varepsilon) \tag{12}$$

Maximizing the mutual information with respect to q , $\ln(2)\,dI(X;Y)/dq$

$$=1+\ln(1-q)-\varepsilon-\varepsilon\ln\varepsilon q-(1-\varepsilon)-(1-\varepsilon)\ln(1-\varepsilon)q+\varepsilon+\ln\varepsilon+(1-\varepsilon)\ln(1-\varepsilon)$$
$$=\ln(1-q)-\varepsilon\ln q-(1-\varepsilon)\ln q=\mathrm{n}(1-q)-\ln q=\ln\left[(1-q)/q\right]=0 \tag{13}$$

which implies that $q=1/2$.

Therefore, the capacity of the channel is

$$C=-0.5\log(0.5)-0.5\varepsilon\log0.5\varepsilon-0.5(1-\varepsilon)\log0.5(1-\varepsilon)+0.5\varepsilon\log\varepsilon+0.5(1-\varepsilon)$$
$$\log(1-\varepsilon)=-0.5\log0.5+0.5\varepsilon\log2+0.5(1-\varepsilon)\log2=0.5+0.5=1\text{ . }\textbf{QED}. \tag{14}$$

We can analysis capacity of the the Jong Nang Binary Erasure multiple access channel (MAC), in which four senders send the information to 3 receivers. This channel has binary inputs, $X_1, X_2 \in \{000,100,101,111\}$ and a ternary output Y_e, Y_1, Y_0 .

Proof. Note that $p(x=000)=p(x=100)=p(x=101)=p(x=111)=1/4$. We can represent 000, 101, 110 and 111 by 00, 10, 01, 11 repectively. For the give information $X_1, X_2 \in \{000,100,101,111\}$ and ternary outputs $Y\in\{1,e,0\}$. We can get the channel capacity

$$C_1 < I(X_1;Y\,|\,X_2),\quad C_2 < I(X_2;Y\,|\,X_1).$$

we also get

$$C_1 = C_2 = Max\{I(X_1;Y \mid X_2)\}$$
$$= \sum_{X_1}\sum_{X_2}\sum_{Y} p(X_1)p(X_2)p(Y \mid X_1, X_2)\log\frac{p(Y \mid X_1, X_2)}{\sum_{X_1} p(X_1)p(Y \mid X_1, X_2)} \qquad (15)$$

$=1$, as shown in Fig 6(a).

Then the channel capacity of the combined channel is

$$C_{12} = Max\, I(X_1, X_2; Y)$$
$$= \sum_{X_1}\sum_{X_2}\sum_{Y} p(X_1)p(X_2)p(Y \mid X_1, X_2)\log\frac{p(Y \mid X_1, X_2)}{\sum_{X_1}\sum_{X_2} p(X_1)p(X_2)p(Y \mid X_1, X_2)} \qquad (16)$$

It is easy to see that

$$P(Y=1) = P(X=000) = K,\ P(Y=\varepsilon) = P(X=100)\varepsilon + P(X=101)\varepsilon + P(X=111)\varepsilon$$
$$= [P(X=000) + P(X=101) + P(X=111)]\varepsilon = 3K\varepsilon, \text{ and}$$
$$P(Y=0) = [P(X=000) + P(X=101) + P(X=111)](1-\varepsilon) = 3K(1-\varepsilon) \qquad (17)$$

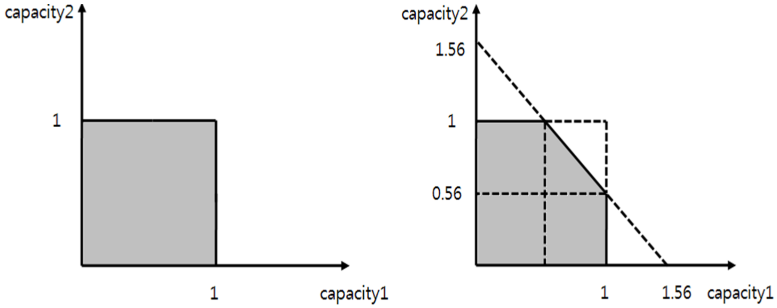
Consequently,

$$H(Y) = -[K\log K + 3K\varepsilon\log 3K\varepsilon + 3K(1-\varepsilon)\log K(1-\varepsilon)]. \qquad (18)$$

Since $K = 1/4,\ \varepsilon = 1/2$ and $H(Y \mid X_1, X_2) = 0$, we have

$$H(Y) = -\left[\frac{1}{4}\log\frac{1}{4} + 3\times\frac{1}{4}\times\frac{1}{2}\log\frac{3}{8} + 3\times\frac{1}{4}\times\frac{1}{2}\log\frac{3}{8}\right] = 1.56 \qquad (19)$$

and $Max(C_{12}) \approx 1.56$, which is shown in Fig 6(b).



(a) Independent (b) Dependent

Fig. 6. Capacity region for the Jong Nang channels

In case Fig 6(a), the inputs of Jong Nang channel are dependent, there is interference between the senders. In case Fig 6(b), the inputs of Jong Nang Erasure multiple access channel are independent, there is no interference between the senders.

5 Deterministic Model for Jong Nang Channel as Mobile Backhaul

In this section, we introduce a deterministic model for Jong Nang channel. From Fig. 2 & 6, we propose a deterministic model of Jong Nang channel in Fig 7. We notice in this simplified model there is no noise any more and we call it a deterministic model like [10, 11].

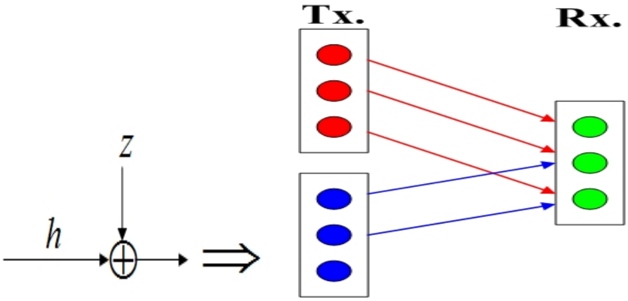


Fig. 7. A deterministic model of Jong Nang channel

This section directly illustrates the connections between the capacity regions of the Gaussian two-user MAC and the deterministic Jong Nang MAC. The capacity region of the MAC is

$$\begin{aligned} R_1 &\le \log\left(1+SNR_1\right) \approx \log\left(SNR_1\right) \\ R_2 &\le \log\left(1+SNR_2\right) \approx \log\left(SNR_2\right) \\ R_1+R_2 &\le \log\left(1+SNR_1+SNR_2\right). \end{aligned} \tag{20}$$

The capacity region of the deterministic Jong Nang MAC is

$$R_1 \le n_1, \quad R_2 \le n_2 \text{ and } R_1+R_2 \le \max\left(n_1,n_2\right). \tag{21}$$

Comparing with the capacity region of the Gaussian MAC and the deterministic Jong Nang MAC, we can make the correspondence $n_1 = \log SNR_1, n_2 = \log SNR_2$. This region is plotted in Fig 8. The simple two-user multiple-access channel case illustrates the idea that one can attempt to reduce the Gaussian problem to a deterministic one by proving a constant gap between the capacity regions of the two models.

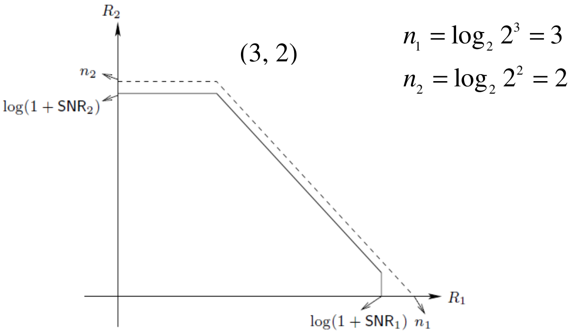


Fig. 8. Capacity region of Gaussian MAC. (Solid line). Capacity region of deterministic Jong Nang MAC. (Dashed line)

The traffic in the Mobile backhaul can be separated either based on services or location into tagged frames and untagged frames. There is a critical requirement to migrate the mobile backhaul network to technologies that can support quality of service to separate traffic streams, timing synchronization, lower packet loss, and high availability along with maintaining low operational expenditure.

6 Conclusion

The capacity of the Jong Nang channel which means the wooden gate in Korea Jeju island dialect, had three wooden rafters placed on Jong-Ju-Mok (two large vertical stones with three holes) to convey the family's whereabouts was calculated and compared to traffic signal codes with digital logic 1's complementary. A modified code for the Jong Nang is proposed in case that errors occurs because timbers falls down from its place by natural or manmade sort of event. Even though the error is introduced, the capacities of the Jong Nang channel and the noisy Jong Nang channel are one, which means transmission may be accomplished without error based on the Jong Nang binary erasure multiple access channel. Jong Nang is deterministic model channel, so it can be applied backhaul mobile base station and wireless relay network.

References

1. Heath, F.G.: Origins of the binary code. Scientific American 227, 76–83 (1972)
2. Aschoff, V.: The early history of the binary code. IEEE Communication Magazine, 4–10 (1983)
3. Lee, M.H.: Jong Nang. In: EXPO 1993 Information & Telecom. Pavilion poster (1993)
4. Cover, T.M., Thomas, J.A.: Elements of Information Theory. John Wiley & Sons, New York (1991)
5. Lee, M.H.: The History of Information and Communication. Kimyeong-Sa, Seoul (1994)
6. Lee, M.H.: Jong Nang: The symbol of digital communication and Ying and Yang. Telecom. 9(1) (1993)
7. Lee, M.H.: Jong Nang System. Patent, no. 133285 (1998)

8. Lee, M.H.: The History of Jeju Jong Nang Binary Code. IEEE VTS News 50(1) (2003)
9. Shannon, C.E.: A Mathematical Theory of Communication. Bell System Technical Journal 27, 379–423, 623–656 (1948)
10. Marseh, P., Fettweis, G.P.: Coordinated Multi-Point in Mobile Communications, ch. 12, pp. 277–310. Cambridge Press (2011)
11. Salman, A., Suhas, D., Tse, D.: A Deterministic approaches to wireless relay networks. In: ISIT (2007)

Multiuser Transmitter Preprocessing Aided Downlink Communications in Correlated Frequency-Selective Channels

Nithin Srinivasan, Sriram Ravichandran, Shruthi Ravichandran,
and Prabagarane Nagaradjane

Department of Electronics and Communication Engineering, SSN Institutions, Chennai, India-603110
{nithin.181990,srirampsbb,rshruthir90}@gmail.com,
prabagaranen@ssn.edu.in

Abstract. In this treatise, we investigate and present the performance of zero-forcing multiuser transmitter preprocessing (ZF-MUTP) in the context of multiuser multiple-input multiple-output (MIMO) systems based on space division multiple access (SDMA) in correlated frequency-selective channel, as well as in downlink (DL) direct sequence code division multiple access (DS CDMA) systems. The ZF-MUTP approach adopted in this treatise creates an orthogonal subspace to the interfering subspace in contrast to the conventional approach of pseudo inverting the channel state information (CSI) matrix. This approach completely removes the multiuser interference in downlink (DL) communications. Further, we investigate two power allocation policies based on maximum signal to noise ratio (MSNR) and equal power in the context of achievable capacity and symbol error rate (SER) in the DL system. It is shown through simulation that the achievable capacity in SDMA systems in terms of the number of bits per second per channel use is better in the case of MSNR power allocation policy than equal power allocation scheme. Also, simulation results demonstrate that MSNR results in better achievable SER than equal power allocation policy. Further, this ZF-MUTP based on creating orthogonal subspace to interfering subspace results in the same performance as that of the conventional ZF-MUTP.

Keywords: Multiple-input multiple-output, multiuser transmitter preprocessing, orthogonal subspace, direct sequence code division multiple access.

1 Introduction

Information theoretic results have shown that multiple-input multiple-output (MIMO) systems can provide very high capacity without calling for an increase in transmission bandwidth and power in rich scattering environments at high signal to noise ratios [1]. So far, intense research has been carried out in the analysis of single user (SU) MIMO systems. As a design alternative, space division multiple access (SDMA) systems have been proposed for multiuser (MU) MIMO systems. In a single user (SU) MIMO system, users experience only multi-antenna interference (MAI), also called

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 563–573.
springerlink.com

multi-stream interference (MSI). However, in MU-MIMO systems, users conflict multiuser interference (MUI) in addition to MAI. To deal with these interferences, multiuser detectors (MUDs) can be employed at the mobile station (MS), but they impose too much computational complexity in signal detection. Further, such MUDs are not practically implementable even at the base station (BS) where complexity is acceptable.

Of late, multiuser transmitter preprocessing (MUTP) that completely removes the multiuser interference has received widespread attention among researchers. The MUTP results in low complexity signal detection by carrying out the preprocessing at the transmitter. Furthermore, this MUTP can result in highly power-efficient mobile stations (MSs). In [3-7] the zero-forcing multiuser transmitter preprocessing (ZF-MUTP) that completely eliminates the downlink (DL) MUI has been studied. Further, the most widely adopted method in deriving the ZF based preprocessing matrix is to pseudo invert [3], [4], [6], [9], [10] the small scale fading matrix or channel state information (CSI) matrix. Alternatively, singular value decomposition (SVD) can be exploited to construct the preprocessing matrix that constitutes an orthogonal subspace to the interfering users' subspace [7], [8].

Motivation for this work stems from a recent study by Lie-Liang Yang [2] where the ZF preprocessing matrix has been derived by an alternate approach with the aid of the CSI associated with the interfering users. The investigations of [2] were primarily carried out for flat fading channels. However, frequency-selectivity can severely degrade the system performance in terms of achievable error rate. Also, lack of angle spread and rich scattering environment will cause fading to be correlated and thus severely affect the performance of the MIMO system. Hence, in this work, we present the performance of the ZF-MUTP aided SDMA system by adopting the conventional and proposed ZF-MUTP approaches of [2] in the context of multiuser scenarios for correlated frequency-flat and frequency-selective channels. Furthermore, we present the achievable capacity of the ZF-MUTP based DL system by exploiting equal power and MSNR allocation regimes.

The rest of the paper is organized as follows. Sections 2 and 3 describe the SDMA and DS CDMA system models with MUTP. Section 4 describes the power allocation regimes, section 5 elucidates the performance results and in section 6, conclusions are drawn.

2 Preprocessing Assisted SDMA System

Figure 1 elucidates a typical SDMA DL system. Here we consider a single base station (BS) with N_t transmit antennas supporting K MSs, each with $N_k, (k = 1,2...., K)$ receive antennas. Further, we assume that the CSI matrix connecting the kth MS and BS to be both frequency-flat and frequency-selective. Furthermore, it is assumed that accurate and prompt CSI is available at the transmitter. As shown in the figure, after MUTP, the received $(KN_k \times 1)$ component vector constituting the decision variables of the K DL users is given by

$$\boldsymbol{y} = \mathcal{H}\mathcal{P}\boldsymbol{x} + \boldsymbol{n} \tag{1}$$

where, $\mathcal{H} = [\boldsymbol{h}_1; \boldsymbol{h}_2; \ldots; \boldsymbol{h}_K]$ constitutes the $(KN_k \times N_t)$ component small scale fading matrix connecting the BS and all the K MSs, $\mathcal{P}$ is the preprocessing matrix of dimension $(N_t \times KN_k)$ and $\boldsymbol{x} = [x_1, x_2, \ldots, x_K]^T$ is a $(KN_k \times 1)$ component symbol vector transmitted to all the K MSs by the BS. Further, $\boldsymbol{h}_k = [h_{j_1}, h_{j_2}, \ldots, h_{j_{N_t}}]$, $(j = 1, 2 \ldots, N_k)$ defines the $(N_k \times N_t)$ small scale fading matrix connecting the kth MS and BS and $\boldsymbol{n}$ denotes the $(KN_k \times 1)$ length noise observation vector which, is assumed to be Gaussian distributed with zero mean and covariance $E\left[\boldsymbol{n}\boldsymbol{n}^H\right] = \frac{\sigma^2}{2} I_{KN_k}$. Further, in the context of a frequency-selective channel, the impulse response from the jth receiving antenna to the ith transmit antenna is given by [13]

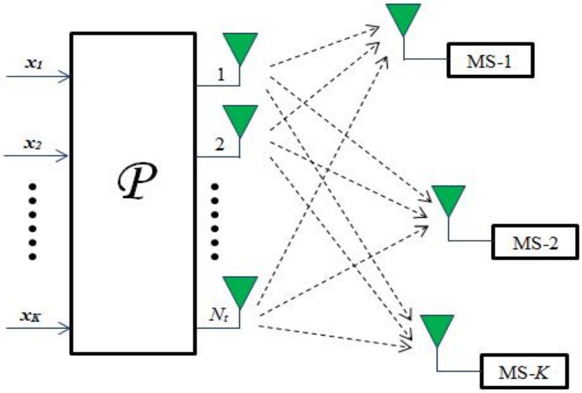


Fig. 1. Block diagram of transmitter preprocessing assisted SDMA systems

$$h_{ji}(t) = \sum_{l=1}^{L} h_{ji}^{l} \delta(t - \tau_l) \tag{2}$$

where, h_{ji}^{l} is a complex zero mean Gaussian random process with variance $p(\tau_l)$ in accordance with [13]. L is the total number of paths between the ith transmit antenna and jth receive antenna. In (1), the preprocessing matrix is so designed, that it completely eliminates the MUI for each of the K MSs. Assuming $N_t \geq KN_k$, the MUI can be fully removed when the DL preprocessing matrix is chosen to satisfy

$$\mathcal{H}\mathcal{P} = \boldsymbol{\varphi} \tag{3}$$

where, $\boldsymbol{\varphi} = \text{diag}\{\psi_1, \psi_2, \ldots, \psi_{KN_k}\}$ is a diagonal matrix denoting the power allocation regime. These power normalization coefficients represent the constraints on the transmit power. In order to satisfy (3), $\mathcal{P}$ can be set to

$$\mathcal{P} = \mathcal{H}^{+}\psi = \tilde{\mathcal{P}}\psi \tag{4}$$

In the above expression, $(.)^{+}$ denotes the pseudo inverse of the matrix and $\tilde{\mathcal{P}} = [\mathcal{H}]^{+} = \mathcal{H}[\mathcal{H}^{H}\mathcal{H}]^{-1}$. Further, $\mathcal{P}$ in (4) refers to the conventional ZF - MUTP in the context of the DL SDMA system. Alternatively, ZF-MUTP can also be derived by the method proposed in [2]. Invoking (1), the *k*th MS's decision variable can be expressed as

$$\boldsymbol{y}_k = \boldsymbol{h}_k \boldsymbol{p}_k \boldsymbol{x}_k + \sum_{j \neq k}^{K} \boldsymbol{h}_k \boldsymbol{p}_j \boldsymbol{x}_j + \boldsymbol{n}_k \, , k = 1, 2, \ldots, K \tag{5}$$

The ZF-MUTP can be arrived at, if we satisfy the following conditions for any $k \in \{1,2,...,K\}$

$$\boldsymbol{h}_k \boldsymbol{p}_k > 0 \text{ and } \boldsymbol{h}_k \boldsymbol{p}_j = 0 \text{ for any } j \neq k \tag{6}$$

In order to satisfy the condition (6), the preprocessing matrix of the *k*th MS, $\boldsymbol{p}_k$ should be designed in such a way that it lies in a sub space that is orthogonal to the sub space determined by the CSI matrix of the remaining $(K-1)$ MSs. In other words, the preprocessing matrix is derived with the aid of the $[N_t \times (K-1)N_k]$ CSI matrix associated with the interfering users. Denoting $\overline{\mathcal{H}}_k$ as the CSI matrix for deriving the preprocessing matrix for the *k*th MS, $\overline{\mathcal{H}}_k$ is defined as

$$\overline{\mathcal{H}}_k = [\boldsymbol{h}_1^T, \ldots, \boldsymbol{h}_{k-1}^T, \boldsymbol{h}_{k+1}^T, \ldots, \boldsymbol{h}_K^T] \tag{7}$$

Exploiting the interference matrix of (7), $\boldsymbol{p}_k$ can be realized as follows [2], [11]

$$\boldsymbol{p}_k = \psi_k [\boldsymbol{I}_{N_t} - \overline{\mathcal{H}}_k^{*} (\overline{\mathcal{H}}_k^{T} \overline{\mathcal{H}}_k^{*})^{-1} \overline{\mathcal{H}}_k^{T}] (\boldsymbol{h}_k^{*})^T , k = 1, 2, \ldots, K \tag{8}$$

where, (.)* represents the conjugate of the argument. ψ_k designates the power normalization coefficients that should satisfy the constraint

$$\text{Trace}(\boldsymbol{p}_k^H \boldsymbol{p}_k \psi_k^2) \leq N_k \tag{9}$$

Furthermore, it can be shown that (8) satisfies the conditions of (6). The preprocessing matrix $\mathcal{P} = [\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_K]$ realized based on $\boldsymbol{p}_k$ of (8) results in the ZF solution. Alternatively, the preprocessing matrix can be derived with the aid of eigen analysis. Applying singular value decomposition (SVD) on (7), $\overline{\mathcal{H}}_k$ can be expressed as

$$\begin{aligned} \overline{\mathcal{H}}_k &= [\overline{\mathcal{U}}_{ks} \mid \overline{\mathcal{U}}_{kn}] \begin{bmatrix} \Sigma_{ks}^{1/2} \\ 0 \end{bmatrix} \overline{\mathcal{V}}_k^H \\ &= \overline{\mathcal{U}}_{ks} \Sigma_{ks}^{1/2} \, \overline{\mathcal{V}}_k^H \, , \, k = \boldsymbol{1}, \boldsymbol{2}, \ldots, K \end{aligned} \tag{10}$$

where, $\overline{\mathcal{U}}_k$ is an $(N_t \times N_t)$ component unitary matrix, $\overline{\mathcal{V}}_k$ is an $[N_k(K-1) \times N_k(K-1)]$ component unitary matrix and Σ_{ks} is an $[N_k(K-1) \times N_k(K-1)]$ component diagonal matrix constituting the non-zero eigen values of $\overline{\mathcal{H}}_k \overline{\mathcal{H}}_k^H$. $\overline{\mathcal{U}}_{ks}$ is an $[N_t \times N_k(K\text{ -}1)]$ matrix constituting the $[N_k(K-1)]$ columns of $\overline{\mathcal{U}}_k$ that correspond to the non-zero eigen values of $\overline{\mathcal{H}}_k \overline{\mathcal{H}}_k^H$ and $\overline{\mathcal{U}}_{kn}$ is an $[N_t \times (N_t \text{ - } N_k(K\text{ -}1))]$component matrix constituting the $[N_t - N_k(K-1)]$ columns of $\overline{\mathcal{U}}_k$ that correspond to the null space of $\overline{\mathcal{H}}_k \overline{\mathcal{H}}_k^H$.

Upon substituting (10) into (8) and exploiting the property $\overline{\mathcal{U}}_k^H \overline{\mathcal{U}}_k = \boldsymbol{I}_{N_t}$ it can be readily shown that $\boldsymbol{p}_k$ can be expressed as

$$\boldsymbol{p}_k = \psi_k (\overline{\mathcal{U}}_{kn}^H \overline{\mathcal{U}}_{kn}) \boldsymbol{h}_k^*, \; k = 1, 2, \ldots, K \tag{11}$$

3 Preprocessing Assisted DS CDMA System

Here, we consider K users arbitrarily distributed in a single cell DS-CDMA system. The BS and the MSs are equipped with single antenna. In the context of the multiuser DS CDMA DL system, users experience multiuser interference (MUI). To fully remove this DL MUI, the preprocessing based on ZF addressed in the previous section in the context of SDMA can be exploited. Let $\mathcal{C}$ denote the code matrix of all the users which is given by

$$\mathcal{C} = [\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_K] \tag{12}$$

where, $\boldsymbol{c}_k$ is the *k*th user's orthogonal spreading code of length N given by

$$\boldsymbol{c}_k = \frac{1}{\sqrt{N}} [c_{k0}, c_{k1}, \ldots, c_{k(N-1)}]^T \text{, where } \quad k = 1,2,\ldots,K \tag{13}$$

In the context of MUTP aided DS-CDMA system, the $(K \times 1)$ received vector can be expressed as

$$\boldsymbol{y} = \tilde{\mathcal{H}} \tilde{\mathcal{P}} \boldsymbol{x} + \boldsymbol{n} \tag{14}$$

where $\tilde{\mathcal{H}}$ is a $(K \times N)$ component matrix constituting the small scale fading effects, $\tilde{\mathcal{P}} = [\tilde{\mathcal{H}}]^+$ is the preprocessing matrix and $\boldsymbol{n}$ is the $(K \times 1)$ vector whose elements are Gaussian distributed random variables with zero mean and variance 1/2 per dimension. Also, here, we consider that the spreading operation and the preprocessing are carried out jointly [2]. In order to carry out the preprocessing, $\tilde{\mathcal{P}}$ can be expressed as

$$\tilde{\mathcal{P}} = [\tilde{\boldsymbol{p}}_1, \tilde{\boldsymbol{p}}_2, \ldots, \tilde{\boldsymbol{p}}_K] \tag{15}$$

where, $\tilde{\boldsymbol{p}}_k$ is given by

$$\tilde{\boldsymbol{p}}_k = \psi_k [I_{N_t} - \overline{\mathcal{H}}_k^* (\overline{\mathcal{H}}_k^T \overline{\mathcal{H}}_k^*)^{-1} \overline{\mathcal{H}}_k^T](\boldsymbol{h}_k^*)^T, k = 1, 2, \ldots, K \qquad (16)$$

where, $\boldsymbol{h}_k$ is the kth column of $\tilde{\mathcal{H}}^T$. Further, $\overline{\mathcal{H}}_k$ is $[N \times (K-1)]$ component matrix obtained by removing the kth column of $\tilde{\mathcal{H}}^T$. Now, the decision variable for the K DL users can be expressed as

$$\boldsymbol{y} = \tilde{\mathcal{H}}\tilde{\mathcal{P}}\boldsymbol{x} + \boldsymbol{n} \qquad (17)$$

4 Power Allocation Regime

The power allocation policy considered in this work is based on [12] and is implemented under the constraint of

$$E\left[\left\| \tilde{\mathcal{P}}\boldsymbol{x}^2 \right\| \right] \leq E\left[\left\| \boldsymbol{x}^2 \right\| \right] = KN_k \qquad (18)$$

This implies that, in the context of DL transmission, the total transmission power should not exceed the original transmission power without preprocessing. Also, to impose the above constraint, we assume $\mathrm{E}\left[\left\| \boldsymbol{xx}^H \right\|^2 \right] = \boldsymbol{I}_{KN_k}$. Upon substituting (3), (17) can be expressed as

$$\mathrm{Trace}(\tilde{\mathcal{P}}^H \tilde{\mathcal{P}} \psi^2) \leq KN_k \qquad (19)$$

Let $p_{ii} > 0 \ (1 \leq i \leq KN_k)$ be the diagonal elements of the matrix $\tilde{\mathcal{P}}^H \tilde{\mathcal{P}}$. Therefore, (19) can be expressed as

$$\sum_{i=1}^{KN_k} p_{ii} \psi_{ii}^2 \leq KN_k \qquad (20)$$

Alternatively, the power allocation policy can be implemented for individual users under the constraint of

$$E\left[\left\| \mathcal{P}_k x_k^2 \right\| \right] \leq E\left[\left\| x_k^2 \right\| \right] \qquad (21)$$

where, $\mathcal{P}_k = \tilde{\mathcal{P}}_k \psi_k$. Hence, by substituting $\mathcal{P}_k$, the above constraint can also be expressed as

$$\mathrm{Trace}(\tilde{\mathcal{P}}_k^H \tilde{\mathcal{P}}_k \psi_k^2) \leq N_k \qquad (22)$$

Denoting $\{\mathcal{P}_{k_{ii}}\}$ $(1 \leq i \leq N_k)$ as the diagonal elements of the matrix $\tilde{\mathcal{P}}_k^H \tilde{\mathcal{P}}_k$, the constraint can be written as

$$\sum_{i=1}^{N_k} \mathcal{P}_{k_{ii}} \psi_{k_{ii}}^2 \le N_k \tag{23}$$

In the context of MSNR power allocation regime, the overall DL information can be maximized if the BS maximizes φ_k given by

$$\varphi_k = \sum_{i=1}^{N_t} \frac{1}{\omega_{ki}} = \sum_{i=1}^{N_t} \frac{\sigma^2}{\lambda_{ki}\psi_{ki}^2} \tag{24}$$

where, ω_{ki} is the SNR of the ith antenna. Further, the power allocated to the ith data stream is then given by

$$\psi_{ki}^2 = N_t \left(\sum_{j=1}^{N_t} \sqrt{\frac{1}{\lambda_{kj}}} \right)^{-1} \frac{1}{\sqrt{\lambda_{ki}}}, \quad i=1,2,\ldots,N_t, k=1,2,\ldots,K \tag{25}$$

and ψ is chosen to satisfy the power allocation constraint of (20). Hence, the overall capacity normalized by the total number of receive antennas is given by

$$\bar{C} = \frac{1}{KN_k} \sum_{i=1}^{KN_k} \log_2 \left[1 + \frac{\lambda_{ii}\psi_{\text{ii}}^2}{\sigma^2} \right] \tag{26}$$

where, $\bar{C}$ denotes the average capacity.

5 Performance Results

In this section, the simulation results elucidating the performance of the SDMA and DS-CDMA systems employing the conventional ZF-MUTP, ZF-MUTP of [2] and eigen based ZF-MUTP in flat and frequency-selective fading channels are presented. It is assumed that, perfect CSI is present at the BS. The modulation technique considered is 64QAM (quadrature amplitude modulation) for the SDMA system and BPSK (binary phase shift keying) for the DS-CDMA system. Further, the achievable capacity per channel use of the MUTP assisted SDMA system using ZF-MUTP of [2] when invoking the two power allocation regimes namely, equal power regime and MSNR regime is analyzed. Furthermore, the performance of MUTP assisted SDMA system in terms of achievable SER is investigated by implementing two power allocation regimes based on the constraint of (19).

Figure 2 shows the performance of the SDMA system in the context of correlated frequency-selective channel for the scenario considered in Figure 1. It can be observed from the figure that the eigen based approach and ZF-MUTP of [2] result in the same performance. Furthermore, the performance of the system degrades with increasing number of users. In order to achieve a target SER of 10^{-3} for $K = 10$ users, the SDMA system requires 17 dB less compared to the case with K=20 users. Figure 3 illustrates the performance of the SDMA system when invoking the two power allocation regimes in correlated flat and frequency-selective channel for the scenario with N_t = 10 and K=4. The MSNR power allocation regime tries to allocate more power to the sub-channel with less eigen value. Further, for benchmarking, we have considered equal power allocation regime in the context of SDMA DL system. It is seen from the simulation results that, performance of the system employing MSNR is superior to that of equal power allocation regime. In addition to this, performance of the system in flat fading channel is better than the performance of the system in frequency-selective channel.

Figure 4 shows the performance of the DS CDMA system when communicating over frequency-selective channel. In our simulations, we have considered random spreading sequences with a spreading gain of 32. Simulation results show that ZF-MUTP can completely remove MUI and can result in better performance. Further, it is observed that as the number of users increases in the system, the achievable error rate decreases. Figure 5 demonstrates the achievable capacity in terms of number of bits per second per channel use in the context of correlated frequency-selective channel for the SDMA system with two power allocation regimes. The simulation results show that, the SDMA system with MSNR power allocation regime results in higher capacity compared to equal power allocation regime. This is mainly because MSNR allocates more power to the sub channel with the least eigen value under the

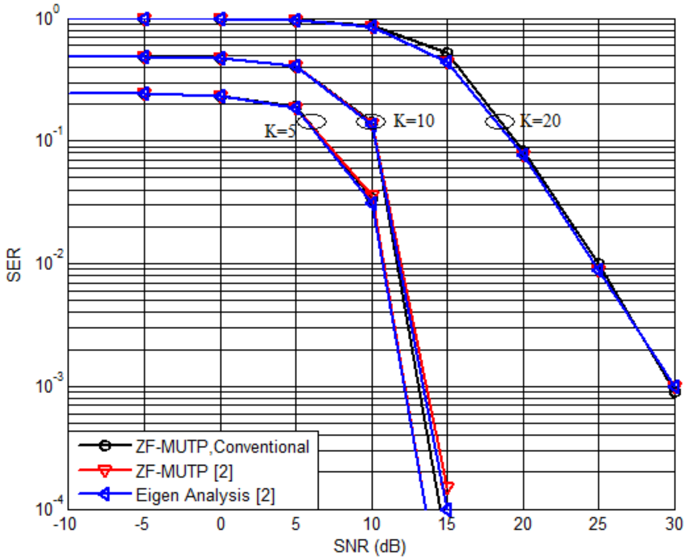


Fig. 2. SER vs SNR performance of the ZF-MUTP assisted DL SDMA system for correlated frequency-selective fading with 64 QAM modulation using 20 transmit antennas and 1 receive antenna per mobile station for K=5, 10 and 20 users. The simulations are based on the preprocessing matrix of (8) and (11) for ZF-MUTP and Eigen analysis respectively. (correlation coefficient=0.7)

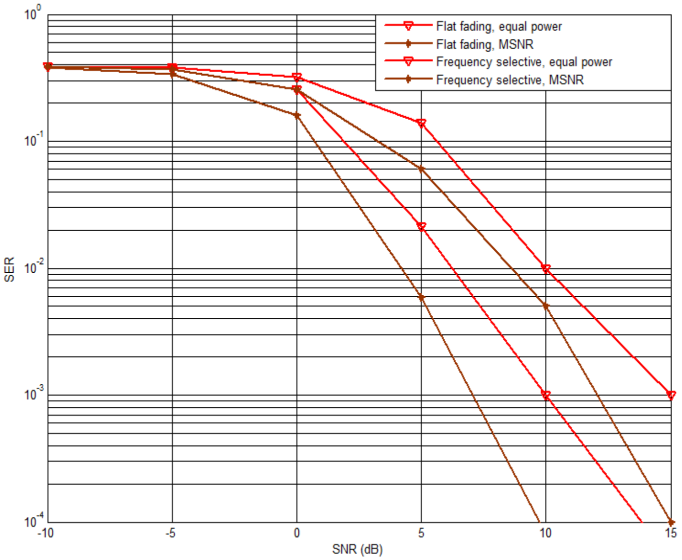


Fig. 3. SER vs SNR performance of the ZF-MUTP assisted DL SDMA system with N_t=10 and K=4 users in correlated flat and frequency-selective fading channel with equal power allocation and MSNR (correlation coefficient=0.7)

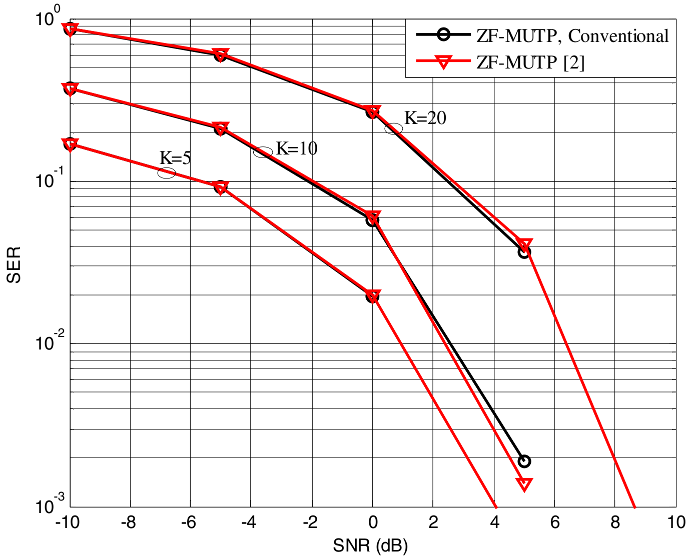


Fig. 4. BER vs SNR performance of the DL DS-CDMA system with N=32 in frequency-selective fading channel. The simulations are based on the preprocessing matrix of (8) for ZF-MUTP.

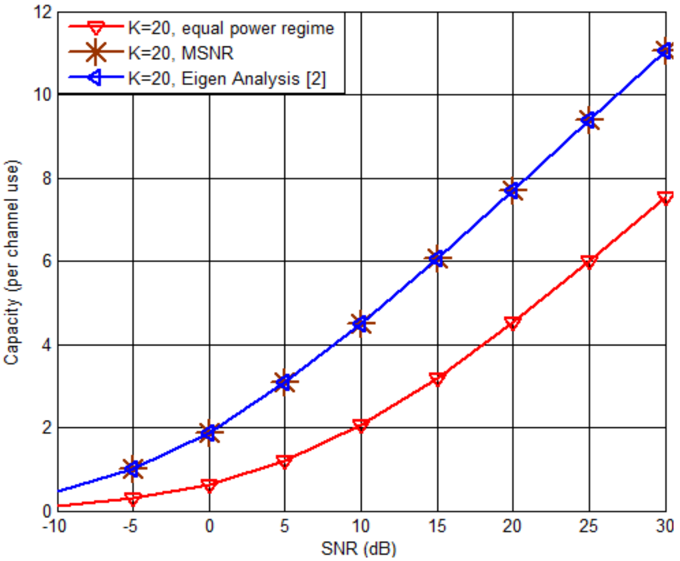


Fig. 5. Capacity vs SNR of the SDMA system with N_t=20 and K=20 users in correlated frequency-selective channel with equal power allocation regime and MSNR. The simulations are based on the preprocessing matrix of (11) for Eigen analysis.

constraint of (23). Further, for an SNR of 30 dB, MSNR results in 11 bits/sec/channel use as against 7.5 bits/sec/channel use in the case of equal power allocation strategy.

5 Conclusion

In this treatise, we investigated the performance of a single cell multiuser MIMO system based on SDMA in correlated flat and frequency-selective channels for DL communications with the aid of conventional ZF-MUTP, ZF-MUTP approach that creates an orthogonal subspace to the interfering subspace and ZF-MUTP based on eigen analysis. Our study shows that, all the three ZF-MUTP techniques exhibit similar performance in terms of achievable SER by completely eliminating the DL MUI. Further, MUTP with MSNR power allocation regime shows better performance in terms of achievable SER and capacity than equal power allocation regime.

References

1. Wolniansky, P.W., Foschini, G.J., Golden, G.D., Valenzuela, R.A.: V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel. In: International Symposium on Signals, Systems and Electronics, September 29-October 2, pp. 295–300 (1998)
2. Yang, L.-L.: A Zero-Forcing Multiuser Transmitter Preprocessing Scheme for Downlink Communications. IEEE Trans. on Commun. 56(6), 862–865 (2008)

3. Gerlach, D., Paulraj, A.: Adaptive transmitting antenna arrays with feedback. IEEE Signal Processing Lett. 1, 150–152 (1994)
4. Vojcic, B., Jang, W.: Transmitter preprocessing in synchronousmultiuser communications. IEEE Trans. Commun. 46, 1346–1355 (1998)
5. Choi, L.-U., Murch, R.D.: Transmit-preprocessing technique with simplified receivers for the downlink of MISO TDD-CDMA systems. IEEE Trans. Veh. Technol. 53, 285–295 (2004)
6. Yi, S.J., Tsimenidis, C.C., Sharif, B.S.: Transmitter precoding in downlink MC-CDMA systems over frequency-selective Rayleigh fading channels. IEE Proc. Commun. 152, 952–958 (2005)
7. Choi, L.-U., Murch, R.D.: A transmit preprocessing technique for multiuser MIMO systems using a decomposition approach. IEEE Trans. Wireless Commun. 3, 20–24 (2004)
8. Liu, J., Krzumien, W.A.: A space constraint based block Tomlinson-Harashima precoding technique for the multi-user mimo downlink. In: Proc. IEEE. PACRIM, pp. 61–64 (August 2005)
9. Xia, X.-G.: New precoding for intersymbol interference cancellation using nonmaximally decimated multirate filterbanks with ideal FIR equalizers. IEEE Trans. Signal Processing 45, 2431–2441 (1997)
10. Li, L., Gu, G.: Design of optimal zero-forcing precoders for MIMO channels via optimal full information control. IEEE Trans. Signal Processing 53, 3238–3246 (2005)
11. Trees, H.L.V.: Optimum Array Processing. Wiley Interscience (2002)
12. Morelli, M., Sanguinetti, L.: A novel prefiltering technique for downlink transmissions in TDD MC-CDMA systems. IEEE Trans. Wireless Commun. 4, 2064–2069 (2005)
13. Prabagarane, N., Ashwina, Y., Sabrish Karthik, V., Muralidharan, P.: Perrformance of transmitter preprocessing assisted DSTTD over frequency-selective wireless communication channels. In: Proc. IEEE WCNC 2011, Wireless Communications and Networking Conference, WCNC, March 28-31, pp. 1298–1303 (2011)

Multi-dimensional Performance Characterization of Directional Antennas for Applications in Energy Efficient Ad-Hoc Network

C.H. Sandhya, Deepali R. Borade, Rinki Sharma, and Govind R. Kadambi

Department of Computer Engineering, M. S. Ramaiah School of Advanced Studies,
Bangalore, India

Abstract. A simple and elegant mathematical formulation for the analysis of relative improvement of the performance metrics of ad-hoc networks with omnidirectional and directional antenna is presented. Through extensive numerical simulations, the multi dimensional desirable performance attributes of wireless link such as improved range, improved RSS, reduced RF transmit power and consequent reduced consumption of battery power have been analyzed keeping the directional gain of the antenna as a variable parameter. A formulation to compute the required Battery Energy taking into account the data pertaining to the power efficiency of the associated transceiver design as well as the specified link performance parameters is also discussed. Through a case study involving the specifications of a typical transceiver operating in the 2.4 GHz band, the desirable impact of higher gain of a directional antenna in the reduction of RF transmitter power is illustrated. The consequential reduced battery power consumption while still retaining the specified performance parameters of the ad-hoc network like range and Received Signal Strength (RSS) is also demonstrated. This paper also addresses the importance of alignment of beam peaks of directional antennas of a link and the quantification of additional RF power in lieu of Beam Pointing Angle (BPA) error in ad-hoc network. The profile of improved range with directional gain as an independent variable exhibits much sharper feature than an exponential function. The relationship between improvement in the RSS and higher directional gain bears linear characteristics and typical results reveal that for a dB increase in gain ratio, the corresponding improvement in RSS is 2 dBm

1 Introduction

An ad-hoc network is established with multiple mobile nodes, coming in range of each other and exchanging data without an access point. The ability of participating nodes to move around gives rise to Mobile Ad hoc Networks (MANETs). Such networks are very useful for military applications, emergency and rescue operations, health care, home networking and other commercial and educational applications. In a MANET, nodes are highly interdependent for exchange of information with each other. The nodes must cooperate to routing and other services, making it crucial to maximize the network lifetime. The participating nodes in a MANET (laptops, PDA's and sensors) have limited battery power; therefore energy efficient communication approach is critical for the longevity and efficiency of the network. To support

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 575–585.
springerlink.com

energy-efficient communication in an Ad-hoc network, it is important to apply energy-efficient design at multiple layers of the network protocol stack.

In this paper, we concentrate on the physical layer techniques used for energy-efficient communication in an attempt to illustrate the enhanced performance of Ad-hoc networks with directional antenna. Analysis to link the significant design parameters of directional antenna to wireless link performance is of practical significance to appreciate the system design considerations of ad-hoc networks. Through simple and elegant mathematical formulation, relative improvement of the performance metrics of ad-hoc network is analyzed when high gain directional antenna is replaced with conventional low gain omnidirectional antenna. Resource utilization particularly that of electric power of battery and the associated energy is of paramount importance since it has significant say in the outage condition of a wireless link or network. Through a case study involving the specifications of a typical transceiver operating in the 2.4GHz band, the desirable impact of higher gain of a directional antenna in the reduction of RF transmitter power is illustrated. This in turn results in the reduced consumption of battery power consumption, while still retaining the specified performance parameters of the ad-hoc network such as range and Received Signal Strength (RSS). Although, directional antenna offers potentially many desirable performance improvement attributes, it is also associated with additional constraint or requirement of precise or accurate alignment of the directive or main beam of the Transmitter and Receiver. The mismatch of angular alignment of the Transmitter and Receiver results in Beam Pointing Angle (BPA) error, which leads to the degradation in the range. If one desires to regain the desirable ideal network performance despite the presence of BPA error, it would call for additional RF Transmitter power ultimately culminating in the consumption of extra power of the battery that support the RF operation. This paper also facilitates the quantification of additional RF power in lieu of BPA error in ad-hoc network.

To realize an energy efficient ad-hoc network is to minimize the RF transmit power for conservation of battery power. In view of this, this paper emphasizes the application of directional antenna to realize the reduction in transmitter power and yet maintaining the desired link performance or network performance.

2 Directional Antenna in Mobile Ad Hoc Networks (MANETs)

A representative generic MANET with three mobile nodes is shown in Fig. 1 with the dotted circles, implying the communication range of each node. Each of the nodes is assumed to be design configured with a RF transceiver as their network interface. If Node A wishes to communicate with Node C and if Node C is not in communication range of Node A, Node B will serve as an intermediate node resulting in 2 hop communication.

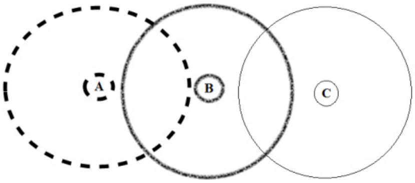


Fig. 1. Multi Hop Communication in MANET

A transceiver of a MANET is configured with an antenna which will serve as its RF front end to establish a wireless communication link with its intermediate (participating) nodes. Basically, the antenna can be of two types namely omnidirectional and directional. Fig. 2 shows the difference between radiation patterns of directional and omnidirectional antenna. It can be seen that the radiation pattern of omnidirectional antenna exhibits uniform angular distribution while the radiation pattern of directional antenna is concentrated in the preferred angle (direction) of communication. In Fig. 2, S and D denote the source and destination nodes respectively. As illustrated in Fig. 2, the communication range exhibited by directional antenna will be higher than an omnidirectional antenna. In view of this, it is implicit that the transceivers of nodes in Fig. 1 are associated with omnidirectional antennas. In the past, omnidirectional antennas did find extensive applicability in MANET system design. With the rapid advancement of antenna technology conjunctured with miniaturization in size and its proven system utility in subscriber end of cellular communication, increasingly one finds rapid progression of directional antenna in ad-hoc networks/MANETs. The characterization of directional antenna to realize performance improvements in ad-hoc networks/MANETs wherein mobility is a key factor causing potentially dynamic and vast variations in communication range, constitutes the core theme of this paper.

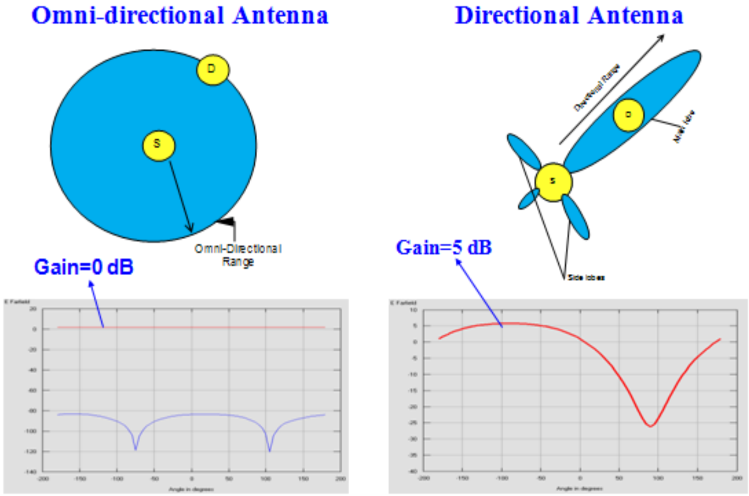


Fig. 2. Radiation Patterns of Omnidirectional and Directional Antennas

3 Related Work on Directional Antenna in MANETs

Most of the papers available in literature propose MAC layer or network layer protocols for ad-hoc networks, using directional antenna. Su Yi et.al [1] provide a good insight on the improvement in capacity of wireless network with the usage of directional antenna. This paper also provides results for throughput improvement with combination of omnidirectional and directional antennas in Transmit (Source) and Receive (Destination) nodes. In [2] Ram Ramnathan provides a good insight on beamforming antennas, and their use in ad-hoc networks. This paper also provides a rough comparison of relative lesser interference potential with: only beamforming, only power control and when both beamforming and power control are used together. The research in [3,4,5] suggests the use of directional antennas for increasing the throughput and enhancing the performance of wireless networks.

However, the above cited research papers do not address a detailed formulation to link the relative improvement in Received Signal Strength (RSS), range and reduction in transmit power when high gain or moderate gain directional antenna replaces the conventional low gain omnidirectional antenna in a MANET system. This paper illustrates the potential of directional antenna to ensure the performance improvement of the communication link relative to omnidirectional antenna. Further, with the directional antennas establishing the link, it is all the more necessary to ensure that the direction of beam (peak gain) of the antenna of the transmit node is aligned with the direction of peak gain of the antenna of the receive node. In case of mismatch in antenna beam pointing angle between the two communicating nodes, there shall be a consequent degradation in the performance of communication link. This paper provides a quantitative analysis of the degradation in performance of the communication link in the presence of misalignment (angular error) between the beam pointing angles of transmit and receive nodes. An estimate of the required increase in the transmit power of the link to avoid the performance degradation and to maintain link quality as with ideal zero beam pointing error case is also provided.

4 Analysis and Relative Comparison of Link Performance with Directional and Omni Directional Antennas

From the joint perspectives of antenna engineering and its relevance to system performance of the links established through participating nodes, the communication range, RSS, Transmitter power to maintain the link performance and resource utilisation particularly with respect to Battery power emerge as more critical parameters. This section provides extensive analysis of the enhancement of the parameters listed above with the use of directional antenna when compared to that of omnidirectional antenna.

4.1 Improvement in Range due to Gain of Directional Antenna

Friis transmission formula is used in wireless communication to calculate the power received by the antenna at the receiver section under idealized conditions. Friis transmission formula is defined through

$$P_r = P_t \times G_t \times G_r \times \left(\frac{\lambda}{4\pi R}\right)^2 \tag{1}$$

Where, P_t, P_r,G_t,G_r,R and λ represent Transmit Power, Receiver Power, Transmit antenna gain, Receive antenna gain, Range and Wavelength respectively.

Let, R_{omni}, RSS_{omni}, P_{Tomni}, λ, G_{Tomni} and G_{Romni} represent Range of Omnidirectional antenna, Transmit power with omnidirectional antenna, Wavelength in meters, Gain of Transmit omnidirectional antenna and Gain of receive omnidirectional antenna respectively.

Then we have,

$$RSS_{omni} = P_{T\,omni} \times G_{T\,omni} \times G_{R\,omni} \times \frac{\lambda^2}{4^2\pi^2 {R_{omni}}^2}$$

$${R_{omni}}^2 = P_{T\,omni} \times G_{T\,omni} \times G_{R\,omni} \times \frac{\lambda^2}{4^2\pi^2 RSS_{omni}}$$

$$R_{omni} = \frac{\lambda}{4\pi} \sqrt{\frac{P_{T\,omni}\, G_{T\,omni}\, G_{R\,omni}}{RSS_{omni}}} \qquad (2)$$

The increase in the communication range ΔR due to directional antenna can be related to the term directional gain ratio g_r, where

$$g_r = G_{DA}/G_{OMNI}$$

$$\Delta R = R_{omni}\,(g_r - 1) \qquad (3a)$$

Effective range $\quad \mathrm{R_{effective}} = \mathrm{R_{omni}} + \Delta\mathrm{R} \qquad (3b)$

In Equation (3), g_r involves $\mathrm{G_{DA}}$ and $\mathrm{G_{omni}}$ where,

$\mathrm{G_{DA}}$ = Gain of Directional Antenna ($\mathrm{G_{DA}} = \mathrm{G_{TDA}} = \mathrm{G_{RDA}}$)

$\mathrm{G_{omni}}$ = Gain of Omnidirectional Antenna ($\mathrm{G_{omni}} = \mathrm{G_{T\,omni}} = \mathrm{G_{R\,omni}}$)

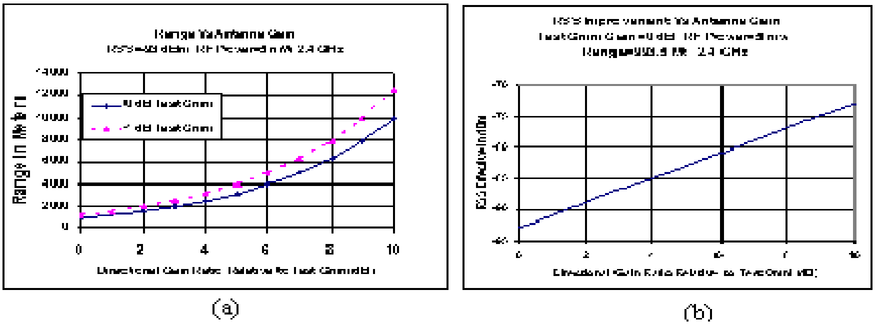


Fig. 3. Influence of Directional Antenna Gain on Communication Range and Improvement of RSS $_{\text{effective}}$ with Directional Gain

Fig. 3(a) illustrates the improvement in the communication range solely due to the gain of the directional antenna relative to omnidirectional antenna keeping the transmit power $\mathrm{P_T}$=5 mW and RSS=-93 dBm to be the same in both cases of directional as well as omnidirectional antennas with the link operating at 2.4 GHz. The profile of increased gain variation exhibits much sharper feature than an exponential function.

4.2 Improvement in Received Signal Strength (RSS) due to Directional Antenna

If the parameter Pr in Equation 1 is considered as RSSomni , with GT omni and GR omni as omnidirectional antenna gain, then we have,

$$RSS_{omni} = P_{T\,omni} \times G_{T\,omni} \times G_{R\,omni} \times \left(\frac{\lambda}{4\pi R_{omni}}\right)^2 \tag{4}$$

The improvement in the RSS with the use of directional antenna can be written as,

$$\Delta_{RSS} = RSS_{omni}\,[(g_r)^2 - 1] \tag{5}$$

The effective RSS at the receiver of the link with the directional antennas replacing the omnidirectional antennas is

$$RSS_{effective} = RSS_{omni} + \Delta_{RSS} \tag{6}$$

Fig.3(b) depicts the variation of realizable improved RSS $_{\text{effective}}$ as a function of the directional gain ratio. Apart from linear nature of variation, the result depicted in Fig. 3(b) also reveals that for a dB increase in gain ratio, the corresponding improvement in RSSeffective is 2 dBm.

4.3 Reduction in Transmitter Power with the Use of Directional Antenna

In this section, we show the reduction in required transmission power with the use of directional antenna, when compared to that required for omnidirectional antenna. Based on the Friis Transmission formula given by Equation 1, we have,

$$P_{T\,omni} = \frac{RSS_{omni}4^2\pi^2R_{omni}^2}{G_{T\,omni}G_{R\,omni}\lambda^2} \tag{7a}$$

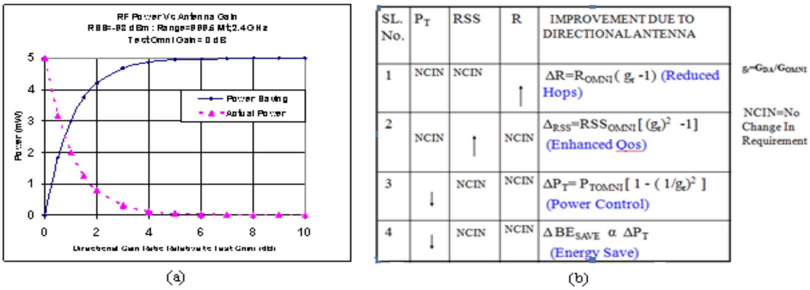
The reduction in the Transmitter power ΔP_T because of higher gain of directional antenna can be expressed as

$$\Delta P_T = P_{T\,omni}\left[1 - \left(1/g_r\right)^2\right] \tag{7b}$$

The definition of Gain Ratio remains the same as in earlier sections. The actual transmit power required by the directional antenna to retain the range and RSSomni of the omnidirectional antenna is given by

$$P_{T\,effective} = P_{T\,omni} - \Delta P_T \tag{8}$$

Fig. 4(a) shows the reduction in the RF Transmitter power with the use of directional antenna keeping the RSS as well as range parameters the same in both the directional and omnidirectional cases. Drastic reduction in transmitter power is evident with smaller initial increase in directional gain and then exhibiting asymptotic nature with higher directional gain.



(a)

SL. No.	P_T	RSS	R	IMPROVEMENT DUE TO DIRECTIONAL ANTENNA
1	NCIN	NCIN	↑	$\Delta R = R_{OMNI}(g_r - 1)$ (Reduced Hops)
2	NCIN	↑	NCIN	$\Delta_{RSS} = RSS_{OMNI}[(g_r)^2 - 1]$ (Enhanced Qos)
3	↓	NCIN	NCIN	$\Delta P_T = P_{TOMNI}[1 - (1/g_r)^2]$ (Power Control)
4	↓	NCIN	NCIN	$\Delta BE_{SAVE}\ \alpha\ \Delta P_T$ (Energy Save)

$g_r = G_{DA}/G_{OMNI}$

NCIN=No Change In Requirement

(b)

Fig. 4. Reduction in RF Transmitter Power with Directional Gain and Multifaceted Advantages of Directional Antenna

4.4 Multidimensional Versatility of Directional Antenna

Fig. 4(b) summarizes the highlights the advantages of using directional antenna. In Fig. 4(b), ΔBE_{Save} denotes the reduction in the energy of battery with the directional antenna to retain or maintain the link performance as in the case of omnidirectional antenna.

With increased communication range (R), the number of hops required for data exchange in an ad-hoc network can be reduced thus reducing the end-to-end latency. Higher RSS increases the signal quality at the receiver, thus providing enhancement in Quality of Service (QoS). With the reduction in transmitter power, the battery energy consumed at each node decreases, thus providing longer battery life, which further increases the network lifetime. Power control also reduces interference among neighboring nodes thus increasing the efficiency of the network.

4.5 Estimation of Battery Energy for Given Data Transmission

Research publications [6,7,8,9] address the issue of need and challenges in the realization of energy efficient ad-hoc/MANET. However, a formulation or a detailed formulation with clearly illustrated steps to compute the battery energy required for a specified data transmission seems to have not been addressed in the literature even though this is very significant from system design perspective. This section summarizes the detailed steps through which one could compute the required battery energy and the formulation shall take into account the data pertaining to the power efficiency of the associated transceiver design as well as the specified link performance parameters.

From the classical relationship between Power, Energy and Time, one can draw analogy to relate the requirement of Battery Power (P_{Bat}) and Battery Energy (BE_{Data}) for transmission of given data.

$$BE_{Data} = P_{Bat} x T_{ON} \tag{9}$$

Where T_{ON} = Transmitter ON Time.

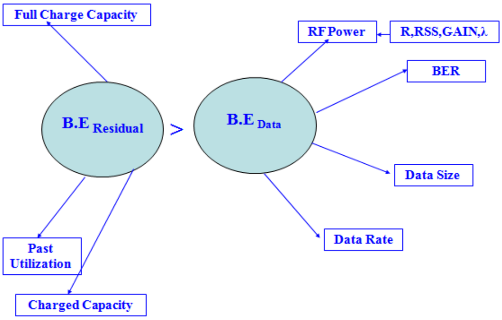


Fig. 5. Requirement of Battery Energy for Data Transmission

T_{ON} is dependent on the size of data, data rate as well as channel condition between the Transmit and Receive ends of the link. At an instant of time for data transmission, the following inequality should be satisfied.

$$BE_{Residual} > BE_{Data} \tag{10}$$

Where, $BE_{Residual}$ denotes the available Battery Energy at that instant for operation. Many research publications dwell with energy efficient routing protocols in ad-hoc/MANET. However, it is difficult to find a formulation to estimate or compute BE_{Data} that has a generic or empirical appeal and which facilitates a better appreciation from system design considerations. In this section, a generic procedure is outlined to systematically compute the BE_{Data}. As illustrated in Fig.5, the Battery Energy is dependent on communication range (R), RF power of transmitter, gain of the transmitting as well as receiving antennas, the expected or threshold RSS at the receive node and wavelength.

For a specified range, RSS, Gain of Transmit and Receive antennas and wavelength, the RF Transmit Power can be computed using equation (11):

$$P_T = \frac{RSS\, 4^2\, \pi^2 R^2}{G_T\, G_R \lambda^2} \tag{11}$$

The relationship between P_T and Battery Power P_{Bat} is very specific to the RF design of the Transceiver module of the ad-hoc/MANETs as well as the type of battery used in it. However, the data sheet of RF transceiver usually provides the relationship between P_T and P_{Bat} in tabular form. Either a look up table or an algebraic expression deduced from the tabulated data is required to estimate the P_{Bat} for a specific PT determined through Equation (11). Such recourse of computing the Battery Power P_{Bat} shall further facilitate built in provision for computation of Battery Energy BE_{Data} through Equation (9). Fig.6 illustrates the least square curve fitting equation obtained to relate the RF Power P_T and Battery Power P_{Bat} for a typical RF transceiver widely used in 2.4 GHz band. Satisfactory correlation between the results of derived functional relation and the reference data is noticed in Fig.6.

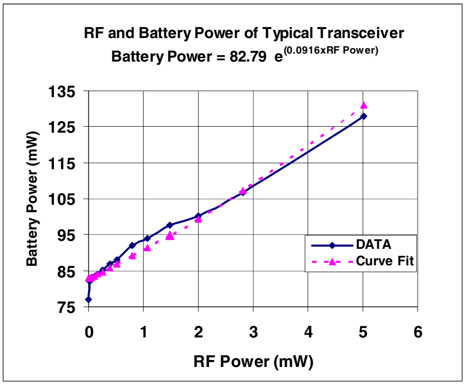


Fig. 6. Derivation of Functional Relation between RF and Battery Power

4.6 Effect of Beam Pointing Angle Error of Directional Antenna on Network Performance

In order to retain all the network performance enhancements highlighted thus for, one has to also address the importance of alignment of direction of peak gain of the transmitting beam of the source node with the corresponding receiving beam of the destination node. Misalignment of the beam peaks of the source and destination nodes result in a parameter defined as a Beam Pointing Angle (BPA) Error or Beam Pointing Error (BPE) as shown in Fig. 7. In many of the aperture antennas with moderate gain, the shape of the radiation pattern up to -10 dB points is assumed to follow functional distribution $e^{-p\theta^2}$.

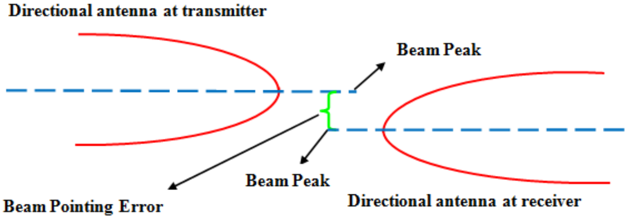

Directional antenna at transmitter
Beam Peak
Beam Peak
Beam Pointing Error
Directional antenna at receiver

Fig. 7. Illustration of Beam Pointing Angle Error of Directional Antenna

The value of the constant 'p' can be derived with the plot of the radiation pattern at say -3 dB or -10 dB points relative to the beam maximum angle. This procedure has been adopted to analyse the effect of BPA or BPE on the link performance metrics such as range and the RF power requirement as well as Battery Power. The degradation in the range performance as a function of BPA is shown in Fig.8 (a). As expected, the degradation becomes more pronounced with increase in BPA.

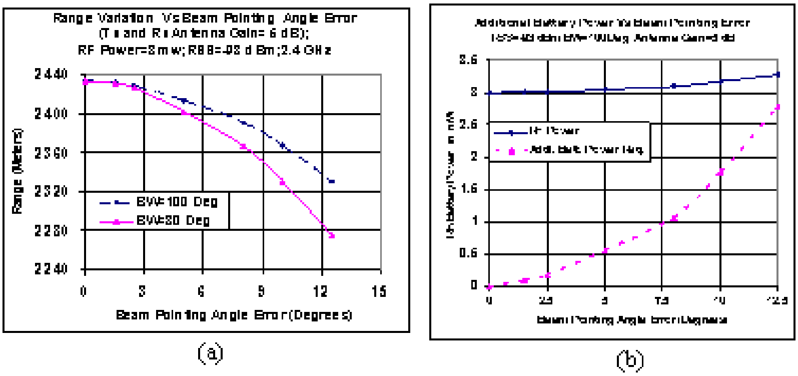


Fig. 8. Influence of Beam Pointing Angle Error on Network Performance and Effect of Beam Pointing Angle Error on RF and Battery Power

If one desires to regain the ideal link performance despite the presence of BPA or BPE, it can be realized only at the expense of increased RF transmit power. This in turn would result in additional consumption of battery power resulting in reduced operational time of the link. Fig. 8(b) explains the undesirable influence of BPA on the link transmitter power and which in turn has a negative effect on the required battery power to sustain the link performance.

5 Conclusion

A relative improvement of the performance metrics of ad-hoc network with omnidirectional and directional antennas is presented. The multiple desirable performance attributes of wireless link such as improved range, improved RSS, reduced RF transmit power and consequent reduced consumption of battery power have been analyzed, keeping the directional gain of the antenna as a variable parameter. A detailed formulation is described for specified data transmission as well as importance of alignment of beam peaks of the Transmit and Receive antennas to harness the potential advantages of directional antenna has also been demonstrated. The contributions derived out of this paper will be of potential utility to address the realistic design challenges and issues pertaining to wireless network in general and energy efficient ad-hoc network in particular.

References

[1] Yi, S., Pei, Y., Kalyanaraman, S.: On the Capacity Improvement of Ad Hoc Wireless Networks using Directional Antennas. In: Mobihoc, pp. 108–116 (June 2003)

[2] Ramnathan, R.: On the Performance of Ad Hoc Networks with Beamforming Antennas. In: Mobihoc, pp. 95–105 (2001)

[3] Nasipuri, A., Li, K., Sappidi, U.R.: Power Consumption and Throughput in Mobile Ad Hoc Networks using Directional Antennas. In: IEEE International Conference on Computer, Communication and Networking (2003)

[4] Spyropoulos, A., Raghavendra, C.S.: Energy Efficient Communications in Ad Hoc Networks Using Directional Antennas. In: INFOCOM 2002 (November 2002)

[5] Arora, A., Krunz, M., Muqattash, A.: Directional Medium Access Protocol (DMAP) with Power Control for Wireless Ad Hoc Networks. In: Global Telecommunication Conference (2004)

[6] Yahya, B., Ben-Othman, J.: Robust and Energy Efficient Multipath Routing Protocol for Wireless Networks. In: Global Telecommunication Conference, November 30, pp. 1–7 (2009)

[7] Vidhyapriya, R., Vanathi, P.T.: Energy Efficient Adaptive Multipath Routing for Wireless Sensor Networks. International Journal of Computer Science, 1–9 (May 2006)

[8] Yu, C., Lee, B., Youn, H.Y.: Energy Efficient Routing Protocols for mobile ad hoc Networks. In: Wireless Communications and Mobile Computing, September 14, pp. 959–973 (2003)

[9] Olagbegi, B.S., Meghanathan, N.: A Review of Energy Efficient and Secure Multicast Routing Protocols for Mobile Ad Hoc Networks. International Journal on Applications of Graph Theory in Wireless ad hoc Networks and Sensor Networks, 1–15 (June 2010)

Reliability Enhanced Routing Protocol for Wireless Mesh Networks

Rakesh Matam and Somanath Tripathy

Department of Computer Science and Engineering
Indian Institute of Technology Patna
Patna, Bihar-800013
India
{m.rakesh,som}@iitp.ac.in

Abstract. Wireless mesh network (WMN) has become a popular technology to provide broadband Internet access mainly due to its low up-front cost and less administration overhead. A variety of routing protocols are proposed for WMN focussing on higher throughput and performance. On the other hand, determining reliability of an established route is a critical issue because of varying channel conditions, interference and byzantine node-behavior, which has not been considered in most of the existing routing protocols. This leads to degradation in performance. In this paper, we propose a mechanism to enhance the reliability of a routing protocol in WMN evaluating the quality of links. The proposed mechanism is simulated over HWMP (the default path-selection protocol for WMN) and its performance is evaluated.

1 Introduction

Wireless mesh networks have emerged as a promising technology for next-generation wireless networking, capable of providing wireless broadband Internet access to its clients. WMN significantly lower the deployment cost and administrative overhead by replacing majority of the wired infrastructure there by forming a desirable solution to build networks in a campus, office and community [1]. A typical WMN consists of a set of mesh routers (MR) and mesh clients (MC). Mesh routers are wireless routers equipped with one or more wireless interfaces with very less or no mobility. A group of MRs along with few MR's that are assigned gateway functionality to connect to the Internet form the backbone of WMN. The main aim of the mesh backbone is to provide access services to its clients with the help of routing protocols. A mesh client can be any device capable of networking which can be either stationary or mobile.

WMN need to support reliable and high throughput Internet applications including audio and video streaming to achieve the desired level of success. Majority of these applications require reliable transfer of data. So, routing protocols play an important role in meeting these requirements. Unlike routing protocols for wired networks, WMNs have to deal with open-air shared communication medium, dynamic and time-varying operating environment and limited resource

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 587–596.
springerlink.com

availability. There are various factors that cause unreliability in WMNs. Consequently, it is desirable and necessary to develop efficient routing protocols that are capable of recovering from such adverse conditions and ensure reliability.

The existing routing protocols for WMN are based on different link-layer metrics that aim at capturing various such properties of a link that usually account for majority of packet losses in wireless network. Protocols based on metrics such as expected transmission count (ETX)[2], expected transmission time (ETT) [3] and many other such metrics such as WCETT [4], ATLM [5], per-packet RTT [6] were proposed replacing hop-count. Majority of the above specified metrics depend on active probing of neighbors to measure the quality of wireless links. Broadcast probe packets does not always generate same quality measurement as data transmissions due to PHY layer settings (e.g. Modulation). Moreover they are further susceptible to packet losses resulting from the communication grey zone problem [9] and unaccounted node behavior. Therefore, there is a necessity for a mechanism to ensure reliability in data communication.

This paper proposes a mechanism to increase the reliability of the existing routing protocol HWMP (Hybrid wireless mesh Protocol). We define reliability as the acceptable user performance under network failures in dynamic conditions, i.e. no disruption of network service and low packet loss ratio. The proposed mechanism heightens the reliability. Our focus of work is to develop a mechanism that accounts for packet losses due to various factors such as errors induced by the channel, multipath fading, shadowing, interference and independent byzantine node-behavior. The proposed mechanism depends on the ability of a node to verify the link conditions and passively measure the level of reliability offered by the network. Based on the obtained results, the routing protocol can further select better routes in terms of reliability, bandwidth and end-to-end delay.

The rest of the paper is organized as follows. Section 2 presents the related work on routing in WMN. Section 3 presents our network model and design considerations. We present our proposed mechanism to enhance the reliability of a routing protocol for WMN in Section 4. Section 5 presents the working of reliability enhanced HWMP. Section 6 presents the analysis of the proposed reliability mechanism. Section 7 presents the simulation results and finally Section 8 concludes the paper.

2 Related Work

Recently, lot of research has been carried out to improve the performance of routing protocols for WMN. The main design goal of these routing protocols is throughput maximization. Therefore, majority of the research has been carried out in design of routing metrics that increase the overall throughput offered by the network [2,3,4,5,6,7]. Metrics such as expected transmission count (ETX), expected transmission time (ETT), weighted cumulative ETT (WCETT), per-hop round-trip time (RTT) and airtime link metric (ATLM) have been designed to exploit various characteristics of a wireless link. The computation of these metrics involve active probing of neighbors to measure the quality of links. As an example, to compute ETX, each node broadcasts fixed size probe packets periodically (for e.g., one second). A node remembers the number of probes received during the last window of w seconds. Based on the obtained probe information, a node determines the ETX value of a link.

Even though active probing by broadcasting incurs less overhead, it has its own set of limitations. Mainly, it does not always generate the same quality measurements due to different PHY setting (e.g., modulation) [8]. The other major limitation is the generation of bi-directional links. This is due to the use of identical type of probing mechanism in both the directions which results in ignoring asymmetric behavior of a link. Broadcast active probing involving Hello messages which introduces the problem of communication gray zone [9]. Unicast probing achieves better results over a broadcast mechanism, as it resembles unicast data transmission, but incurs very high overhead. Even though the above mentioned link-layer metrics achieve much better performance over hop-count metric, they still suffer from packet loss resulting from unaccounted factors for eg. communication gray zones.

The other important factor that cannot be accounted through probing mechanism is the byzantine behavior of nodes. Such nodes may co-operate during metric computation but may not involve in forwarding process or intentionally drop packets to disrupt network operations. Few works [15,16] have also been done to particularly address the selfish behavior of nodes in WMN by employing trust/reputation in routing process. AODV-REX [10] is a routing framework that extends the functionality of ad hoc on-demand routing protocol with a reputation model proposed in [11].

Expected forwarding counter (EFW) is a cross layer routing metric that addresses the problem of selfish participants in WMN [12]. It integrates the forwarding behavior of a node with the ETX metric to derive a new metric called EFW. Kone et.al. propose a quality of service (QoS) routing protocol for WMN that ensures the selected paths to satisfy minimum bandwidth and maximum end-to-end delay that can be tolerated by an application [13]. Sen et. al. propose a similar kind of protocol that satisfies the desire QoS requirements in terms of bandwidth and end-to-end delay [14]. Both these protocols depend on control packets to measure reliability of a link that does not capture actual link behavior to employ in data transmissions.

Majority of the existing work aims to achieve better performance in terms of throughput by designing different kinds of routing metrics. Even though these metrics assure higher throughput, they are still constrained by other factors such as dynamic link conditions and byzantine node behavior. Protocols proposed by Kone et.al. and sen aim at improving reliability of selected routes as part of QoS requirements. These protocols depend on control packets to measure the reliability of a link that gives inaccurate results [9]. Therefore, there is necessity for a mechanism to verify the actual performance of selected routes and ensure that the selected routes are in fact reliable.

3 Network Model and Design Considerations

We consider a typical WMN architecture, where a set of MR's form the backbone of the WMN, out of which few MR's designated as gateways are connected to the Internet. MC's are typical wireless clients connected to specific MR's with access point functionality. As, hybrid wireless mesh protocol (HWMP) has been chosen as the mandatory path-selection protocol in IEEE 802.11s draft standard [5], we consider it as the candidate routing protocol operating on the mesh backbone.

3.1 Design Considerations

The following design considerations are taken into account in the proposed scheme to enhance reliability.

Determining Reliability. We use a two-hop reporting mechanism that allows each node to determine the reliability of individual links along a path with the help of report packets generated by a two-hop neighbor. The report packet contains the actual number of data packets received through an intermediate node for a particular transmission, which is used to compute the reliability of a link. The reliability factor at instance $t+1$ denoted by $R(t+1)$ is computed using exponential weighted cumulative average (EWMA) as

$$R_{t+1} = \alpha * R_t + (1-\alpha) * R_{t-1}. \tag{1}$$

A node maintains the reliability value for each of its neighbors and updates it at the end of a transmission session. This reliability value is employed in the route selection process by integrating it with the underlying routing metric (airtime in case of HWMP). The final route selection decision is based on a cumulative metric value that is a combination of measured reliability and advertised airtime. This integration process results in more robust routing metric that guarantees desired level reliability over a selected route.

Estimating End-to-End Delay. We estimate the end-to-end delay of a routing path with the help of employed air-time metric. The airtime or channel access time C_a is the amount of channel resources consumed for transmitting a test frame on a particular link, which can be computed as

$$C_a = [O + \frac{B_t}{r}]\frac{1}{1-e_f} \tag{2}$$

Where O is the channel access overhead, which includes frame headers, training sequences, access protocol frames etc. B_t is the number of bits in the test frame, it has a constant value of 8192 bits. The rate 'r' represents the data rate at which the mesh STA would transmit a frame of standard size B_t based on current conditions. Its estimation is dependent on local implementation of rate adaptation. The frame error rate e_f is the probability of transmission error when a frame of standard size B_t is transmitted at the current transmission bit rate 'r'. Its estimation is also a local implementation choice. No additional overhead is incurred in estimating the end-to-end delay of a routing path, as it is computed during normal metric computation.

Identifying Malicious Nodes. A malicious node can selectively or completely drop packets to preserve resources or disrupt network services. The reliability scheme is designed to avoid such malicious nodes during route selection process. Any kind of selfish or malicious packet dropping activity of a node is reported by its neighbor through a two-hop report packet. Since, the behavior (and performance) of each node is passively monitored by its preceding and succeeding node, any malicious packet drop can be easily tracked. This in turn is reflected in the computation of reliability that influences the route selection process.

4 The Proposed Reliability Enhancement Mechanism

The proposed mechanism operates in two phases to enhance the reliability of a routing protocol. They are discussed below.

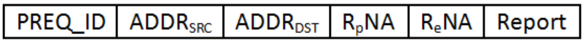
4.1 Maintaining Two-Hop Neighborhood Information

A beacon frame is extended to facilitate gathering and maintaining of two-hop neighborhood information. In addition to usual beacon contents, the additional neighborhood information is appended to the beacon frame and transmitted. A set flag is used to indicate the presence of two-hop neighborhood information. Whenever a node receives a beacon, the two-hop information is retrieved and added it to the two-hop neighbor table to facilitate accurate processing of report messages.

4.2 Estimating Reliability

Reliability of a route is estimated with the help of a report packet, generated by each node on an active route and propagated towards the source. The REPORT packet contains the fields as shown in Fig. 1. The $PREQ_{ID}$, $ADDR_{SRC}$ and $ADDR_{DST}$ fields are used to uniquely identify a REPORT packet. The $ADDR_{SRC}$ and $ADDR_{DST}$ fields are the addresses of source and destination. The major contents of the REPORT packet are addresses of node on which the report packet is generated R_eNA, address of the reporting node R_pNA and the Report field containing the number of packets received in the current active routing session. To prevent endless traversing of the REPORT packets, its TTL value is set to 2.

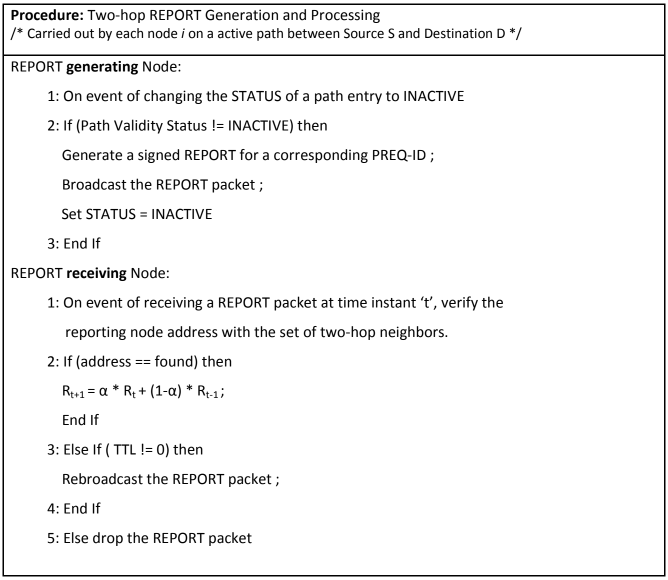
Each node generates a REPORT packet immediately after the routing session expires (5.02 sec according to the IEEE 802.11s draft standard). The two-hop report generation and processing mechanism is shown in Fig.2. On receiving a



PREQ_ID	$ADDR_{SRC}$	$ADDR_{DST}$	R_pNA	R_eNA	Report

Fig. 1. Report Packet Element

REPORT packet, a node first verifies the address of the sender with the set of two-hop neighbors. If the REPORT is generated by an unregistered node, it ignores such REPORT. The information received through REPORT packet is used to determine the reliability of a link by computing the percentage of packets successfully delivered to the percentage of packets actually sent. The obtained information is employed in computing the reliability of a link at time instance *t* by combining it with computed reliability information at instance *t-1* using an exponential weighted cumulative average (EWMA) mechanism. This computed reliability information is later employed to enhance the reliability of routes selected by a routing protocol.



```
Procedure: Two-hop REPORT Generation and Processing
/* Carried out by each node i on a active path between Source S and Destination D */

REPORT generating Node:
    1: On event of changing the STATUS of a path entry to INACTIVE
    2: If (Path Validity Status != INACTIVE) then
         Generate a signed REPORT for a corresponding PREQ-ID ;
         Broadcast the REPORT packet ;
         Set STATUS = INACTIVE
    3: End If
REPORT receiving Node:
    1: On event of receiving a REPORT packet at time instant 't', verify the
          reporting node address with the set of two-hop neighbors.
    2: If (address == found) then
         $R_{t+1} = \alpha * R_t + (1-\alpha) * R_{t-1}$ ;
         End If
    3: Else If ( TTL != 0) then
         Rebroadcast the REPORT packet ;
    4: End If
    5: Else drop the REPORT packet
```

Fig. 2. Two-hop report generation and processing

5 Reliability Enhanced HWMP (RE-HWMP)

The above proposed mechanism is employed on HWMP the default routing protocol for WMN, to enhance its reliability. The path selection process of RE-HWMP is similar to that of HWMP. The metric computation process is modified to include the reliability of individual links along a path. The proposed mechanism depends on HWMP to accurately determine the address of two-hop reporting node (the node from which a report packet is expected). In a reactive protocol like HWMP, where nodes maintain next-hop information to identify routes, a node usually does not have any information about the two-hop neighbor involved in the routing process. But, this information can be obtained by a node operating in promiscuous mode to overhear at least one of the transmitted data packets, using which it can exactly determine the address of two-hop reporting node.

A REPORT packet is identified with the help of (the fields PREQ-ID and address of source and destination STA's) the PREQ. The major contents of the report packet are addresses of node on which the report packet is generated, reporting node address and the Report field containing the number of packets received in the current active session. To prevent endless traversing of the REPORT packets, the TTL value is set to 2. Each node generates only one report packet for an entire communication session, that is differentiated with the help of a unique PREQ-ID, and is transmitted immediate after expiring of the routing session (default value is 5.02 sec).

5.1 Ensuring Reliability of the Routing Protocol

Consequent upon receiving a REPORT packet, each node determines the reliability of a link by computing the percentage of packets successfully delivered with respect to the packets actually sent. The reliability value of a link is integrated with the airtime metric to determine the effective metric of a link (ERM). We employ an integration mechanism similar to that proposed in [10]. The modified metric is incorporated in the path request (PREQ). This metric computation process is repeated by each intermediate node along a path to obtain a cumulative path metric which is a combination of reliability and actual airtime metric. The routing decision is based on the PREQ that offers a better overall metric. This integration process ensures that the links with high reliability are preferred.

6 Efficiency of the Proposed Mechanism

The performance of the proposed reliability enhancement mechanism mainly depends on the ability of a node to accurately generate, transmit and process report packets. Each node generates and transmits a REPORT packet to report the performance of its immediate neighbor (on the reverse path) towards the source. As, the proposed mechanism cumulatively accounts for losses incurred from various sources collectively, the content reported in the report packet heavily influences the routing decision. Here in, we show how the proposed mechanism handles abnormal situations in handling and processing report packets.

– **Fabricated Reports:** Nodes explicitly sign the report packets to prevent alteration of its content. This, explicit signing of report packets prevents fabricating report messages.

- **Falsified Reports:** Malicious nodes involving in data forwarding process can generate false reports to degrade the reliability of a link. Fortunately, the honest reports generated by other nodes in the network balance the overall reliability.
- **Receipt of Multiple Report Packets:** Malicious nodes can generate fake report packets to disrupt the smooth computation of reliability. For example, a malicious node X can generate a report packet on a node Y even though it is not a forwarder on the selected path. Hop-by-hop routing protocols like HWMP cannot distinguish fake to genuine report packets as they maintain only next-hop information. To avoid such a scenario, a node can monitor data transmissions of a neighbour and learn the address of the next hop node involved in the data forwarding process. This monitoring process also allows nodes to identify malicious activity in their neighbourhood.
- **Intentional Dropping of Report:** Malicious nodes can drop report packets to prevent nodes from updating the reliability of a node. This intentional dropping is effectively handled by the self-healing feature of WMN. The successful delivery of a REPORT packet is mainly attributed to the employed broadcast mechanism and the existence of multiple paths between a pair of nodes.
- **Non-generation/Lost Report Packets:** To handle the case of genuine loss of report packets resulting from various network factors, we follow no action approach for a particular transmission session. The reliability of a link is not updated for a particular session to ensure the stability of the network. Malicious nodes involved in the data forwarding process can also not generate REPORT packets, but an active adversary trying to disrupt network stability does not gain by avoiding report generation.

7 Simulation Results

Performance of RE-HWMP is simulated in ns-3 discrete event simulator [17]. We considered a WMN backbone consisting of 24 MR's that are randomly placed over a 600m by 200m area. The IEEE 802.11s MAC protocol is employed with a channel data rate of 54Mbps. The data packet size considered was 512 bytes. Source and destinations were randomly chosen and total simulation time was set to 500 seconds. UDP traffic was chosen to simulate the performance of our proposed reliability enhanced HWMP (RE-HWMP) along with HWMP.

Performance of RE-HWMP is compared with HWMP, in terms of packet delivery ratio (PDR). PDR is calculated as the ratio of packets received by the destination to the packets sent by the source. The Fig. 3 shows that RE-HWMP performs better than HWMP. As, co-operative behavior of nodes is assumed in this experiment, in certain situations both RE-HWMP and HWMP yield identical performance.

Further, we repeated the same experiment by randomly assigning selfish and malicious behavior to few of the nodes. With the increasing number of nodes exhibiting selfish and malicious behavior, the PDR of HWMP falls consistently,

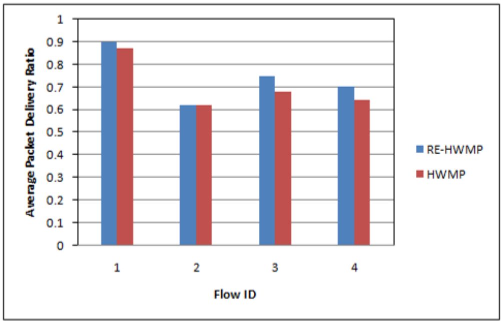


Fig. 3. Packet Delivery Ratio of Different Flows

where as RE-HWMP achieves better performance even in such adverse conditions. Fig.4 shows the behavior comparison of RE-HWMP with HWMP. The reason behind this enhanced performance is that the proposed reliability scheme ensures to avoid selfish and malicious nodes from the route selection process by preferring links with higher reliability. Any kind of selfish or malicious activity of a node is reported by its neighbor through a two-hop report packet. In a way the behavior and performance of each node is passively monitored by a preceding and succeeding node and any malicious packet drop can be easily measured. Any kind of malicious packet drop is reflected in the computation of reliability which influences the route selection.

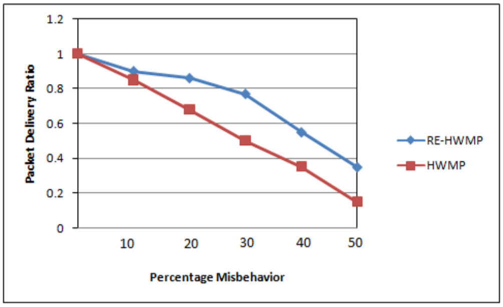


Fig. 4. Packet Delivery Ratio in Presence of varying Number of Malicious Nodes

8 Conclusion

In this paper, we proposed a routing protocol called RE-HWMP to enhance the reliability of selected routes using HWMP. The proposed mechanism uses two-hop REPORT packets to determine the reliability of the links along a selected route. The computed reliability metric is then employed in the route selection process to ensure higher reliability routes. The simulation results confirm the improved performance of RE-HWMP over HWMP that chooses unreliable links, grey zones and erratic node behavior.

References

1. Akyildiz, I.F., Wang, X., Wang, W.: Wireless mesh networks: A survey. Computer networks and ISDN systems (2005)
2. De Couto, D., Aguayo, D., Bicket, J., Morris, R.: A High-throughput path metric for multi-hop wireless routing. In: Proc. of 9th Annual International Conference on Mobile Computing and Networking, MobiCom (2003)
3. Draves, R., Padhye, J., Zill, B.: Comparision of Routing Metrics for Static Multi-Hop Wireless Networks. In: Proc. of ACM SIGCOMM, pp. 133–144 (2004)
4. Draves, R., Padhye, J., Zill, B.: Routing in Multi-Radio, Multi-Hop Wireless Mesh Networks. In: Proc. of 10th Annual International Conference on Mobile Computing and Networking, MobiCom (2004)
5. IEEE P802.11s/D5.0 Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, Amendment 10: Mesh Networking
6. Yang, Y., Wang, J., Kravets, R.: Designing Routing Metrics for Mesh Networks. In: IEEE Workshop on Wireless Mesh Networks (September 2005)
7. Subramanian, P., Buddhikot, M.M., Miller, S.C.: Interference Aware Routing in Multi-Radio Wireless Mesh Networks. In: IEEE Workshop on Wireless Mesh Networks (September 2006)
8. Kim, K.-H., Kang, G.S.: On accurate Measurement of Link Quality in Multi-Hop Wireless Mesh Networks. In: Proc. of MobiCom 2006, pp. 246–257 (2010)
9. Lundgren, H., Nordstrom, E., Tschudin, C.: THe Gray Zone Problem in IEEE 802.11b Based Ad Hoc Networks. Proc. of MC2R 6(3), 104–105 (2002)
10. Oliviero, F., Romano, S.: A reputation-based metric for secure routing in wireless mesh networks. In: Proc. of IEEE GLOBECOM 2008, New Orleans, LA (December 2008)
11. Oliviero, F., Peluso, L., Romano, S.P.: Refacing: an autonomic approach to network security based on multidimensional trustworthiness. The International Journal of Computer and Telecommunications Networking 52(4) (October 2008)
12. Paris, S., Rotaru, C.N., Martignon, F., Capone, A.: EFW: A Cross-Layer Metric for Reliable Routing in Wireless Mesh Networks with Selfish Participants. In: Proc. of IEEE INFOCOMM 2011, Shanghai, China (2011)
13. Kone, V., Das, S., Zhao, B.Y., Zheng, H.: Quorum: Quality of Service in WIreless Mesh Networks. Journal of Mobile Networks and Applications 12(5), 358–369 (2007)
14. Sen, J.: An Efficient and Reliable Routing Protocol for Wireless Mesh Networks. In: Taniar, D., Gervasi, O., Murgante, B., Pardede, E., Apduhan, B.O. (eds.) ICCSA 2010. LNCS, vol. 6018, pp. 246–257. Springer, Heidelberg (2010)
15. Liu, K., Deng, J., Varshney, P.K., Balakrishnan, K.: An acknowledgment based approach for the detection of routing misbehavior in MANETs. IEEE Transactions on Mobile Computing 6(5), 536–550 (2007)
16. Yau, P., Mitchell, C.: 2HARP: A Secure Routing Protocol for Mobile Ad Hoc Networks. In: Proc. of the 5th World Wireless Congress, WWC 2004 (2004)
17. The NS-3 Discrete Event Simulator, http://www.nsnam.org

DARIH: Distributed Adaptive Routing via Information Highway in Sensor Network

Monomita Mazumdar, Srimanta Halder, Prasenjit Chanak,
and Indrajit Banerjee

Department of Information Technology
Bengal Engineering and Science University, Shibpur, Howrah, India
{mazumdar.monomita,srimanta86,prasenjit.chanak}@gmail.com,
ibanerjee@it.becs.ac.in

Abstract. Wireless sensor network is one of the most important fields of research in this century. This technology is used in environmental monitoring, health monitoring, vehicle tracking system, military surveillance etc. Sensor nodes are able to sense the environment, perform computations and communicate in a limited range. In this paper, we are proposing distributed adaptive routing via information highway in sensor network (DARIH) where data reaches to base station via traffic nodes. The proposed algorithm uses dynamic traffic node selection method to distribute load among nodes and balances the energy dispersion of sensor nodes. The adaptive routing protocol selects routing path based on weight of the path. By this way it prolongs the lifetime of the sensor network in large scale. The routing scheme is fault tolerant as it replaces faulty node quickly with a fresh node. This algorithm also takes special care for the nodes near base station. For multi-hop routing, nodes near base station deplete their battery sooner due to relay traffic. DARIH provides traffic group of smaller size near base station to solve this problem.

Keywords: Wireless sensor network (WSN), cluster head (CH), base station (BS), traffic group, traffic node.

1 Introduction

A huge number of low-power sensors deployed randomly create Wireless Sensor Networks (WSNs). Each sensor monitors environmental condition like temperature, pressure, sound, vibration etc and passes the sensed data through the network to base station (BS) [1]. Sensor nodes have limited energy and it is impossible to recharge their battery because most of the time they are deployed in harsh environment [2]. Wireless sensor networks do not have any fixed network infrastructure and they are dynamic most of the time. The topology of the network changes very frequently due to node failures, damages, energy depletion etc. So our main goal is to provide an energy efficient, distributed and adaptive routing technique. Our proposed algorithm works on the platform of cluster-based network. Cluster based protocols are very energy efficient. Here, cluster head accumulates data from its cluster nodes, aggregates and compresses the data and then sends it to base station [3]. In this paper,

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 597–607.
springerlink.com

modified cyclic cellular automata scheme is used for cluster formation and management which is done without message passing [4]. In Cyclic Cellular Automata (CCA) [5], [6], every cell changes it states depending on its own state and state of its neighbor cell. In modified cyclic cellular automata, state of a cell changes depending on state of its nine neighbor cell [4]. DARIH introduces traffic group that works on cluster based topology. A traffic group has a leader node that is known as traffic node. Traffic node routes data to base station [4]. Here, we propose dynamic traffic node selection method to distribute load among sensors. As we are using multi-hop routing, traffic nodes near BS get loaded with huge relay traffic and tend to die faster (i.e., the hot spots problem [7]). DARIH proposes smaller traffic groups near BS to solve the problem. Smaller traffic groups consume less energy during intra traffic group communication, thus it can save some energy for relay traffic. Data packets are routed to BS by forwarding data from upper level traffic node to lower level. The routing path is selected adaptively based on distance and residual energy of the path.

The paper is organized as follows: Section 2 covers related work; section 3 describes DARIH framework; section 4 present DARIH algorithm; section 5 reports simulation results and finally section 6 concludes the paper.

2 Related Work

In popular hierarchical routing protocols, CHs are responsible for intra cluster data processing and inter cluster traffic forwarding as well. So they deplete their battery rapidly. In traffic node routing scheme, CHs hand over the data to traffic nodes. Now traffic node transmits the data to base station via information highway. Information highway is constructed by traffic nodes present in the network. Hence lifetime of CHs increase. In [4], energy efficient routing via information highway (EERIH) is first investigated as traffic node routing. Here, a traffic node is selected for every pair of cluster heads. In EERIH, traffic node is selected at a position which is minimum distance from a pair of CH and base station as it expects direct transmission most of the time. But direct transmission is not preferable for long-distance transmissions. It consumes huge energy compared to multi-hop transmission and results faster death of traffic nodes far away from BS. But multi-hop network has weakness of hot spots problem. For multi-hop routing, shortest path should not be the only criteria. The nodes in shortest paths drain their power quickly. Considering paths with maximal residual energy also as a criterion, helps to maximize network lifetime. The objective of DARIH is to design unequal traffic group formation, dynamic traffic node selection and multi-hop adaptive routing. Unequal traffic group formation will solve the hot spots problem whereas dynamic traffic node selection and multi-hop routing technique will extend network lifetime.

3 DARIH Framework

In WSN, data transmission energy loss per bit is represented by E_t [4]:

$$E_t = (\alpha_1 + \alpha_2 \gamma^n)\beta \tag{1}$$

Where α_1 is energy consumption by transmitter circuit, α_2 is amplifier energy, γ is distance between sender and receiver node, n is the power index for channel path loss of antenna and β is number of bits transmitted. So transmission energy loss (E_t) is proportional to distance between transmitter and receiver (γ) i.e. $E_t \propto \gamma$. Data receiving energy loss is represented by E_r [4]:

$$E_t = \alpha_3 \beta \tag{2}$$

Where α_3 is receiver circuit energy loss and β is number of bit received.

DARIH assumes the following idea:

i) The sensor network consists of large number sensor nodes that are deployed randomly and sensors transmit data to base station.
ii) All the sensor nodes are static and initially they all have same energy.
iii) A node can act as a traffic node, cluster head or cluster node. Responsibility changes over time.

Definition 1: Neighbor: The 1-hop neighbor set of sensor i is defined as [8],

$$N(i)= \{j \in \aleph | \, d(i, j) \leq R, i \in \aleph, j = i\},$$

where ℵ represents the set of nodes in the sensor network, d(i, j) denotes the distance between sensor i and sensor j, and R is the radius of the sensing range.

Traffic node selection scheme

1. **For traffic group of two CHs:** A and B are two neighbor nodes. Traffic node is located at a point P, where P is the midpoint of straight line AB.
2. **For traffic group of three CHs:** A, B and C are three neighbor nodes in the sensor network. A, B and C form a triangle and traffic node is located at point P. We have to find the point P such that, sum of the distance from P to A, B and C is a minimum. This point is known as Fermat point.\

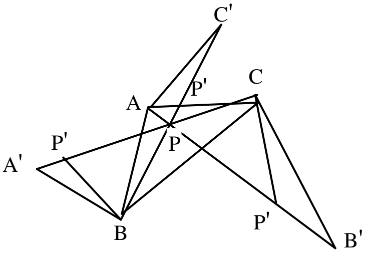


Fig. 1. Fermat point solution

We select a point P in triangle ABC (Fig. 1) and connect three vertices with P. Then rotate triangle APB 60° around vertex B. Distances are preserved under a rotation. So, AP = A'P' and BP = B'P' and ∠PBP' = 60°. Triangle PBP' is equilateral as ∠PBP' = 60° and BP = B'P'. We form a straight line A'C using A'P' + PP' + CP = AP + BP + CP. We will have the minimum distance when the segments form a straight line. Thus the point that will create this minimum distance must lie somewhere on this line. In the same way we can rotate triangle APC and BPC and get BC' and AB'. A'C, BC' and AB' intersect in a point P. So P is the Fermat point such that AP + BP + CP is a minimum.

Three major steps in DARIH protocol:

i) **Leveling [9]:** BS broadcasts LVL=0 message. All neighbor node of BS receives this message and make their LVL=1.The nodes who have already set their LVL, broadcast LVL message further to set the LVL of other nodes.

ii) **Clustering and traffic group formation:** Clustering is done using modified cyclic cellular automata [4]. A traffic group is formed by including two/three clusters that are closer to each other. Each cluster node reports to its CH. CH transmits data to the traffic node of its traffic group. Our proposed model creates traffic group of unequal size to solve the hot spots problem. Traffic groups near BS are of smaller size than those that are far away from BS (Fig. 2). For level-1 nodes, a traffic group includes two clusters whereas it includes three for other levels. Traffic nodes near base station consume lower energy during intra traffic group data processing and preserve more energy for inter traffic group data forwarding. Density of traffic node also increases near BS than the region far away from BS (Fig. 3). Thus huge load near BS can be distributed properly among traffic nodes.

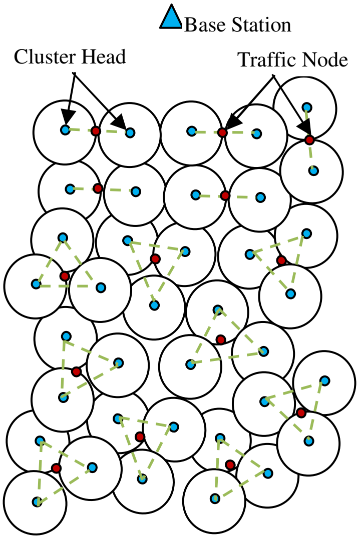


Fig. 2. Traffic node position is selected by DARIH

iii) **Data routing:** Since only traffic nodes are involved in inter traffic group communication, our algorithm forms a graph with traffic nodes. A multi-hop routing technique is used to forward traffic from upper level to lower level (Fig. 3).

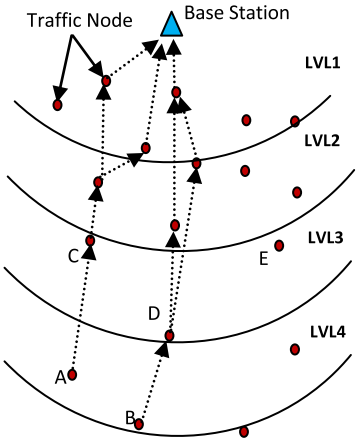


Fig. 3. One of the multiple paths is selected based on routing algorithm

4. DARIH Algorithm

DARIH algorithm has three parts: traffic group formation, traffic node selection and data routing from traffic nodes to BS.

1. Traffic group formation and updation

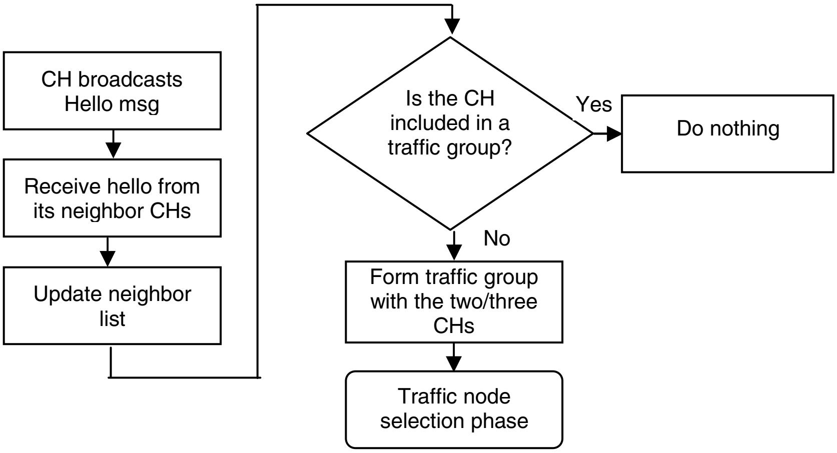

CH broadcasts Hello msg
Receive hello from its neighbor CHs
Update neighbor list
Is the CH included in a traffic group?
Yes
Do nothing
No
Form traffic group with the two/three CHs
Traffic node selection phase

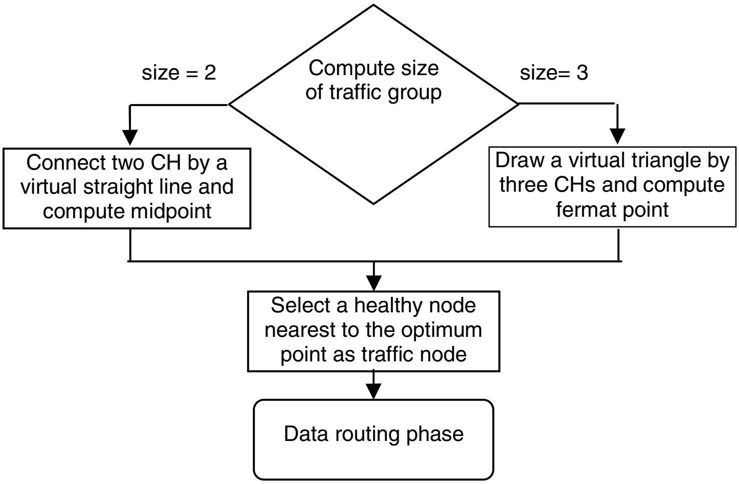
Fig. 4. Workflow of traffic group formation process

Initially traffic groups are formed (Fig. 4) as discussed in Algorithm 1. Members of group are updated when cluster head changes. A situation may occur where a CH cannot find a companion to form a traffic group. In this case, the CH will contact a neighbor traffic node to relay data.

Algorithm 1: Traffic group formation algorithm

```
FOR each CH DO
    Broadcast Hello msg with ID and position to 1-hop neighbor CHs set
    Receive hello from its neighbor CHs
    Update neighbor list and estimate distance from the neighbor CHs
    IF CH is not included in a traffic group THEN
        Compute level of the CH
        IF Level = 1 THEN
            Search the closest neighbor CH of LVL1 (which is not in
            a group) from the neighbor list
        ELSE IF Level > 1 THEN
            Search two closest neighbor CHs of same LVL (which
            are not in a group) from the neighbor list
        END IF
        IF one/two neighbor found which are not in a group THEN
        Form traffic group with the two/three CHs
            Set status for the CHs as already in a group
        END IF
    END IF
END FOR
```

2. Dynamic Traffic Node Selection



Compute size of traffic group
size = 2
size= 3
Connect two CH by a virtual straight line and compute midpoint
Draw a virtual triangle by three CHs and compute fermat point
Select a healthy node nearest to the optimum point as traffic node
Data routing phase

Fig. 5. Workflow of dynamic traffic node selection process

After traffic group formation, traffic node is selected (Fig. 5). Healthy node is the one whose energy is beyond threshold energy. DARIH checks periodically if all the traffic nodes are healthy. If not then new traffic nodes are selected dynamically. The old traffic nodes become cluster member node. Traffic nodes consume their energy faster for long data transmission. The proposed algorithm distributes the load of traffic nodes among sensor nodes by periodical selection of new traffic nodes.

3. Data routing from traffic node to base station

Traffic node forwards the data to its lower level traffic nodes using routing process shown in Fig. 6. To initialize routing table, every traffic node advertises its node id, level id and its residual energy (RE) to its neighbors. Traffic node receives advertisement from all the neighbors. Traffic nodes form routing table with help of distance vector routing where distance from neighbor (d) and the residual energy of neighbor (RE) are used as routing metrics. Neighbor's weight is calculated by : k1.d + k2.RE. So path length is not the only criteria, path with higher residual energy will also get priority.

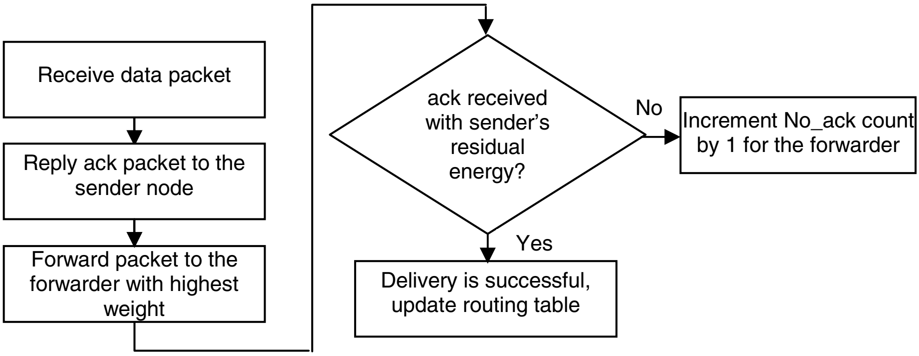

Receive data packet
Reply ack packet to the sender node
Forward packet to the forwarder with highest weight
ack received with sender's residual energy?
No
Increment No_ack count by 1 for the forwarder
Yes
Delivery is successful, update routing table

Fig. 6. Workflow of multi-hop, adaptive traffic node routing process

The proposed multi-hop, adaptive traffic node routing is discussed next.

According to equation of energy consumption due to data transmission (equation. 1), shortest path should be selected for minimum communication cost. Routing algorithm also considers the path with maximal residual energy. It will balance the energy consumption among nodes. Considering shortest path only as routing metric may result faster death of some nodes due to heavier relay traffic. On the other hand, considering residual energy only may unnecessarily waste total energy of network. As multiple paths are available from a source to BS, if any node fails in the optimal routing path, then another path is selected dynamically.

Algorithm 2: Multi-hop, adaptive traffic node routing algorithm

```
FOR each traffic node DO
    IF data packet received to the traffic node THEN
        Reply ack packet with its energy to the sender
    END IF
    IF traffic node available at lower level THEN
        Find neighbor with highest weight from lower level
    ELSE IF traffic node available at the same level THEN
        Find neighbor with highest weight from the same level
    END IF
    IF neighbor found THEN
        Forward packet to the neighbor
        Wait for ack packet from forwarder
    IF received ack packet with residual energy THEN
            Update routing table accordingly
        ELSE
            Increment No_ack count by 1 for the forwarder
            IF No-ack count is 2 THEN
                Delete the entry
            END IF
        END IF
    END IF
END FOR
```

5 DARIH Result

This section shows how much improvement we get using DARIH Routing. All the simulation is done by C programming in Linux platform. Each sensor has the following attributes: an identification number that uniquely identifies a node in the sensor network, a level number, state of the node that indicate if the node is alive or dead, residual energy of the node. The node is declared as dead when its residual energy is zero. The simulation parameters are given in Table 1.

Data travelling time: Time required to send a packet from source to BS.

Packet delivery ratio: Ratio of the number of packet received by base station to the number of packet sent from different sensors.

Table 1. Simulation Parameters

Parameter	Value
Network Size	150×150
Number of nodes	22500
Data Packet Size β	800 bit
Initial Energy	0.5 J
Energy consumed in the transmitter circuit α_1	50 nJ/bit
Energy consumed in the amplifier circuit α_2	10 pJ/bit

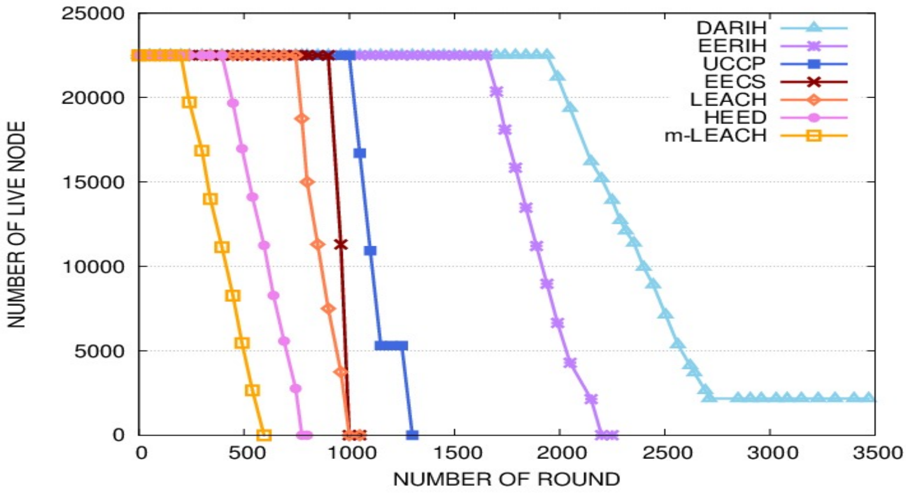


Fig. 7. Number of live node over time

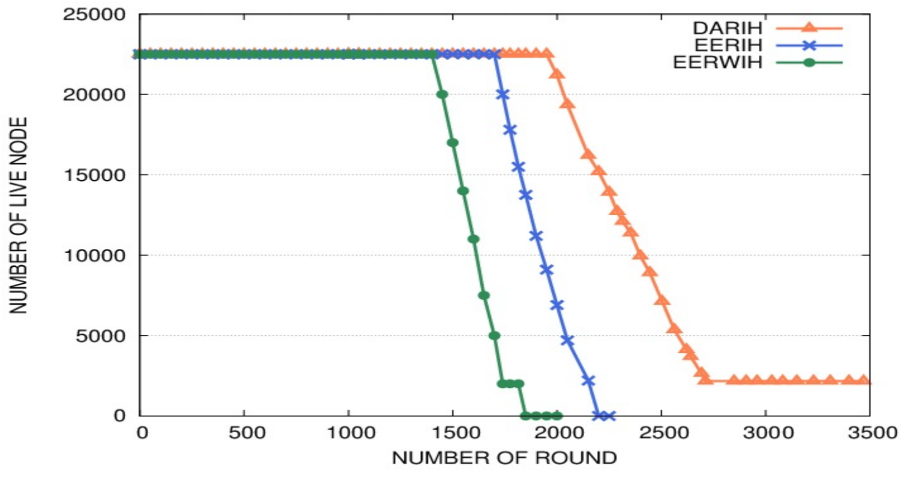


Fig. 8. Network lifespan for DARIH, EERIH and EERWIH

The simulation result in Fig. 7 shows lifespan of the sensor network. DARIH maximizes the lifetime of sensor network by delaying death of nodes using proper load balancing with help of dynamic traffic node selection. In DARIH the time until the first node dies is increased by 1.89 times compared to UCCP [10], 2.56 times compared to LEACH[11], 4.45 times compared to HEED [12] and 8.29 times compared to m-LEACH[3]. Comparison with EERIH [4] shows that, in DARIH the time of first node death is increased by 1.17 times where as network lifespan (the time until the last node dies) is increased by 1.23 times. After round 2700 the sensor network becomes disconnected though some nodes are still alive.

Fig. 8 shows the lifespan of senor network using DARIH, EERIH and without traffic node (EERWIH). The network lifetime is increased by 22.73% compared to EERIH [4] and 45.94% compared to EERWIH.

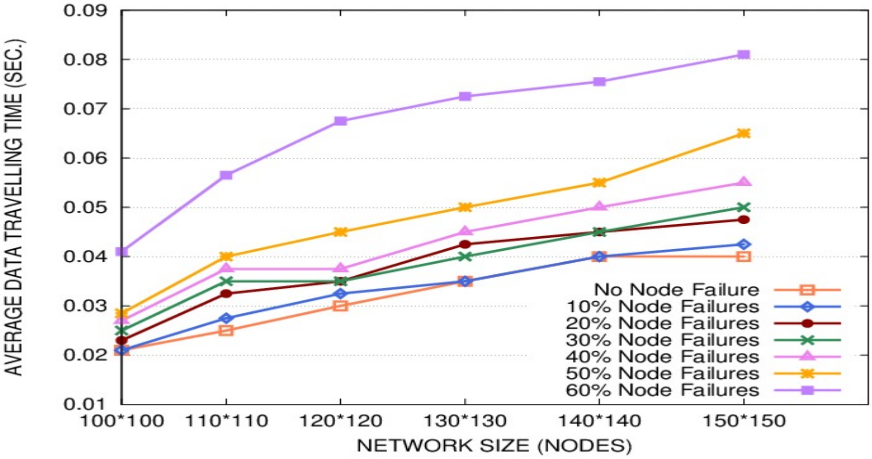


Fig. 9. Average data travelling time according to node failure

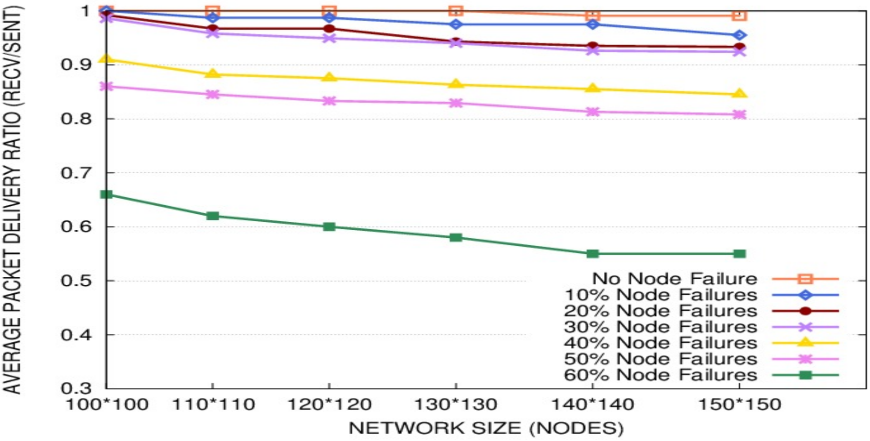


Fig. 10. Average packet delivery ratio with node failure

Fig.9 shows that data traveling time for different network size as well as different percentage of node failure. As DARIH scheme can tolerate node failure, so transmission time increases slowly with increment of node failure. DARIH technique can tolerate node failure up to 50%. However, data delivery time is hugely increased when node failure is more than 50%.

Fig.10 shows the packet delivery ratio for different percentage of node failure. Average packet delivery ratio does not affect much with 50% node failure for DARIH technique. But the simulation result shows that for 60% node failure, only 60% transmitted data is received by base station.

6 Conclusion

This paper gives idea of distributed, adaptive routing using traffic node. As most of the data transmission is done by traffic nodes, CHs have very small responsibility in data transmission. Thus CHs save their energy and increase their lifespan. Again dynamic traffic node selection method replaces an old traffic node with new one when its energy becomes lower than threshold energy. So information highway is constructed dynamically. DARIH provides energy efficient, multi-hop, adaptive routing technique to select routing path depending on weight metric. The simulation result shows that DARIH improves the network lifetime compared to EERIH.

References

1. Banerjee, I., Chanak, P., Sikdar, B.K., Rahaman, H.: EER: Energy Efficient Routing In Wireless Sensor Network. In: IEEE International Technology Symposium 2011, IIT kharagpur, India, January 14-16 (2011)
2. Hsieh, H.-C., Leu, J.-S., Shih, W.-K.: A fault tolerant scheme for an autonomous local wireless sensor network. Computer Standards & Interfaces 32(4), 215–221 (2010)
3. Nguyen, L.T., Defago, X., Beuran, R., Shinoda, Y.: An Energy Efficient Routing Scheme for Mobile Wireless Sensor Networks. IEEE 978-1-4244-4/08 (2008)
4. Banerjee, I., Chanak, P., Sikdar, B.K., Rahaman, H.: EERIH: Energy Efficient Routing via Information Highway in Sensor Network. In: IEEE International Conference on Emerging Trends in Electrical and Computer Technology, Kanyakumari, India, March 23-24 (2011)
5. Alfonseca, M., Orttega, A.: Representation of some cellular automata by means of equivalent L-systems. Complexity International 7 (1999)
6. Shalizi, C.R., Shalizi, K.L.: Quantifying Self-Organisation in Cyclic Cellular Automata. Proceedings of SPIE, vol. 5114 (2003)
7. Chen, G., Li, C.F., Yeo, M., Wu, J.: An Energy-Efficient Unequal Clustering Mechanism for Wireless Sensor Networks. In: IEEE International Conf. Mobile Adhoc and Sensor Systems, p. 8 (November 2005)
8. Tian, D., Georganas, N.: A coverage-preserved node scheduling scheme for large wireless sensor networks. In: Proc. of First Intl' Worshop on Wireless Sensor Networks and Applications, pp. 32–41 (2002)
9. Banerjee, I., Rahaman, H., Sikdar, B.K.: UDDN: Unidirectional Data Dissemination via Negotiation. In: IEEE International Conference on Information Networking 2008, Pusan, Korea, January 23-25 (2008)
10. Aslam, N., Phillips, W., Robertson, W.: A Unified Clustering and Communication Protocol for Wireless Sensor Network. IAENG International Journal of Computer Science, 35, 3, IJCS_35_3_01 (August 01-21, 2008)
11. Heinzelman, W.R., Chandrakasan, A., Balakrishnan, H.: Energy Efficient Communication Protocol for Wireless Micro sensor. In: IEEE Proceedings of the 33rd Hawaii International Conference on System Sciences (2000)
12. Younis, O., Fahmy, S.: Distributed Clustering in Ad-hoc Sensor Networks: A Hybrid. Energy-efficient Approach (September 2002)

A Comparative Study of Cache Replacement Policies in Wireless Mobile Networks

Preetha Theresa Joy and K.Polouse Jacob

Cochin University of Science and Technology,
Kochi, Kerala, India
preetha@mec.ac.in, kpj@cusat.ac.in

Abstract. Data caching can remarkably improve the efficiency of information access in a wireless ad hoc network by reducing the access latency and bandwidth usage. Cache replacement policy plays a vital role to improve the performance in a cached mobile environment, since the amount of data stored in a client cache is small. In this paper we have made a comparative study of the existing cache replacement algorithms for wireless mobile networks .Cache replacement policies proposed for cooperative caching in ad hoc networks are also reviewed and attempts to classify existing replacement policies for ad hoc networks based on the replacement decision. In addition, this paper suggests some alternative techniques for cache replacement. Finally, the paper concludes with a discussion on future research directions.

Keywords: Cache Replacement, MANET, Cooperative caching Wireless Networks.

1 Introduction

Wireless mobile communication is a fastest growing segment in communication industry. It has currently supplemented or replaced the existing wired networks in many places. The wide range of applications and new technologies simulated this enormous growth. The new wireless network will support heterogeneous traffic, consisting of voice, video and data. There exists two different ways of configuring a mobile network, infrastructure based and ad hoc based. The former type is most prominent, as it is used in wireless LANS and global wireless networks. An infrastructure based wireless network uses fixed network access points with which mobile terminals interact for communication and this requires the mobile terminal be in the communication range of a base station. The ad hoc based network structure alleviates this problem by enabling mobile terminals to cooperatively form a dynamic network without any pre existing infrastructure. It is much convenient for accessing information available in local area and possibly reaching a WLAN base station, which comes at no cost for users.

Mobile terminals available today have powerful hardware, but the capacity of the batteries goes up slowly and all these powerful components reduce battery life. Therefore adequate measures should be taken to save energy. Communication is one of the major sources of energy consumption. By reducing the data traffic energy can

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 609–619.
springerlink.com

be conserved for longer time. Data caching has been introduced as a techniques to reduce the data traffic and access latency. By caching data, the data request can be served from the mobile clients without sending it to the data source each time. It is a major technique used in the web to reduce the access latency. In web, caching is implemented at various points in the network. At the top level web server uses caching, and then comes the proxy server cache and finally client uses a cache in the browser.

1.1 Cache Replacement

When the cache is full, an object has to be removed from the cache to make room for the data that has to be brought in .While it would be possible to pick a random object to replace when cache is full, system performance will be better if we choose an object that is not heavily used. If a heavily used data item is removed it will probably have to be brought back quickly, resulting in extra overhead. So much work has been done on the subject of cache replacement.

Caching in wireless environment has unique constraints like scarce bandwidth, limited power supply, high mobility and limited cache space. Due to the space limitation, the mobile nodes can store only a subset of the frequently accessed data. The availability of the data in local cache can significantly improve the performance since it overcomes the constraints in wireless environment. A good replacement mechanism is needed to distinguish between the items to be kept in cache and that is to be removed when the cache is full. The extensive research on caching for wired networks can be adapted for the wireless environment with modifications to account for MT limitations and the dynamics of the wireless channel. These limitations include the MT's limited battery life and its small cache size.

This paper provides a general comparison of the cache replacement policies in wireless mobile networks based on the criteria used for evicting documents. We reviewed the various replacement policies for wireless networks with more focus on function based and location based policies. The different policies used in ad hoc networks are also reviewed. The topic of caching in ad hoc networks is rather new, and not much work has been done in this area. We classified the replacement policies for MANETs in to two groups uncoordinated and coordinated. In coordinated replacement policy the mobile nodes which forms cooperative cache collectively takes the replacement decision. In the later case the data item to be evicted is determined independently by each node based on its local access information. Alternative techniques for cache replacement are also proposed.

1.2 Performance Metrics

Caching in wireless networks deals with data items of different costs and sizes. Performance measures used should consider this non uniformity. A definite performance ranking of the different replacement strategies is not possible as there is no best strategy for different workload situations. It is not possible to identify a best replacement policy as different schemes uses different optimization strategy.

The most typical measures used to analyze the cache replacement policy are hit ratio ,byte hit ratio and delay savings ratio[18]. Let S_i be the size of the data item i, C_i the cost of fetching the data item i into the cache, R_i the total number of references made to data item i, H_i the number of hit references made to data item i and D_i is the delay time to fetch the data item i from the original data source to cache. Cache hit ratio defines the number of references made from the cache over the total number of references. It is a metric used in traditional caching systems like operating systems and database which handles data of uniform size, which may not be reliable metric for data items with varying size and cost. Byte hit ratio represents the number of bytes saved from retransmission by using the cache over the total number of bytes referenced. Delay savings Ratio represents the reduced latency by using a cache over the total latency when cache is not used. Due to the inconsistency in download time due to traffic variations, performance results based on this metric may vary.

$HR = \sum Hi / \sum Ri$
$BHR = \sum Si.\ Hi \,/ \sum Si.\ Ri$
$DSR = \sum Di\ Hi / \sum Di.\ Ri$

2 Cache Replacement Policies in Wireless Networks

Efficient replacement schemes for wireless mobile environments should consider different parameters like data access pattern, access costs, mobility pattern, connectivity, bandwidth, update rates, location dependence of the data. Most of the replacement algorithms form a value function by combining these parameters and evicts the data with minimum value. This section discusses some of the function based replacement policies in wireless environment.

Yin and Cao [1] proposed a generalized cache replacement policy for mobile environment. The value function they proposed can be used for different performance metrics and they considered minimum query delay and minimum download traffic as the target. The value function was based on parameters like probability of reference, cost of fetching data item, cost of validation, probability of invalidating cached data item and cost of getting updated data item to the cache. Based on these parameters the algorithm replaces a data item with min Value (i) / S_i, where S_i is the size of the data item. Here a strong consistency model is assumed. Xu and Lee [2] proposed a gain based replacement policy SAIU, for on demand broadcasts. The gain function for each data item is calculated as gain (i) = L_i. A_i / S_i. U_i where L_i is data retrieval delay, A_i is the access rate, S_i is the size of the data item and U_i is the update frequency.

Another algorithm proposed by Zeitunlian and Haraty [3] uses a least unified value cache replacement for SACCS, scalable asynchronous cache constituency scheme. Here the replacement is based on the reference information of the object, fetch cost and size. They considered the complete reference history for finding the probability of reference in the future. The book keeping involved in this method is too high. Chem. and Xiao [4] presented a cache replacement policy called on bound selection which used both data access and update information for replacement decision. The above mentioned schemes uses a function based policy. Since the relative importance of these parameters can vary from one type of request to another, some policies are needed to adjust the weights dynamically to achieve the best performance. Table 1 gives the summary of function based replacement policies.

2.1 Location Based Cache Replacement Policies

In Location Dependent information services (LDIS) the value of the data item depends on the location and varies as the user changes his location. The factors that are considered in a location aware replacement policy are the valid scope area, distance and direction of client movement. The area under which the data item is valid is the valid scope area. Distance is the distance between mobile node's current location and the valid scope area. When the data is distant from the valid scope area, it will have a lower chance to become useful. Direction indicates the direction of data movement from the valid scope area. The data that are moving in the opposite direction of the valid scope area will be irrelevant after sometime.

The cache replacement policy that supports location dependent services was early proposed by [5] (Manhattan) .Here the replacement was based on the Manhattan distance, which is the distance between the location of each cached data item's origin location and a mobile client's current location .The data items having the highest Manhattan distance are replaced. The only parameter considered for replacement is the distance.

The FAR (Farther Away Replacement) [6] replacement policy considers the current location and direction of the mobile client to make the replacement decision. The replacement strategy is based on the fact that the data which are not in the moving direction and farthest away from the user won't be visited in the near future. Based on the direction of movement, the data is arranged as two sets, In –Direction and Out-Direction. Whenever we want to replace data the Out Direction set is considered first, when it is empty the furthest segment in the In Direction set will be replaced. FAR considers only the spatial properties for cache replacement and the temporal properties are not taken.

In [7] two cache replacement policies PA and PAID are proposed. In this replacement policy a cost function is formed by considering the parameters access probability, valid scope area and data distance. Valid scope area refers to the geometric area of the valid scope of a data value. When this area is broad there is a higher chance that the client will request the data. In PA the cost function is formed as the product of access probability and valid scope. In PAID in addition to the above mentioned parameters data distance is also considered. The data with low access probability, a small valid scope area, and a long distance is evicted first.

K .Lai at el designed and implemented [8] mobility aware replacement scheme (MARS) which uses a cost function which consists of a clients location, movement of direction and access probability. The data item with lowest value for cost function is removed first. They also proposed an extension to this, The MARS+ tries to keep the clients movement patterns and from this history the future location of the client can be predicted. This is incorporated in to the replacement cost function and more accurate replacement decisions are made.

A network distance based cache replacement policy (ND – CRP) introduced by [9] considers the network distance which is the shortest path from current location of the mobile client (P) to a point of interest Pi for data eviction. Access probability and network density are the other factors considered in the replacement policy. This algorithm assumes that when the network density is high there is more chance to remain in that area for a long time. Dijkstra's algorithm is used to find the shortest path from the single source to single destination. The policy would choose the data with less access probability, less network density and greater network distance for eviction.

Prioritized Predicted Region based Replacement Policy (PPRRP) [10] tried to get the benefit of both temporal and spatial property in one unified scheme. In their scheme the distance is calculated based on a predicted region, where the client can be in the near future. In this policy instead of taking the direction of client's movement they predict an area in which the client will be in the near future. The data item cost is calculated based on the access probability, valid scope area, data size in cache and distance of data based on the predicted region. Table 2 summarizes the various location based replacement policies.

3 Cache Replacement Policies in Ad Hoc Networks

Data caching in MANET is mostly proposed as cooperative caching. In cooperative caching the local cache in each node is shared among the adjacent nodes and they form a large unified cache. So in a mobile cooperative caching environment, the mobile hosts can obtain data items not only from local cache but also from the cache of their neighboring nodes. This aims at maximizing the amount of data that can be served from the cache so that the server delays can reduced which in turn decreases the response time for the client. In many applications of MANET like automated highways and factories, smart homes and appliances, smart class rooms, mobile nodes share common interest. So sharing cache contents between mobile nodes offers significant benefits.

Cache replacement algorithm plays a central role in response time reduction by selecting suitable subset of data for caching. The available cache replacement mechanisms for ad hoc network can be categorized in to coordinated and uncoordinated depending on how replacement decision is made. In uncoordinated scheme the replacement decision is made by individual nodes. In order to cache the incoming data when the cache is full, replacement algorithm chooses the data items to be removed by making use of the local parameters in each node. Effective caching schemes in mobile environments should ideally consider proper cache admission control, consistency maintenance and replacement. Cache admission control decides whether the incoming data is cacheable or not. Substantial amount of cache space can be saved by proper admission control, which can be utilized to store more appropriate data, thereby reducing the number of evictions. If a node doesn't cache the data that adjacent nodes have it can cache more distinct data items which increase the data availability. Another feature of coordinated replacement is that the evicted data may be stored in neighboring nodes which have free space.

Table 1. Summary of Function based Cache Replacement Policies

Algorithm	Parameters Considered	Eviction	Performance measure	Advantage	Disadvantage
Target Based	Reference Probability, cost of fetching data validation cost, probability of invalidating cached data item, cost of getting updated data.	Value is calculated using the parameters considered and replaces data with min value by size.	Average delay, Average downlink traffic	Can be used for multiple targets. Considered data updations.	Too many parameters to consider. How to select the target is not specified.
SAIU	Data retrieval delay, access probability, size,update frequency.	Low access rate, low delay and maximum sized data	Cache Hit Ratio, Strech	Uses a new performance metric	Parameters considered are not easily available
LUV -SACCS	Access frequency, recency, fetch cost, size	Smaller size, low access frequency, low cost	Cache hit ratio, Total Delay	Relates cache replacement with consistency	Book keeping is high, usage of a fixed parameter
On Bound Selection	Access frequency, update frequency	Low access frequency, high update frequency	Cache Hit Ratio, Communication cost	Stale documents are evicted increases hit ratio	Not useful for short term access

Table 2. Summary of Location based Cache Replacement Policies

Algorithm	Parameters Considered	Eviction	Performance measure	Advantage	Disadvantage
Manhattan	Manhattan distance	Lowest Distance	Response time, Network traffic	Supports location dependent queries	Single parameter. Difficult to find estimated weights
FAR	Distance and movement direction of clients	Data in the out direction set is evicted first then the farthest in the indirection	Average Response time	Considers the direction of client motion and future movements	Not taken temporal properties. Ineffective when client changes its direction frequently.
PA	Access probability and valid scope area	Low access probability, minimum valid scope area	Cache Hit Ratio	Considers temporal property	Objects close to the client are often replaced as their valid scope area is smaller
PAID	Distance between the current location and valid scope, Access Probability, valid scope area	Low access probability, minimum valid scope area, maximum distance	Cache Hit ratio	Considers temporal and spatial property	Considers only the clients current movement direction
MARS	Client location, movement direction, access probability, update and query rate	Low temporal score and spatial score	Cache Hit ratio	Temporal and spatial properties are taken along with update frequency	Fails to recognize regular client movement patterns

In the following section we discuss various uncoordinated cache replacement policies for mobile ad hoc networks.

LRU

LRU (Least Recently Used) is based on the observation that data that have been heavily used recently will probably be heavily used again in the future. Conversely, data that have not been used for ages will probably remain unused for a long time. In LRU when cache is full the data item that has been unused for the longest time has been thrown out. It is a widely used algorithm in cache replacement. Logically, the cache consists of a list with most recently referenced data being in the front of the list. When a data is referenced it is moved from its existing position to the front of the list. When a new data comes in it is placed on the top of the list and the data at the back end is removed. LRU doesn't take in to account the non uniformity in the size of data, which is an important factor in mobile communication as the cost to fetch the data depends on size.

LRU Min

LRU Min [11] is a variant of LRU that tries to minimize the number of documents replaced. It is similar to LRU in implementation but will consider size of the data during replacement. In this scheme the data is arranged on the basis of access time and if a data item of size S needs to be cached it will search for items least recently accessed with size greater than S. If there is'nt any data in cache with size S, we start removing the items with size greater than S/2 and then objects of size S/4 until enough cache space is created. LRU Min policy will increase the hit ratio of smaller sized data items.

SXO

This is a local replacement policy[12] which considers the parameters data size and access frequency for replacement .Here larger sized data items are removed first as they occupy more cache space. More cache space can be made available by replacing bigger objects. The second parameter considered is order(di) which gives the frequency of access of data .Here replacement is done by combining the two parameters as value(di)=S*order(di).The advantage of this scheme is that the parameters used are easily available. But recently accessed data are not given any privilege.

LUV

A cache replacement policy based on least utility value (LUV) has been used in [13] For computing the LUV of a data item the access probability (Ai), size of the data item (Si), coherency which can be known by TTLi field and distance (δ) between the mobile client and data source were considered. Eq. for *utility*i function for a data item (di) is:

$$utilityi = Ai \,.\, TT\,Li\; \delta i\,/Si$$

3.1 Coordinated Cache replacement Policies

TDS

The cache replacement [14] is based on two parameters distance (D) which is measured as the number of hops and access frequency. As the network is mobile the value of distance (d) may become obsolete .So the value is chosen based on the time at which it is last updated. The T value is obtained by the formula $1/t_{cur}$- t_{update}. Distance is updated by looking at the value of T. Based on how the distance and time is selected three different schemes are proposed TDS_D,TDS_T and TDS_N.TDS_D considers distance as the replacement criteria. If two data items have the same distance least value of (D+T) is replaced. In TDS_T the replacement decision is made by selecting the data with lowest T value. In the third scheme product of distance and access frequency is considered. In these algorithms TDS_D has the lower success rate and TDS_T has the higher hit ratio.

LUV Mi

This replacement scheme [15], has two parts replacement and migration. The replacement decision is based on a utility value formed by combining the parameters access probability, distance, Size and Coherency. In the migration part the replaced data is stored in the neighboring nodes which have sufficient space. For migration the data with highest utility value is given preference. Here even though the replacement decision is made locally migration is a coordinated operation. In order to save the cache space the data item is cached based on the location of the data source. If it is from the same cluster the data is not cached. The limitation of this scheme is that no checking is done whether the data is already present in the migrating node.

ECORP

Energy efficient Cooperative cache Replacement Problem (ECORP) [16] is an energy efficient cache replacement policy used in ad hoc networks. They considered the energy cost for each data access. For this, they considered the energy for in zone communication, energy for sending the object, energy for receiving and energy cost for forwarding the object. Based on this they proposed a dynamic ECORP DP and ECOPR _greedy algorithms to replace data. The neighboring nodes will not cache the same data item in its local cache which reduces the redundancy and increases hit ratio.

Count Vector

In this scheme [17], each data item maintains a count which gives the number of nodes having the same data. Whenever the cache is full data item with maximum count is removed first as this will be available in the neighboring nodes. Whenever a data item is removed from the cache the access count will be decremented by one. Initially when the data is brought in to cache the count is set to zero.

4 Discussion and Future Work

Most of the replacement algorithms used in ad hoc networks are LRU based which uses the property of temporal locality. This is favorable for MANET which is formed for a short period of time with small memory capacity. Frequency based algorithms will be beneficial for long term accesses. It is better if the function based policies can adapt to different workload condition. In these schemes if we are using too many parameters for finding the value function, which are not easily available the performance can be degraded. Most of the replacement algorithms mentioned above uses cache hit ratio as the performance metric. In wireless network the cost to download data item from the server may vary. So in some cases this may not be the best performance metric. Schemes which improve cache hit ratio and reduce access latency should be devised. In cooperative caching coordinated cache replacement is more effective than local replacement since the replacement decision is made by considering the information available in the neighboring nodes. The area of cache replacement in cooperative caching has not received much attention. Lot of work needs to be done in this area to find better replacement policies.

Location dependent services are becoming popular in ad hoc networks. Replacement policies which consider location dependent parameters should be devised for cooperative caching in ad hoc networks. Another area of research in ad hoc networks is semantic caching in which the query is served from the cache based on the semantic description and results of previous queries. Cache admission control also plays role in improving the performance of cooperative cache. Value based admission control can be incorporated to minimize the number of replacements. Cache replacement based on Quality of Service (QOS) parameters can be explored. An alternative to cache replacement is that the data items that have their Time to Live (TTL) expired can be removed as the data becomes stale and cannot be used. So periodical checks can be done to delete the data items with TTL expired.

5 Conclusions

In this paper we made a general comparison of the major replacement policies in wireless networks and summarized the main points. Numerous replacement policies are proposed for wireless networks, but a few for cooperative caching in ad hoc networks. We also summarized the operation, strengths and drawbacks of these algorithms. Finally we provided some alternatives for cache replacement and identified topics for future research.

References

[1] Yin, L., Cao, G., Cai, Y.: A Generalized Target-Driven Cache Replacement Policy for Mobile Environments. In: Proceedings of SAINT, pp. 14–21 (2003)

[2] Xu, J., Hu, Q.L., Lee, L.: SAIU: An efficient cache replacement policy for wireless on demand broadcastsin. In: Proceedings of the 9th ACM International Conference on Information and Knowledge Management (CKIM 2000), McLean, VA, USA, pp. 46–53 (November 2000)

[3] Zeitunlian, A., Haraty, R.A.: An Efficient Cache Replacement Strategy for the Hybrid Cache Consistency Approach. World Academy of Science, Engineering and Technology 63 (2010)

[4] Chen, H., Xiao, Y.: On-bound selection cache replacement policy for wireless data access. IEEE Transactions on Computers 56(12), 1597–1611 (2007)

[5] Dar, S., Franklin, M.J., Jonsson, B.T., Srivatava, D., Tan, M.: Semantic Data Caching and Replacement. In: Proceedings of the 22nd VLDB Conference, India, pp. 330–341 (1996)

[6] Ren, Q., Dhunham, M.: Using semantic caching to manage location dependent data in mobile computing. In: Proc. of ACM/IEEE MobiCom, vol. 99, pp. 210–221 (2000)

[7] Zheng, B., Xu, J., Lee, D.L.: Cache invalidation and replacement strategies for location-dependent data in mobile environments. IEEE Trans. on Comp. 51(10), 14–21 (2002)

[8] Lai, K., Tari, Z., Bertok, P.: Mobility aware cache replacement for location dependent information services. Technical Report T R- 04-04 (RMIT School of CS & IT) (2004)

[9] Magdalene, M., Jane, F., Nouh, Y., Nadarajan, R.: Network Distance Based Cache Replacement Policy for Location-Dependent Data in Mobile Environment. In: Proceedings of the 2008 Ninth International Conference on Mobile Data Management Workshops, IEEE Computer Society, Washington, DC (2008)

[10] Kumar, A., Sarje, A.K., Misra, M.: Prioritised Predicted Region based Cache Replacement Policy for location dependent data in mobile environment. Int. J. Ad Hoc and Ubiquitous Computing 5(1), 56–67 (2010)

[11] Denko, M.K., Tian, J.: Cross-Layer Design for Cooperative Caching in Mobile Ad Hoc Networks. In: Proc. of IEEE Consumer Communications and Networking Conf. (2008)

[12] Yin, L., Cao, G.: Supporting cooperative caching in ad hoc networks. IEEE Transactions on Mobile Computing 5(1), 77–89 (2006)

[13] Chand, N., Joshi, R.C., Misra, M.: Efficient Cooperative Caching in Ad Hoc Networks Communication System Software and Middleware (2006)

[14] Lim, S., Lee, W.C., Cao, G., Das, C.R.: A novel caching scheme for internet based mobile ad hoc networks. In: Proc.12th Int. Conf. Computer Comm. Networks (ICCCN 2003), pp. 38–43 (October 2003)

[15] Chand, N., Joshi, R.C., Misra, M.: Cooperative Caching Strategy in Mobile Ad Hoc Networks Based on Clusters. International Journal of Wireless Personal Communications Special Issue on Cooperation in Wireless Networks 43(1), 41–63 (2007)

[16] Li, W., Chan, E., Chen, D.: Energy- efficient cache replacement policies for cooperative caching in mobile ad hoc network. In: Proceedings of the IEEE WCNC, pp. 3349–3354 (2007)

[17] Heng, B.Z., Xu, J., Lee, D.: Cache invalidation and replacement strategies for location dependent data in mobile environments. IEEE Transactions on Computers 51(10), 1141–1153 (2002)

[18] Aggarwal, C., Wolf, J.L., Yu, P.S.: Caching on the World Wide Web. IEEE Trans. Knowledge and Data Eng. 11(1), 94–107 (1999)

FTLBS: Fault Tolerant Load Balancing Scheme in Wireless Sensor Network

Srimanta Halder, Monomita Mazumdar, Prasenjit Chanak, and Indrajit Banerjee

Department of Information Technology
Bengal Engineering and Science University, Shibpur, Howrah, India
{srimanta86,mazumdar.monomita,prasenjit.chanak}@gmail.com,
ibanerjee@it.becs.ac.in

Abstract. In wireless sensor network (WSN) hundreds or thousands of sensor nodes perform their sensing and transmitting tasks independently. Due to adverse environment, the probability of fault occurrence is very high in wireless sensor network. Higher frequency of fault occurrence decreases the performance as well as lifetime of WSN. The ability of fault tolerance is a primary metric of good wireless sensor network. Energy is an imperative issue in WSN. The sensor nodes include very small battery power and once the nodes are deployed they cannot be recharged or replaced. In this paper, we propose a fault tolerant load balancing scheme (FTLBS) to increase fault tolerability and lifetime of sensor network. The proposed scheme organizes the entire sensor network into groups and levels. In this technique we propose multipath data transmission technique for fault tolerance and the transmission load is balanced by varying group size. It dynamically selects a route based on fitness of the nodes. This approach delivers data efficiently with minimum delay even in faulty network. The simulation results establish that our proposed work gives better performance compared to existing fault tolerant methods.

Keywords: Wireless sensor network (WSN), load, fault tolerance, base station (BS), group, group head.

1 Introduction

Wireless sensor network (WSN) is a collection of thousands of tiny smart electronics sensing devices [1]. Sensor nodes are deployed in a remote environment where it senses data without any human interface. The sensor nodes send the sensed data to the base station by single hop or multi-hop data forwarding technique. Sensor nodes are comprised of low power processing unit, a trans-receiver system and battery module. The main research challenges of WSN are limited battery power, low processing capability and very small size of memory. Today WSN is largely used in environment monitoring, military application, machinery fault diagnosis [2] etc. Due to different deployment policy and environmental barrier, probability of fault occurrence is very high in WSN [3]. Limited power decreases life time of the network. Therefore, extension of life time demands a high performance protocol for data transmission. On the other hand, automatic fault tolerance technique improves network performance as well as increases the applicability of sensor network in a vast range.

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 621–631.
springerlink.com

There are mainly two strategies for fault tolerant in WSN. Those are retransmission [4] of lost data and multipath data transmission or replication [5]. In retransmission based fault tolerant technique, the sender sends the data to the base station and waits for a certain time. When the base station receives the data it sends an acknowledge signal to the sender node for the confirmation of receiving of data. In this approach the transmission of acknowledgement causes routing delay and a huge buffer memory is required to keep sending data until acknowledgement is received. On the other hand in multipath data transmission technique, same data is sending in multiple paths to the base station. When path fault occur then data reaches through duplicate path to the base station. Therefore, in this technique, a large number of nodes are involved for duplicate data transmission. Due to large duplication, the described technique suffers from message overhearing problem.

Different type energy conservation techniques such as Low Energy Adaptive Clustering Hierarchy (LEACH) [6], Hybrid Energy-Efficient Distributed Clustering (HEED) [7], UCCP [8], TL-LEACH [9] etc. have been proposed earlier. However these techniques are unable to balance load throughout the network. In Group Based Fault Tolerant Scheme (GBFTS) [10], fixed routing path causes small network lifetime. In GBFTS when data are transmitted from upper level to lower level, data duplication is done into 2^n ratio. Here smaller size group increases the data delivery delay. Our proposed scheme FTLBS works preciously for a long time and node fault affects hardly the performance of the network. To achieve this we have partitioned the network into several groups. In every group there is a group head. All the group members in a group are connected to group head via multi-hop communication. According to [11], multi-hop data transmission is more energy efficient than single hop data transmission. In network configuration phase, every node is assigned a level depending on the hop distance from BS. To forward data to group head, we have used multipath data transmission which involves minimum nodes. Simulation shows that, the proposed scheme provides higher performance than the existing GBFTS [10], PEQ [12] and DD [13].

The rest of paper arranged as follows: Proposed scheme briefly elaborated in section 2. The section 3 describes the implementation of the propose work. The simulation result is described in section 4. Finally section 5 concludes the propose work.

2 Fault Tolerant Load Balancing Scheme

In this section we are elaborating the energy efficient load balancing technique which is called as FTLBS. The FTLBS mainly focuses on transmission load distribution of sensor nodes near region to base station. This algorithm is an improvement of GBFTS and is influenced by Application Level Framework [14], SPIN [15] and UDDN [16] protocol.

2.1 Key Concepts of FTLBS

i. At the time of network construction, the sensor nodes form groups, and in every group there is a group-head. The group-head receives data from the nodes of the group. The group head aggregates the data and then retransmits to the nodes of the next group.

ii. The size of a group depends upon the local node density where group is to be formed. Near the base station where the densities of nodes are relatively high the group size is larger. However, in the remote area where density of nodes are low, the group size also small.

iii. For every sensor node there is a table. The table keeps count how many times current node sends its data to its neighbor nodes.

The various types of messages are used to communicate in the initialization stage:

LVL-Level determination message: This message is used at the time of network configuration to determine the level of nodes.

LVLREQ- Level Request message: This message is used to get level and id of neighbor node before transmission data.

LVLRPLY- Level Reply message: This message is used to reply the LVLREQ message.

3 Implementation of FTLBS

3.1 Energy Model Used in FTLBS

In this section we are describing the mathematical equation used in FTLBS for energy calculation [17]. To transmit $l$bit data the energy loss is

$$E_{Tr} = (l\alpha_1 + l\alpha_2 d^n)\text{.................} \qquad \text{......(1)}$$

Here α_1 is the energy consumed by internal electronics circuit of the sensor. d is the distance which the signal travels. n is an index which depends on antenna current .

Energy depletion to receive l bit of data is denoted by E_{Re} and which is a constant value for a particular electronics circuit of sensor node.

$$E_{Re} = l\alpha_1\text{.............................} \qquad (2)$$

Before data transmission in FTLBS three primary steps are required to ready the network:

1. Leveling of nodes.
2. Neighbor node discovery and node table initialization.
3. Group formation

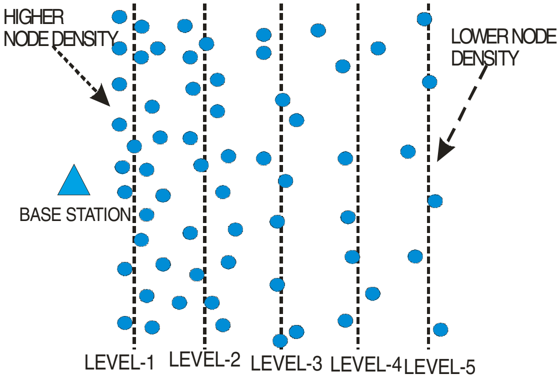


Fig. 1. Lower levels have higher number of nodes

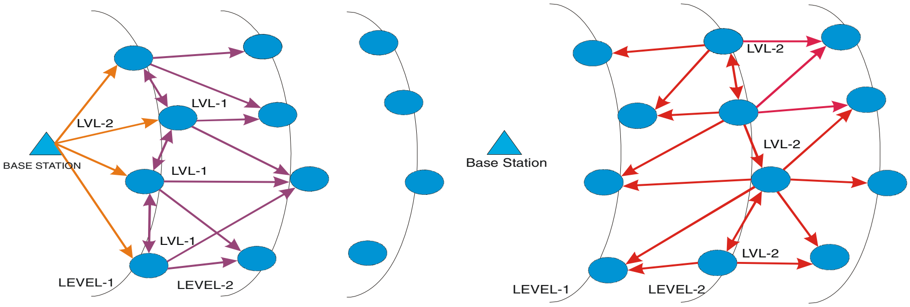
In FTLBS network, the node distribution is done by decreasing density of nodes with increasing distance from base station. Fig. 1 shows there are higher number of nodes in lower level than upper level.

The first stage is leveling of nodes. The Leveling is done using Algorithm 1, described below [16].

Algorithm 1: Level Determination

```
FOR all active node
    SET level of nodes as LEVEL=INFINITE
END FOR
Base Station Broadcasts LVL=0
WHILE there is no node remains DO
    FOR nodes receive LVL message
        IF (LVL<LEVEL of receiving node) THEN
        LVL=LVL+1
        SET level of current node LEVEL=LVL
        Broadcast LVL
        END IF
    END FOR
END WHILE
```

Initially all the sensor nodes are initialized to level value infinite (LEVEL=INF). First base station sends message LVL=0 to the nodes adjacent to it (Fig. 2a). All neighbor nodes receive message LVL=0, check whether the value of LVL message is less than its current level. If LVL value is less, then it increments the received LVL value by 1. The incremented value is set as its level. Then the current node retransmits the message as LVL=1. Thus level 2 is determined (Fig. 2b) and so on.



a) Determination of level 1

b) Determination of level -2

Fig. 2. Level determination of nodes

The next step is neighbor node discovery and count table initialization. The count table is a table which keeps information about the level of the neighbor nodes and a count of how many times data has been forwarded to each neighbor nodes. Initially the count value of each neighbor node is set to zero as no data has forwarded through any neighbor node. In this step any node, who wants to initialize its count table, sends message LVLREQ to its neighbor nodes (Fig. 3). When a neighbor node receives this message, it sends its level and node ID to the requesting node with LVLRPLY message. On receiving LVLRPLY message the node saves the level and node ID of its neighbor nodes in its count table.

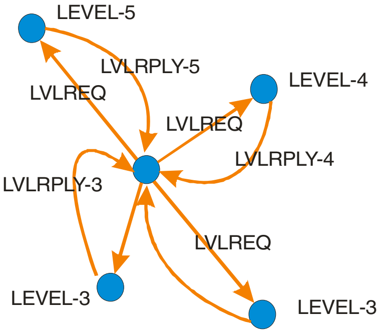


Fig. 3. Neighbour node table initialization

After level determination and neighbor node table initialization, group formation is done. The number of nodes in each group is same and it depends on the total number of deployed node in the network. Large number of nodes in a group generates huge number of replicated data at the time of data forwarding. On the other hand, smaller group size increases data delivery delay.

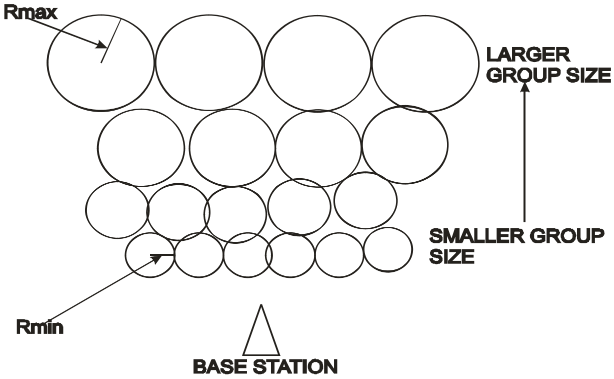


Fig. 4. Variable group size

To overcome this problem we use variable group size. We maintain group size near to base station is relatively smaller than the group size far away from base station (Fig. 4). So data duplication in lower groups is lesser than the data duplication in higher groups. This solves the large data duplication problem generated in GBFTS [10]. In FTLBS, for every group there is a group head. Among the nodes of upper level of lower group, one node acts as group head. The group head is selected based on highest energy level.

Theorem 1: The radius of largest group R_{max} and radius of smallest group R_{min} are $\sqrt{\frac{A_1 x}{\pi n k}}$ and $\sqrt{\frac{A_k x}{\pi n}}$

Proof: Let N be the total number of nodes. Traffic generated by each node is T_i. Traffic send from node j to i is P_{ij} Load of each node is L_i. Let the area under i^{th} level is A_i. For uniform node distribution number of nodes in each level is n. L_i can be written as

$$L_i = T_i + \sum_j L_j P_{ij}$$

There is k number of levels. So the total load in k^{th} level is

$$TL_k = n_k \left(T_i + \sum_j L_j P_{ij}\right)$$

For $(k-1)$ level, load is

$$TL_{k-1} = n_{(k-1)} \left(T_i + \sum_j L_j P_{ij}\right) + TL_k$$

Thus for first level total load is

$$TL_1 = n_1 \left(T_i + \sum_j L_j P_{ij}\right) + \sum_{m=2}^{k} TL_m$$

Individual load of level 1 is $\frac{n_1 \left(T_i + \sum_j L_j P_{ij}\right) + \sum_{m=2}^{k} TL_m}{n_1}$

Individual load of k_{th} level is $\frac{n_k \left(T_i + \sum_j L_j P_{ij}\right)}{n_k}$

To keep individual load of first level nodes same as k^{th} level nodes the ratio of $n_1/n_k = k$

Node density of level 1 is $\frac{k n_k}{A}$, and node density of k is $\frac{n_k}{A}$

Let x number of nodes are there in every group.

$$R_{min} \text{ will be } \sqrt{\frac{A_1 x}{\pi n k}}$$

$$R_{max} \text{ will be } \sqrt{\frac{A_k x}{\pi n}}$$

3.2 Data Forwarding Technique in FTLBS

When node senses any event from the environment, it checks if the event generates new data. Then the data is send to the group head by multipath connection. Group head collects data from group members, then it sends data to next lower group head (Fig. 5). When a group member of upper level receives data it sends data to two nodes of lower level. The two nodes of lower level are selected by the Algorithm 2.

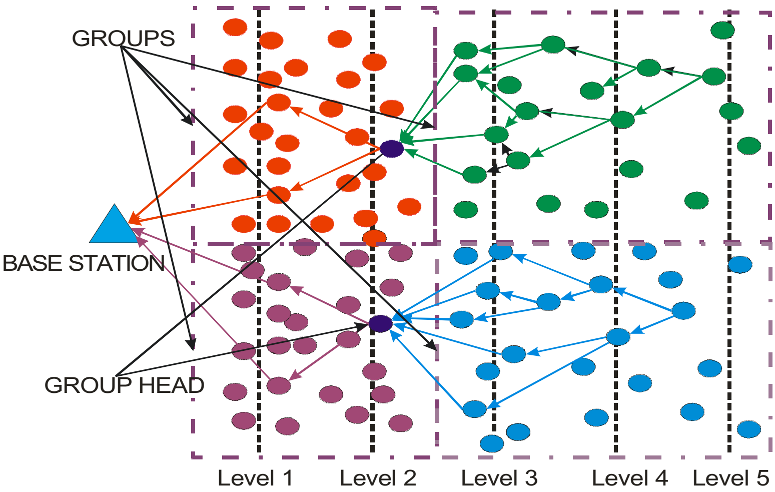


Fig. 5. Data Forwarding Technique

Algorithm 2: Node selection for data forwarding

```
SET sum=0
FOR all neighbor nodes
    Sum = Sum + (count of this neighbor node * A) + (level of this neighbor
    node * B)
END FOR
FOR all neighbor nodes
    Probability = {(count of this neighbor node * A) + (level of this neighbor
    node*B)}/sum
END FOR
Arrange probabilities in ascending order
Number = random number between 0 and 1
            // a random number is generated between 0 and 1
FOR every neighbor nodes
    IF (number >Probability AND number <next Probability) THEN
        Current node is selected for data forwarding
    END IF
END FOR
```

If any node wants to forward data to next level, it checks its count table. Count signifies how many times it has forwarded data to the nodes. Depending on the level value and count value probability value is generated. Based on the probability value a roulette wheel based selection is done among the nodes. Applying this algorithm twice the two nodes are selected and data is forwarded to those nodes. After data forwarding, the count value of these two nodes are incremented by 1. In the above algorithm, A and B are two constants for a particular sensor network. Higher value of A decreases data delivery delay. On the other hand, higher value of B increases fault tolerability of network. But it also increases the data delivery delay. Tuning these two values, network performance can be improved.

4 Performance Evolution of FTLBS

In this section, we have shown the simulated result of FTLBS to evaluate the performance. To obtain this, we have used C programming based simulation technique. The various constrains that we consider in simulation procedure is summarized as Table 1. We use DD [13], PEQ [12] and GBFTS [10] algorithms to measure the quality of service of FTLBS.

Data Delivery Ratio: It is the ratio of number of event generated in the network to the number of successful data received at the base station.

Data Delivery Delay: It is the time elapsed between an event is sensed in the network and the time when the data is successfully received at the base station.

Table 1. Simulaiton Assumpptions

Network Size	200X200
Number of Nodes	100-800
Transmission Packet Size	800
Initial Node Energy	0.5 J
α_1Energy required due to node's internal electronics circuit.	50nJ/bit
α_2 Energy required for node's antenna current.	10pJ/bit/m^2

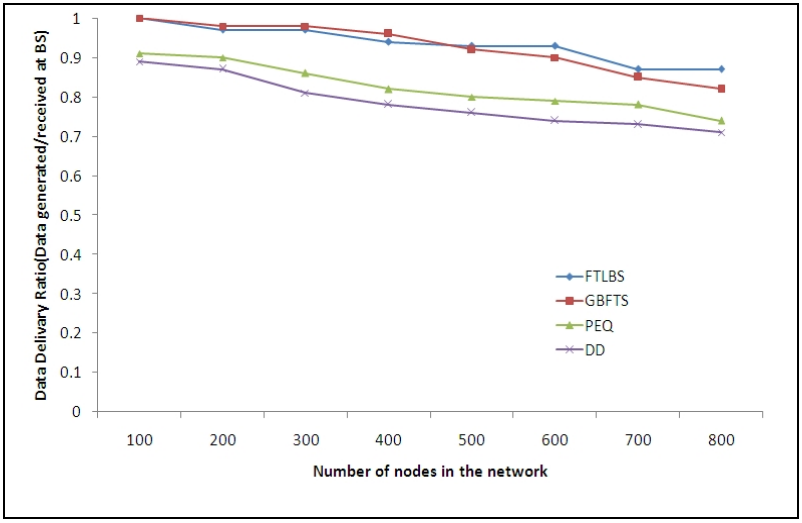


Fig. 6. Data delivery ratio

Fig. 6. shows that, FTLBS has high rate of data delivery ratio. The ratio is higher than DD[13], PEQ[12] and GBFTS[10]. The Fig. 6 also shows that network size has very less impact on data delivery ratio. That means for larger network, the ratio of total number of event sensed and total number of data received at base station remains almost same.

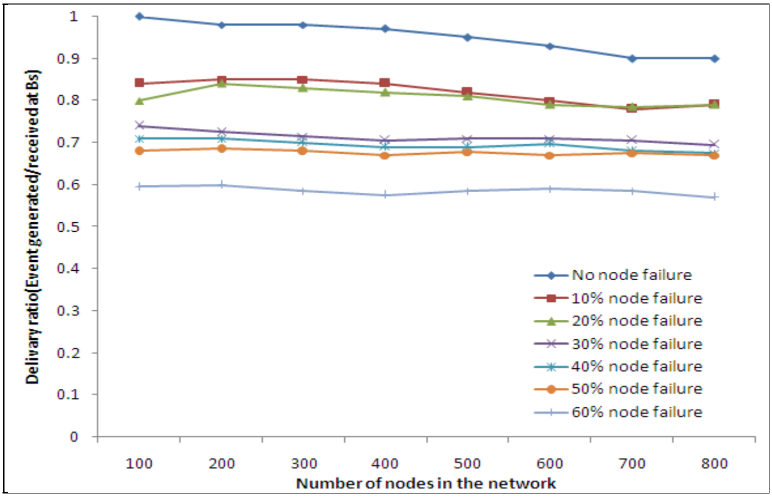


Fig. 7. Delivery ration under node failure

Fig. 7 shows that, in FTLBS the number of node failure has very less impact on delivery ratio. Up to 50 percent of node failure, there is very less fall in delivery ratio. The figure also shows that, the delivery ratio is almost independent of network size. The network is unable to manage more than 50% node failure. There is a sharp change in delivery ratio when the node fault is 60%.

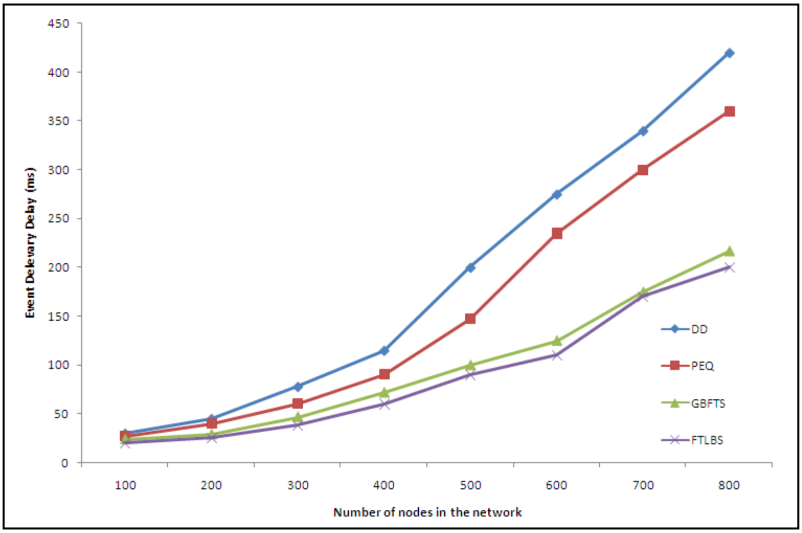


Fig. 8. Data Delivery Delay

Fig. 8 shows that, when network size increases, the time elapsed between an event is sensed and successful reception of the data at BS also increases rapidly for DD and PEQ. But FTLBS has data delivery delay as GBFTS.

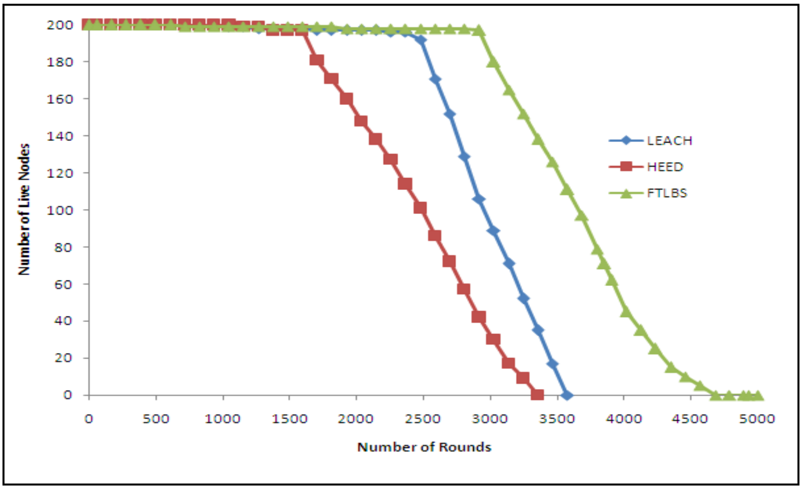


Fig. 9. Number of Dead nodes per Round

Fig.9. shows a comparison of network lifetime between fault tolerant load balancing scheme (FTLBS), LEACH and HEED. The result shows that, the number of rounds data transmission before first node fault in FTLBS is 1.23 times than LEACH and 1.88 times than HEED.

5 Conclusion

In this paper we have proposed a fault tolerant load balancing scheme. This scheme increases network lifetime by load balancing and efficient energy utilization. In FTLBS the node to which data is to be forwarded is selected dynamically based on the fitness of neighbor nodes. This reduces network failure. In FTLBS when data moves towards base station the cost of data also increases. To provide more fault tolerability near base station the data is replicated with 2n order as data moves towards the base station. In FTLBS data delivery delay is low, and occurrence of fault has very less effect on data delivery delay.

References

1. Banerjee, I., Chanak, P., Sikdar, B.K., Rahaman, H.: EER: Energy Efficient Routing In Wireless Sensor Network. In: IEEE International Technology Symposium 2011, IIT kharagpur, India, January 14-16 (2011)
2. Gao, B., Xiong, S., Xu, Z.: The Application of Wireless Sensor Networks in Machinery Fault Diagnosis

3. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: A Survey on sensor networks. IEEE Communications Magazine 40(8), 102–114 (2002)
4. Intanagonwiwat, C., Govindan, R., Estrin, D.: Directed diffusion: A Scalable and Robust Communication Paradigm for sensor Networks. In: ACM Intl., Conf. on Mobile Computing and Networking, pp. 56–67 (2000)
5. Kim, S., Foneseca, R., Culler, D.: Reliable Transfer on Wireless Sensor Networks. In: The First IEEE International Conference on Sensor and Ad hoc Communication and Networks, pp. 449–459 (October 2004)
6. Heinzelman, W.R., Chandrakasan, A., Balakrishnan, H.: Energy-Efficient Communication Protocol for Wireless Micro sensor. In: IEEE Proceedings of the 33rd Hawaii International Conference on System Sciences (2000)
7. Younis, O., Fahmy, S.: HEED: A Hybrid, Energy-Efficient, Distributed Clustering Approach for Ad Hoc Sensor Networks. IEEE Transactions on Mobile Computing 3(4), 366–379 (2004)
8. Aslam, N., Phillips, W., Robertson, W.: A Unified Clustering and Communication Protocol for Wireless Sensor Networks. IAENG International Journal of Computer Science 35(3), 249–258 (2008)
9. Loscri, V., Morabito, G., Marano, S.: A two-levels hierarchy for low-energy adaptive clustering hierarchy (TL-LEACH). In: Vehicular Technology Conference, pp. 1809–1813 (2005)
10. Banerjee, I., Chanak, P., Rahaman, H., Das, N.: GBFTS: Group Based Fault Tolerant Scheme in Wireless Sensor Networks. International Journal of Information and Electronics Engineering 2(2), 179–184 (2012)
11. Heinzeman, W.R., Chandrakasn, A., Balakrishnan, H.: Energy efficient communication protocol for wireless microsensor networks. In: Procl. of IICSS (January 2000)
12. Vu, H., Nguyen, T., Mittal, N., Venkatesan, S.: PEQ: A Privacy-Preserving Scheme for Exact Query Evaluation in Distributed Sensor Data Networks. In: 28th IEEE International Symposium on Reliable Distributed Systems, SRDS 2009, pp. 189–198 (2009)
13. Boukerche, A., Nelem Pazzi, R.W., Araujo, R.B.: Fault-tolerant wireless sensor network routing protocols for the supervision of context-aware physical environments. Journal of Parallel and Distributed Computing (2005)
14. Clark, D., Tennehouse, D.: Architectural Consideration for a New Generation of Protocols. In: Proc ACM SiGCOMM (September 1990)
15. Kulik, J., Heinzelman, W.R., Balakrishnan, H.: Negotiation-based protocols for disseminating information in wireless sensor networks. Wireless Netwroks 8, 169–185 (2002)
16. Banerjee, I., Rahaman, H., Sikder, B.: UDDN: Unidirectional Data Dissemination via Negotiation. In: International Conference on Information Networking, Pusan Korea, January 23-25 (2008)
17. Banerjee, I., Chanak, P., Sikdar, B.K., Rahaman, H.: EERIH: Energy Efficient Routing via Information Highway in Sensor Network. In: IEEE International Conference on Emerging Trends in Electrical and Computer Technology, Kanyakumari, India, March 23-24 (2011)

Effective Resource Allocation Technique for Wireless Cellular System

Banda Sreenivas[1], S. Ramesh Babu[2], S. Prabhakar[3],
and D. Karunakar Reddy[4]

[1] Department of ECE, Jyothishmathi Institute of Technology & Science, India
[2] Department ECE Nigama Engineering College, India
[3] Department of ECE, RRSCET, India
[4] Department of ECE, JNTU-H, India
{targetsrinu8,karunakar.jntuh}@gmail.com,
ramesh_sag@rediffmail.com,
saggurthi_p@yahoo.co.in

Abstract. An efficient resource allocation is one of the greatest challenges in wireless cellular communication. The resource allocation schemes avoid wastage of resources by allocating resources to a mobile terminal over a short period of time, providing quality of service over wireless networks is the most stressing point for service providers. In general a high degree of sharing is efficient, but requires service protection mechanisms to guarantee the Q o S for all services. In this paper we address the multi cell interference on overall radio resource utilization and propose a new strategy for resource allocation in multi cell systems. We also propose a joint management of interference within and between cells for allocation of radio resources, Simulation results are showing that there is a significant improvement in the resource utilization so that overall network performance.

1 Introduction

Wireless communication is one of the most vibrant areas in the communication field today. With growing demand for wireless communications, advanced mobile cellular systems have evolved in many countries. The major challenge in supporting multimedia content and real-time services over wireless network is the QoS. Future wireless communications will be a major move toward ubiquitous wireless communication system and seamless high quality services resource allocation methods must be developed. The objective of the allocation scheme is to maximize total network utility can be utilized for optimal resource allocation. [1]

In this paper, we discuss inter cell interference problem of scheduling process by introducing load matrix concept and using HSUPA system to prove it. Section 2 describes effect of interference on HSUPA in terms of user terminals, is limited by total received power at the base station limits the uplink capacity. In section 3 gives the uplink resource allocation in both single cell and multi cell cases to achieve maximum capacity. The load matrix concept is detailed in Section 4. The performance of the LM scheduling can be observed in the simulation results provided in section 5 and finally section 6 provides conclusion for this paper.[4][6]

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 633–640.
springerlink.com

2 Inter Cell Interference

A cellular system consists of many cells with channels (timeslots, bandwidth, or codes) reused at spatially separate locations. Due to the fundamental nature of wireless propagation, transmissions in a cell are not limited to within that cell, and thus there is inter cell interference between users and base stations, that use the same channels. The majority of current systems are interference limited rather than noise limited. Interference is part of every mobile cellular communications system, and it constitutes a limitation to both radio network capacity and quality of service provided to users [3]. Inter cell interference is managed via averaging of the effects of multiple interferers. It is more effective in the uplink than in the downlink. Interference averaging also allows statistical multiplexing of bur sty users, thus increasing system capacity.

Resource allocation schemes in the uplink are of two categories, distributed and centralized. The objective of distributed allocation is to reduce the complexity to the Radio Network Controller (RNC). This scheme does not know the channel conditions of adjacent cells. Where as in case of centralized schemes, the network controller is responsible for allocating the resources in every cell.On the forward link, the data is split by the RNC to a number of base stations and the received data is combined by the mobile terminal. On the reverse link, the participating base stations forward the received data to the serving RNC to combine. [12]

In interference limited systems, the uplink capacity is limited by the total transmitted power at the base station and this power was limited by uplink capacity. Inter cell interference calculation is done by multiplying the number of users in a cell by the average interference offered in this cell, this kind of calculation, being suitable for real-time interference simulations based on the number of users, their path loss, slow fading, and the cell area. But in uplink, inter cell interference density analysis is performed by assuming perfect power control. The number of users is taken into account, as well as the received signal power and the activity factor according to the user's service calculates the average inter cell interference per cell. [7]

2.1 Interference Model

The performance of a high SINR user is dominated by intra cell (or inter-user) interference, increasing number of users also results in increased intra cell interference and an orthogonal access benefits by eliminating the intra cell interference. We know that as the other-cell to own-cell signal ratio increases, the performance is dominated by other-cell interference rather than only by intra cell interference. [8] Therefore, in case of heavy inter-cell interference, the gains of orthogonal access over non-orthogonal access go down.

2.2 Intra - Cell Model

The intra cell interference limits the maximum achievable data rates and limits the capacity of the uplink. This model calculates the average inter-cell interference per cell, being necessary to use a user distribution in the cell area. The calculation of the intra-cell interference in Down link, on MT $_i$ is given by equation 1

$$I_{\downarrow}(\mathrm{Intra}, j)^{\uparrow}\mathrm{DL} = [(P]_{\downarrow}(\mathrm{Total}, BS) - P_{BS-MT_i}) \times^{\alpha \times} L_{BS-MT_i} \;[W] \qquad (1)$$

Where $P_{Total, BS}$ is the total power transmitted by the Base Station (BS), $P_{BS \rightarrow MTi}$ is the power transmitted by the BS to the (Mobile Terminal) MT in which interference is being calculated, and $L_{BS \rightarrow MTi}$ is the propagation loss between BS and MT. The ortho gonality factor α can take values between 0-1. In Uplink Interference is given by equation 2

$$I^{UL}_{\mathrm{Intra},j} = \sum_{g=1}^{G} P_{BSj \leftarrow MT} \times \eta_g \times N_{j,g} \;[W] \qquad (2)$$

Where $P_{BSj \leftarrow MT}$ is the power received at BS $_j$ from an MT, η_g is the activity factor of service g, $N_{j,g}$ is the number of MTs using service g on the cell of BS $_j$, and G is the total number of services used.

2.3 Inter-Cell Model

Power control on the down link has less impact on inter cell interference than on the uplink because the downlink transmissions all originate from the cell center. Whereas Uplink transmissions can come from the cell boundaries. Hence need to focus more on the effect of power control on the uplink. In DL, the model used for inter-cell interference, in an MT using a service g is given by,

$$I^{DL}_{\mathrm{Inter},j} = \sum_{j=2}^{N_{BS}} P_{\mathrm{Total},BSj} \times^{\alpha} r_j^{-a} \times 10^{\frac{\Delta L_j}{10}} \;[W] \qquad (3)$$

Where P_{total},BS_j is, BS j's total transmitted power, N_{BS} is the number of interfering BS's, r_j^{-a} is the path loss, a representing path loss exponent, ΔL_j is associated to slow fading, following a statistical distribution with zero mean and a certain standard deviation, and r_j represents the distance between the interfering BS $_j$ and the MT.

$$I^{UL}_{\mathrm{Inter},j} = \sum_{K=1,K\neq j}^{N_{BS}} \sum_{g=1}^{G} P_{BSj}\, T_{MK} \times \eta_g \times N_{k,g} \times A[W] \qquad (4)$$

$P_{BSj} \leftarrow T_{Mk}$ is BS_j power received from the MT that is being covered by an adjacent cell k, $N_{k,g}$ is the number of users using service g in interfering cell k. A larger interference reduces SINR and hence increases user BER. Inter cell interference can be reduced by separating Cells operating on the same channel by a large distance.

Good cellular system designs are interference limited that is the interference power is much larger than the noise power. Figure.1 shows the R o T fluctuation due to inter cell interference in a typical cell. The R o T of a cell dramatically increases to above the threshold and rapidly decreases to below the threshold while allocating resources to users. But for ideal performance in terms of interference management this R o T should be close to threshold (R o T $_{target}$). As the uplink load increases user terminals have to their transmit power substantially to overcome the increased interference level at the base station. Due to the fact that interference cell, the transmit power of user terminals is limited by total received power at the base station actually limits the uplink capacity. [5]

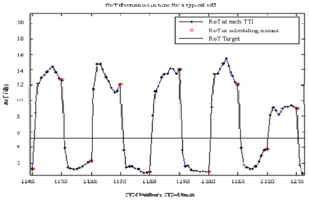


Fig. 1. RoT fluctuation in multi cell

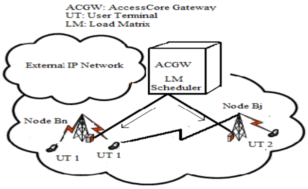


Fig. 2. Centralize LM Scheduling in a 3G LTE System

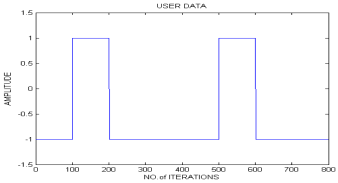


Fig. 3. User data

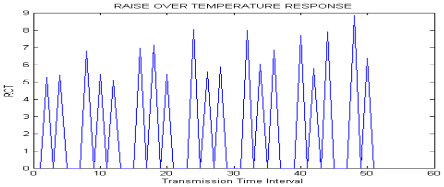


Fig. 4. BPSK Modulated Signal

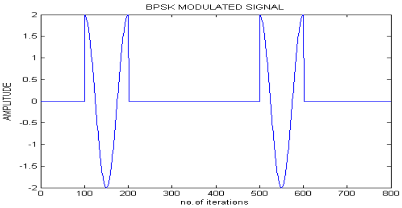


Fig. 5. ROT Response

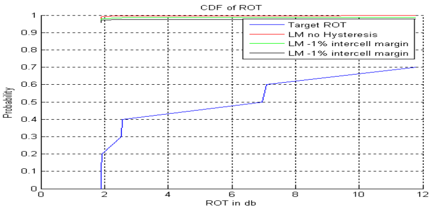


Fig. 6. CDF of RoT for Inter cell Interferen

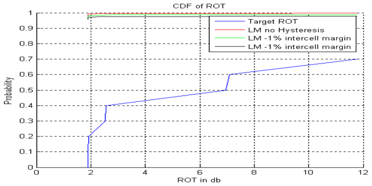


Fig. 7. CDF of RoT for Intra cell Interference

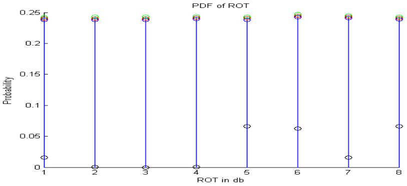


Fig. 8. PDF of ROT

3 Resource Allocation

In next generation networks a variety of services with different requirements, like real time communications, broadband Internet access, email services are expected.

Consequently, packet scheduling mechanisms and resource allocation techniques for QoS guarantees will play a key role. The Radio resource allocation is a challenging problem in wireless networks due to different channel conditions of user and the main aim of resource allocation is to assign radio resources to individual users in order to achieve maximum capacity while meeting the required quality of service. A contiguous resource allocation scheme is defined for both the uplink and the downlink.

In uplink the distributed and centralised allocation schemes reduce the complexity of network. The resource allocation problem in these systems causes inefficient use of radio spectrums and to utilize multiple and maximize the system capacity, but they have to consider admission and access control in conjunction with resource allocation mechanism, sub carriers in wireless systems such as OFDM (orthogonal frequency division multiplexing)[9][14]. Allocating different number of sub carriers intelligently, the inefficiency issue can be handled. In order to provide various choices of scheduling performance and signalling overhead, multiple resources allocation types are defined.

In multi user OFDMA systems, multi user diversity can be easily achieved by the allocation of sub channels to users, and these channels are independent for each user, with this the resource allocation problem for multi user OFDMA systems has been extensively investigated. The quality of resource allocation can be assessed by overall throughput and fairness. In a wireless network environment the trade-off between throughput and fairness is important for scheduling. Due to the nature of resource allocation (time and frequency), transmissions suffer no interference from within the cell and further see minimal interference from neighbouring cells.[11]

4 Load Matrix Concepts

The Load Matrix (LM) concept has the facility to joint management of interference within and between cells while allocating radio resources to users and this concept proposed intakes the inter cell interference information into account in order to avoid R o T outage. In a multi cell system one of the main challenges in resource allocation is the control of inter cell interference. LM is a centralized scheduler, uses a database containing the load contribution of all active users in the network and it assigns radio resources to all active users in the net work. The basic problem in the uplink scheduler is to assign appropriate transmission rate and time to all active users, result maximum radio resource utilization across the network while satisfying the QoS requirements of all the users. [13]

The important factor in the resource allocation is the users transmit power. The constraints to be satisfied for a network of M users and N cells are

Constraint1: This constraint states that the maximum user power $P_{i,max}$. For each active user i in the network, its transmit power P_i must be maintained in an acceptable region defined as

$$0 \leq P_i \leq P_{max} \quad i \in \{1,\ldots.., M\} \tag{5}$$

Constrain2: The total received power at base station should be kept below a certain threshold for all N base stations in the network it uses Rise over Thermal noise (RoT) to represent the interference constraints.

$$RoTj \leq RoT_{target} \quad j \in \{1,\ldots\ldots\ldots, N\} \tag{6}$$

RoTj is the total in band received power fixed target value to maintain uplink interference level at the base station j (BS_j) over thermal noise. The RoTj for M active users in the network given below is used to estimate RoT of cells, can be written as

$$ROT_j = \left(N' + \sum_{i=0}^{M} P_i G_{ij} \right) / N' \tag{7}$$

Constraint3 **:** The signal to noise plus interference ratio required at the serving base station j if rate k is being assign to the user to achieve a given frame error rate is $SINR_{target,k}$. For each user, depending on its channel type and speed, each rate k has a minimum required SINR called $SINR_{target,k}$. This constraint satisfies only by considering $SINR_{target,k}$ as SINR.

$$SINR_{i,j} \geq SINR_{target,k} \quad i \in \{1,., M\},\ k \in \{1,\ldots,K\} \tag{8}$$

A centralized scheduler assigns radio resources to all the M users and N cells in the network, $LM_{i,j}$ is the load factor contribution by user i at BS j defined as

$$LM_{i,j} = \frac{P_i G_{ij}}{N' + \sum_{m=1}^{M} P_m G_{mj}} \tag{9}$$

Where G_{ij} is the channel gain from user i to $BS_{j\ averaged}$ over scheduling period, N' is the thermal noise and P_i is the transmitted power. The $LM_{i,j}$ values stored in column j of LM database, R o T of cell j is

$$ROT_j = \frac{1}{1 - \sum_{i=1}^{M} LM_{ij}} \tag{10}$$

$SINR_{i,j}$ can be written as

$$SINR_{i,j} = \frac{P_i G_{ij}}{N' RoT_j - P_i G_{ij}} \tag{11}$$

The required transmitted power for user i at rate k is,

$$P_{i,k} = \frac{N' RoT_{target}}{G_{ij}} \frac{\text{SINR target},k}{1 + \text{SINR target},k} \tag{12}$$

If above all constraints are satisfied then only power $P_{i,k}$ is acceptable and user i will be scheduled for transmision. After that LM elements are updated and RoT is calculated for each cell using [10]. The performance of the LM scheduling has the best RoT over other algorithms because this scheduler significantly reduce the probability of the RoT exceeding its target. Priority functions are used to rank the uers in the scheduling process and make a balance between cell throughput and fairness. Commonly used priority functions shown in table 1.

A priority function in LM scheduling is introduced based on users load vector that includes intra and inter cell impact on the network and it tries to maximise the network capacity through these interference managements. In each base station the LM allocation process simultaneously increases allocated resources to avoid interference imbalance among the cells. A user with higher channel gain has highest priority it is evident that giving priority to a user with better channel condition increases the cell throughput but in a multi cell network could have severe impact on the throughput of other cells. Hence a new priority approach to load matrix is the Global Proportional Priority (GPP) considers interference contribution of the user to other cells is defined as

Table 1. Comparison between different type schedulers

Parameter	Traditional Schedulers				Load Matrix
	Round Rabin	Max C/I	Proportional fair	Score- based	
Throughput	Low	High	Medium	Moderate	High
Fairness	High	Low	Medium	Moderate	High
Performance	Less due to Low TP	Less due to low fairness	Trade-off between TP&F	Moderate compare to PF	Efficient than traditional schedulers

$$\text{Priority}_{i}=\frac{G_{ij}}{\sum_{n=1,n\neq j}^{M} G_{i,n}}\ \forall i\in\{1,\ldots M\} \qquad (13)$$

Where $G_{i,j}$ is the total channel gain from under i to BS_j averaged over the scheduling period. LM process uses Capacity Check (CC) process to assign rates to the highest priority user in each cell to update LM elements and while performing capacity checking it maintains overall performnace of the network. The CC operates on small margin around RoT target instead of a fixed RoT target threshold result much improved interference outage performance and higher resource utilization.[15]

5 Conclusion

In this paper we have evaluated system capacity and fairness performance of several transmission schemes with LM scheduling. LM concept is presented specifically to provide an efficient resource allocation by jointly considering inter cell and intracellular interferences before allocating radio resources. A novel approach towards efficient uplink scheduling is presented We have developed a system level simulator for HSUPA system based on the proposed simulation conditions. The effect on the scheduling performance can be observed in the simulation results provided and these results indicate that selection of RoT as well as transmit power significantly affect the performance.

References

[1] Wu, L., Tian, Z.: Capacity-Maximizing Resource Allocation for Data-Aided Timing and Channel Estimation. In: IEEE International Conference on Acoustics, Speech and Signal Proceeding, Montreal, Canada, vol. 4, pp. 525–528 (May 2004)

[2] 3GPP TR 25.896. Feasibility Study of Enhanced Uplink for UTRAFDDV6.0.0 (March 2004), http://www.3gpp.org

[3] 3GPP, TS 36.211, Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation (Release 8)

[4] Holma, H., Toskala, A.: WCDMA for UMTS: Radio Access for Third Geeration Mobile Communications, 2nd edn. John Wiley & Sons (2002)

[5] Hassibi, B., Hochwald, B.M.: High-rate codes that are linear in space and time. IEEE Transactions on Information Theory 48(7), 1804–1824 (2002)

[6] Khan, F.: A time-orthogonal CDMA high speed uplink data transmission scheme for 3G and beyond. IEEE Communication Magazine, 88–94 (February 2005)

[7] Tse, D., Viswanath, P.: Fundamentals of wireless communication. Cambridge University Press, Cambridge (2005)

[8] Goldsmith, A.: Wireless communications, Cambridge (2005)

[9] Khan, F.: LTE for 4G mobile broadband. Cambridge Unive. Press, Cambridge (2009)

[10] Moshavi, S.: Multi-user detection for DS-CDMA communications. IEEE Communication Magazine 34, 124–136 (1993)

[11] Varanasi, M.K., Aazhang, B.: Multistage Detection in Asynchronous Code Division Multiple access communications. IEEE Trans. Commun. 38, 509–519 (1990)

[12] Kim, J.B., Honig, M.L.: Resource allocation for Multiple classes of DS-CDMATraffic. IEEE Transactions on Vehicular Technology 49(2), 506–518 (2000)

[13] Abedi, S.: Efficient Radio Resource Management for Wireless Multimedia Comms. IEEE Transactions Wireless Communications 4(6), 2811–2822 (2005)

[14] Saraydar, U., Mandayam, N.B.: Pricing and Power Control in a Mul cell Wireless Data Network. IEEE Journal on Communications 19(10), 1883–1892

[15] Ekstrom, H., Furuskar, A., Parkvall, S.: Technical Solutions for the 3G Long- Term Evolution. IEEE Communications Magazine, 38–45 (March 2006)

Performance Analysis of AODV and DSDV Protocols Using RPGM Model for Application in Co-operative Ad-Hoc Mobile Robots

Rajesh P. Barnwal and Arnab Thakur

CSIR-CMERI, MG Avenue, Durgapur, W.B., India
{barnwal.r,arnab.nitd}@gmail.com

Abstract. With the advancement in communication technology, the robotic researchers started to think in the direction of developing the co-operative mobile robot. These robots can co-operate with each other in the group or sometimes outside the group in ad-hoc manner for successful completion of the mission. Ad-hoc routing protocol provides mechanisms for ad-hoc communication in infrastructureless and dynamic environment, where these robot teams are suppose to be deployed. For the reliability of ad-hoc communications, it is very much essential to have an idea of the relative performance of different available routing protocols, which can be applied for the purpose. In this paper, the authors discuss about the cooperative ad-hoc mobile robot communication and different types of ad-hoc network protocols applicable for the said situation. Performance of two ad-hoc routing protocols i.e., DSDV and AODV have been discussed and adjudged for their efficiency in Reference Point Group Mobility (RPGM) scenario having multiple groups of communicating nodes.

1 Introduction

The advent of mobile ad-hoc network technology widens the path for development of co-operative ad-hoc mobile robots. These robots are supposed to mimic the behavior of animal and human beings in real sense during job handling process. They also demonstrate the capability of managing more complex task with the help of other team members by the means of communication. The potential application areas of these types of mobile robots ranges from disposal of toxic waste, nuclear power processing, fire fighting, military or civilian search and rescue missions, planetary exploration, security, surveillance, to reconnaissance tasks. The wireless communication technology provides low cost solutions for addressing the flexible and efficient communication requirements of these robots [1].

The conventional wireless communication solutions require static infrastructure such as base stations to perform communications. The requirement of centralized infrastructure or pre-defined topology poses constraints in deployment of the mobile robots in dynamic environment and restricts their mobility during the mission. Unlike the legacy wireless communication technology, MANET (Mobile Ad-hoc Network) [2] provides many standard protocols, which can be used for the communication purpose in ad-hoc manner. In MANET, node also acts as router and thus can forward packet for other

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 641–649.
springerlink.com

nodes. Due to this property of ad-hoc network, the robot teams can easily be deployed at such places, where group can move in desired direction irrespective of their administrative point. Also, any group member can leave or join the mission in the midway without affecting others. Co-operation is however very difficult to maintain, if the performance of the underlying protocol is poor. Therefore, there is a need to perform a study for performances of some exisiting standard protocols, which can be readily used for communication among co-operative mobile robot.

The paper analyzes performance of the two standard ad-hoc routing protocols each from two different categories (Proactive Routing and Reactive Routing) [3]. In the present study, Reference Point Group Mobility (RPGM) [5] has been used to simulate mobility of co-operative robots. Simulations, under varying number of nodes, have been carried out for comparison of the performance of DSDV [6] and AODV [7] protocols.

2 Co-operative Mobile Robot Communication

Co-operative multiple mobile robots need to communicate among themselves for completing the assigned tasks in a successful manner. The wireless network provides the networking infrastructure to support the quality of service (QoS)(bandwidth, latency and reliability) requirements of robotic communications [1]. The prime requirement of the networking infrastructure is to support quick reconfiguration, mobility management and QoS. Figure 1 gives a layered model of mobile robot networking.

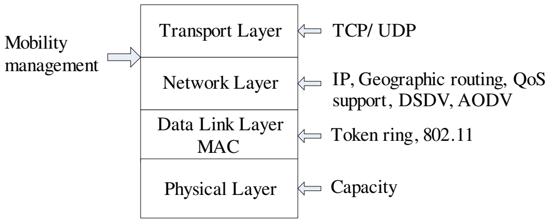


Fig. 1. Mobile robot networking layered model [1]

The model comprised of network, data link and physical layer. The success of the assigned mission to the team of cooperative mobile robot greatly depends upon the performance of each layer of the communication model.

2.1 Ad-Hoc Network for Co-operative Mobile Robot

Traditionally, robotic researchers have proposed the use of centralized communication networks for robots, where the members of a team of robots communicate with a central controller (base station) over a wireless medium [9]. However, in several situations, it is very difficult and sometimes impossible for a central base station to simulataneously cope up with a number of robots and their sensors in terms of communication resources and performance. This situation also restricts the mobility of the robots, which is basically guided by the communication capability of the base station.

Application of ad-hoc network for communication among cooperative robots provides great flexibilities to the team members. It allows the dynamic topology of interconnections and also supports changes in population, connectivity and local constraints. Multi-hop based ad-hoc network system demonstrates robustness to local failures and facilitates better scalability.

2.2 Ad-Hoc Routing Protocols

Ad-hoc routing protocols can broadly be divided into two categories i.e., proactive routing protocol and reactive routing protocol [3].

2.2.1 Proactive Routing Protocols

Proactive routing is also often termed as table-driven routing. In this type of routing protocols, fresh lists of destinations and their routes are maintained by periodic distribution of routing tables throughout the network. This category of protocol always strives to maintain consistent and updated routing information at each node [11]. Destination-Sequenced Distance-Vector Routing Protocol (DSDV) [6] and Link-State Routing (LSR) [12] are the two common proactive routing protocols.

2.2.2 Reactive Routing Protocols

Reactive routing is often known as on-demand routing or source-initiated routing protocol. The main advantage of reactive protocols is that it imposes less overhead due to route messages on the network but at the same time, it is also facing high latency time in route finding process. Sometimes excessive flooding of the communication packets may lead to network clogging. Ad-hoc On-Demand Distance Vector Routing (AODV) [6], Dynamic Source Routing (DSR) [7] and Temporally Ordered Routing Algorithm (TORA) [5] are some of the examples of reactive routing protocol.

2.3 Destination-Sequenced Distance-Vector Routing

The DSDV [6] routing protocol works on the basis of Bellman-Ford algorithm for shortest paths and avoid any loop in the routing table. Figure 2 shows the effect of movement of mobile host MH_1 in an example ad-hoc network. Due to movement of MH_1, the new neighbours (MH_7 and MH_8) broadcast the incremental routing updates, so that other can updates their routing table.

2.4 Ad-Hoc On Demand Distance Vector Routing

AODV [8] is basically based on distance vector algorithm. It works on the basis of single path routing protocol even though multiple routes can be detected due to routing discovery. This protocol enables multihop routing between participating mobile nodes, that wish to establish and maintain an ad-hoc network [13].

Figure 3 describes the message exchanging strategy of the AODV protocol. In this figure, there are four nodes. S is the source node, D is the destination node and rests are

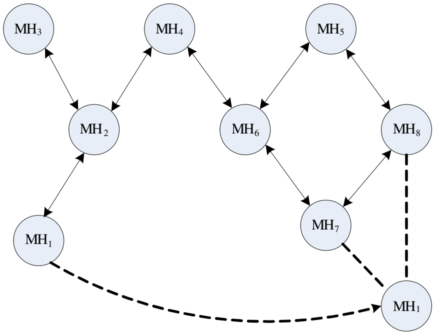


Fig. 2. Movement of mobile hosts in ad-hoc network [6]

intermediate nodes. The algorithm uses *HELLO* messages to detect and monitor links to neighbors. When *HELLO* messages are used, each active node periodically broadcasts a *HELLO* message to its neighbors. If a node fails to receive several messages, then node detects the situation as link break.

2.5 Mobility Models

Mobility model is one of the key parameters that affects the performance of a particular protocol. It describes the movement pattern of mobile user and the change of their location, velocity and acceleration over time [10]. For mobile networks, there are many mobility models suggested by the researchers like Random Waypoint Model (RWP) [14], Random Walk Mobility Model [15], Random Direction Mobility Model [13], Reference Point Group Mobility Model (RPGM)[10] etc. However, due to its simplicity, reliability and closeness to the movement behavior of team workers, RPGM has been chosen as the mobility model for the present study.

2.6 Reference Point Group Mobility Model

In the areas like battlefields, disaster relief, or other mission critical situations, where co-operative mobile robots have got their applications, each team members of the group have to follow the group leader for the successful completion of the mission. Thus, the mobility of the robotic team members are affected by their neighbouring robots. In this scenario, collaborative interaction among the mobile team members in group is best modeled by the reference point group mobility model [5]. In this model, the group has a logical centre or a group leader along-with the other members. The mobility of the group is determined by the movement of the group leader or *reference point*. The movement of the group leader at any instant of time t is represented by the motion vector V^t_{group}. The mobility of the cooperating member of the group is assigned with a reference point that follows the group movement. The motion vector V^t_i of group member i at time t is described in [17] as Equation 1.

$$V^t_i = V^t_{group} + RM^t_i \tag{1}$$

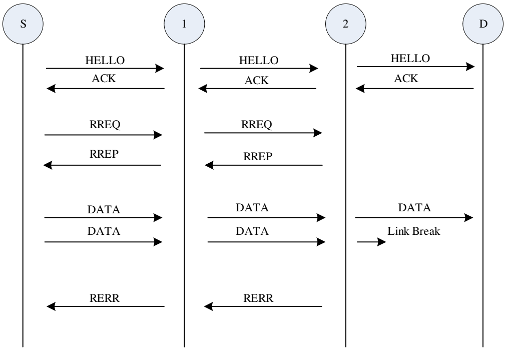


Fig. 3. Message Exchanging in the AODV Protocol

Where, RM_i^t is an independent, identically distributed random process, whose value lies in [0, r_{max}], where r_{max} is the maximum distance deviation allowed and whose direction is uniformly distributed in the interval $[0, 2\pi]$. The advantage of this model is minimization of possibility of link breakage thus enhancing the performance [4]. In the present work, the size of the group is kept constant i.e., 10, irrespective of number of communicating nodes. This has been done with a view to get better performance from cooperative robot teams by avoiding very large group size.

3 Simulations

The simulations in present study are performed with $ns-2$ network simulator [18]. The experiments are carried out with varying number of nodes in multiples of 10. The total time for each simulation is set to 100 seconds. Table 1 provides a summary of the simulation parameters used during the experimentations. The simulation results are analyzed for assessing the performance of DSDV and AODV protocols based on three main performance metrics namely Packet Delivery Fraction, Normalized Routing Load and Average End-to-End Delay.

Table 1. Simulation Parameters

S. No.	Parameters	Value
1	Area Size	1000 x 1000 m^2
2	Mobility Model	RPGM
3	Network Interface	Phy/WirelessPhy
4	Bandwidth	10 Mbps
5	Traffic Type	TCP
6	Max Packet in Queue	1000
7	MAC Protocol Type	MAC/802.11
8	Packet Size	512 Bytes
9	Time of Simulation	100 sec
10	Av. Number of Nodes per Group	10
11	Max. Number of Connections	10

3.1 Packet Delivery Fraction (PDF)

Packet Delivery Fraction is the percentage of total number of packets successfully received by the destination nodes with respect to the number of packets sent by the source nodes during the simulation. PDF is defined as in Equation 2.

$$PDF = \frac{n_{pr}}{n_{ps}} \times 100 \tag{2}$$

Where, *PDF* is packet delivery fraction, n_{pr} is total number of successful packets received by the destination and n_{ps} is total number of packets sent by the source nodes.

PDF indicates the efficiency of the protocol in successful delivery of the packets to the application layer. Higher packet delivery fraction indicates better efficiency of the protocol.

3.2 Normalized Routing Load (NRL)

Normalized Routing Load is the ratio of *number of routing packets transmitted* and *number of data packets actually received*. *NRL* is defined as in Equation 3.

$$NRL = \frac{np_{rout}}{n_{pr}} \tag{3}$$

Where, *NRL* is normalized routing load, np_{rout} is total number of routing packet sent and n_{pr} is total number of data packets received.

NRL estimates the efficiency of routing protocol in maintaining the updated routing information. Higher value of NRL indicates the lower efficiency of the routing protocol.

3.3 Average End to End Delay (AED)

Average End-to-End Delay is defined as the ratio of *end-to-end delay in transmission of data packets* and *number of data packets received*. *AED* is calculated using Equation 4.

$$AED = \frac{\sum_{i=0}^{n}(t_i(r) - t_i(s))}{n_{pr}} \tag{4}$$

Where, AED is average end to end delay, $t_i(r)$ is the receiving time of packet i by the destination node, $t_i(s)$ is the sending time of packet i by the source node and n_{pr} is total number of data packets received.

A higher value of end-to-end delay signifies the higher congestion in the network and thus indicates the lower efficiency of the protocol.

4 Results and Discussions

In the simulation, we varied the number of nodes from 10 through 50, keeping the average number of nodes per group as 10. The simulation was carried out using RPGM mobility model scenario. The same sets of experiments were performed for AODV and DSDV. The simulation was conducted for a period of 100 sec.

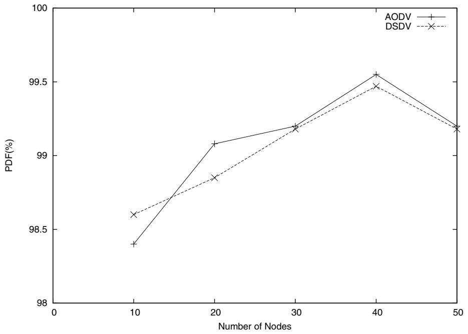


Fig. 4. Performance Comparison between AODV and DSDV in terms of Packet Delivery Fraction

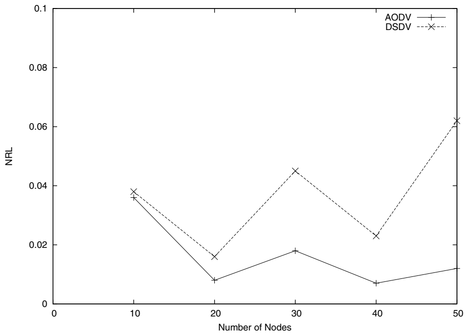


Fig. 5. Performance Comparison between AODV and DSDV in terms of Normalized Routing Load

From the simulation results, as shown in Figure 4, it can be observed that AODV outperforms the value of DSDV in terms of PDF at most of the locations. This shows that the packet delivery efficiency of the AODV is better than DSDV in the given scenario except for the case, when numbers of working groups are restricted to one (number of nodes=10). This might be due to the fact that in this particular case, since there is only one group is present, therefore it eliminates the possibilities of node movement between different groups and hence lessen the packet loss during transmissions.

However, as seen from Figure 5, the value of NRL is more in the case of DSDV compared to AODV at each point. The differences between the values of NRL for the two protocols are also increasing with increase in the number of nodes in the network. This is because, unlike AODV, DSDV has to periodically broadcast its routing table to the other nodes of the network. When the number of nodes increases, this phenomena contributes a substantial number of routing overhead packets. Whereas, in the case of AODV, the route is determined purely on ad-hoc basis as and when required.

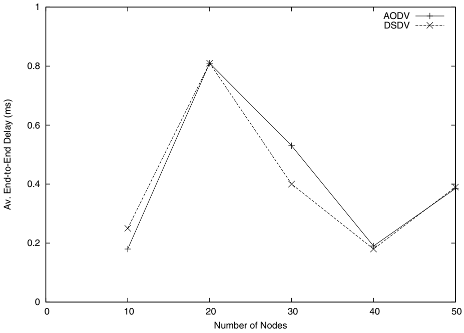


Fig. 6. Performance Comparison between AODV and DSDV in terms of Average End-to-End Delay

Figure 6 shows the performances of AODV and DSDV in terms of average end-to-end delay. From the results, it can be seen that the values of average end-to-end delay for AODV and DSDV are almost comparable with each other at most of the points. This means that the average end-to-end delay in both the cases are almost equal, even if the average number of communicating nodes per group is constant.

The above results shows that AODV exhibits better performances than DSDV, when the number of mobile robot groups are more than 1 (group consists of 10 nodes each). However, if the robots are moving in a single group, the DSDV demonstrates better results.

5 Conclusion

Co-operative ad-hoc mobile robot communication needs mobility, flexibility, reliability and efficiency to perform any mission critical job in successful manner. These requirements of the co-operative ad-hoc robot communication can only be addressed with the help of self-forming, self-healing and self-organizing multi-hop communication network. Mobile ad-hoc wireless network technology with efficient routing protocol is capable of providing the right solution in efficient way. With the help of simulation, the performance analysis of two standard protocols (AODV and DSDV) have been carried out in terms of Packet Delivery Fraction, Normalized Routing Load and Average End-to-End Delay. The results exhibit that AODV protocol outperforms DSDV in RPGM scenario for larger number of groups. However, DSDV shows better efficiency, when the numbers of nodes are less and there is a minimum or no possibility of movement of nodes across the groups. This concludes that use of AODV protocol for the inter-node communication between co-operative ad-hoc mobile robots will provide a better communication performance compared to DSDV in case of multiple teams of robots. But DSDV can be preferred, when the group of mobile robots is restricted to one.

References

1. Wang, Z., Liu, L., Zhou, M.C.: Protocols and applications of ad hoc robot wireless communication networks: An overview. Int. J. Intell. Contr. Syst. 10 (4), 296–303 (2005)
2. Corson, S., Macker, J.: EMobile ad hoc networking (MANET): Routing protocol performance issues and evaluation considerations. IETF RFC 2501 (January 1999), http://www.ietf.org/rfc/rfc2501.txt
3. Royer, E., Toh, C.: A review of current routing protocols for ad hoc mobile wireless networks. IEEE Personal Communication, 46–55 (April 1999)
4. Kulkarni, S.A., Rao, G.R.: Mobility and Energy-Based Performance Analysis of Temporally Ordered Routing Algorithm for Ad Hoc Wireless Network. IETE Technical Review 25(4), 223–227 (2008)
5. Camp, T., Boleng, J., Davies, V.: A Survey of Mobility Models for Ad-hoc Network Research. Wireless Communication and Mobile Computing [WCNC]: Special Issue on Mobile Ad-hoc Networking: Research, Trends and Applications 2(5), 483–502 (2002)
6. Perkins, C.E., Bhagwat, P.: Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for Mobile Computers. In: SIGCOMM 1994, pp. 234–244 (August 1994)
7. Perkins, C.E., Royer, E.M.: Ad Hoc On-Demand Distance Vector Routing. In: 2nd IEEE Workshop on Mobile Computing Systems and Applications (WMCSA 1999), pp. 90–100 (February 1999)
8. Perkins, C.E., Royer, E.M., Das, S.: Ad Hoc On Demand Distance Vector (AODV) Routing, IETF RFC 3561 (2003), http://www.ietf.org/rfc/rfc3561.txt
9. Winfield, A.F.T., Holland, O.E.: The Application of Wireless Local Area Technology to the Control of Mobile Robots. Microprocessors and Microsystems Series 23/10, 597–607 (2000)
10. Saad, M.I.M., Zukarnain, Z.A.: Performance Analysis of Random-Based Mobility Models in MANET Routing Protocol. European Journal of Scientific Research 32(4), 444–454 (2009)
11. Shah, R.C., Rabaey, J.: Energy Aware Routing for Low Energy Ad Hoc Sensor Networks. In: IEEE Wireless Communications and Networking Conference (WCNC 2002), vol. 1, pp. 350–355 (2002)
12. Clausen, T., Jacquet, P.: Optimized link state routing protocol, IETF RFC 3626 (2003), http://www.ietf.org/rfc/rfc3626.txt
13. Royer, E.M., Perkins, C.E.: An Implementation Study of the AODV routing protocol. In: IEEE Conference Wireless Communications and Networking Conference (WCNC 2000), vol. 3, pp. 1003–1008 (2000)
14. Lee, S.J., Su, W., Gerla, M.: On-demand multicast routing protocol in multihop wireless mobile networks. Mobile Networks and Applications 7(6), 441–453 (2002)
15. Jabbari, B., Zhou, Y., Hellier, F.: Random walk modeling of mobility in wireless networks. In: 48th IEEE Vehicular Technology Conference (VTC 1998), vol. 1, pp. 639–643 (1998)
16. Zonoozi, M.M., Dassanayake, P.: User mobility modeling and characterization of mobility pattern. IEEE Journal on Selected Areas in Comm. 15(7), 1239–1252 (1997)
17. Sarkar, S.K., Basavaraju, T.G., Puttamadappa, C.: Ad Hoc Mobile Wireless Networks: Principles, Protocols and Applications, pp. 267–273. Auerbach Publications (2008)
18. The Network Simulator-ns-2, http://www.isi.edu/nsnam/ns/

Enhanced SAFER+ Algorithm for Bluetooth to Withstand Against Key Pairing Attack

Payal Chaudhari[1] and Hiteishi Diwanji[2]

[1] Lecturer, Computer Engg. Department, LDRP – ITR, Gandhinagar, India
ce_payalchaudhary@yahoo.com
[2] Assistant Professor, I.T. Department, LD College of Engg., Ahmedabad, India
hiteishi@hotmail.com

Abstract. The SAFER+-128 algorithm is the basis of all block ciphers used in Bluetooth security environment. Though the SAFER+ with key size 128 bits has yet not any practical attack on it, the pairing mechanism of bluetooth which uses this algorithm founds to be vulnerable to the brute force attack upon it. In the enhanced SAFER+ algorithm the shuffling parameter of SAFER+ has been customized in order to provide it strength against the key pairing attack. The results of the experiments and other statistical details obtained for this change are shown at the end of this paper.

Keywords: Pairing, Unitary Matrix, Primitive Roots.

1 Introduction

The Bluetooth uses two types of encryption algorithms : Block cipher and stream cipher. Block cipher algorithms E_1, E_{21}, E_{22} and E_3 all are used to generate keys and to provide mutual authentication, while E_0 which is a stream cipher is actually used to encrypt the data. All block ciphers are based on "SAFER+" with the key size 128 bit, which is the improved version of one of the AES candidate algorithm - SAFER.

1.1 Use of SAFER+ in Bluetooth

The security provided by the Bluetooth core is built upon the use of symmetric key cryptographic mechanisms for authentication, link encryption, and key generation. A number of different key types are used in connection with these mechanisms. The various key generation algorithms are based upon the SAFER+ structure.

SAFER+ uses a round construction consisting of pseudo-Hadamard transforms, substitution tables, and subkey insertion. An important improvement in SAFER+ is the introduction of the so called Armenian Shuffle permutation, which boosts the diffusion of single-bit modifications in the input data. ([1])

2 Current SAFER+

SAFER+ has two subsystems: the encryption subsystem and the key scheduling subsystem. It shares this setup with many other block cipher algorithms.

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 651–660.
springerlink.com

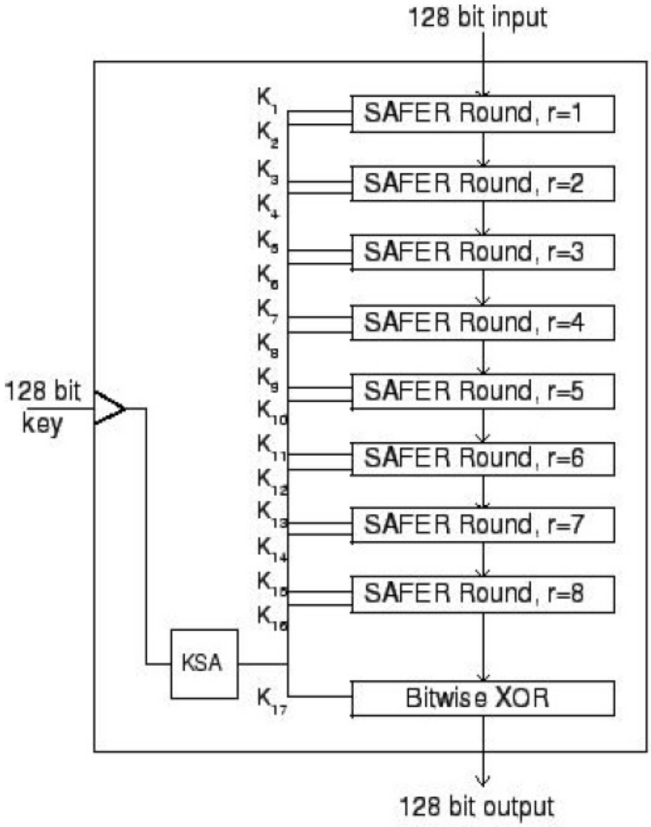


Fig. 1. SAFER+ Design (Source : [2])

2.1 Key Scheduling

The task of key scheduling is to provide key material, called a round key, for each of the encryption rounds in the encryption subsystem. Each round key consists of two vectors of 16 octets. Each byte is cyclic-rotated left by 3 bit positions, and 16 bytes (out of 17) are selected for the output subkey. The last round key, K_{17}, is a single vector of 16 octets that are "added" to the output of the last round.

Each of the 16-octet vectors $K_i = (K_i[0], K_i[1], . . ., K_i [15])$, except K_1, are offset by a bias $B_i = (b_i[0], b_i[1], . . ., b_i[15])$, i = 2, 3, . . ., 17 using modulo 256 addition. The bias vectors are defined by

$$b_i[j] = \left[\left[(45)^{\left(45^{17i+j+1} mod 257\right)} mod\ 257\right) mod\ 256\right]\ for\ j = 0,1, \ldots, 15 \qquad (1)$$

The round keys are fed into SAFER+ round mechanism where they are added into the round data. The addition is done by intertwined modulo 256 and XOR additions.[1]

2.2 Encryption Subsystem

Figure 1 shows the encryption subsystem which consists of 8 identical rounds and at last an output transformation which is implemented as a XOR between the output of the last round and the last subkey.

Single Round Structure. The SAFER+ has block size of 128 bits. All bits are grouped into 16 bytes which are then operated as shown in figure 2.

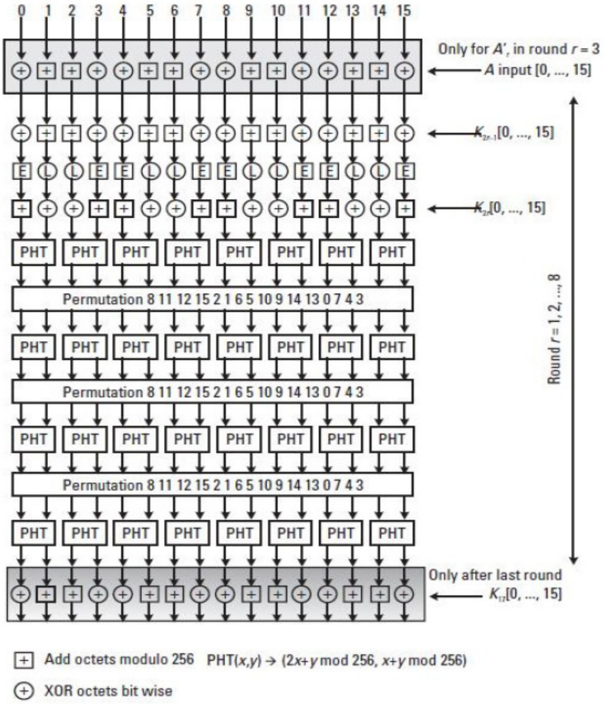


Fig. 2. One Round of SAFER+ with Bluetooth Adoption(Source : [1])

The SAFER+ uses two tables, referred to as E and L, that implement the mappings:

$$E(x) = [(45]^{x} \bmod 257) \bmod 256 \tag{2}$$

$$L(x) = y \text{ such that } x = E(y) \tag{3}$$

These two mappings introduce nonlinearity. Figure 2 also shows the modification of SAFER+ used in the Bluetooth A'r algorithm. A'r has a noninvertible mapping. [1]

Evaluation of Paramters. *Base of E and L Parameter* The 45 has been chosen as a base value on which the operations are carried out to get the substitution tables. 45 is one of the primitive root of 257 which is the prime number nearest to 255 and 255 is the maximum value represented by 128 bits.

PHT Parameter. The PHT parameter provides a 2×2 matrix which when multiply with 2 octets then give as a result other 2 octets. This is a multiplication of a vector with a matrix.

The PHT parameter used is [2 1 ; 1 1]. The matrix used is in fact a unitary matrix. So, theoretically any other unitary matrix can be set in the algorithm. But as the SAFER+ algorithm is going to be used for handheld devices, the time and space complexity puts a restriction on the values of matrix. However without violating the constraints some alternate solutions have been suggested as in [3 - 4].

3 Key Pairing Mechanism of Bluetooth

The Bluetooth pairing operation is crucial in the process of establishing a secure connection (link) between two Bluetooth devices. The procedure starts with establishing an ACL connection.[1]

The pairing procedure consists of the following steps:

- Generating an initialization key
- Generating a link key
- Link key exchange
- Authentication

The key generation procedure is shown in figure 3. At each of the stage a different algorithm is used which is derived from SAFER+ except the last two stages of Constraint key generation and Payload key generation. These last two phases do not use SAFER+ based algorithms.

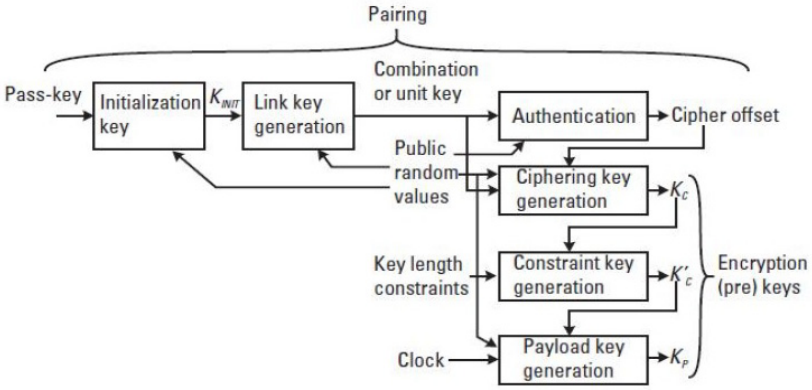


Fig. 3. Overview of Key Management in Bluetooth Pairing (Source : [1])

The various phases of key generation and their corresponding algorithms are shown in table 1.

Table 1. Algorithms for Key Generation

Initialization Key Generation	E_{22}
Link Key Generation	E_{21}
Authentication	E_1
Ciphering Key Generation	E_3
Payload Key Generation	E_0

3.1 Authentication

In the authentication process, a device will take either the role of claimant or verifier. In case of a mutual authentication, the roles will be interchanged in the process.

Suppose device A is the verifier and device B is the claimant. Then A challenges device B by sending the random 128-bit value AU_RAND and expects from B the response

$$SRES = E_1(K, AU\_RAND, BD_{ADDR_B}) \quad (4)$$

Besides the peer authentication, the Bluetooth authentication procedure also results in the creation of the authenticated ciphering offset (ACO). The ACO is used when computing the ciphering key.

4 Attack on Pairing

Two types of attacks are described in [2]. One of which is Primary attack, discussed here.

4.1 Primary Attack

The following is the list of messages transmitted during the key generation and pairing process. Suppose A and B are two devices which are going to establish a connection between them.

1. A random number IN_RAND with size 128 bits is transmitted from A to B in plaintext format.
2. A random number LK_RAND_A with size 128 bits transmitted from A to B, XORed with $K_{init.}$
3. A random number LK_RAND_B with size 128 bits transmitted from B to A, XORed with $K_{init.}$
4. A random number AU_RAND_A with size 128 bits is transmitted from A to B as a challenge for authentication in plaintext format.
5. A signed response SRES with size 32 bits is transmitted from B to A in plaintext format.
6. A random number AU_RAND_B with size 128 bits is transmitted from B to A as a challenge for authentication in plaintext format.

7. A signed response SRES with size 32 bits is transmitted from A to B in plaintext format.

The attacker can now use a brute force algorithm to find the PIN used. The attacker enumerates all possible values of the PIN. Knowing IN_RAND and the BD_ADDR, the attacker runs E_{22} with those inputs and the guessed PIN, and finds a hypothesis for K_{init}. The attacker can now use this hypothesis of the initialization key, to decode messages 2 and 3. Messages 2 and 3 contain enough information to perform the calculation of the link key K_{ab}, giving the attacker a hypothesis of K_{ab}. The attacker now uses the data in the last 4 messages to test the hypothesis: Using K_{ab} and the transmitted AU_RAND_A (message 4), the attacker calculates SRES and compares it to the data of message 5. If necessary, the attacker can use the value of messages 6 and 7 to re-verify the hypothesis Kab until the correct PIN is found [2]. Figure 4 describes the entire process of PIN cracking.

The attack, as described, is only fully successful against PIN values of under 64 bits. If the PIN is longer, then with high probability there will be multiple PIN candidates, since the two SRES values only provide 64 bits of data to test against. But as the users typically have a habit to use a short pin which will have the length less then 64 bits, the probability of above attack remains high.

The primary attack is only applicable if the attacker has eavesdropped on the entire process of pairing and authentication. This is a major limitation since the pairing process is rarely repeated. Once the link key K_{ab} is created, each Bluetooth device stores it for possible future communication with the peer device To make the primary attack possible the opponent must have to reinitiate the pairing process. This is known as repairing attack which is described in [2].

5 Solution to Withstand against Bruteforce Attack

One optimized solution within the boundaries of mobile environment constraints is to randomize the shuffling parameter. However this randomness should be known and predicted at both ends of communication. The simplest shuffling randomness we can provide is using a secure number suppose given a name comm_id. The value of comm_id is added into the position of bytes while doing the shuffling. For example, during shuffling suppose a byte at position i is permuted at position j then we can change the shuffling sequence in the following way.

```
Place byte at position[comm_id + i]
              To position[comm_id + j]
```

The value of comm_id should be incremented with each subsequent session or otherwise with each subsequent transaction. Since the value of comm_id is secure, the bruteforce attack will include more difficulty and complexity. The pin cracking procedure described in section 4.1 will be time consuming now, because now each step in the figure should be repeated at most 16 times.

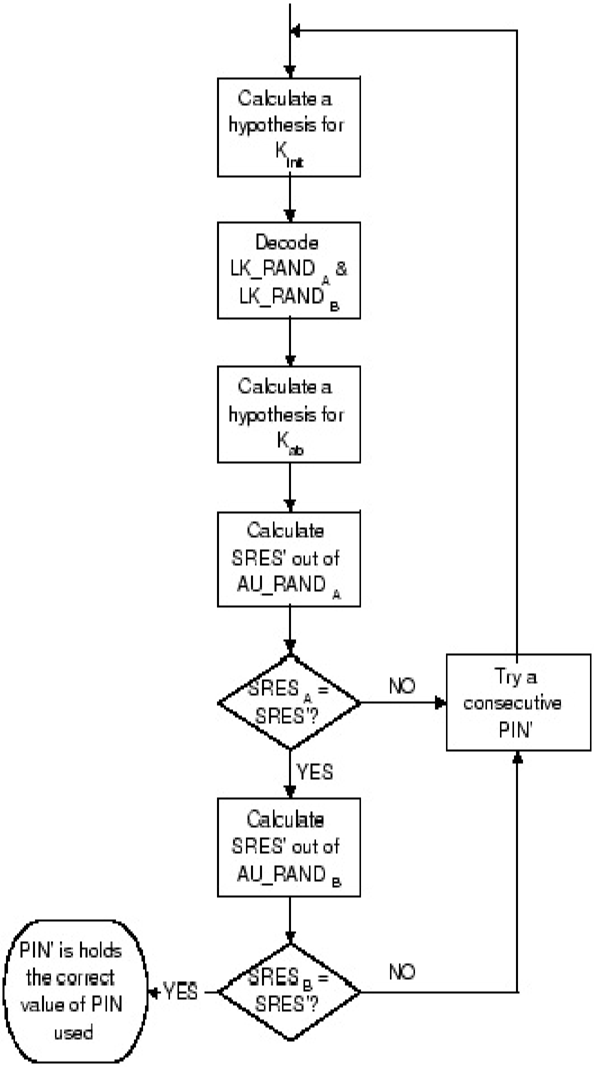

Calculate a hypothesis for K_{init}
Decode LK_RAND $_A$ & LK_RAND $_B$
Calculate a hypothesis for K_{ab}
Calculate SRES' out of AU_RAND $_A$
SRES $_A$ = SRES'?
NO
Try a consecutive PIN'
YES
Calculate SRES' out of AU_RAND $_B$
SRES $_B$ = SRES'?
NO
YES
PIN' is holds the correct value of PIN used

Fig. 4. The Basic Attack Structure (Source:[1])

5.1 Value of comm_id

Primary Solution. In its simplest form the value of comm_id can be based on the length of actual pin. The length of original pin should be given to a formula which can produce any value from 0 to 255. At the first glance we may produce it with either of the following formulas:

$$\text{comm\_id} = \text{pin\_length} \tag{5}$$

or

$$comm\_id = 2^{pin_{length}} mod\ 256 \quad (6)$$

However as the users have the habit to use a pin with 4 to 7 digits, the opponent can try to guess the initial comm_id with static formulas. Hence, a dynamic strategy as shown below can be taken as a solution. Let a number defined with the first two bits and last two bits from the original pin, and give the name to this number as secure id. Then comm_id can be defined as

$$comm\_id = (secure\_id^{pin\_length}) mod\ 256 \quad (7)$$

Since the value of actual value of pin is not known to the attacker, this method can be counted as secure.

5.2 Results

Strength of Customized Algorithm. In the original scheme the Bruteforce attack has effort of O(2^{128}), since the pin length is maximum 128 bits. But in the new approach, it is equal to the effort of O(2^{128} × 256), since 256 values of comm_id for shuffling needs to be considered.

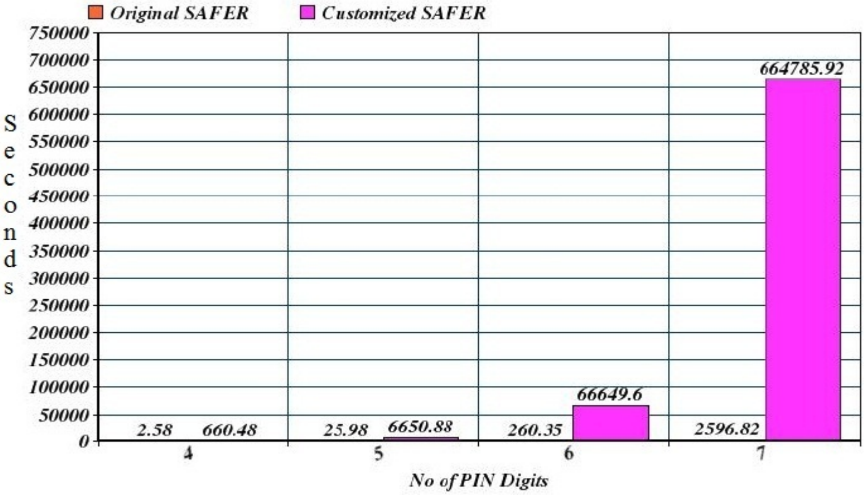


Fig. 5. Time required to crack the algorithm

As per the results of [2] when the pin consists of 7 digits (28 bits), then the pin cracking algorithm as shown in section 4.1 takes the time of 2596.82 seconds (43.28 minutes). According to these results if the approach described in this paper is applied, then the time to break the algorithm will be equivalent to 54332719236956200.00 seconds even with the 7 digit pin, which is considerably large.

The figure 5 shows the comparison of time required to crack the algorithms with the pin digits 4,5,6 and 7. The data is shown in table 1.However even if the opponent can somehow guess the correct PIN', he will not be able to deduce the further keys, as the comm_id is incremented periodically.

Space Complexity. The customized algorithm needs only one more byte to store the value of comm_id, hence the space complexity of both the algorithms will be quite same.

Table 2. Time to crack the Algorithm

No. of Pin Digits	Original SAFER+	Customized SAFER+
4	2.58 sec.	660.48 sec
5	25.98 sec.	6650.88 sec
6	260.35 sec.	66649.60 sec
7	2596.82 sec.	664785.92 sec

Time Complexity. The original and customized both the algorithms are implemented and tested in desktop as well as in mobile handset. Both have the execution time of 0 milliseconds.

Operation Complexity. In the original SAFER+ scheme used in bluetooth has 320 operations in each round**.** With the above stated approaches the additional number of operations to be included is total 98 (33 addition operations, 32 and operations, 32 storage operations).

If the frequency of change of comm_id is low, at least one comm_id value for each execution of full algorithm, then this will create negligible overhead.

6 Conclusion

Among all the versions of SAFER+, SAFER+ - 128 (with key size 128) is found to be most secure and with less implementation complexity. The E and L functions provide nonlinearity. The PHT is used to provide diffusion in output. Alternative matrices for PHT parameter have been already discussed in [3][4].

Though there are not practical attacks available on SAFER+ till today, but the key pairing attack is possible on pairing mechanism which is a bruteforce attack that try to deduce the pin from the ongoing conversation. The randomness in shuffling parameter of SAFER+ gives a good strength to withstand against this attack. The randomness is provided using a comm_id. Two approaches are presented to get comm_id. The simplest approach is to use a pin length as an initial comm_id. The second approach is to use a secure id based on pin value.

The strength as well as the space, time and operational complexity of existing algorithm have been compared which proves the customized algorithm better than the existing SAFER+ algorithm.

References

1. Gehrmann, C., Persson, J., Smeets, B.: Bluetooth Security. Artech House, London (2004)
2. Yaniv, S., Avishai, W.: Cracking the Bluetooth PIN. In: 3rd International Conference on Mobile Systems, Applications, and Services, pp. 39–50. ACM, New York (2005)
3. Sharmila, D., Neelaveni, R.: A Proposed SAFER Plus Security algorithm using Fast Walsh Hadamard transform for Bluetooth Technology. International Journal of Wireless & Mobile Networks (IJWMN) 1(2) (2009)
4. Sharmila, D., Neelaveni, R.: A Proposed SAFER Plus Security algorithm using Fast Psuedo Hadamard Transform (FPHT) with Maximum Distance Separable code for Bluetooth Technology. International Journal of Wireless & Mobile Networks (IJWMN) 1(2) (2009)
5. Eric, G.: A man-in-the-middle attack using Bluetooth in a WLAN interworking environnment. In: 3GPP TSG SA WG3 Security 32, UK (2004)
6. Jonathan, K., Rafail, O., Moti, Y.: Efficient Password-Authenticated Key Exchange using Human-Memorable Passwords. Springer (2001)
7. Markus, J., Susanne, W.: Security Weaknesses in Bluetooth. Springer (2001)
8. Kelsey, J., Schneier, B., Wagner, D.: Key Schedule Weaknesses in SAFER+. In: Second AES Candidate Conference (1999)
9. James, M.: On the Optimality of SAFER+ Diffusion. In: Second AES Candidate Conference. National Institute of Standards and Technology, pp. 22–23 (1999)

Super Peer Deployment in Unstructured Peer-to-Peer Networks

R. Venkadeshan[1] and M. Jegatha[2]

Assistant Professor / CSE Department,
Chettinad College of Engineering & Technology, Karur,
Tamilnadu, India
{venkadengg,vpragatha}@gmail.com

Abstract. Two-layer hierarchy unstructured peer-to-peer (P2P) systems, comprising an upper layer of super-peers and an underlying layer of ordinary peers, are commonly used to improve the performance of large-scale P2P systems. A perfect difference graph has desirable properties to satisfy the above design rationale of super-peers overlay network. This paper proposes a two-layer hierarchical unstructured P2P system in which a perfect difference graph (PDG) is used to dynamically construct and maintain the super-peer overlay topology. In addition, the broadcasting performance of the P2P system is enhanced through the use of a PDG-based forwarding algorithm which ensures that each super-peer receives just one lookup query flooding message. The theoretical results show that the proposed system improves existing super-peer hierarchical unstructured P2P systems in terms of a smaller network diameter, fewer lookup flooding messages, and a reduced average delay and the experimental results show that the proposed two-layer hierarchy P2P system performs very well in the dynamic network environment.

Keywords: Unstructured peer-to-peer system, super-peer, perfect difference graph, forwarding algorithm.

1 Introduction

Peer-to-peer (P2P) overlay networks are massively-distributed ad-hoc computing systems in which the participating peers directly distribute their tasks and share their resources without any form of hierarchical organization or centralized control [1-4]. Such networks offer numerous advantages, including a robust wide -area routing architecture, an efficient search capability, anonymity, excellent fault tolerance, a massive amount of redundant storage, and so forth. Furthermore, since each peer in the system is not only a client, but can also per-form the role of a server, the capacity and scalability of P2P systems are far higher than those of traditional client-server systems. Although various P2P overlay networks have been proposed in recent years, decentralized, unstructured P2P systems such as Gnutella [1] and KaZaA [2] are the most commonly used in current Internet-based applications. In contrast to structured networks, content placement in P2P networks is unrelated to the overlay topology, and thus such networks are better equipped to deal with the problem of high-churn

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 661–669.
springerlink.com

peer populations. KaZaA and the newest version of Gnutella (Gnutella v0.6) both create a two-layer hierarchical unstructured P2P system comprising an upper layer of “super-peers” (KaZaA) or “ultra-peers” (Gnutella) and an underlying layer of ordinary peers. In both systems, the super (or ultra) peers are chosen from amongst the participating nodes having a fast Internet connection and cannot be blocked by a firewall.

2 Related Work

In recent years, various hierarchical two-layer unstructured P2P systems have been proposed as a means of scaling up conventional unstructured P2P systems. Such systems, of which Gnutella vs. 6 [1] and KaZaA [2] are the most widely used, comprise super-peers and ordinary peers and have a number of key advantages for the execution of large-scale distributed applications, including a higher search efficiency and the ability to harness the power and resources of multiple heterogeneous nodes.

Gia improved the performance of unstructured P2P systems by using a dynamic scheme to select appropriate super-peers and to construct the topology around them in an adaptive manner. Furthermore, a search-based random walk mechanism was proposed for directing the lookup messages issued by the ordinary peers towards the high-capacity nodes in the system. In the worst case scenario, the random walk search mechanism either gives up without finding a match or may have to traverse a very long path.

Pyun presented a protocol designated as SUPs for constructing the super-peer overlay topology of scalable unstructured P2P systems using a random graph method. The results showed that SUPs was not only more computationally straightforward than the scheme presented in , but also was much compatible with existing system and was likely be adopted.

Xiao *et al.* [10] presented a workload model for establishing the optimal size ratio between the super-layer and the leaf-layer, and proposed an efficient dynamic layer management (DLM) scheme for super-peer architectures. In the proposed approach, the DLM algorithm automatically selects the peers with larger lifetimes and capacities as super-peers and designates those with shorter lifetimes and capacities as leaf peers.

3 Super-Peer Overlay Networks and Broadcasting Protocols

Since super-peers have a fast Internet connection, they can accommodate a high traffic demand. The topology for super-peers can be modeled by a graph with higher degree, in which vertices represent individual super-peers while undirected edges stand for connections between super-peers. Since all of the super -peers are regarded as being of equal importance in terms of their ability to route traffic, it is suitable to construct the topology of super-peers into a regular graph, which the degree of each vertex is the same, to easily achieve load balancing. Besides balancing the load within the P2P system, it is also desirable to minimize the diameter of the super-peer overlay topology in order to limit the length of the paths which a lookup query generated by any super-peer must traverse to reach the other super-peers in the network. Finally, the degree of the super-peers in the overlay topology should be such that the P2P system is both practical and scalable.

Table 1 summarizes the vertex degree and graph diameter of various well known graph methods. As shown, the complete graph models a regular n-vertex network in which the vertex degree is $d= O\ (n)$ and the diameter is $D=1$. (Note that the diameter indicates the maximum number of hops in the path between the source-destination vertices in the graph.) Although the complete graph provides a simple approach for modeling a net-work, it is impractical for large n and lacks the scalability required to support network growth. Therefore, it is generally preferable to relax the maximum hop-count parameter to D=2 for practical large-scale systems and to model the network using a perfect difference graph (PDG).

Each vertex in a PDG has a degree $O(n)$, and thus the network is significantly more scalable than that modeled by a complete graph (i.e. $O\ (n)$). Furthermore, even though the vertices in the PDG have a lower degree than those in the complete graph, the performance of a PDG-based network is similar to that of a complete graph-based system. In addition, Table 1 shows that the other common graph methods have both a lower vertex degree than the PDG method and a greater diameter. Thus, the PDG-like graph is an ideal solution for the dynamic super-peer overlay construction scheme presented in this study.

3.1 Perfect Difference Graphs

PDGs [8], based on the mathematical notion of perfect difference sets (PDSs), are undirected graphs of degree d=2δ (where δ is the number of elements in the PDS) and diameter D=2.

Table 1. Comparison Of Vertex Degree And Graph Diameter

Vertex Degree	Graph Diameter	Example
O(n)	1	Complete – Graph
O(√n)	2	Perfect Difference Graph
Ω(n ln n)	$\Theta\left(\frac{\ln n}{1\ n \ln n}\right)$	Random Graph
O(log n)	log n	Binary Tree, Hypercube
O(l)	n/2	Ring

Definition 1: A PDG is an undirected inter-connection graph with $n= \delta^2 + \delta + 1$ vertices, numbered 0 to n-1. In the PDG, each vertex i is connected via undirected edges to vertices $(i \pm s_j)(\bmod\ n)$ for $1 \le j \le \delta$, where s_j is an element of the PDS $\{s_1, s_2, \ldots, s_j\}$of order δ .

Table 2 illustrates the number of vertices, the order and the number of elements in the first ten PDSs. Furthermore, the PDS has a degree of 2δ , and thus each vertex has four undirected edges leading to neighboring vertices. For example, vertex 0 has undirected edges leading to vertices (0±1) mod7 and (0±3) mod 7 . In other words, vertex 0 has undirected edges to vertices 1, 3, 4 and 6. For convenience, the following terms are adopted when discussing the PDG methodology in the remainder of this paper:

- Ring edge: the edge connecting consecutive vertices i and $i \pm s_1$ (mod n), where $s_1 = 1$.
- Chord edge: the edge connecting non-consecutive vertices i and $i \pm s_j$ (mod n), $2 \leq j \leq \delta$.
- Forward edges: for vertex i, the forward edges include the chord edge connecting vertices i and $i \pm s_j$ (mod n) and the ring edge connecting vertices i and $i \pm s_1$(mod n) .
- Backward edges: for vertex i, the backward edges include the chord edge connecting vertices i and $i - s_j$ (mod n) and the ring edge connecting vertices i and $i - s_1$(mod n) .

For example, in Fig. 1, the forward edges of vertex 0 are the edges connecting vertex 0 to vertices 1 and 3, respectively, while the backward edges are the edges connecting vertex 0 to vertices 4 and 6, respectively.

Proposition 1. If G=(V,E) is a graph consisting of a set of vertices V and a collection of edges E connecting pairs of vertices in *V,* then $\sum_{v \in V(G)} d(v) = 2e(G)$, where $d(v)$ represents the degree of vertex v in a graph G and $e(G)$ represents the number of edges in G .

Table 2. Correlation between Number Of Vertices, Super-Peer Order And Perfect Difference Sets

n	δ	Perfect difference sets
7	2	1,3
13	3	1,3,9
21	4	1,4,14,16
31	5	1,3,8,12,18
57	7	1,3,13,32,36,43,52
73	8	1,3,7,15,31,36,54,63
91	9	1,3,9,27,49,56,61,77,81
133	11	1,3,12,20,34,38,81,88,94,104,109
183	13	1,3,16,23,28,42,76,82,86,119,137,154,175
273	16	1,3,7,15,63,90,116,127,136,181,194,204,233,238,255

Proof. Summing the degree of vertices counts each edge twice, since each edge has two ends and contributes to the vertex degree at each endpoint [5].

Lemma 1. The total number of edges in a PDG is equal to $n.\delta= (\delta^{2} + \delta + 1) .\delta$.
Proof. Since the connectivity of the PDG leads to a degree $d = 2\delta$, the total degree of vertices equals $\sum_{v \in V(G)} d(v) = n.2\delta$. By Proposition 1, $n.2\delta$ is equal to $2e$. Therefore, the total number of edges is equivalent to $n . \delta = (\delta^{2} + \delta + 1) .\delta$.

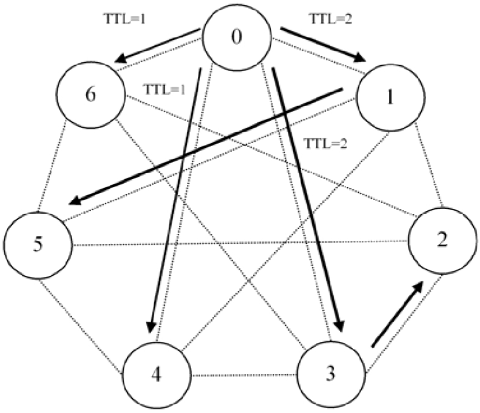


Fig. 1. PDG-based forwarding algorithm

Figure 1 presents a schematic illustration of the pro-posed PDG-based forwarding algorithm for a super-peer overlay network forming a PDG with an order of $\delta = 2$. In this example, it is assumed that super-peer 0 wishes to flood a lookup message to all the other super-peers in the network. In accordance with the two- step procedure described above, super-peer 0 sends a flooding message with TTL= 1 along its backward edges to neighbors 4 and 6, respectively. Since the TTL value is reduced to zero following its decrement upon receipt at these nodes, neighbors 4 and 6 take no further action. Meanwhile, super-peer 0 also sends a flooding message with TTL=2 along its forward edges to neighbors 1 and 3, respectively. Following the receipt of these messages, the TTL value is reduced to 1, and thus both neighbors forward a copy of the message along all their backward edges other than the edge on which they received the original message. In other words, neighbor 1 duplicates the message to node 5, while neighbor 3 copies the message to node 2. Nodes 2 and 5 obtain a value of TTL=0 when decrementing the TTL parameter, and therefore take no further action.

4 System Architecture and Construction

4.1 Extension of Topology to Accommodate New Super-peers

In accordance with the request process algorithm, any peer with a fast Internet connection to enter the P2P net-work as a super-peer issues a joining request with its bandwidth description and IP to the BS server. After identifying the connectivity quality, the BS server accepts the peer as a super-peer, and assigns the new peer the appropriate forward and backward connections. When the number of super-peers is larger than the value $(\delta^{2} + \delta + 1)$, it represents that all the positions in the PDG are already filled with active super-peers. If the re-questing peer is qualified to become a super-peer, the BS server designates the peer as the role of a redundant super-peer, and is allowed to connect to the network by accessing a super-peer with a minimal response time selected from a list randomly compiled by the BS server. When the number of super peers and redundant super-peers increases to threshold value, that can be given as $½[(\delta^{2}+ \delta+1) + (\ell^{2} + \ell +1)]$, there are a number of redundant super-peers existing in the system. In order to utilize the bandwidth capability of the redundant super-peers and increase system scalability, the order of the cur-rent PDS is enlarged to that of the successor PDS and the super-peer overlay topology is extended accordingly. Thus, the BS server first assigns the new joining peer a new vertex ID and the peer IP into the super-peer table. It then assigns the status of 1 to the new joining peer and all of redundant super-peers. Next, the BS server calculates and updates new forward and backward connections based on the new order δ in the super-peer table for these active super-peers.

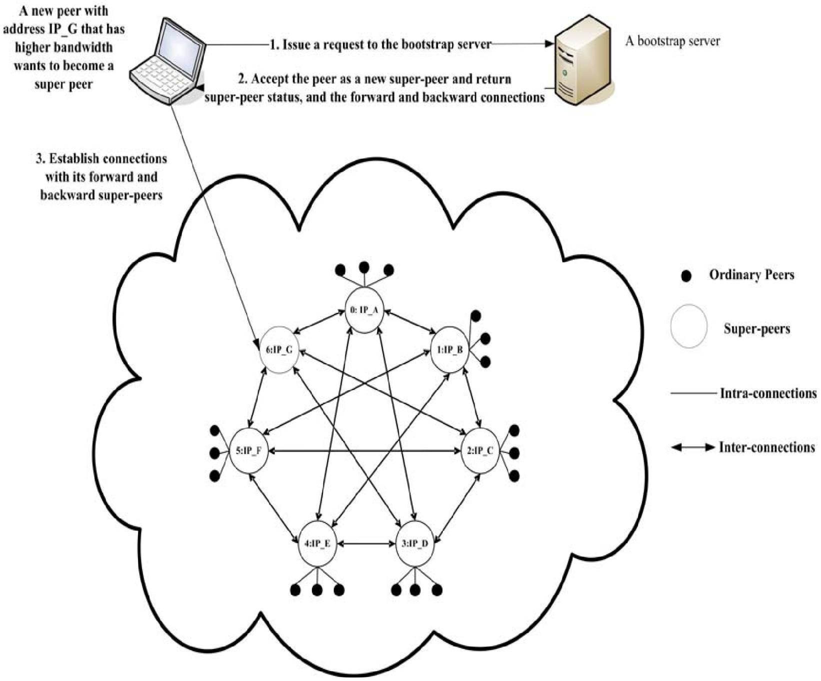

A new peer with address IP_G that has higher bandwidth wants to become a super peer
1. Issue a request to the bootstrap server
A bootstrap server
2. Accept the peer as a new super-peer and return super-peer status, and the forward and backward connections
3. Establish connections with its forward and backward super-peers
0: IP_A
1:IP_B
2:IP_C
3:IP_D
4:IP_E
5:IP_F
6:IP_G
Ordinary Peers
Super-peers
Intra-connections
Inter-connections

Fig. 2. Schematic illustration showing a new peer joining the super-peer overlay networks

Finally, the BS server sends the new joining peer information, such as the status, the for-ward connections and the backward connections. The BS server also notifies redundant super-peers about the status, the forward connections, and the backward connections and informs the original active super-peers about the new forward and backward connections.

We illustrate an example to describe the overlay topology extension. In the initial set-up phase (i.e. no super-peers have yet been identified), the BS server adopts a low-order PDS (i.e. an order of 2) to construct an initial super-peer overlay network for a maximum of 7 super-peers. Assume that there are 10 new peers wishing to be-come super- peers. Since the number of new peers exceeds the number of available spaces in the overlay network, the former 7 peers are assigned as super-peers, and the remaining peers temporarily designated as redundant peers. Later, when a new peer wishing to become a super- peer enters the system, it will result in the number of peers, including active super-peers, redundant super-peers, and the new joining peer, exceeding a threshold 10(= (7+13)/2). The BS server according to the request process algorithm extends the super-peer overlay topology using a PDS with an order of 3, thereby allowing space for a maximum of 13 super-peers. Thus, the redundant super-peers and the new joining peer are assigned as new super-peers and are informed about the IP addresses of their forward and backward connections by BS server. At this point, 11 active super-peers participate in the new enlarged topology.

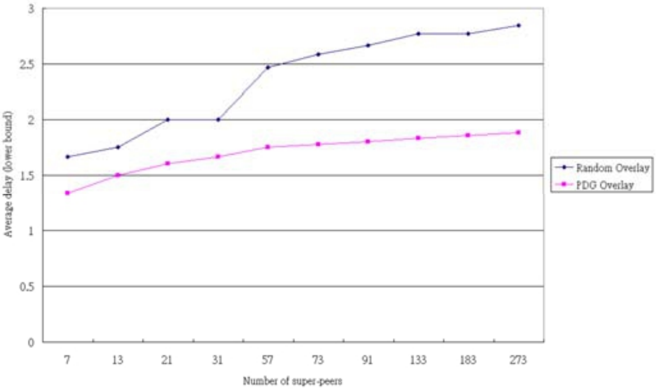


Fig. 3. Comparison of Super-peer degree in random-based and PDG-based overlay networks

In general, the super-peer degree provides an indication of the cost incurred in maintaining the connections of the super-peers in the overlay topology. Thus, Fig.3 shows that the maintenance costs of the super-peers in the proposed PDG-based network are higher than those of the super-peers in the random-based overlay network. Although, formulations for the diameters of a random-based overlay network and a PDG-based overlay network, respectively, the diameter of the random overlay network cannot be precisely determined by the number of super-peers. Therefore, Fig. 3 compares the lower bound of the random-based overlay network diameter with the diameter of the PDG-based network. Although the random-based overlay network diameter represents the best-case scenario for this particular type of network, it can be seen that the diameter of the PDG-based overlay network is significantly smaller at all values of n equal to or greater than 13. The results presented above confirm that the PDG-based forwarding algorithm proposed in this study out-performs the SNC forwarding algorithm used in a conventional random-based super-peer overlay topology in terms of a reduced number of broadcast messages and a lower average hop-count delay.

5 Implementation

To evaluate the file transfer performance of the proposed two-layer hierarchical unstructured P2P system using a perfect difference graph (PDG), we implemented a prototype super-peer and BS server incorporating the request process algorithm presented in Section 4 on our tested. This work presents a series of experimental results to benchmark the performance of the proposed two-layer hierarchical unstructured P2P system against that of a Gnutella hierarchical unstructured P2P system. The initial super-peer overlay topology is constructed by 91 nodes on the testbed with a bandwidth capacity 100 Megabits/sec to form a Gnutella P2P and a PDG-based overlay. The PDG-based overlay topology makes use of PDS with an order of 9 described in Section 3.1, thereby allowing space for a maximum of 91 super-peers. The system performance of the two schemes is quantified by hit rate. The hit rate is defined as the total number of discoveries over the total number of queries. A lookup query can result in multiple discoveries, which are copies of the same files stored at distinct nodes. We allow the system to run several rounds on condition that the number of super-peers equals the shrinking threshold value (e.g. 10). In the beginning of each round, each super-peer issues lookup queries to search files not stored in its local space. Lookup queries are flooded by the PDG-based forwarding algorithm in the PDG-based algorithm in the Gnutella P2P overlay topology with TTL=2, respectively. When each round terminates on the condition that each search request is serviced, one randomly selected super-peer leaves the system and the other active super-peers then enter next round to issue new lookup queries. In the first round, since each overlay topology is a complete and connected graph for these overlay topologies, the hit rate achieves a highest value. Moreover, since the lookup queries on the PDG-based overlay can be efficiently flooded to each super-peer, the total number of discoveries is more than that on the Gnutella P2P overlay. Therefore, the hit rate of the PDG-based overlay is better than the Gnutella P2P overlay.

6 Conclusion

This paper has presented an efficiency technique for constructing and maintaining the super-peer overlay topology of a two layer hierarchical P2P system using a perfect difference graph (PDG) – based method. In addition, a PDG-based forwarding algorithm is proposed for enhancing the efficiency of the lookup process. The performance of the proposed super-peer overlay topology based on a perfect difference graph has been benchmarked against a super-peer overlay topology based on a random graph using SNC forwarding algorithm. The theoretical results have grown that the PDG-based construction scheme and the forwarding algorithm yield a lower network diameter, a reduced number of lookup flooding messages, and a lower average hop-count delay. Through experimental results on our testbed, the proposed PDG-based two-layer hierarchy overlay is an efficient P2P solution in the dynamic network environment

References

1. Gnutella -A protocol for Revolution, http://rfc-gnutella.sourceforge.net.com/
2. KaZaA, http://www.kazaa.com/
3. Overnet/edonkey2000 (2000), http://www.edon-key2000.com/
4. Bittorrent (2003), http://bitconjurer.org/BitTorrent/
5. West, D.B.: Introduction to Graph Theory. Prentice-Hall, Inc. (1996)
6. Bollobás, B.: Random Graphs. Academic Press, London (1985)
7. Kurose, J.F., Ross, K.W.: Computer Networking: A Top-down Approach Featuring the Internet, 3rd edn. Addison Wesley
8. Parhami, B., Rakov, M.: Perfect Difference Networks and Related Interconnection Structures for Parallel and Distributed Systems. IEEE Trans. on Parallel and Distributed Systems 16(8), 714–724 (2005)
9. Parhami, B., Rakov, M.: Performance, Algorithmic, and Robustness Attributes of Perfect Difference Networks. IEEE Trans. on Parallel and Distributed Systems 16(8), 725–736 (2005)
10. Xiao, L., Zhuang, Z., Liu, Y.: Dynamic layer management in superpeer architectures. IEEE Trans. on Parallel and Distributed Systems 16(11), 1078–1091 (2005)
11. Dalal, Y., Metcalfe, R.: Reverse Path Forwarding of Broadcast Packets. Communications of the ACM 21(12), 1040–1048 (1978)
12. Stoica, I., Morris, R., et al.: Chord: A Scalable Peer-to-Peer Lookup Protocol for Internet Applications. IEEE/ACM Trans. on Net. 11(1), 17–32 (2003)
13. Gallager, R.G., Humblet, P.A., Spira, P.M.: A Distributed Algorithm for Minimum Weight-Spanning Trees. ACM Trans. on Programming Languages and Systems, 66–77 (January 1983)
14. Gartner, F.C.: A Survey of Self-Stabilizing Spanning-Tree Construction Algorithms. Technical Report IC/2003/38, Swiss Federal Institute of Technology. School of Computer and Communication Sciences (June 10 2003)
15. Yan, J., Yang, Y., Raikundalia, G.K.: A SwinDeW p2p-based Decentralized Workflow Management System. IEEE Trans. on Systems, Man and Cybernetics, Part A 36(5), 922–935 (2006)
16. Baumert, L.D.: Cyclic Difference Sets. Lecture Notes in Mathematics, vol. 182. Springer (1971)
17. Kirkman, T.P.: On the Perfect r-Partitions of r^2+r+1. Trans. Historical Soc. of Lancashire and Cheshire 9, 127–142 (1857)
18. Guy, R.K.: Unsolved Problems in Number Theory, 2nd edn., pp. 118–121. Springer (1994)
19. Rowstron, A., Druschel, P.: Pastry: Scalable, Decentralized Object Location and Routing for Large-Scale Peer-to-Peer Systems. In: Guerraoui, R. (ed.) Middleware 2001. LNCS, vol. 2218, pp. 329–350. Springer, Heidelberg (2001)
20. Zhao, B.Y., et al.: Tapestry: A Resilient Global- Scale Overlay for Service Deployment. IEEE JSAC 22(1), 41–53 (2004)
21. Lua, K., Crowcroft, J., Pias, M., Sharma, R., Lim, S.: A survey and comparison of peer-to-peer overlay network schemes. IEEE Communications Surveys & Tutorials (2005)
22. Lv, C., Cao, P., Cohen, E., Li, K., Shenker, S.: Search and replication in unstructured peer-to-peer networks. In: ICS (2002)
23. Ratnasamy, S., et al.: A Scalable Content Addressable Network. In: Proc. ACM SIGCOMM, pp. 161–172 (2001)

Efficient Path Selection to Propagate Data Message for Optimizing the Energy Dissipation in WSN

Subrata Dutta[1], Nandini Mukherjee[2], Monideepa Roy[2], and Sarmistha Neogy[2]

School of Mobile Computing and Communication, Jadavpur University
Dept. Of Computer Sc. and Engg, Jadavpur University, Kolkata-700032, India
{subrataduttaa,monideepa.roy}@gmail.com,
{nmukherjee,sneogy}@cse.jdvu.ac.in

Abstract. Original Directed diffusion algorithm chooses the shortest path to transmit data from source node to sink node. Thus, a particular set of nodes are used more frequently leading to energy hole problem. If the message transmission load is distributed considering remaining energy and the remaining path length of a node, then the above problem can be solved. In this paper we suggest a scheme for reducing energy consumption in WSN. The scheme is an extension of the concept introduced in [1].

Keywords: Wireless Sensor Network, Directed Diffusion, Uniform Energy Dissipation.

1 Introduction

Directed Diffusion [2] is a well-known routing algorithm in Wireless Sensor Networks (WSN). The algorithm chooses best path (shortest path) to transmit data from the source node to the destination (sink) node. Thus, a particular set of nodes are used frequently for message transmission. The energy of these nodes may get exhausted and energy hole may be created. This paper explores alternative techniques to avoid such problems. The objective is to balance between minimal energy consumption for data transmission and maximizing network lifetime by using path selection algorithm. Energy dissipation during transmission should be uniformly distributed among the nodes on different routes from source to destination. Here we propose an efficient path selection approach in case of Directed Diffusion algorithm and prove the efficiency of the approach using mathematical logic. The work presented in this paper is motivated by the solution given by An Kyu Hwang et. al [1]. They proposed an algorithm for uniform energy dissipation based on maximum remaining energy in the paths. Our research work proves the efficiency of path selection techniques based on remaining energy in the nodes and provides theoretical basis for choosing the values of different parameters used in the approach. The solution, as presented in [1], provides mathematical model for selection of maximum-remaining-energy path. We name this approach as Uniform Energy Dissipation Directed Diffusion Algorithm (UEDDD). Next, we provide a mathematical model for selection of a minimum remaining-hop-count path from the maximum-remaining-energy group. This approach is named as Approximated Uniform Energy Dissipation Directed Diffusion Algorithm (AUEDDD).

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 671–683.
springerlink.com

Initially, Section 2 points out the limitations of the classical/ original directed diffusion algorithm. Section 3 describes MREDD algorithm as proposed in [1]. UEDDD is implemented as part of MREDD and its advantages are described in Section 4. Section 5 point out the limitations of UEDDD and an approximated UEDDD algorithm (AUEDDD) is described in Section 6. Section 7 lists the advantages of the AUEDDD algorithm. Section 8 describes the expected results in terms of theoretical analysis considering different limiting cases for the above mentioned algorithms.

2 Limitation of Classical Directed Diffusion Algorithm

In case of the Classical Directed Diffusion (CDD) algorithm, the path for data transmission is established on the basis of the shortest hop count. Thus, the repetitive use of some particular nodes may cause those nodes running out of energy very fast. In this way some energy holes may be created and sensor nodes become unable to collect event data in these areas. The network may also become partitioned. As a result, the effective lifetime of a WSN reduces.

Let us assume that the total energy dissipation by the nodes on the optimal path for q number of packets of a message transmission (in case of the CDD algorithm) is E_T^{cl}. We have already shown [3] that the value of E_T^{cl} can be calculated as:

$$E_T^{cl} = E_G + E_{RI} + E_{DATA}$$

$$\text{or, } E_T^{cl} = (c_{s\_th} \sum_{i=2,j=1}^{H_{opt},H_{opt}-1} d_{ij}^n / k + c_r H_{opt})(q+3) \tag{1}$$

where c_{s_th} is the amount of threshold energy of a signal at the receiving end, k is the permittivity constant, n is the path loss coefficient, c_r is the energy consumption for receiving a packet, H_{opt}^{j} is the optimum hop count from the j^{th} node to the destination node, q is the total number of packets constituting each message. The number of nodes on the path with hop count H_{opt}^{j} is $(H_{opt}^{j} + 1)$. Here also E_G is the energy dissipation of the optimum path at the time of gradient setting, E_{RI} is the energy dissipation at the time of reinforcement of path and E_{DATA} is the energy dissipation at the time of data transmission. Let us assume that the total energy dissipation for q number of packet transmissions per node through j^{th} node is e_j^{cl}.

$$e_j^{cl} = \{(c_{s\_th} \sum_{i=2,j=1}^{H_{opt}^{j}H_{opt}^{j}-1} d_{ij}^n / k + c_r H_{opt}^{j})(q+3)\}/(H_{opt}^{j} + 1) \tag{2}$$

Let us assume that the total energy capacity per node is $e_{\max}$ and the remaining energy of node j is e_{rem}^{j}. Thus we can write

$$e^{j}_{rem} = e_{\max} - e^{cl}_{j}$$

$$e^{j}_{rem} = e_{\max} - \{(c_{s\_th} \sum_{i=2,j=1}^{H_{opt},H_{opt}-1} d^{n}_{ij}/k + c_r H^{j}_{opt})(q+3)\}/(H_{opt}+1) \tag{3}$$

Let time for one packet to reach from the source node to the sink node be denoted by T_{wsn} and the hop count measurement from the source to the destination be denoted by H_{wsn}. If we ignore other constraints (such as route congestion of any node) we can say that T_{wsn} is directly proportional to the H_{wsn} [4]. So, $T_{wsn} = H_{wsn} t$ [Since $T_T \alpha H_{wsn}$ and t is the time required for one data packet to traverse over one hop]. So we can say that the time required for the data packet to traverse the hop count H^{j}_{opt} is $H^{j}_{opt} t$. If the rate of decrease of energy per node on the optimal path for transmitting a message which contains q number of packets through node j is $e^{cl}_{rate\_j}$ then

$$e^{cl}_{rate\_j} = e^{cl}_{j} /\{H^{j}_{opt} t\} \tag{5}$$

$$e^{cl}_{rate\_j} = (c_{s\_th} \sum_{i=2,j=1}^{H^{j}_{opt},H^{j}_{opt}-1} d^{n}_{ij} / k + c_r H^{j}_{opt})(q+3) /\{(H^{j}_{opt} + 1) H^{j}_{opt} t\} \tag{6}$$

If q is large, then the total energy of the path with path length H^{j}_{opt} will run out soon and an energy hole will be created.

3 A Modified Directed Diffusion Algorithm for Uniform Energy Use

A solution to the problem is provided by An Kyu Hwang et. al [1]. The probability that a specific node j selects node i as a next hop is given by

$$P_{ij} = \{1/C_{ji}\} /\{\sum_{l \in FGT_j} 1/C_{jl}\} \tag{7}$$

where C_{ji} is the cost of message transmission between node j and node i and FGT_j is the Forward Gradient Table(Table 1) of node j. Here Figure 1 represents the scenario,

$$C_{ji} = \{e_{ji}\}^{\alpha} \{e^{i}_{rem}\}^{\beta} \tag{8}$$

where e_{ji} is the amount of energy consumed to transfer one packet from node j to node i and e^{i}_{rem} is the remaining energy of node i, and α , β are constants. The Forward Gradient Tables are established at the time of Gradient setting.

Table 1. Forward Gradient Table

Next Node ID	Remaining Energy of Corresponding Next Node	Average Remaining Hop-Count	Distance Between Current Node to Next Node ID

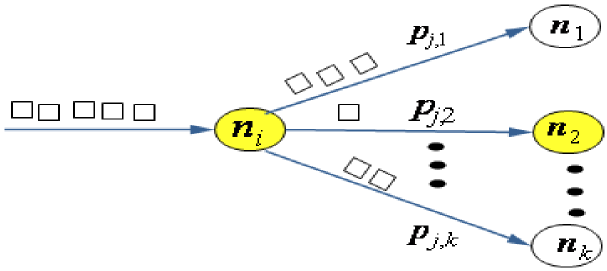


Fig. 1. [1]: Probabilistic selection of next hop in the EAR algorithm

If the distance between node i and node j increases then the energy dissipation for transmitting a packet between node i and node j will also increase. Since the energy dissipation e_{ji} is directly proportional to the cost C_{ji} therefore from (8) we can say $\alpha > 0$. Since e^{i}_{rem} is the remaining energy of any node i from the Forward Gradient Table (Table 1) j, then according to the algorithm if the remaining energy of a particular node from the Forward Gradient Table (Table 1) is high then the cost of sending a packet between node i and node j will also be low. Thus e^{i}_{rem} is inversely proportional to the cost of transmission through the path from j to i and for that if we consider the value of $\beta > 0$ then the equation (8) will be

$$C_{ji} = \{e_{ji}\}^{\alpha} / \{e^{i}_{rem}\}^{\beta} \text{ where } \alpha, \beta \geq 0 \tag{9}$$

According to the modified Directed Diffusion algorithm proposed by An Kyu Hwang et. al[1], the probability of choosing the next node for packet transmission is calculated as per (7) and the cost of transmission of a packet is calculated as per (8). The Maximum Remaining Energy Constrained Directed Diffusion Routing (MRE-DD) algorithm actually selects a least energy consuming path among the paths formed by nodes with highest remaining energy. However as the least energy consuming path is chosen from among them, therefore it does not always guarantee that the maximum remaining energy nodes are always chosen. They claim to make a tradeoff between the uniform energy dissipation of every node in the network and total energy dissipation of the whole network for a particular message transmission. In short we can say that in case of the modified Directed Diffusion algorithm no route is predefined for packet transmission. The next node for packet transmission is chosen by each node, based on probability (7).

4 Advantage of the UEDDD Algorithm over the Classical Directed Diffusion Algorithm

To traverse q number of data packets, more than one route can be followed, and among these paths, each path can be used multiple number of times. Suppose the q number of packets in a data message follow p number of routes while travelling from the current node j to the destination, where $p \leq q$. These set of routes can be denoted as

$$S^{j}_{route} = \{R_{j1}, R_{j2},..., R_{ji},...., R_{jp}\} \text{ where, } 1 \leq i \leq p \tag{10}$$

and R_{ji} indicates the set of all possible routes from node j to the destination node, through node i, where the average remaining hop count is H^{i}_{j} and v_i is the number of times that a packet has traversed through R_{ji} routes. Thus j being the current node and having i as its neighbor node, we define H^{i}_{j} as the average remaining hop count from node j where i is the next node. Then, $H^{i}_{j} = \sum_{k=1}^{n_i} (H^{k}_{i} + 1) / n^{F}_{i}$ where H^{k}_{i} is the remaining hop count from the i^{th} node to the destination node through node k and the k^{th} node is the next selected node from node i. The i^{th} node is the neighbor node of the j^{th} node and n^{F}_{i} is the total number of entries in the Forward Gradient Table (Table 1) of node number i. We assume that any i^{th} node is chosen v_i times while sending q messages and $q = \sum_{i=1}^{p} v_i$ where v_i is the total number of times that node j selects node i as the next node. Let H^{j}_{opt} be the minimum path length from node j to the destination node. Let the average length of all the paths from node j to the destination, via the i^{th} node during this transmission be H^{j}_{avg}. So,

$$H^{j}_{avg} = \sum_{u=1}^{p} v_u H^{u}_{j} / \sum_{u=1}^{p} v_u \tag{11}$$

From (10) and (11) we can write $H^{j}_{avg} \geq H^{j}_{opt}$.

$$\text{Let } \gamma = H^{j}_{avg} / H^{j}_{opt} \text{ where } \gamma \geq 1 \tag{12}$$

Let N^{j} is the number of distinct nodes who participated in the q number of packet transmissions from node j to destination node where

$$H^{j}_{avg} + 1 \leq N^{j} \tag{13}$$

$$\text{Let } \psi = N^{j} / (H^{j}_{avg} + 1) \text{ then } \psi \geq 1 \tag{14}$$

Let e^{cl}_j and e^{uni}_j be the respective average total energy dissipations per node for the CDD algorithm and Uniform Energy Dissipation Directed Diffusion (UEDDD) algorithm, for sending a total of q number of packets through different paths starting from node j to the destination node. Therefore the equation for the e^{uni}_j will be

$$e^{uni}_j = \{(c_{s\_th} \sum_{i=2,j=1}^{H^j_{avg}+1, H^j_{avg}} d^n_{ij} / k + c_r H^j_{avg})(q+3)\} / N^j \tag{15}$$

If we consider the that sensor nodes are distributed uniformly over the network region and the distance between any two nodes are the same,

then we can write $d_{ij} = d_{uv}; \forall i, j, u, v$ where $i, j, u, v \in \{1,2,3,...N\}$ (16)

Therefore we can replace d^n_{ij} with a constant K_d and rewrite equation (2) as

$$e^{cl}_j = (K_d c_{s\_th} / k + c_r)(q+3) H^j_{opt} / (H^j_{opt} + 1) \tag{17}$$

Therefore equation (15) can be rewritten as

$$e^{uni}_j = (K_d c_{s\_th} / k + c_r)(q+3) H^j_{avg} / N^j \tag{18}$$

It can be proved that $H^j_{opt} / (H^j_{opt} + 1) \le H^j_{avg} / N^j$. Thus, $e^{cl}_j > e^{uni}_j$. Now if we UEDDD algorithm is applied, then the total energy dissipation per node and the rate of energy dissipation per node are less than in case of the CDD algorithm. For q number of packet transmissions, let the cumulative averages of the total energy dissipation per node in case of the CDD and the UEDDD algorithms be e^{cl}_T and e^{uni}_T respectively. Therefore the expression for e^{cl}_T and e^{uni}_T will be

$$e^{cl}_T = \sum_{j=1}^{H^j_{opt}+1} e^{cl}_{T\_j} / (H^j_{opt} + 1) \text{ and } e^{uni}_T = \sum_{j=1}^{N^j} e^{uni}_{T\_j} / N^j$$

As j is any node in the message transmission path in the WSN so the above relation is true for all j. Since $H^j_{opt} \le H^j_{avg}$, $H^j_{avg} + 1 \le N^j_{route}$ and $e^{cl}_j \ge e^{uni}_j$ therefore we can say that:

$$\sum_{j=1}^{H^j_{opt}+1} e^{cl}_{T-j} / (H^j_{opt} + 1) \ge \sum_{j=1}^{N^j} e^{uni}_{T\_j} / N^j$$

$$\text{or, } e^{cl}_T \ge e^{uni}_T \tag{19}$$

Let us denote $e^{cl}_{rate\_j}$ to be the rate of energy dissipation (from node j to the destination) per node in case of the CDD algorithm and $e^{uni}_{rate\_j}$ to be the rate of energy dissipation per node in case of the UEDDD. As per (5) we can write

$$e^{cl}_{rate\_j} = e^{cl}_j / \{H^j_{opt}\ t\}$$

Similarly the expression for (20)

$$e^{uni}_{rate\_j} \text{ will be } e^{uni}_{rate\_j} = e^{uni}_j / \{N^j t\} \tag{21}$$

Since

$$1/H^j_{opt} \geq 1/N^j \text{ and } e^{cl}_j \geq e^{uni}_j \tag{22}$$

Therefore from (20), (21) and (22) we can write

$$e^{cl}_{rate\_j} \geq e^{uni}_{rate\_j} \tag{23}$$

Let the cumulative averages of the rates of energy dissipation per node in case of the CDD algorithm and the UEDDD algorithm be e^{cl}_{rate} and e^{uni}_{rate} respectively for q number of packet transmissions. In a way similar to equation (23) it can be proved that $e^{cl}_{rate} \geq e^{uni}_{rate}$. From (24) we can say that if the UEDDD algorithm is used then the energy dissipation rate per node will decrease and hence the time of first failure of any sensor node will also increase.

5 Limitation of UEDDD Algorithm

From Section 3 we came to know that the hop count from the source node to the destination node of the UEDDD algorithm is greater than the hop count for the CDD algorithm. Let us assume that when q number of packets are transmitted from node j to the destination, the total energy dissipation using the CDD and the UEDDD algorithms will be E^{cl}_j and E^{uni}_j respectively.

$$E^{cl}_j = e^{cl}_j (H^j_{opt} + 1)$$

From (17) we can say

$$E^{cl}_j = (K_d c_{s\_th} / k + c_r)(q+3)(H^j_{opt}) \tag{24}$$

$$E^{uni}_j = e^{uni}_j N^j$$

From (18) we can say

$$E^{uni}_j = (K_d c_{s\_th} / k + c_r)(q+3) H^j_{avg}$$

$$\gamma = E^{uni}_j / E^{cl}_j \text{ (since from (10) } \gamma = H^j_{avg} / H^j_{opt} \text{)} \tag{25}$$

Thus we can say that if

$$H^j_{avg} >> H^j_{opt} \text{ then } \gamma >> 1 \text{and } E^{uni}_j >> E^{cl}_j \tag{26}$$

Let the cumulative averages of the total energy dissipation of WSN in case of the CDD algorithm and the UEDDD per node be E^{cl}_T and E^{uni}_T respectively for q number of packet transmissions. Like (23) we can say $E^{uni}_T \geq E^{cl}_T$. From the above analysis it could be said that if the value of H^j_{avg} is much larger than H^j_{opt} then the energy dissipation will be more with respect to the CDD. Intuitively also we can say that as the minimum path is not taken in the UEDDD algorithm, the total energy dissipation in the entire WSN will increase, although the energy will dissipate uniformly in the WSN.

6 Approximate UEDDD (AUEDDD) Algorithm

Usually, in WSN, all the nodes have sensing capability, which means that every node can act as a source node. Thus, if we can minimize the average path length (H_{avg}^{j}), then the total energy dissipation to transmit q number of data packets will reduce. In this section, we propose a technique for reducing H_{avg}^{j}. Suppose, when the probabilities of selection of the next node are calculated for all the nodes, in the Forward Gradient Table (Table 1) of node j, let the maximum probability from among them be $p_{j}^{\max}$. Let us assume that if the node x with probability $p_{j}^{\max}$ is selected, then the average remaining hop count from node j to the destination will be H_{avg}^{x}. If the Forward Gradient Table (Table 1) of j contains more than one node, then there will exist another node y whose probability of being selected is $p_{j}^{'}$. The average remaining hop count from node j to the destination, if node y would have been selected is H_{avg}^{y}. If $H_{avg}^{y} << H_{avg}^{x}$, then choosing the path through y is a much more logical decision than choosing x as the next node of j, although the probability of choosing y is slightly less than the probability of choosing x. In the above case, we propose that if the difference between the above two probabilities is below a threshold value, the second path may be chosen. We define a threshold value ε, and within the range $p_{j}^{\max}$ and $p_{j}^{\max} - \varepsilon$, there may exist r number of nodes $N_{y}\,(1 \le y \le r)$, which may be selected if $H_{avg}^{y} < H_{avg}^{x}$. Any node which has the probability of choosing the next node to be less than $p_{j}^{\max} - \varepsilon$, will not be considered.

Let,

$$H_{avg}^{x} - H_{avg}^{y} = \rho \tag{27}$$

If the data message follows the path with length H_{avg}^{y} instead of the path with length H_{avg}^{x}, where $(H_{avg}^{x} > H_{avg}^{y})$, then the energy that can be saved for transmitting q number of data packets is E_{save}^{j}. For different data packets the value of ρ might be different. We assume $\varphi = \sum_{k=1}^{q} \rho_{k}$ where $\rho_{k} \ge 0$ for all $k \in \{1,2....q\}$

Hence

$$E_{save}^{j} = (c_{s\_th} \sum_{i=2,\,j=1}^{\varphi+1,\,\varphi} d_{ij}^{n} / k + c_{r}\varphi) \tag{28}$$

Suppose after sending $(k - 1)$ number of data packets along the path length $H_{uni}^{j\_next}$ the probability difference is exactly ε. So after sending of the k^{th} data packet, the probability difference exceeds the value of ε such that $(p_{uni} - p_{uni}^{j} \geq \varepsilon)$. In that case to preserve the concept of uniform energy usage we have to choose another path within the probability range $p_{uni}^{j} < p < p_{uni}^{j} - \varepsilon$.

7 Advantages of the Approximate UEDDD Algorithm

As discussed in section 3 the transmission cost of one data packet from node j to node i is given by equation (8) $C_{ji} = \{e_{ji}\}^{\alpha} / \{e_{rem}^{i}\}^{\beta}$. We assume that the average distance between the two neighboring nodes of any two paths are the same. Then we can write that the energy consumption for transmitting one packet from node j to node i is constant, i.e. $e_{ji}^{\alpha} = k_c$ where k_c is a constant. Therefore,

$$C_{ji} = k_c / \{e_{rem}^{i}\}^{\beta} \tag{29}$$

From (7) and (29)we can write

$$P_{ji} = \{e_{rem}^{i}\}^{\beta} / \sum_{l = FGT_j} (e_{rem}^{l})^{\beta} \tag{30}$$

Let us assume that Figure 2 is an example of a subset of a network topology of a WSN where the thick lines represent the line of data transmission for a particular case among all the possible cases. We consider a node which transmits data along with its Forward Gradient Table (Table 1) nodes to constitute the sub component of the network. Here, in Figure 2 the node j transmits data and its sub component also named j consists of nodes $\{1,2,3,j,i\}$. From the Forward Gradient Table (Table 1), node j only knows about the neighbor nodes, the remaining energy of the neighbor nodes and the approximate hop count from node j to the destination node. When node j computes the probability of choosing the next node to transfer data, then it considers the summation of the factors $(e_{rem}^{l})^{\beta}$ where $l \in FGT_j$ of the neighbor nodes of, j which has a fixed value at any particular instance of time, with respect to the remaining energy of each individual node of that Forward Gradient Table (Table 1).

Thus, $P_{ji} = \{e_{rem}^{i}\}^{\beta} \Psi^{j}$ where $\Psi^{j} = 1 / \{ \sum_{l = FT_j} (e_{rem}^{l})^{\beta} \}$ (31)

So the value of Ψ^{j} remains unchanged at any particular instant of time with respect to each numerator value representing the remaining energy of any individual node from the Forward Gradient Table (Table 1). The value of this constant (Ψ^{j}) may vary from one node to another node. As per the previous discussion, in any particular instance of time the value of Ψ^{j} is relatively fixed with respect to the remaining energy of the neighbor nodes of node j.

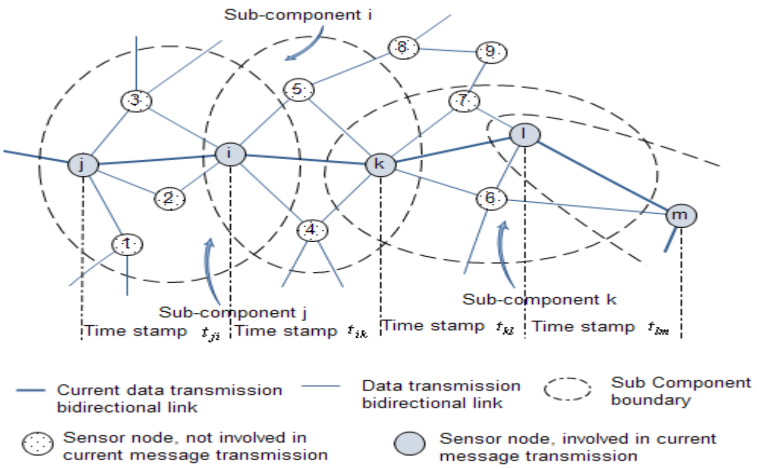

Sub-component i
3
5
8
9
7
l
j
i
k
2
6
m
1
4
Sub-component j
Sub-component k
Time stamp t_{ji}
Time stamp t_{ik}
Time stamp t_{kl}
Time stamp t_{lm}
Current data transmission bidirectional link
Data transmission bidirectional link
Sub Component boundary
Sensor node, not involved in current message transmission
Sensor node, involved in current message transmission

Fig. 2. Packer transmission from one sub components of network to another

Suppose we choose two different values of ε , for a particular node j , given by ε_j^{low} and ε_j^{high} where $\varepsilon_j^{low} \le \varepsilon_j^{high}$. The total number of choices for selecting the next node to which the current data packet is to be sent is r_{low}^j when $\varepsilon = \varepsilon_j^{low}$. Similarly, when $\varepsilon = \varepsilon_j^{high}$ then the total number of choices for selecting the next node to which the current data packet is to be sent is r_{high}^j . The set of the different routes from node j to the destination is $S_{route}^j = \{R_{j1}, R_{j2}, ..., R_{ji}, ..., R_{jr}\}$. When $\varepsilon = \varepsilon_j^{low}$ then $r = r_{low}^j$, and for that case we denote $S_{route}^j = S_{low}^j$. Similarly when $r = r_{high}^j$ then $\varepsilon = \varepsilon_j^{high}$ and for that case we denote $S_{route}^j = S_{high}^j$. It is clear that the number of possible routes for larger threshold values will be greater than the number of possible routes for smaller threshold values. Therefore we can say that $S_{low}^j \subseteq S_{high}^j$. Let $H_{low}^{j\_\min}$ is the minimum remaining hop count from node j to the destination node when $\varepsilon = \varepsilon_j^{low}$ and similarly $H_{high}^{j\_\min}$ is the minimum remaining hop count from node j to the destination node when $\varepsilon = \varepsilon_i^{high}$. Now $H_{low}^{j\_\min} = \min\{H_u^j\}$ where $1 \le u \le r_{low}^j$ and $H_{high}^{j\_\min} = \min\{H_u^j\}$ where $1 < u < r_{high}^j$. As $S_{low}^j \subseteq S_{high}^j$ so there is a greater chance that a route with an even lower hop count may be included in S_{high}^j which is not in S_{low}^j

Therefore,

$$H_{high}^{j\_\min} \le H_{low}^{j\_\min} \tag{32}$$

Let

$$\rho_{high}^j = H_{uni}^j - H_{high}^{j\_\min} \text{ where } S^j = S_{high}^j \tag{33}$$

and

$$\rho_{low}^{j} = H_{uni}^{j} - H_{low}^{j\_\min} \text{ where } S^{j} = S_{low}^{j} \tag{34}$$

The amount of total energy saved for choosing a route with lower hop count (for threshold value ε_{j}^{high}) is $E_{high}^{j\_save}$

$$E_{high}^{j\_save} = (c_{s\_init} + c_{r})\rho_{high}^{j} \tag{35}$$

The amount of total energy conserved for choosing a route with lower hop count (for threshold value ε_{j}^{low}) is $E_{low}^{j\_save}$.

$$E_{low}^{j\_save} = (c_{s\_init} + c_{r})\rho_{low}^{j} \tag{36}$$

From (32), (33) and (34) we can write

$$\rho_{high}^{j} \geq \rho_{low}^{j} \tag{37}$$

From (35), (36) and (37) we can say

$$E_{high}^{j\_save} \geq E_{low}^{j\_save} \tag{38}$$

From the above discussion we can say that since

$$\varepsilon_{j}^{high} > \varepsilon_{j}^{low} \text{ then } E_{high}^{j\_save} \geq E_{low}^{j\_save} \tag{39}$$

Therefore we can say that if we choose a greater threshold value of probability difference, then the total energy saved could be more.

7.1 Conformance of UEDDD Algorithm with AUEDDD Algorithm

We denote the probabilities of choosing nodes x and y as the next nodes for sending data from node j as p_{j}^{x} and p_{j}^{y} respectively. We also denote e_{rem}^{x} and e_{rem}^{y} to be the remaining energy of the nodes x (and the probability of choosing that node to be p_{uni}^{x}) and y (and the probability of choosing that node to be p_{uni}^{y}) respectively. As mentioned earlier, both the nodes x and y are the members of the Forward Gradient Table (Table 1) of j (FGT_{j}) . From (39) we can say that a greater value of ε implies that more energy will be saved. Therefore from the previous assumption we can write

$$p_{uni}^{x} - p_{uni}^{y} \leq \varepsilon \tag{40}$$

$$(e_{rem\_uni}^{jx})^{\beta} - (e_{rem\_uni}^{jy})^{\beta} \leq \varepsilon \sum_{k \in FT_{j}} \{e_{rem}^{jk}\}^{\beta}$$

Hence, $\{e_{rem\_avg}^{j}\}^{\beta} = \sum_{x=1}^{r} \{e_{rem}^{jx}\}^{\beta} / r$ where $e_{rem\_avg}^{j}$ is the power average of $\{(e_{rem}^{jx})^{\beta}\}$ where $1 \leq x \leq r$

$$(e^{jx}_{rem\_uni})^{\beta} - (e^{jy}_{rem\_uni})^{\beta} \leq r\varepsilon(e^{j}_{rem\_avg})^{\beta} \tag{41}$$

When $e^{jy}_{rem\_uni} \approx 0$ and $(e^{jx}_{rem\_uni})^{\beta} - (e^{jy}_{rem\_uni})^{\beta} \leq \varepsilon r(e^{j}_{rem\_avg})^{\beta}$ then we can say

$$(e^{ji}_{rem\_uni})^{\beta} \leq r\varepsilon(e^{j}_{rem\_avg})^{\beta} \tag{42}$$

It is obvious that as the energy in individual neighboring nodes decreases, the average remaining energy also decreases. As the value of $e^{jy}_{rem\_uni}$ is almost zero, so the value of $e^{j}_{rem\_avg}$ will decrease. Also if the value of ε is very small with respect to the other parameters of equation (42) then the product of ε and $r(e^{j}_{rem\_avg})^{\beta}$ will become low. So, it can be said that the value of e^{jx}_{rem} is small when the value of $e^{jy}_{rem\_uni} = 0$. Therefore if the value of ε increases then the total amount of energy conservation will increase but the uniform energy dissipation rule will be violated to some extent. Thus there should be an optimum solution for choosing ε and also α and βin terms of known parameters to get the optimum energy dissipation.

8 Conclusion

When an application in a Wireless Sensor Network implements Directed Diffusion routing algorithms after setting the gradient, many paths are established between the source node and the sink node. In case of the classical Directed Diffusion algorithm the optimum path (the path with the minimum path length) is reinforced. If the same set of nodes is selected repeatedly for the packet transmission then the remaining energy of those nodes would run out very easily leading to the partitioning of the network. There is a high chance that some set of nodes would run out of energy first, whereas some other sets of nodes would still have enough energy. So the effective lifetime of the WSN would decrease. For that problem Kyu Hwang et. al [1] suggested an efficient algorithm based on the Directed Diffusion so that every sensor node consumes energy uniformly and also tried to find out the shortest possible path. But Kyu Hwang et. al [1] have just suggested a method but there was no proof provided in the continuous domain(mathematical proof). They took some variables like α and β, but did not specify any way to find out the values of those variables. We prove the concept described by Kyu Hwang et. al [1] in continuous domain assuming any kind of network topology . Based on the proof and modeling of the AUEDDD algorithm in future we will able to estimate the value of α , β and ε to optimize the energy dissipation per node and the total energy dissipation to maximize the life time of a WSN.

Acknowledgements. This work was partially funded under the “UGC Research Fellowship in Science for Meritorious Students” scheme.

References

[1] Hwang, A. K., Lee, J. Y., Kim, B. C.: Design of Maximum Remaining Energy Constrained Directed Diffusion Routing for Wireless Sensor Networks. In: Proceedings of the International Conference, UK, pp. 788–795 (May 2006)

[2] Intanagonwiwat, C., Govindan, R., Estrin, D., Heidemann, J., Silva, F.: Directed Diffusion for Wireless Sensor Networking. IEEE/ACM Transaction on Networking (TON), 2–16 (February 2003)

[3] Dutta, S., Mukherjee, N., Neogy, S., Roy, S.: A Comparative Study on Different Wireless Sensor Routing Algorithms. International Journal of Information Processing, 1–9

[4] Dutta, S., Mukherjee, N., Neogy, S., Roy, S.: A Comparison of the Efficiencies of Different Wireless Sensor Network Algorithms with Respect to Time. In: Proceedings of NeCoM, pp. 602–618 (July 2010)

[5] Heinzelman, W.R., Chandrakasan, A., Balakrishnan, H.: Energy-Efficient Communication Protocol for Wireless Microsensor Networks. In: Proceedings of Hawaii International Conference on System Science, January 4-7, pp. 1–10 (2000)

[6] Khude, N., Kumar, A., Karnik, A.: Time and Energy Complexity of Distributed Computation of a Class of Functions in Wireless Sensor Networks. IEEE Transaction on Mobile Computing, 617–632 (May 2008)

[7] Nghiem, T.P., Kim, J.H., Lee, S.H., Cho, T.H.: A Coverage and Energy Aware Cluster-Head Selection Algorithm in Wireless Sensor Networks. In: Huang, D.-S., Jo, K.-H., Lee, H.-H., Kang, H.-J., Bevilacqua, V. (eds.) ICIC 2009. LNCS, vol. 5754, pp. 696–705. Springer, Heidelberg (2009)

[8] http://www.eol.ucar.edu/rtf/facilities/isa/internal/CrossBow/DataSheets/mica2.pdf

[9] Chitte, S., Dasgupta, S.: Distance Estimation From Received Signal Strength Under Log-Normal Shadowing: Bias and Variance. IEEE Signal Processing Letters 16(3), 216–218 (2009)

Peak to Average Power Ratio Reduction in OFDM System over PAM, QAM and QPSK Modulation

Gaurav Sikri[1] and Rajni[2]

[1] Assistant Professor
LLRIET, Moga India
[2] Assistant Professor
SBSCET, Ferozepur, Inida
er.gaurav19@gmail.com, rajni_c123@yahoo.co.in

Abstract. Peak to Average Power Ratio (PAPR or PAR) is major problem of Orthogonal Frequency Division Multiplexing. The Classical Clipping method is used in this paper for reduction of PAPR using 4-PAM, 4-QAM and QPSK. Through the Analysis, it is shown that Clipping on 4-PAM is better than QPSK and 4-QAM with 64 subcarriers.

Keywords: Peak to Average Power Ratio, Classical Clipping, Orthogonal Frequency Division Multiplexing (OFDM), QAM, QPSK, PAM.

1 Introduction

Orthogonal Frequency division Multiplexing (OFDM) has been considered as one of the strong standard candidates for the next generation mobile radio communication systems. OFDM technique is spectrally efficient and very robust to wireless multipath fading environment. Therefore it has been adopted as many standards of DAB/DVB (digital audio/video broadcasting) IEEE 802.11x, 3G LTE, and WiMAX systems. One of the main drawbacks of OFDM is its high Peak to Average Power Ratio (PAPR) because it is inherently made up of so many subcarriers. The subcarriers are added constructively to form large peaks. High peak power requires High Power Amplifiers (HPA), A/D and D/A converters. Peaks are distorted nonlinearly due to amplifier imperfection in HPA. If HPA operates in nonlinear region, out of band and in-band spectrum radiations are produced which appears as the adjacent channel interference. Moreover if HPA is not operated in linear region with large power backs-offs, it would not be possible to keep the out-of-band power below the certain limits. This further leads to inefficient amplification and expensive transmitters. To prevent all these problems, power amplifiers has to be operated in its linear region [1].

There are many methods on PAPR reduction such as Clipping, Coding [2], Selective Mapping (SLM), Interleaving [3,4], Nonlinear Companding Transform[5,6], Hadamard Transform [7], Partial Transmit Sequence(PTS) [2] etc. The simple and widely used method is clipping the signal to limit the PAPR below a threshold level, but it is the nonlinear method which further distorts the OFDM signal. Clipping at Nyquist sampling rate will cause all the clipping noise to fall in band and suffers considerable peak regrowth after digital to analog conversion (D/A) conversion. The

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 685–690.
springerlink.com

out-of-band radiation is produced by Filtering. Filtering causes peaks to regrow. Iterative clipping and filtering (ICF) works in recursive way to achieve less PAPR. Its modified version such as Simplified Clipping and Filtering (SCF) and one Time Iteration and Filtering is proposed in [5].

The strength of Clipping and Filtering method is based on total degradation (TD) and results show that it degrades the system performance instead of an improvement**.** This method is still considered as a good choice in 60 GHz CMOS radio transceivers because of its simple implementation and effective PAPR reduction with small degradation [6].

2 System Description

An OFDM System consists of N subcarriers. The OFDM symbol x(t), $0 \le t \le T$, consist of N complex baseband data $X_0, X_1,\ldots,X_{N-1}$ carried on N subcarriers, chosen to be orthogonal with constant spacing Δf as shown in Fig (1). The OFDM symbol x(t) is

$$x(t) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_k e^{jk2\pi\Delta ft} \quad , \quad 0 \le t \le T \tag{1}$$

The Bandwidth of OFDM symbols is *B=Δf.N* and symbol time *T=1/Δf.*X_k is the complex baseband data modulating the *k*-th subcarrier for *x(t).* The PAPR of OFDM symbol may be defined as [8]

$$\xi = \frac{\max t\varepsilon[0,T) \mid x(t) \mid^2}{P_{av}}, \tag{2}$$

Where P_{av} is the average power of the transmitted symbol and maximum sought over the symbol duration defined as $P_{av}=E\{|x(t)|^2\}$. Where *E*{.} is the expectation operator. The value of ξ can be as large as *N* for Quadrature Phase Shift Keying (QPSK), Quadrature amplitude modulation (QAM) and Pulse amplitude modulation (PAM). However large PAPR occurs very less. The PAPR can be best marked by its statistical parameter, Complementary Cumulative Distribution Function (CCDF). For proper values of PAPR oversampling is necessary. L is the oversampling factor. L=1 determines discrete-time signal sampled at Nyquist rate, whereas L=4 gives sufficient samples to capture continuous-domain signal peaks. The oversampled signal can be obtained by (L-1)*N* zero-padding in the middle of the original input vector and converting frequency domain signal into time domain. The OFDM signal sampled at time instant $t=n\Delta t$ is then expressed as

$$x(n) = x(n\Delta t), \quad n = 0,\ldots, LN - 1 \tag{3}$$

3 Clipping and Filtering Method

The Clipping based techniques clips the time domain signal to predefined level [9]. The method of Clipping and Filtering can be described with three modulation

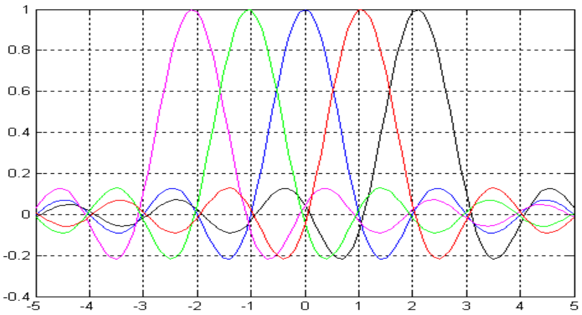


Fig. 1. Orthogonal subcarriers

techniques, Quadrature Phase Shift Keying (QPSK) Quadrature Amplitude Modulation (QAM) and Pulse Amplitude Modulation (PAM). The OFDM signal contains high peaks so it is transferred from the clipping block shown in Fig (3b). In this when amplitude crosses the threshold or cut off level, the amplitude is clipped off shown in Fig (2), while saving the phase. The clipped sample is given by

$$x(n) = \begin{cases} |\, x(n) \,| & if \ \ |\, x(n) \,| \leq C(threshold) \\ C & if \ \ |\, x(n) \,| \leq C(threshold) \end{cases} \tag{4}$$

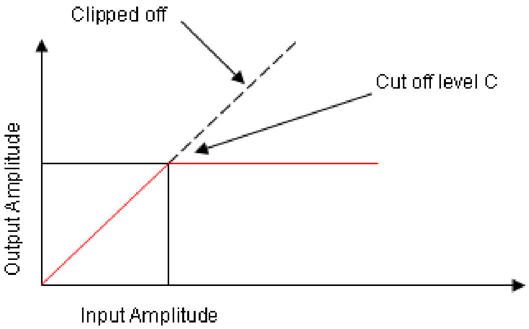


Fig. 2. Clipping method

The out-of-band radiations occurred without filtering due to non linearity. To reduce the interference to neighboring channels, out-of-band components must be reduced with a band limiting filter [1]. The peak growth becomes small after filtering the oversampled signal. The repeated clipping and filtering can reduce the peak regrowth and increases the system cost. So there has been a tradeoff between PAPR and system cost.

The Modulated data can be of any type 4-QAM, QPSK or 4-PAM. In this paper we are trying to show the effect of clipping and filtering between the modulated data using constellation mapping of three modulations on 64 subcarriers.

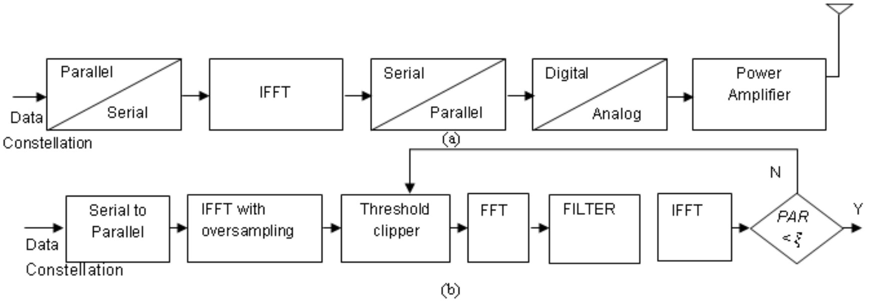


Fig. 3. Block Diagram of (a) Original OFDM system (b) Clipped using threshold

4 Results and Simulations

We use the computer simulations to evaluate the performance of the proposed PAPR reduction technique over different types of modulated data. As a performance measure for proposed technique, we use the CCDF of the PAPR. Performances of the proposed system are first compared without clipping and filtering to OFDM for a multicarrier system with QAM and PAM symbols modulated on N=64,128,256 subcarriers and then with QAM and QPSK symbols modulated on N=64,128,256 subcarriers. 10000 random OFDM blocks were generated to obtain the CCDF. Fig (4) shows the CCDF of PAPR of QAM signals is better than PAM and is given in Table (1). Fig (5) shows the CCDF of QPSK signals is better than QAM without clipping and filtering. The increase in the number of subcarriers results into more PAPR as given in Table (2). Fig (6) shows the effect of clipping and filtering over the CCDF of PAPR of QAM, PAM and QPSK signals with N=64. The decrease in PAPR is 7.89 dB over QAM, 9.73 dB over PAM and 7.68 dB over QPSK due to the effect of classical clipping. A comparison of QPSK, QAM and PAM with N=64 shows the difference of 1.12 dB as given in Table (3).

Table 1. Comparison of 4-QAM and 4-PAM using 64,128 and 256 subcarriers without clipping

Modulations	64 subcarriers	128 subcarriers	256 subcarriers
QAM	11.35 dB	11.97 dB	12.27 dB
PAM	11.73 dB	13.59 dB	12.76 dB

Table 2. Comparison of 4-QAM and QPSK using 64,128,256 subcarriers without clipping

Modulations	64 subcarriers	128 subcarriers	256 subcarriers
QAM	11.63	12.02	12.40
QPSK	11.58	11.41	12.04

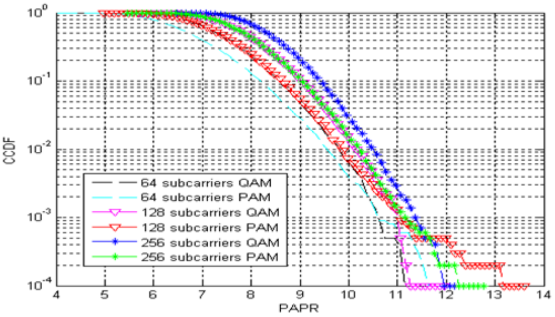


Fig. 4. PAPR of 4-QAM and 4-PAM using 64,128 and 256 subcarriers without clipping

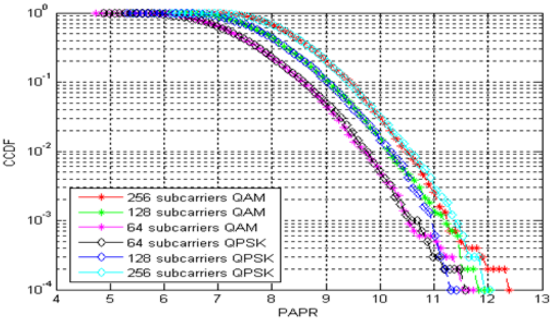


Fig. 5. PAPR of 4-QAM and QPSK using 64,128 and 256 subcarriers without clipping

Table 3. Comparison of 4- QAM, 4-PAM and QPSK using 64 subcarriers with and without clipping

Modulations	Without Clipping	With Clipping
4-QAM	11.10	3.21
4-PAM	11.82	2.09
QPSK	10.89	3.21

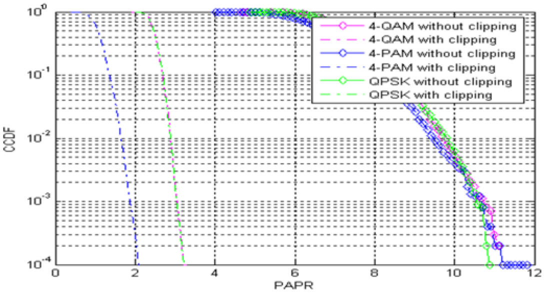


Fig. 6. PAPR of 4-QAM, 4-PAM and QPSK using 64 subcarriers with and without clipping

5 Conclusion

In this paper, a Classical clipping and filtering technique is introduced to reduce the PAPR in multicarrier system applying 4-QAM, 4-PAM and QPSK with N=64 subcarriers. The PAPR of three different modulation techniques is compared with each other. Results show that PAM modulated with N=64 by clipping and filtering is better than QAM and QPSK.

References

[1] Ryu, H.G.: Combination of PAPR reduction and linearization for the OFDM communication system. Wirel. Commun. and Mob. Computing, 46–52 (2010)

[2] Jiang, T., Wu, Y.: An Overview: Peak-to-Average Power Ratio Reduction Techniques for OFDM Signals. IEEE Trans. on Broadcast, 257–268 (2008)

[3] Jiang, T., Imai, Y.: An overview: peak-to-average power ratio reduction techniques for OFDM signals. IEEE Trans. on Wirel. Commun. 56–57 (2008)

[4] Han, S.H., Lee, J.H.: An overview of peak-to-average power ratio reduction techniques for multicarrier transmissions. IEEE Trans. on Wirel. Commun., 56–65 (2005)

[5] Wang, L.Q., Tellambura, C.: A Simplified Clipping and Filtering Technique for PAR Reduction in OFDM systems. IEEE Signal Process. Lett. 12, 453–456 (2005)

[6] Gurung, A.K., Fawaz, S., Qahtani, A.L., Sadik, A.Z., Hussain, Z.M.: Power Savings Analysis of Clipping and Filtering Method in OFDM Systems. In: IEEE ATNAC, pp. 204–208 (2008)

[7] Park, M., Heeyong, J., Cho, N., Hong, D., Kang, C.: PAPR reduction in OFDM transmissions using Hadamard transform. In: IEEE Int. Conf. of Commun., vol. 1, pp. 430–433 (2000)

[8] Wang, L., Tellambura, C.: An Overview of Peak-to-Average Power Ratio Reduction Techniques for OFDM systems. In: IEEE Int. Symp. on Signal Process and Inf. Technol., pp. 840–845 (2006)

[9] Li, X., Cimini, L.J.: Effects of clipping and filtering on the performance of OFDM. IEEE Commun. Lett. 2, 131–133 (1998)

MERCC: Multiple Events Routing with Congestion Control for WSN

Ayan Kumar Das[1] and Rituparna Chaki[2]

[1] Department of Information Technology, Kolkata, India
ayandas24114057@yahoo.co.in
[2] Department of Computer Science & Engineering, Kolkata, India
rituchaki@gmail.com

Abstract. Wireless sensor networks consisting of many small sensor nodes with limited power resources, are useful in gathering data in different environment. Some of the algorithms focus on setting up an event path and the nodes get the information by sending the queries to the event path. On the other hand some algorithms flood the event information from source node to its neighbors and from neighbors to the neighbors of neighbors and so on. The first type involves extra delay, and the second type consumes more power. In this paper we introduce a new protocol 'MERCC' which can handle multiple events at a single instance with power effective methodology. MERCC tries to retain the performance even in case of network congestion.

Keywords: Event categorization, Congestion Control, Event Priority, Event Number, Network longevity and Flooding.

1 Introduction

The sensor networks of future will be collaborative, dynamic and distributed computing communicating systems and self organizing. There are wide ranges of promising applications for these types of networks, which can identify different adverse situations. Energy efficiency and congestion less communication is a great challenge in designing such networks.

Most of the algorithms which are being presently used in the field of sensor networking highly concentrate on sending signals to a base station through an event path. Some researchers focus on event-centric approach, so as to detect the source of natural events. The major problems encountered are—depletion of energy in the chosen paths. This leads to unstable uniform power of the whole system leading to system failure. The query nodes may face longer delays in receiving the event information if it sends its query irregularly. Otherwise, if the query node sends the query regularly then the power consumption will be very high for that node.

This paper proposes The Power balanced routing with Multiple Event Handling and Congestion Control (MERCC) algorithm for handling multiple events and to reduce the congestion of the network by balancing the power of the overall network, so that the longevity of the network may increase.

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 691–698.
springerlink.com

The remaining part of this paper is organized as follows: Section 2 deals with the review of state of the art, section 3 gives a description of the proposed methodology, section 4 contains the simulation reports and section 5 is the concluding part.

2 Review

2.1 Power Aware Routing Algorithms

In Energy Efficient Routing for Single Destination Flow redirection algorithm (FR) [4], the path with largest battery capacity and less energy consumption per bit transmission than all other nodes in network and the path with minimum battery capacity and higher energy consumption per bit transmission than all other nodes in network are used for communication. This algorithm chooses the two paths from source node to the destination which are to be involved in redirection. Maximum Residual Energy Path Routing (MREP) [10] is designed to augment the flow on the path whose minimum residual energy after the flow augmentation will be longest. This algorithm works with static networks only. The algorithm used fixed information generation rates and required a priori knowledge of future information generation. A Novel Power-Balancing Routing Scheme for WSN [2] is designed to detect the source of event in the network. The source node floods the event information along an event path and the other nodes send the queries at a regular time interval. When the query discovers a node belonging to the event path, then it detects the source node along the event path. The algorithm guarantees that the query packet will not get into an infinite loop by status checking mechanism and thus increasing the longevity of the overall network. Intelligent Energy Efficient Routing for Ad-Hoc Sensor Network by Designing QCS Protocol [3] uses signals of three different types for regular information, irregular Information, and devastating information. Load-Balanced Minimum Energy Routing (LBMER) [5] algorithm proposes to maximize the network lifetime. The algorithm uses a mixture of energy balance and traffic balance to solve the problem of congestion of WSNs.

2.2 Cluster Based Routing Algorithm

In clustered network, nodes are clustered in the form of hierarchical structure. The advantage of cluster based method is improving routing efficiency, scalability, supporting QOS and saving power consumption in the nodes. Generally clustering transforms a physical network into a virtual network which has interconnected clusters.

Low Energy Adaptive Clustering Hierarchy (LEACH) [11] is also cluster-based protocol, which includes distributed cluster formation. LEACH randomly selects a few sensor nodes as cluster heads (CHs) and rotates this role to evenly distribute the energy load among the sensors in the network. In LEACH, the cluster head (CH) nodes compress data arriving from nodes that belong to the respective cluster, and send an aggregated packet to the base station in order to reduce the amount of information that must be transmitted to the base station. LEACH uses a TDMA/CDMA MAC to reduce inter-cluster and intra-cluster collisions. However, data collection is centralized and is performed periodically. Therefore, this protocol is most appropriate when there is a need for constant monitoring by the sensor network. A user may not need all the data immediately. Hence, periodic data transmissions are unnecessary which may drain the limited energy of the sensor nodes. After a given interval of time, a randomized rotation of the role of the CH is conducted so that uniform energy dissipation in the sensor network is obtained. The authors showed that only 5% of the nodes need to act as cluster heads.

Hybrid Energy-Efficient Distributed (HEED) routing [13] is also a clustering approach, which is one of the most recognized energy-efficient clustering protocols. It extends the basic scheme of LEACH by using residual energy and node degree or density. In HEED, the initial probability for each node to become a tentative cluster head depends on its residual energy, and final heads are selected according to the intra-cluster communication cost. The clustering process is divided into a number of iterations, and terminates with in a constant number of iterations. HEED achieves fairly uniform distribution of cluster heads across the network.

2.3 Query Driven Routing

In Information-Driven Sensor Querying (IDSQ) [16], the querying node is capable of determining which node can provide the useful information with the advantage of balancing the energy cost. However, IDSQ does not specifically define how the query and the information are routed between sensors and the base station. Therefore, IDSQ can be seen as a complementary optimization procedure. Simulation shows that these approaches are more energy-efficient than directed diffusion where queries are diffused in an isotropic fashion and reaching nearest neighbors first.

2.4 Zone Based Routing

The zone-based routing algorithm [1] generally follows a method in which max-min coordinate system is used. It is used for large scale networks and is fairly good to optimize the lifetime of the network. Zone-base routing is a hierarchical method where the area covered by the sensor network is divided into a small number of zones. Each zone has many nodes and thus a lot of redundancy occurs in routing a message through it. To send a message across the entire area it finds a global path from zone to zone and gives each zone control over how to route the message within itself. A local path for the message is computes within each zone so as to not decrease the power level of the zone too much.

3 Proposed Work

3.1 Basic Concept

In WSN, the main issue of concern is to optimize the power consumption of the entire network, so as to guarantee the delivery of urgent information to the base station with minimum power consumption. Congestion in WSNs also add up to the delay in packet delivery. Keeping in mind these factors, the proposed MERCC Algorithm introduces three new features— Event Categorization, Multiple Event Handling and Congestion Control. The Event Categorization is done to set priority to events that has occurred in the vicinity of the nodes. For example, if an earthquake occurs and its intensity is computed by the node in the range of 7 and above on the Richter scale then treats that as a higher priority event. When the intensity is below 7 in Richter scale, then the event is treated as a lower priority event. The data packets with higher priority event will be sent through the shortest path from source node to base station so that it can take very less time and we are not concerning about the power at that time. The lower priority events are sent through the neighbor nodes having maximum available power. When multiple events occur simultaneously, then the network has multiple source nodes. Every source node will start sending data packets of different priorities. This may cause some nodes to get overloaded, damaging the overall performance of the network. To control the congestion of the network an event queue is maintained. It is assumed that the event queue can receive 2 high priority packets and 4 low priority packets, and it can transmit only one packet at a time interval of every 2ms. The queue will transmit the high priority packet first, when the packets of same priority will be sent on first come first serve basis.

3.2 Data Dictionary

Ep : Event priority value.

arr[]:An array consists of the connections between the nodes.

s_path[ ]:List of all the nodes for shortest path from a source node.

power[node_id][energy]:An array consists of initial power of each node.

status[node_id]:Visited nodes are stored in this array.

Pi[node_id, status]: Neighbors of node i are stored in this array along with their status.

D :The distance between the last and penultimate visited nodes.

Total :Total distance traveled during one simulation.

node_max_pow[ ]:The neighboring nodes with the maximum power are stored in this array.

3.3 Description

3.3.1 MERCC ( )

```
Step 1: Read n, p (total number of nodes, initial power for all the nodes)
Step 2: Create the network with n nodes and set the initial value p for every node.
Step 3: The source node m sense the event and set the priority value according to
event intensity as ep : 2/1
Step 4: If ep=2 then do—
            a) Find the shortest path from source node m to base station and store all
               the nodes of this path in the array s_path[ ].
            b) Call Reduce_Power ( )
        Else
          a) Find the neighboring nodes of source node (m) and store them in the
             array 'p'.
          b) Repeat following steps for n times
                  If status i = visited then delete $i^{th}$ node from Visit[node_id]
```

```
c) If all the neighboring nodes are already visited, then choose any one
   randomly.
   Else
       Find the node among (not visited) nodes with the maximum power
     and store in the array ‘node_max_power’.
d) If more than one node exists with maximum power content, then select
   any of them randomly as the next hop and store it as ‘next_node’.
e) If the chosen next hop node falls on the base station then—
          Call function Reduce_Power ( )
   Else
      Make the next hop node as the source node(m) and continue the
    process from step 6(a).
```

3.3.2 Reduce_Power ( )

```
Step 1: For i=1 to s_node (which is total number of nodes in s_path[ ])
    For j=1 to n repeat
        If s_path[i] = power[1][j], then power[2][j]=power[2][j]-trans_power
Step 5: For i=1 to n set total = total + power[2][i]
Step 6: Set avg_power=total/n
```

3.4 Case Study

Consider the following sensor network of nine nodes. The values along with nodes are representing their powers and that of edges are representing the distance between the nodes.

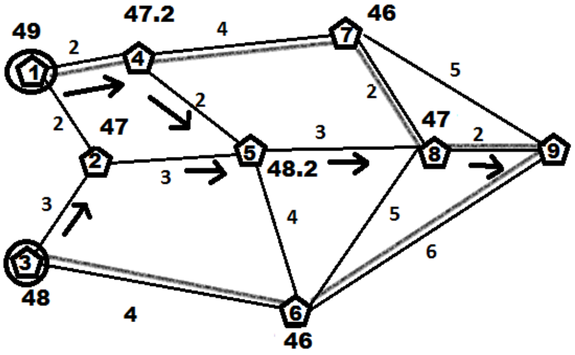


Fig. 1. Routing for multiple event

In the above figure the nodes which are circled are the source nodes. The first one that is node 1 is of the higher priority when the second one that is node 3 is of the lower priority. Thus the information of devastating event from source node 1 will reach at base station (node 9) via the shortest path node 1, node 4, node 7, node 8 and node 9, when the information of non devastating event from source node 3 will search for the neighbor node consisting of maximum power instead of shortest path. Thus this time node 2 will be selected. Now among the neighbor of node 2, node 1 has the maximum power, but is a source node, so node 5 is selected. After that the neighbors of node 5 are node 4, node 6 and node 8. As node 5 already has a message from node 4, it will be rejected. Node 8 will be selected as it has more power than node 6. At last the data packet will be sent to the base station (node 9) from node 8.

4 Simulation Result

To analyze the performance of the algorithm a network with nine sensor nodes has created. The initial power for every node is considered to 50 units. To simulate the algorithm two high priority and eight low priority messages have been sent from two different source nodes to one base station. The simulation parameters are given in the following table.

Table 2. Parameter list

Parameters	Description
Network size	9 nodes
Initial energy	50J per node
MAC Protocol	IEEE 802.15.4
Power consumption	Equivalent to packet size and distance
Number of high priority message	At least 10

After sending each message the average power of the network has measured and a graph of average power of the network vs. number of messages sent has drawn. It is also compared with the concept of sending packets only through the nodes consisting of maximum power.

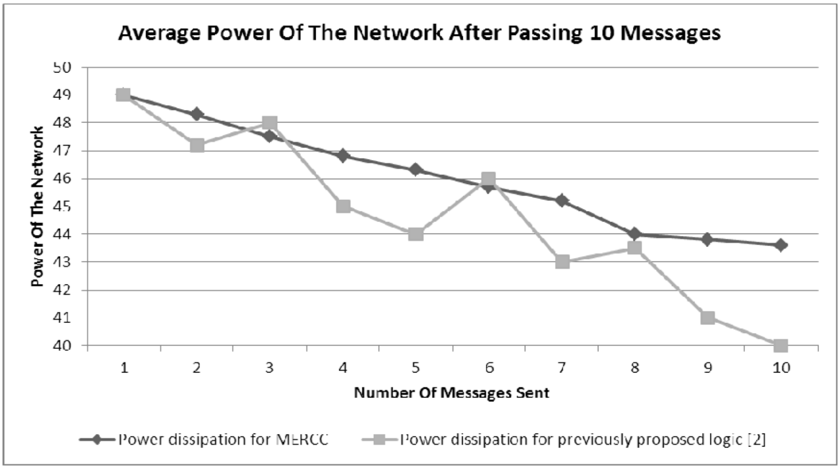


Fig. 2. Average power v/s. Number of messages graph

In the above figure message 4 and message 8 are of high priority and thus propagating in shortest path, while the other messages are of low priority and is propagating considering maximum power. It is clear from the above figure that if the messages are sent only considering the maximum power of the neighbor nodes as stated in the algorithm [2], after passing 10 messages the average power of the network will go down to 40.2 units from 50 units. Where as if we sent packets according to MERCC algorithm then after passing 10 messages the average power of the network is 43.6 units. Therefore the power dissipation of the network according to MERCC is 34% less than other existing algorithm. Thus this algorithm is optimizing the power consumption of the entire network and also guarantying the delivery of the urgent message in minimum time by controlling the congestion.

5 Conclusion

The main objective of any wireless sensor network is to transmit important information to the destination within a very short span of time. The delivery time has to be minimized and the power consumption also should be optimized to serve the basic goal of the WSN. The current state of the art study shows that most of the power saving algorithms fails to take care of timeliness of delivery and the probability of multiple event occurrences at the same instance. This paper proposes MERCC, an event-driven routing methodology to handle multiple events at a single instance with minimum energy drainage. MERCC also takes care of the congestion within the network. The simulation result shows the proposed algorithm increases the lifetime of the network by 34%, as compared with our earlier algorithm [2].

References

[1] Karim, L., Nasser, N., El Salti, T.: Efficient Zone-based Routing Protocol of Sensor Network in agriculture monitoring systems. Communications and Information Technology, ICCIT (2011) ISBN: 978-1-4577-0401-7, Issue Date: March 29-31

[2] Das, A.K., Chaki, R.: A Novel Power-Balanced Routing Scheme for WSN. In: Özcan, A., Zizka, J., Nagamalai, D. (eds.) WiMo 2011 and CoNeCo 2011. CCIS, vol. 162, pp. 24–34. Springer, Heidelberg (2011)

[3] Das, A.K., Ghosh, D., Majumder, P.: Intelligent Energy Efficient Routing For Ad Hoc Sensor Network by Designing QCS Protocol. In: Proceedings of The Second International Workshop on Adhoc, Sensor and Ubiquitous Computing (ASUC 2011). LNCS(CCIS). Springer (June 2011)

[4] Jain, S., Kaushik, P., Singhai, J.: Greedy Heuristic Based Energy Efficient Routing in Wireless Sensor Network. In: Nagamalai, D. (ed.) PDCTA 2011. Part 1, CCIS, vol. 203, pp. 282–292. Springer, Heidelberg (2011), doi:10.1007/978-3-642-24037-9_27

[5] Gong, B.-C., Li, L.-Y., Jiang, T.-Y., Xu, S.-Z.: Distributed Spanning Tree-Based Routing Protocol for Wireless Sensor Networks. Microelectronics & Computer 25(11) (2008)

[6] Siva Kumar, D., Bhuvaneswaran, R.S.: Proposal on Multi agent Ants based Routing Algorithm for Mobile Adhoc Networks. IJCSNS International Journal of Computer Science and Network Security 7(6) (June 2007)

[7] Camilo, T., Carreto, C., Silva, J.S., Boavida, F.: An Energy-Efficient Ant-Based Routing Algorithm for Wireless Sensor Networks. In: Dorigo, M., Gambardella, L.M., Birattari, M., Martinoli, A., Poli, R., Stützle, T. (eds.) ANTS 2006. LNCS, vol. 4150, pp. 49–59. Springer, Heidelberg (2006)

[8] Laxmi, V., Jain, L., Gaur, M.S.: Ant Colony Optimization based Routing on NS-2. In: The Proceedings of International Conference on Wireless Communication and Sensor Networks, WCSN 2006 (2006)

[9] Huang, S.-C., Jan, R.-H.: Energy-aware, load balanced routing schemes for sensor networks. Dept. of Comput. & Inf. Sci., Nat. Chiao Tung Univ., Hsinchu, Taiwan(2004), ISSN: 1521-9097, Issue Date: July 7-9

[10] Qiang Feng, J., Manivannan, D.: Routing protocols for sensor networks. In: First IEEE of Consumer Communications and Networking Conference, CCNC 2004 (2004)

[11] Heinzelman, W.R., Chandrakasan, A., Balakrishnan, H.: Energy Efficient Communication Protocol for Wireless Microsensor Networks. In: Proceedings of the 33rd Hawaii International Conference on System Sciences (2000)

[12] Godfrey, P., Naps, R.D.: Scalable, Robust topology management in wireless ad hoc networks. In: Proceedings of the Third International Symposium on Information Processing in Sensor Networks (2004)

[13] Younis, O., Fahmy, S.: Dept. of Comput. Sci., Purdue Univ., West Lafayette, IN, USA, HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. Appears in IEEE Transactions on Mobile Computing 3(4) (2004) ISSN: 1536-1233 Issue Date: October-December

[14] Gunes, M., Sorges, U., Bouazizi, I.: ARA- the ant colony based routing algorithm for MANET. In: Proc. of the ICPP 2002 (2002)

[15] Handy, M., Haase, M., Timmermann, D.: Low Energy Adaptive Clustering Hierarchy with Deterministic Cluster Head Selection. In: 4th IEEE International Conference on Mobile and Wireless Communication Networks, Stockholm (2002)

[16] Slijepcevic, S., Potkonjak, M.: Power efficient organization of wireless sensor networks. In: IEEE International Conference on Communications, Helsinki, Finland (June 2001)

[17] Heinzelman, W.R., Chandrakasan, A., Balakrishnan, H.: Energy Efficient Communication Protocol for Wireless Microsensor Networks. In: Proceedings of the 33rd Hawaii International Conference on System Sciences (2000)

[18] Intanagonwiwat, C., Govindan, R., Estrin, D.: Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks. In: Proceedings of the Sixth Annual International Conference on Mobile Computing and Networks, MobiCOM 2000 (2000)

Awareness Based Approach against E-Mail Attacks

Gaurav Kumar Tak[1] and Gaurav Ojha[2]

[1] School of Computer Science & Information Technology, Lovely Professional University,
Phagwara, Punjab - 144402, India

[2] Department of Information Technology, Indian Institute of Information Technology and
Management, Gwalior - 474010, India

Abstract. E-mail plays a very important role in modern day communication. It is helpful for personal as well as business correspondences between people, or organizations. The major features of E-mail, which make it a convenient mode of communication, are its speed, efficiency, storage options and search facilities. Due to the high popularity of E-mails, it forms the preferred medium for a large number of web attacks. Spammers usually send spams and phishers send phishing URLs via E-mail. Of the large number of techniques that have already been proposed for the detection of several types of such attacks, quite a few of them provide good results but with higher false positives. In this paper, we are proposing a novel technique, which not only identifies spam but also scam mails, phishing, advertisements, etc. This technique utilizes some intelligence on the part of users, apart from keywords parsing, knowledge base, and token separation methods to detect various E-mail attacks. Implementation of the proposed methodology can help protect E-mail users from a wide range of unwanted E-mails, with increased efficiency and highly reduced number of false positives.

Keywords: Phishing, Scam, Spams, Social networking, Subscription.

1 Introduction

E-mail is considered as the most convenient mode for transferring messages electronically from one person to another, or among a group of people. It can be said that E-mail has become the de-facto standard medium of communication in almost all spheres of human life. Features such as quick, easy, and free access, Instant messaging service, and global acceptance only add to its popularity [1].

People from different age groups utilize E-mail for different purposes, some of which are as follows:

1. **Official purpose:** E-mail is often considered as the official medium of communication in the corporate sector. Notices, announcements, contract papers, and other documents of legal importance are communicated in a group using E-mail.
2. **Entertainment:** College-mates usually share a variety of entertainment material such as jokes, chats, photographs, files, links to videos online, stories, etc. through the medium of E-mail.

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 699–708.
springerlink.com

3. **Creating other accounts:** Various websites use E-mail as a token for their primary login mechanism. This includes websites such as Facebook, Blogger, bharatmatrimony.com, Knowafest.com, among many others. Even with the advent of 'Open Authorization' wherein it is possible to login through one's social networking accounts such as Facebook or Twitter, the primary requirement is an E-mail ID.
4. **Participation in events:** For a variety of online competitions, university fests, and other events, E-mail is the primary medium for submission of one's B-plans, project abstracts, scripts, etc. Research related articles, book chapters, and research papers are often sent in various conferences or journals via E-mail.

Apart from these, banks or organizations for sending notifications, advertisements, and promotional messages also use E-mail.

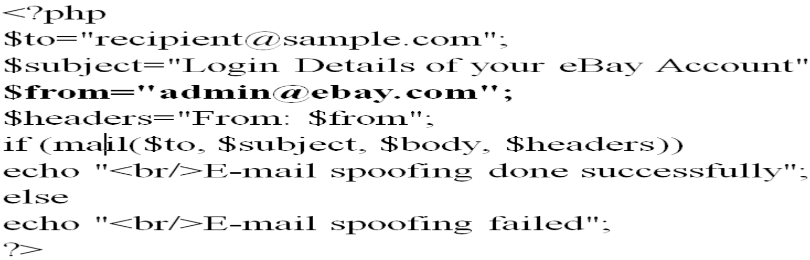
E-mail is also being used for a large number of negative purposes due to its enormous outreach, such as spamming, spreading virus/worms/malware, and other attacks, which waste a user's time and essential network resources such as memory, bandwidth, etc. Following are some the negative uses of E-mails:

1. **Spread of viruses/worms:** Summarily, it can be said that virus/worm attacks over mailboxes are in the form of attachments bearing attractive file names, which may distract a user and force him to open it, causing the virus to spread and attack a multitude of other machines.
2. **Spam:** Usually, the amount of spam received is directly proportional to the age of an E-mail ID, the older an E-mail ID gets, the more spam it is destined to receive, under ordinary usage conditions. Spam refers to any undesired mail, which is useless for the recipient. Spam is usually sent using a script, to a large number of people whose E-mail addresses have been harvested from the World Wide Web. Sometimes, a spam might include just hyper-links, while at times, it may include some text, and senders' address trying to assert that it is not spam. Links, which have been shortened twice or more, often make it difficult to find the final website that they are pointing to, thus serving the purpose of an unidentifiable website.
3. **Scam:** Scam E-mails are meant to lure an unsuspecting user into transactions that are completely bogus. Such scams are believed to have originated from Nigeria in the form of fake banking transactions, lottery, prizes, etc. The name Nigerian Scam was coined for such transactions. Sometimes, they are combined with phishing. The attacker announces a lottery or any such prize of large monetary value, in return of users' personal information. Unaware users gladly provide such information only to be fooled into the trap. The attacker may then ask for some amount as transaction or processing charges for the prize amount, upon receiving which his purpose is served. Any person falling prey to such fake transactions is sure to incur large financial losses, and nobody can be directly blamed for it.
4. **Spreading URLs of phishing sites:** Phishing websites are fake websites, which are almost identical to some popular website such as social networking websites, E-mail websites, Online Banking websites, etc., which require a user to login. Phishing websites then capture the confidential login

credentials of unsuspecting users, which may then be used to perform fraudulent transactions on their behalf.

5. **E-mail Spoofing:** In E-mail spoofing, some important parts of an E-mail message, such as headers including sender's address and other tokens, are morphed in such a way as to imply that the mail was sent by someone else. This is possible because there is very little or no authentication on SMTP. Misusing an E-mail address is like a cakewalk with spoofing, as any kind of mail can be sent bearing an E-mail ID without the knowledge of the person who owns that E-mail ID.

Figure 1 shows a simple script that can be used for E-mail spoofing [5] [6]. There are many other methods using which E-mail spoofing can be performed.



```
<?php
$to="recipient@sample.com";
$subject="Login Details of your eBay Account";
$from="admin@ebay.com";
$headers="From: $from";
if(mail($to, $subject, $body, $headers))
echo "<br/>E-mail spoofing done successfully";
else
echo "<br/>E-mail spoofing failed";
?>
```

Fig. 1. E-mail spoofing of admin account at E-bay

There are many spam detection techniques, which are designed to protect the recipients from spammers and scammers. Most of the techniques detect either spams or scams or other categories of E-mails, but they are not supposed to cover all the attacks. These techniques require a lot of filtering functions, which rely on mathematical operations to differentiate between a spam mail and an ordinary mail. This requires memory, space, and time complexity. Even after such resource-intensive operations, some false positives show up. In this paper, we have discussed novel user awareness based techniques to protect users from E-mail attacks. In the first section, we introduce about the concept of E-mails. Then we go about techniques to detect E-mail attacks in the second section. In the third section, some techniques, which the user can practice to protect him from E-mail attacks, have been described.

2 Related Work

In literature, many discussions have described the techniques of detection of several types of spam or scams. Some techniques are useful, but most of them are implemented on the server side. We list below some of the important works towards detection of E-mail attacks over the web.

In [12], a rule-based approach has been implemented for the detection of spam mails, which is based on some learning process and intelligence. The discussed approach uses the training and testing phases of data for learning as well as for churning out better results. However, the time as well as memory complexity is higher due to the generation of rules, managing data sets and a number of execution operations. E. Damiani et al. have described some basic properties and behavior of spam mails. The use of digests in the proposed approach to identify spam mails in a privacy-preserving way is a fundamental technique for collaborative learning [13]. A social network is designed based on the fact that they are used to exchange a lot of information as well as email address of registered users [14][15]. Spammers are identified by observing misbehavior or abnormalities in the structural properties of the network. Most of the times, email attackers use public social sites to harvest email addresses, (sometimes this task may even be performed by bots) to their mail list database. However, it is a reactive mechanism, which is used to identify spam mails and their behavior since spammers are identified; in this approach, the spam filter uses previous history of spam mails. In [16], a novel and better approach has been described, which creates a Bayesian network out of email exchanges to detect spam [4].

Nitin Jindal et al. discussed an approach of review on content-based spam mails. Review spam is quite different from Web page spam and E-mail spam, and thus requires different detection techniques [17]. Shashikant et al. also proposed separation techniques based on token separation of E-mail contents and some probabilistic approach. Firstly, the email server receives the E-mail content, and then it separates the tokens of E-mail content using some proper operations and analyzes the content based on the requirement or need of users [11]. In [7] [8] [9] [10], some of these filters have been discussed to improve the spam detection and to protect the E-mail users against email attacks. Examples of filtering processes are Checksum based filtering, Bayesian spam filtering, Machine learning based classification, and Memory based filtering. In [1] [2], some techniques have been discussed, which focus on the partial match of attack keywords or spam keywords or spam content. In [1], the proposed approach uses some learning processes to identify E-mail attacks. One effective technique has already been proposed to identify the spam mail viz. ‘Fast Effective Botnet Spam Detection’. It uses the header information of mails to detect the spam mails or other email attacks. It effectively works for both types of spam mails namely, ‘Text based spam’ as well as ‘Image based spam’. It locates the sender’s IP address, sender’s email address; MX records and MX hosts, and analyzes all the recorded information to provide effective results [3].

We are proposing some new techniques to identify such email attacks, which provides efficient results with significantly small number of false positives.

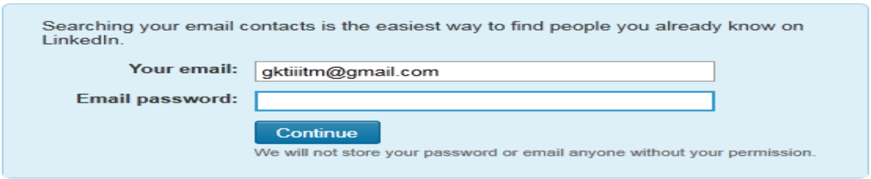
3 Proposed Methodology

Various studies have found that most of the E-mail attacks are carried out due to social networking websites and subscription to various websites. On a stronger note, they are a result of giving out E-mail IDs at the time of registration for various subscriptions and accounts. Having identified the main reason behind the issue, we suggest the following methodology, which can protect E-mail users from a variety of E-mail related attacks.

3.1 Avoid Sharing E-mail ID with Untrustworthy People

E-mail users generally distribute their E-mail address among friends, relatives, and other people concerned. But some of these, especially friends, may try to hack their E-mail to steal personal or official information. This may result in various spam or pornographic advertisements arriving at one's mailbox. Some people sell contact lists to various advertisement companies.

We all know that whenever one joins a social networking site, the site provides a feature to invite one's friends by using contacts in one's E-mail account.



Searching your email contacts is the easiest way to find people you already know on LinkedIn.

Your email: gktiiitm@gmail.com

Email password:

We will not store your password or email anyone without your permission.

Fig. 2. Functionality to invite friends on a social networking site (LinkedIn)

In Figure 2 above, the social networking invitation feature is shown. It is used to add or invite friends with an E-mail contact list.

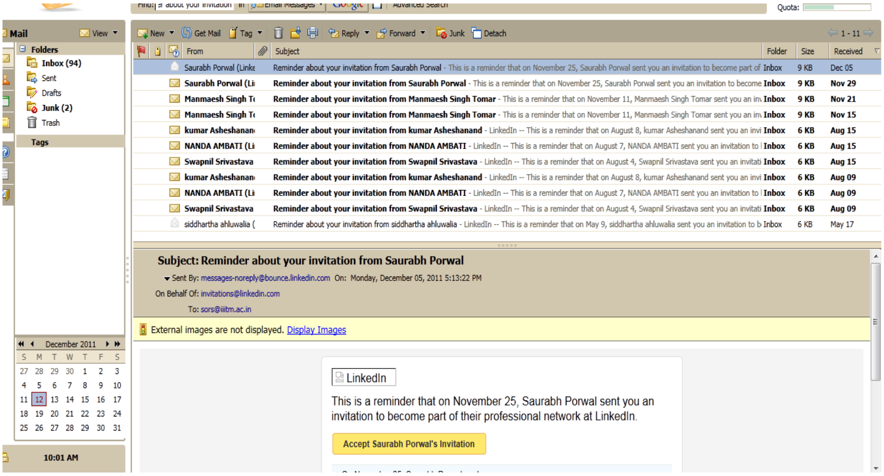
In Figure 3, the invitation mails are shown on an official E-mail ID sors@iiitm.ac.in, which is the E-mail address of the administrator of the Students' Online Record System (SORS), an online portal at Indian Institute of Information Technology and Management, Gwalior (IIITM), which manages information related to registered students. It has been observed that SORS also receives many such invitation mails, advertisements, promotional mails and other spam.

As per this step in the proposed methodology, E-mail users are suggested not to give their official E-mail address to untrusted people or on advertisement groups, etc. If E-mail users follow this simple step, they can protect themselves from E-mail attacks especially spam, very easily.

3.2 Avoid Clicking on Random/New URLs

According to the proposed techniques, E-mail users are suggested not to click on any random or new URLs, as it may not be a trusted URL. Such URLs can be direct download links for viruses, or some phishing URLs, or URLs to pornographic websites.

As per the discussion, most of the malicious URLs are circulated using social networking sites, instant messaging, E-mails, etc.



From	Subject	Folder	Size	Received
Saurabh Porwal (Linke	Reminder about your invitation from Saurabh Porwal - This is a reminder that on November 25, Saurabh Porwal sent you an invitation to become part of	Inbox	9 KB	Dec 05
Saurabh Porwal (Li	**Reminder about your invitation from Saurabh Porwal** - This is a reminder that on November 25, Saurabh Porwal sent you an invitation to become	**Inbox**	**9 KB**	**Nov 29**
Manmaesh Singh T	**Reminder about your invitation from Manmaesh Singh Tomar** - This is a reminder that on November 11, Manmaesh Singh Tomar sent you an inv	**Inbox**	**9 KB**	**Nov 21**
Manmaesh Singh T	**Reminder about your invitation from Manmaesh Singh Tomar** - This is a reminder that on November 11, Manmaesh Singh Tomar sent you an inv	**Inbox**	**9 KB**	**Nov 15**
kumar Asheshanan	**Reminder about your invitation from kumar Asheshanand** - LinkedIn -- This is a reminder that on August 8, kumar Asheshanand sent you an invi	**Inbox**	**6 KB**	**Aug 15**
NANDA AMBATI (Li	**Reminder about your invitation from NANDA AMBATI** - LinkedIn -- This is a reminder that on August 7, NANDA AMBATI sent you an invitation to l	**Inbox**	**6 KB**	**Aug 15**
Swapnil Srivastava	**Reminder about your invitation from Swapnil Srivastava** - LinkedIn -- This is a reminder that on August 4, Swapnil Srivastava sent you an invitati	**Inbox**	**6 KB**	**Aug 15**
kumar Asheshanan	**Reminder about your invitation from kumar Asheshanand** - LinkedIn -- This is a reminder that on August 8, kumar Asheshanand sent you an invi	**Inbox**	**6 KB**	**Aug 09**
NANDA AMBATI (Li	**Reminder about your invitation from NANDA AMBATI** - LinkedIn -- This is a reminder that on August 7, NANDA AMBATI sent you an invitation to l	**Inbox**	**6 KB**	**Aug 09**
Swapnil Srivastava	**Reminder about your invitation from Swapnil Srivastava** - LinkedIn -- This is a reminder that on August 4, Swapnil Srivastava sent you an invitati	**Inbox**	**6 KB**	**Aug 09**
siddhartha ahluwalia (	Reminder about your invitation from siddhartha ahluwalia - LinkedIn -- This is a reminder that on May 9, siddhartha ahluwalia sent you an invitation to b	Inbox	6 KB	May 17

Subject: Reminder about your invitation from Saurabh Porwal

Sent By: messages-noreply@bounce.linkedin.com On: Monday, December 05, 2011 5:13:22 PM

On Behalf Of: invitations@linkedin.com

To: sors@iiitm.ac.in

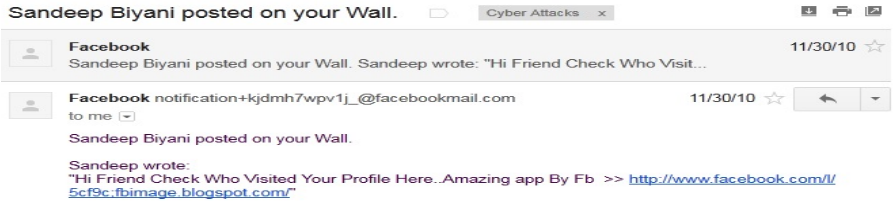
External images are not displayed. Display Images

LinkedIn

This is a reminder that on November 25, Saurabh Porwal sent you an invitation to become part of their professional network at LinkedIn.

Accept Saurabh Porwal's Invitation

Fig. 3. Invitation mails on the official E-mail ID sors@iiitm.ac.in



Sandeep Biyani posted on your Wall.

Facebook 11/30/10

Sandeep Biyani posted on your Wall. Sandeep wrote: "Hi Friend Check Who Visit...

Facebook notification+kjdmh7wpv1j_@facebookmail.com 11/30/10

to me

Sandeep Biyani posted on your Wall.

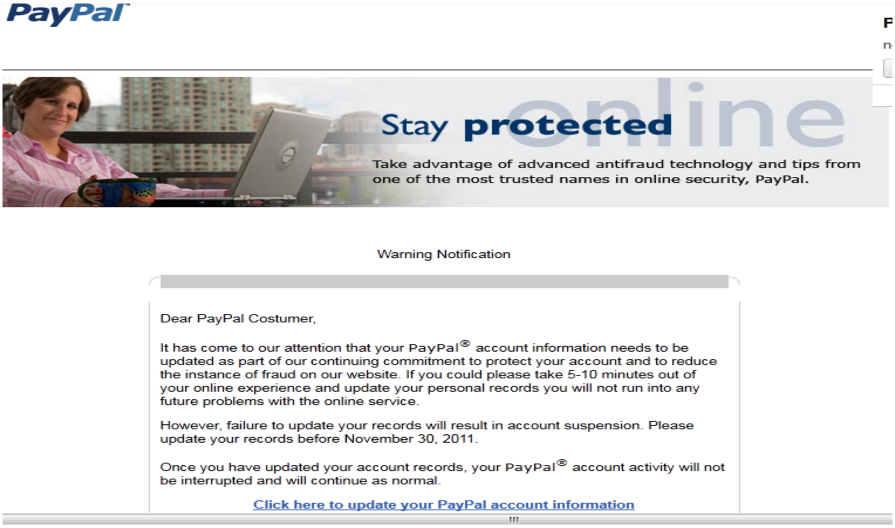
Sandeep wrote:
"Hi Friend Check Who Visited Your Profile Here..Amazing app By Fb >> http://www.facebook.com/l/5cf9c;fbimage.blogspot.com/"

Fig. 4. A malicious link that arrives as an E-mail notification from Facebook

In Figure 4, a malicious URL is shown which has circulated through Facebook, and an E-mail notification has arrived at the mailbox. When an E-mail user clicks on this URL viz. http://www.facebook.com/l/5cf9c;fbimage.blogspot.com/ to know about the new activity, then "Hi Friend Check Who Visited Your Profile Here...Amazing app By Fb >> ", a porn link is spread out to all friends in the friend list without requesting any permission from the email user. There are many such attacks, which can be spread out with URLs by appending/including them in E-mails. In Figure 5, phishing E-mail of PayPal is shown. As per this step of the proposed methodology, E-mail users must avoid clicking on non-trusted URLs, as they may be phishing URLs.

3.3 Do Not Respond to Mails from Untrustworthy People

Many times, E-mail users receive E-mails related to winning of cash prizes, lottery, gifts, or other scams. Sometimes, users gets a mail informing that some changes need to be performed in their online banking accounts, E-bay accounts, etc. for which they need to login using the provided link. Generally, these mails are related to unknown senders who just want to receive users' personal information, account information, or financial details for various fraudulent purposes.


PayPal
Stay protected online
Take advantage of advanced antifraud technology and tips from one of the most trusted names in online security, PayPal.

Warning Notification

Dear PayPal Costumer,

It has come to our attention that your PayPal® account information needs to be updated as part of our continuing commitment to protect your account and to reduce the instance of fraud on our website. If you could please take 5-10 minutes out of your online experience and update your personal records you will not run into any future problems with the online service.

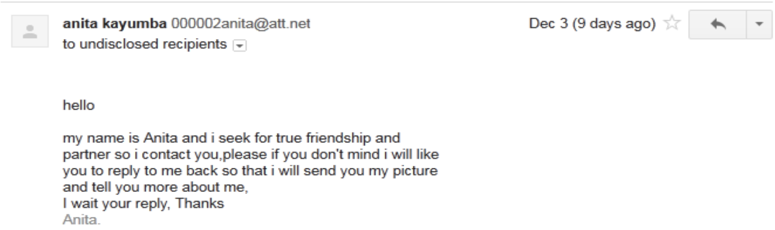
However, failure to update your records will result in account suspension. Please update your records before November 30, 2011.

Once you have updated your account records, your PayPal® account activity will not be interrupted and will continue as normal.

Click here to update your PayPal account information

Fig. 5. A phishing E-mail for Paypal

As per this step in the proposed methodology, users are suggested not to respond to E-mails from unknown senders. Gmail tries to fetch user information for every user on the network, so if user information is unavailable for an E-mail ID, it may be assumed that it is an unknown ID. Users are also suggested not to open any attachments from such mails, as they may be virus or malicious files, which may infect their computers' data or privacy.


anita kayumba 000002anita@att.net
to undisclosed recipients

Dec 3 (9 days ago)

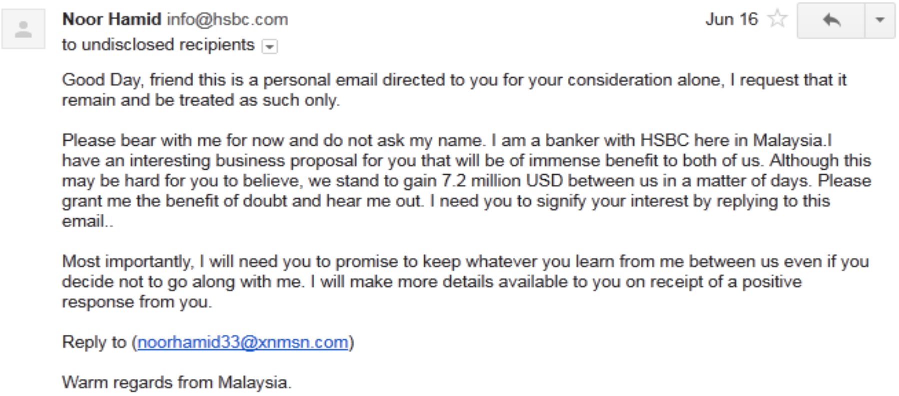
hello

my name is Anita and i seek for true friendship and partner so i contact you,please if you don't mind i will like you to reply to me back so that i will send you my picture and tell you more about me,
I wait your reply, Thanks
Anita.

Fig. 6. An E-mail from unknown person

3.4 Do Not Register or Subscribe to Mailing List Using an Official E-mail ID

At present, if users register on a social networking website using an E-mail ID, they get many notifications about their activity on such websites. This can be in the form of account activity, notifications, messages, or invitations. Due to the large proportion of such mails, users may miss their important or official E-mails. Sometimes, users receive excessive mails from subscriptions that have been willingly accepted.



Noor Hamid info@hsbc.com
to undisclosed recipients

Jun 16

Good Day, friend this is a personal email directed to you for your consideration alone, I request that it remain and be treated as such only.

Please bear with me for now and do not ask my name. I am a banker with HSBC here in Malaysia.I have an interesting business proposal for you that will be of immense benefit to both of us. Although this may be hard for you to believe, we stand to gain 7.2 million USD between us in a matter of days. Please grant me the benefit of doubt and hear me out. I need you to signify your interest by replying to this email..

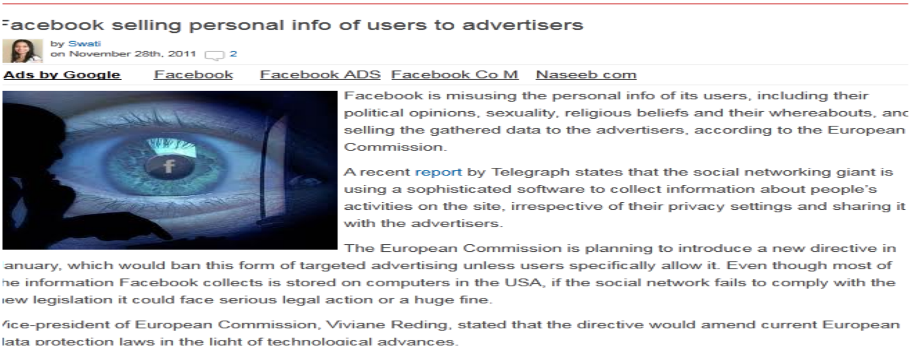
Most importantly, I will need you to promise to keep whatever you learn from me between us even if you decide not to go along with me. I will make more details available to you on receipt of a positive response from you.

Reply to (noorhamid33@xnmsn.com)

Warm regards from Malaysia.

Fig. 7. An E-mail from an untrustworthy person

Many social networking sites are known to sell their users' personal information such as contact details, social preferences and other information to advertisement agencies which eventually fall in the list of spammers. Some sites require premium membership to provide enhanced user experience, which requires the user to pay money. In this way, websites also acquire the financial details of their users. How they will use such details is outlined in the '*Trust and Privacy Policy*' and '*Terms and conditions*' which the users generally neglect. [link: http://www.buzzom.com/2011/11/facebook-selling-personal-info-of-users-to-advertisers/].



Facebook selling personal info of users to advertisers

by Swati
on November 28th, 2011 2

Ads by Google Facebook Facebook ADS Facebook Co M Naseeb com

Facebook is misusing the personal info of its users, including their political opinions, sexuality, religious beliefs and their whereabouts, anc selling the gathered data to the advertisers, according to the European Commission.

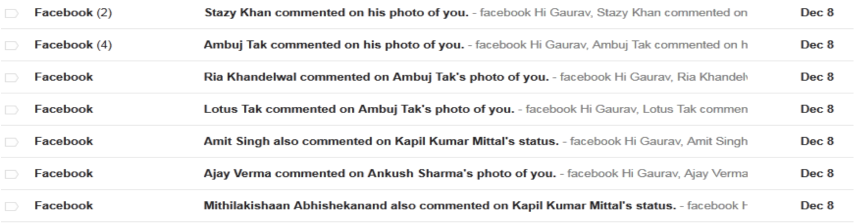
A recent report by Telegraph states that the social networking giant is using a sophisticated software to collect information about people's activities on the site, irrespective of their privacy settings and sharing it with the advertisers.

The European Commission is planning to introduce a new directive in anuary, which would ban this form of targeted advertising unless users specifically allow it. Even though most of he information Facebook collects is stored on computers in the USA, if the social network fails to comply with the ew legislation it could face serious legal action or a huge fine.

ice-president of European Commission, Viviane Reding, stated that the directive would amend current European ata protection laws in the light of technological advances.

Fig. 8. An article showing news that Facebook sells personal information to advertisers

Using this step in the proposed methodology, E-mail users are suggested not to register on social networking sites with their official E-mail IDs, so that they can reduce the rate of E-mails (E-mails per day), thus managing their official mailboxes efficiently.



Facebook (2)	**Stazy Khan commented on his photo of you.** - facebook Hi Gaurav, Stazy Khan commented on	Dec 8
Facebook (4)	**Ambuj Tak commented on his photo of you.** - facebook Hi Gaurav, Ambuj Tak commented on h	Dec 8
Facebook	**Ria Khandelwal commented on Ambuj Tak's photo of you.** - facebook Hi Gaurav, Ria Khandel	Dec 8
Facebook	**Lotus Tak commented on Ambuj Tak's photo of you.** - facebook Hi Gaurav, Lotus Tak commen	Dec 8
Facebook	**Amit Singh also commented on Kapil Kumar Mittal's status.** - facebook Hi Gaurav, Amit Singh	Dec 8
Facebook	**Ajay Verma commented on Ankush Sharma's photo of you.** - facebook Hi Gaurav, Ajay Verma	Dec 8
Facebook	**Mithilakishaan Abhishekanand also commented on Kapil Kumar Mittal's status.** - facebook H	Dec 8

Fig. 9. Notifications from social networking site (Facebook)

4 Analysis and Evaluation

Using the steps in the proposed methodology, users can avoid several types of E-mail attacks as well as other attacks. These steps are helpful for identifying all types of spams, scams and phishing E-mails with very low false positives. Not surprisingly, a famous proverb puts that '*Prevention is better than cure*'.

We have analyzed 10 people's mailboxes for a period of three months (September 2011 to November 2011). Our proposed steps are able to identify all types of attacks over different mailbox. People have also used spam keywords approach or report spam facility to detect all type of spam or scam attacks over their mailboxes. The result is shown is the Table 1, which is given below.

Table 1. Table showing analysis of ten different mailboxes over three months

	Received mails	Identified mails	Efficiency (%)
Total mails	16784	16273	96.95
Social networking sites notifications	3490	3478	99.65
Spam mails or scam mails	4693	4637	98.80
Phishing mails	1034	987	95.45
Official mails	3046	3042	99.86
Advertisements or other mails	4521	4129	91.33

5 Results and Conclusions

We have achieved 96.95% efficiency using all steps. Table 1 categorizes all E-mails accordingly. We have achieved low efficiency in advertisement category of mails, because in the advertisements, email users have to update their own knowledge or managing criteria to find out all the advertisement mails and other attacks.

In future, we are planning to integrate all these steps at mail server with machine learning or artificial intelligence. For the locations of different category mails, we will use proper classifiers for each category and clustering techniques to achieve higher efficiencies.

References

1. Tak, G.K., Tapaswi, S.: Knowledge Base Compound Approach towards Spam Detection. In: Meghanathan, N., Boumerdassi, S., Chaki, N., Nagamalai, D. (eds.) CNSA 2010. Part 2. CCIS, vol. 89, pp. 490–499. Springer, Heidelberg (2010), doi:10.1007/978-3-642-14478-3_49
2. Tak, G.K., Tapaswi, S.: Query Based Approach towards Spam Attacks using Artificial Neural Networks. International Journal of Artificial Intelligence & Applications, IJAIA (2010) ISSN: 09762191, EISSN: 0975900X, Academy & Industry Research Collaboration Center
3. Saraubon, K., Limthanmaphon, B.: Fast Effective Botnet Spam Detection. In: 2009 Fourth International Conference on Computer Sciences and Convergence Information Technology, ICCIT, pp. 1066–1070 (2009)
4. Tak, G.K., Kakkar, A.: Demand based approach to control data load on email servers. In: Tiwari, M.D., Tripathi, R.C., Agrawal, A. (eds.) Proceedings of the First International Conference on Intelligent Interactive Technologies and Multimedia (IITM 2010), pp. 266–270. ACM, New York (2010), http://doi.acm.org/10.1145/1963564.1963611, doi:10.1145/1963564.1963611
5. PHP, AJAX, MySQL and JavaScript Tutorials, http://www.w3schools.com/
6. Naramore, E., Gerner, J., Scouarnec, Y.L., Stolz, J., Glass, M.K.: Beginning PHP5, Apache and MySQL Web Development ISBN: 9780764579660
7. Yang, Y., Elfayoumy, S.: Anti-spam filtering using neural networks and Bayesian classifiers. In: Proceedings of the 2007 IEEE International Symposium on Computational Intelligence in Robotics and Automation, Jacksonville, FL, USA (2007)
8. Pantel, P., Spamcop, D.L.: A spam classification and organization program. Learning for Text Categorization, Papers from the 2006 Workshop, Madison, Wisconsin, AAAI Technical Report (2006)
9. Foster, I., Kesselman, C.: The Grid: Blueprint for a New Computing Infrastructure. Morgan Kaufmann, San Francisco (1999)
10. Sakkis, G., Androutsopoulos, I., Paliouras, G.: A memory based approach to anti-spam filtering. Information Retrieval 6, 49–73 (2003)
11. Rathore, S.K., Jassi, P., Agarwal, B.: A New Probability based Analysis for Recognition of Unwanted Emails. International Journal of Computer Applications 28(4), 6–9 (2011); Published by Foundation of Computer Science, New York, USA (2011)
12. Chiu, Y.-F., Chen, C.-M., Jeng, B., Lin, H.-C.: An Alliance-based Anti-Spam Approach. In: Third International Conference on Natural Computation (ICNC 2007). IEEE (2007)
13. De Capitani, D., Damiani, E., De Vimercati, S., Capitani, P., Samarati, P.: An Open Digest-Based Technique for Spam Detection. In: Proceedings of International Workshop on Security in Parallel and Distributed Systems (2004)
14. O'Donnell, A.J., Mankowski, W., Abrahamson, J.: Using E-mail Social Network Analysis for Detecting unauthorized accounts. In: Third Conference on Email and Anti-Spam, Mountain View, CA (July 2006)
15. Boykin, P.O., Roychowdhury, V.P.: Leveraging Social Networks to fight spam. Computer 38(4), 61–68 (2005)
16. Sahami, M., Dumais, S., Heckerman, D., Horvitz, E.: Bayesian approach to filtering junk E-mail. Learning for Text Categorization, Papers from the 1998 Workshop, Madison, Wisconsin (1998)
17. Jindal, N., Liu, B.: Analyzing and Detecting Review Spam. In: Seventh IEEE International Conference on Data Mining (ICDM), IEEE (2007)

On the Fly File Dereplication Mechanism

Paras Gupta, Manu Vardhan, Akhil Goel, Abhinav Verma,
and Dharmender Singh Kushwaha

Motilal Nehru National Institute of Technology Allahabad,
Department of Computer Science and Engineering,
Allahabad – 211004, U. P., India
{cs1006,rcs1002,cs084084,cs084076,dsk}@mnnit.ac.in

Abstract. Resource replication in distributed environment produces issues of secondary storage. Dereplication of resources is required when replication mechanism is hindered due to lack of secondary storage. This paper introduces dereplication approaches that depend upon last modification time, number of replica available and resource size. Comparative study shows that dereplication can be used to overcome the space overhead issue and reduces the dereplication time. Result shows that in case the space required is same but number of files to be dereplicated varies, dereplication time also varies depending on number of files to be dereplicated. Dereplication time will be more for case having large number of files. Also if file size is increasing by 7 times, increase in dereplication time is by 1.5 times. This shows that dereplication time is decoupled from size of files that are dereplicated on the fly dynamically and does not increase proportionally with respect to file size.

Keywords: Dereplication, Distributed Systems, Replication.

1 Introduction

As the use of computer systems and internet is now becoming the part of our day to day life, requirement for services provided by them increases. To fulfill the requirement of services requested by an individual, service availability is an important issue. Distributed systems will take as the solution by various experts as compare to the centralized systems where services, resources, information are distributed over an environment and can be accessed by the members part of that environment.

A basic definition of distributed system in [1] is that a distributed system is a collection of independent entities that cooperate to solve a problem that cannot be individually solved. A term that describes a wide range of computers, from weakly coupled systems such as wide-area networks, to strongly coupled systems such as local area networks, to very strongly coupled systems such as multiprocessor systems [2].

Replication is a mechanism of service or resource placement to provide their availability in case of unavailability of resources and services. Replication is how to replicate data and request actors using adaptive and predictive techniques for selecting where, when and how fast replication should proceed [3].

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 709–718.
springerlink.com

Dereplication is a mechanism to dereplicate / garbage-collect data or request actors and optimize utilization of distributed storagebased on current system load and expected future demands for the object [3].

Dereplication will be done to optimize the utilization of storage space when a demand for a resource will made. The file to be dereplicated must be carefully taken into consideration of the future demands of a file. File currently being serviced cannot be dereplicated. The number of previously replicated files selected for dereplication can fulfill the requirement for storage space need of the upcoming file to be replicated. Dereplication is considered as a part of resource management process where as replication is considered as a part of resource placement process.

2 Related Work

Globally available various resource management policies and mechanisms represent a step towards efficient and adaptive resource management improving utilization of resources which results in improving the performance of system by reducing several overheads. Venkatasubramanian in [3] discuss about the security and timeliness application requirements using a using a customizable and safe middle ware framework called as CompOSE|Q. N. Venkatasubramanian describes the design and implementation of CompOSE|Q which is a QoS-enabled reflective middle ware framework. Also, to improve the performance of the system in the field of continuous media application, resource management technique is helpful in improving the utilization of resources. In [4], Chou Cheng-Fu et. al. describes various resource management policies on threshold basis in context of continuous media (CM) servers in the area of multimedia application. Venkatasubramanianet. al. in [5] discusses the two replication policies, these are static and dynamic. The division is based upon the number of copies of a file which is termed as degree of replication. In static replication policies, the degree of replication is constant while dynamic replication policies allow it to vary with time.

Santryet. al. in [6] identified four file retention policies for Elephant and have implemented these policies in their prototype. The policies are viz., Keep One, Keep All, Keep Safe and Keep Landmarks. Keep One provides the non-versioned semantics of a standard file system. Keep All retains every version of the file. Keep Safe provides versioning for undo but does not retain any long-term history. Keep Landmarks enhances Keep Safe to also retain a long-term history of landmark versions.

Hurley and Yeap[7]propose a file dereplication method based on β time interval that decides the frequency of invoking the dereplication operation. Over time, all files will eventually be candidates for migration/replication. Although many exist, the one we choose is as follows: every β time units (where β is a uniform time interval which defines the time between dereplication events), storage sites will decide which file qualifies for dereplication. The dereplication policy chosen applies the least recently used concept (i.e., the file selected for dereplication is the file which was not requested for the longest period of time at the storage site). Once the file has been selected, it will be removed from this storage site. Using β, it is possible to create a variety of dereplication policies: the smaller the value of β, the greater the frequency of dereplication, and the larger the value of β, the longer a file copy remains in the system.

3 Problem Definition

During replication when a File Replicating Server (FRS) creates a replica of file on the peer nodes, space management issue arises i.e. whether space is available or not in the secondary storage of the peer nodes on which the file needs to be replicated. If space is available, the file will get copied, but if space is not available dereplication of previously replicated files needs to be done in the secondary storage of that peer node.

Dereplication of files will take place in a manner such that it will fulfill the size requirement of upcoming files. While maintaining the space management overhead, deletion of file should depend on the three criteria which will be discussed in section 3.1.

3.1 Parameters to Be Used

Solution to this problem will be represented on the basis of three parameters of a file which are last modification time of the file, number of replica available of a file and file size.

- *Last Modification Time of a File*:Last modification time is the time at which the file was last modified or last used.
- *Number of Replicas Available of a File*:Number of replicas available of a file is a count on number of copies available for a particular file. Whenever a copy of file is created, it will increase the number of replicas available of a file.
- *Size of a File*:File size is the size of a file required on a disk.

4 Proposed Solution

With everything being lodged on internet, computing paradigm is changing fast to harness this capability. Many information servers and files are resident on various machines and this can be effectively utilized by the users. We present a scenario discussed in section 4.1, although on a smaller scale where geographically disparate clusters interact with each other for information sharing through replication. Each of these cluster are owned by respective Institutes.

In proposed model, we talk about space overhead in replicating file on the storage site. If space is available, the file will get replicated, otherwise dereplication of previously replicated files needs to be done in that directory.

4.1 Architecture Used

One node in each cluster is designated as FRS. FRS can also be replicated on some other node in the cluster for backup and recovery. The scenario presented in the paper is illustrated in figure 1 and is elaborated subsequently.

The proposed architecture consists of loosely coupled systems, capable of providing various kinds of services like replication, storage, I/O specific, computation specific and discovery of resources. Based on the application requirement, the resources are made available to other nodes.

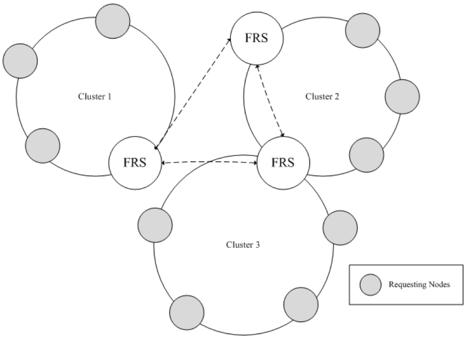


Fig. 1. Architecture

Figure 1 shows a network of three clusters that are connected to each other via intercommunication network. Each cluster consists of a group of trusted nodes and a File Replicating Server (FRS) assigned to these nodes. A FRS can be ‘local’ or ‘remote’. A FRS is assigned to a subset of nodes known as local FRS and FRS positioned outside that cluster, will be called as remote FRS. Each subset of nodes (denoted as requesting nodes) receives the list having IP-address of remote FRS, to increase fault tolerance capability. But the nodes of a cluster will send the file request only to the local FRS. In case of the failure of the local FRS, a node can automatically select a remote FRS from the list and file request will be routed to the selected remote FRS. This makes the model robust and capable of handling crashes in case of local or even remote FRS fails. The system will keep functioning under all circumstances and will never come to halt. Each FRS maintains two tables:

1. File request count table with the following attributes: <file_id, file_name, request_count, meta data>.
2. Peer FRS table with the following attributes: <FRS_IP, FRS_PORT>.

Each FRS is informed whenever a new FRS is added to the network, to updates its peer FRS table. FRS does not monitor and maintains the status of remote FRS, instead FRS request for the current status of remote FRS on-demand. FRS status can either be ‘busy’ or ‘ready’.

Threshold based file replication works as follows:

Each local FRS is responsible for accepting the file request and based on its current status (checks if the number of requests currently serving for a particular file is below the threshold or not), in the following manner:

1. If the status of local FRS is ‘ready’, the local FRS will fulfill the request.
2. If the status of local FRS is ‘busy’, it looks for a remote FRS that can handle the request, by one of the following manner, described as under:

The local FRS contacts the remote FRS that can handle the request by the available copy of the requested file i.e. the status of remote FRS is ready. If not so, the local FRS contacts those remote FRS on which the requested file is not available. In that case file replication will be initiated, by the local FRS of the cluster and the file replica will be created on remote FRS on which the file is not available. For both the cases mentioned above, IP address of the remote FRS that can handle the request will be send to the requesting node. On receiving the IP address, the requesting node will connect to the remote FRS and receives the file, without any user intervention. Thus the overhead of polling and broadcasting is reduced.

4.2 Approaches Proposed for Dereplication

Dereplication of files will take place in a manner such that it willfulfill the size requirement of upcoming files. While maintainingthe space management overhead, three approaches for file dereplication are discussed below.

Last Modification Time Based Approach
In this approach, files are sorted on the usage basis file that was not requested for longest period of time will be selected for dereplication. A drawback of this approach is that if only one requested file is there before deletion, it causes loosing of information. So a check is performed before dereplication which will be done on number of replica available basis approach.

Number of Replicas Available of a File Based Approach
In this approach, files having many copies or the files with more than one replica are dereplicated only when there is not sufficient space available for new replicated files. Files with one replica are not dereplicated to avoid losing information of the file. In this case, before the dereplication of file, a check is performed, whether or not there are other copies of file available or not. If only single copy of file exists in the system, in that case next probable file for dereplication will be selected from the sorted file list on the basis of last modification time.

File Size Based Approach
File size based dereplication approach is used when time required for dereplication considered as important factor. When there is a very little difference in the last modification time of the two files and number of replicas available of both files is more than one, dereplication of file with minimum file size among them will take place to avoid the delay in the process and complete it in the less time.

The proposed approach for dereplication will be described in Figure 2. The detailed description of the number labeled arcs will be describedin sequential manner as follows:

1. *Node A of cluster$_1$* sends connection request to *FRS$_1$*.
2. *FRS$_1$* sends ip addresses of peer FRS and resource list to n*ode A of cluster$_1$*.
3. *Node A of cluster$_1$* sends request for file *f_1* to *FRS$_1$* at time *t_0*.
4. *Node A of cluster$_1$* starts receiving requested file *f_1* from *FRS$_1$*.
5. *Node D of cluster$_1$* sends connection request to *FRS$_1$*.
6. *FRS$_1$* sends IP addresses of peer FRS and resource list to n*ode D of cluster$_1$*.
7. *Node D of cluster$_1$* sends request for same file *f_1* to *FRS$_1$* at time *t_1*.
8. As *FRS$_1$* can fulfill only one request at a time because the value of file threshold is *1* on *FRS$_1$*, so *node D of cluster$_1$* will look for another FRS in the system, here *FRS$_2$*, to fulfill its request.

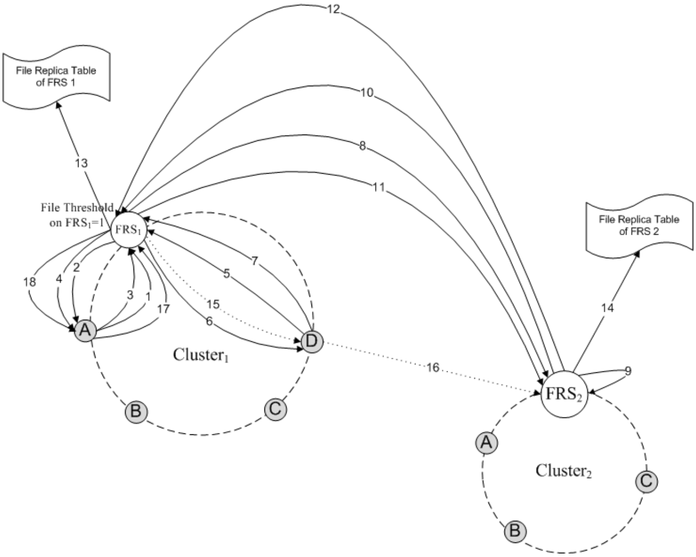

12
10
8
11
File Replica Table of FRS 1
13
File Threshold on FRS$_1$=1
FRS$_1$
18
4
2
3
1
17
7
5
15
6
A
D
Cluster$_1$
B
C
16
File Replica Table of FRS 2
14
9
FRS$_2$
A
Cluster$_2$
C
B

Fig. 2. Proposed Model

9. To fulfill the request of *node D of cluster$_1$* replication of requested files is initiated by *FRS$_1$* as the requested file is not present on *FRS$_2$*. So *FRS$_1$* sends the size of the file to be replicated to *FRS$_2$*.
10. *FRS$_2$* does not accept the file replication request because of space/storage scarcity. *FRS$_2$* initiates dereplication operation on set of previously replicated files. The required amount of space is made available on *FRS$_2$*. If the secondary storage on *FRS$_2$* did not contain any replicated files then user interruption will come, as dereplication of non-replicated file is not allowed.
11. *FRS$_2$* sends message 'ready to receive file f_1' to *FRS$_1$*.
12. *FRS$_1$* starts replicating the file *f_1* to *FRS$_2$*.
13. *FRS$_2$* sends message 'replication of file f_1 to be done successfully' to *FRS$_1$*.
14. *FRS$_1$* updates its file replica table.
15. *FRS$_2$* updates its file replica table.
16. *FRS$_1$* sends IP address and port of *FRS$_2$* to n*ode D of cluster$_1$* informing that the file *f_1* is now available on *FRS$_2$*.
17. Request of *node D of cluster$_1$* for file *f_1* will now be fulfilled by peer FRS, *FRS$_2$*.
18. After some time *node A of cluster$_1$* request same file *f_1* from *FRS$_1$*.
19. In case file with the same name already exists on the n*ode A of cluster$_1$*, file dereplication will be done on that node then the file transfer from *FRS$_1$* to *node A of cluster$_1$* will be initiated.

4.3 Stability Analysis

According to Figure 3, the communication between a requesting node and a FRS (*Source* A and FRS$_1$) is described as follows: *Source* A sends a file request to FRS$_1$through $\overline{M_1}$.FRS$_1$ will receive the request of *Source*A represented as M_1. In return, FRS$_1$ sends file to *Source* A shown by M_3 received on *Source* A using $\overline{M_3}$.

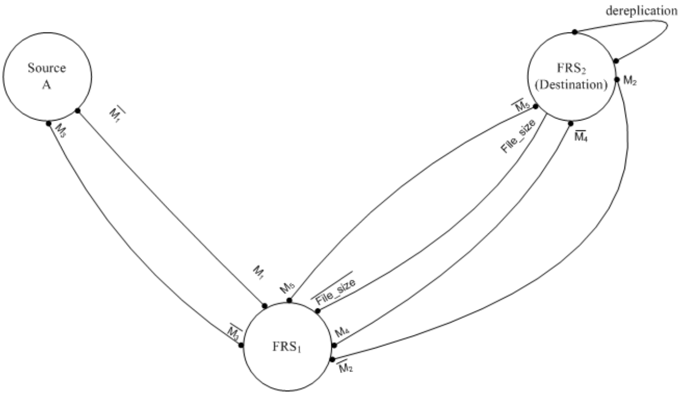


Fig. 3. File Dereplication Model Flow Graph in Process Algebraic Approach

The total communication between requesting node *Source* A and FRS$_1$ with internal actions (τ) will be given by equation 1 as follows:

$$SourceA \overset{\text{def}}{=} \overline{M_1}.M_3.\tau.SourceA \tag{1}$$

Also as shown in Figure 3, communication between the two existing FRS in the architecture (FRS$_1$ and FRS$_2$) is described as follows: FRS$_1$ will send file size of the file to be replicated using $\overline{File\_size}$ which will be received at FRS$_2$ end by $File\_size$. When file size is received by FRS$_2$, it initiates dereplication operation on set of previously replicated files which will be represented by $\tau.derplicate\_file$, which is file dereplication with internal actions (τ). After the successful completion of dereplication operation, the required size for replication will be available on FRS$_2$. Now, FRS$_2$ will send 'ready to receive replicated file' message to FRS$_1$ represented through $\overline{M_4}$. FRS$_1$ received this message using M$_4$. After receiving the message, FRS$_1$ will send the file to be replicated to FRS$_2$ represented by message $\overline{M_2}$. FRS$_2$ will receive the file send by FRS$_1$ represented as M$_2$. When the file will be replicated successfully on FRS$_2$, it will send a message 'successful replication done' to FRS$_1$ by $\overline{M_5}$which was received by FRS$_1$ using M$_5$.

As shown in Figure 3, FRS$_1$ and FRS$_2$ will act as source node and destination node respectively. From this, we can build the definition of FRS$_1$ and FRS$_2$ whichis defined as by the equation 2 and 3 respectively:

$$FRS_1 \overset{\text{def}}{=} M_1.\overline{File\_size}.M_4.\overline{M_2}.M_5.\overline{M_3}.FRS_1 \tag{2}$$

$$FRS_2 \overset{\text{def}}{=} File\_size.\tau.\overline{dereplicate\_file}.\overline{M_4}.M_2.\overline{M_5}.FRS_2 \tag{3}$$

From the equations 1, 2 and 3, we can build the complete system as defined by the equation 4:

$$FDM \overset{\text{def}}{=} Source \,||\, FRS \,||\, Destination \tag{4}$$

5 Results and Discussion

To overcome from the overhead of space management issue, a data structure consisting of a table considered which is described in Table 1.

Table 1. Attributes

Attribute Name	Last Modification Date	Last Modification Time	File Name	File Size	File Replica
Type	yyyy-mm-dd	hh:mm	String	Long	Integer

Replicated files on the storage site will be sorted based on least recently used parameter which will be obtained using the combination of both last modification date and last modification time. The list of replicated files will be sorted in descending order. Example of a data structure of available files maintained at the storage site is described in Table 2.

Table 2. Data structure example for comparison between approaches

Last Modification Date	**Last Modification Time**	**File Name**	**File Size (in MB)**	**File Replica**
2011-12-21	20:08	a.mp3	3	4
2011-12-08	22:48	b.mp3	500	1
2011-11-23	16:36	c.mp3	100	2
2011-11-23	16:03	d.mp3	250	1
2011-11-09	20:11	e.mp3	50	1
2011-11-09	18:47	f.mp3	5	4
2011-11-09	18:43	g.mp3	10	2

The Figure 4 plots efficiency of all the three approaches versus load based on the data shown in Table 2 and the three approaches based on least recently used parameter, replica counts and file size parameters. Efficiency calculated is proportional to the reciprocal of extra memory size vacated during dereplication.

Unlike 2nd and 3rd approaches (i.e. number of replica available of a file basis and file size basis respectively), 1st approach(i.e. last modification time basis) is based only on least recently used parameter and disregards the replica counts and file size parameters. Thus it may even delete the last replica of file present in system. While 2nd approach is based on both least recently used and replica counts parameters and disregards the file size parameter. 3rdapproach is based on all the three parameters, least recently used, replica counts parameters and file size parameter. Percentage efficiency of 2ndand 3rdapproach is always better than 1stapproach while in some cases percentage efficiency of 3rdapproach is also better than 3rdapproach. All the three approaches said to be 100% efficient only when space required before dereplication and after dereplication will be same.

Table 3. Dereplication time in required space

Number of Files dereplicated	Space Required (in MB)	Space Freed (in MB)	Dereplication Time (in msec)
1	6	6.0523	60
2	7.8607	13.1792	75
3	20.0399	21.0399	77
3	36.2634	39.7985	79
5	36.2634	59.7151	96
5	43.9405	51.0140	98

Dereplication time increases, as the number of files not accessed for the longest period and smaller in size, are more as compared to the files that are larger in size.Table 3 shows when the space required is same but the number of files to be dereplicated varies, dereplication time also varies depending on the number of files to be dereplicated. Dereplication time will be more for the case having large number of files.

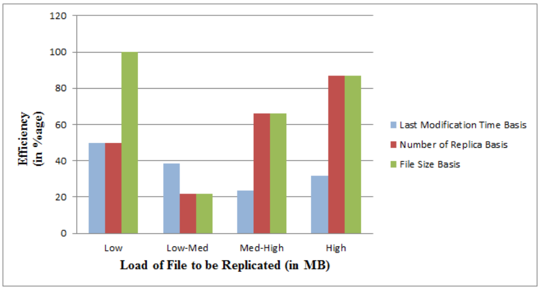


Fig. 4. Comparison of the Three Approaches

Table shows that if file size increases 7 times i.e. from 6 MB to 43.9405 MB, the increase in dereplication time is only 1.5 times i.e. from 60 milisec to 98 milisec. This shows that the dereplication time is decoupled from the size of files that are dereplicated dynamically and does not increase proportionally with respect to the file size.

6 Conclusion

This paper proposes approach that tackles the issue of space overhead in a distributed system environment. Proposed approach resolves this issue of space overhead. Dereplication time increases, as the number of files increases that are not accessed for the longest time period and smaller in size as compared to the files that are larger in size. Result shows that, in case when the space required is same but the number of files to be dereplicated varies, dereplication time also varies depending on the number of files to be dereplicated. Dereplication time will be more for the case having large number of files. If file size increases 7 times, the increase in dereplication time is only 1.5 times. This shows that the dereplication time is decoupled from the size of files that are dereplicated on the fly dynamically and does not increase proportionally with respect to the file size.

References

1. Kshemkalyani, A.D.: Distributed Computing Principles, Algorithms and Systems
2. Gupta, M., Ammar, M.H., Ahamad, M.: Trade-offs between reliability and overheads in peer-to-peer reputation tracking. Computer Networks 50(4), 501–522 (2006)
3. Venkatasubramanian, N.: CompOSE|Q - a QoS-enabled customizable middleware framework for distributed computing, Electronic Commerce and Web-based Applications/Middleware. In: 19th IEEE International Conference on Distributed Computing Systems, pp. 134–139 (1999)
4. Chou, C.-F., Golubchik, L., Lui, J.C.S.: Striping doesn't scale: how to achieve scalability for continuous media servers with replication. In: 20th International Conference on Distributed Computing Systems, pp. 64–71 (2000)
5. Venkatasubramanian, N., Deshpande, M., Mohapatra, S., Gutierrez-Nolasco, S., Wickramasuriya, J.: Design and implementation of a composable reflective middleware framework. In: 21st International Conference on Distributed Computing Systems, pp. 644–653 (April 2001)
6. Santry, D.S., Feeley, M.J., Hutchinson, N.C., Veitch, A.C., Carton, R.W., Ofir, J.: Deciding when to forget in the Elephant file system 33(5), 110–123 (December 1999)
7. Hurley, R.T., Soon, A.Y.: File migration and file replication: a symbiotic relationship. IEEE Transactions on Parallel and Distributed Systems 7(6), 578–586 (1996)

Security Service Level Agreements Based Authentication and Authorization Model for Accessing Cloud Services

Durgesh Bajpai, Manu Vardhan, Sachin Gupta, Ravinder Kumar, and Dharmender Singh Kushwaha

Motilal Nehru National Institute of Technology Allahabad
Computer Science Engineering Department
Allahabad 211004, India
{is1023,rcs1002,cs084054,cs084107,dsk}@mnnit.ac.in

Abstract. Cloud computing is defined as delivering of computing resources as a service. Discovery of reliable resource provider and access control are key components of cloud computing. Service level agreements are negotiated between the service provider and enterprise. This paper proposes authentication interface to access a cloud service. User authentication token is required to validate whether the user is registered employee of enterprise or not. Service authentication token is required to validate the access right of a user for service. Service selection is acquired via monitoring of security measures of services provided by a service provider through security service level agreements at enterprise end. Thereby, completely relieving end user from the nitty-gritty of service providers in comparison to approaches proposed in past. Single sign on mechanisms for user and services is used. Features like Denial of service, man in the middle attack and access control rights of employees are also handled.

Keywords: Authentication, Cloud, Denial of Service attack (DOS), Kerberos, Services, Service Level Agreement, Symmetric Encryption.

1 Introduction

Cloud computing has been envisioned as the next generation architecture of IT enterprise. Cloud consumers face various challenges such as security, privacy and discovery of reliable resource provider with the increase of public cloud providers.

A service level agreement (SLA) is maintained between the service provider and the consumer of the service about the quality parameters of the service which will be delivered by the service provider. In general SLAs consider the terms like packet loss, delay, throughput, etc. The security service level agreement (Sec-SLA) is a specific SLA that deals with metrics related to security instead of the traditional metrics of a service.

In the scenario as shown in Fig. 1 where an enterprise wants to store its data on the cloud, to choose a right service provider is very critical. To eliminate the denial of service due to the traffic on a particular service provider, an enterprise registers with more than one service provider providing the same data storage as the service through negotiating SLAs. A trusted third party plays the role of authentication interface

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 719–728.
springerlink.com

between an enterprise and cloud service provider to access the cloud service. An employee does not directly communicate with a service provider, as it goes through the enterprise via authentication interface. Authentication interface provides the employee authentication token for authenticating a registered employee, a service authentication token for authenticating a employee for accessing the service. Different service providers provide services based on different security measures like different encryption mechanisms viz., AES, DES and RSA to store the data, thus provides different data confidentiality and integrity. To relieve the burden of employee, the record of the security measures of services provided by different service providers should be maintained at the enterprise end. Before providing the service authentication token, security measures of services provided by different service providers, maintained through security service level agreements (Sec-SLAs) should be analyzed and accordingly a service provider should be selected to fulfill the request of the employee. Access rights given to employees for accessing a service according to their roles in an enterprise has to be considered before granting the service authentication token.

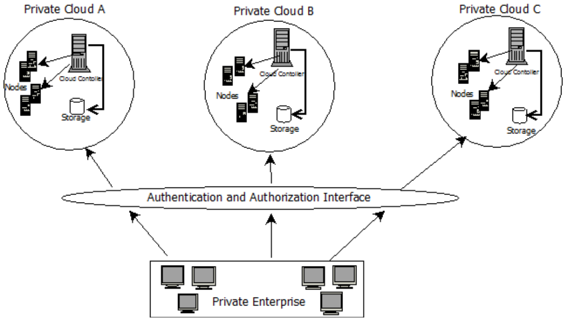


Fig. 1. Cloud computing Scenario

To access a cloud service registered with the enterprise, employees need to be authenticated as well as authorized by the help of employee and service authentication token. Enterprise employee's access rights records should be maintained at the enterprise end only and should not be shared with the authentication interface to eliminate the chances of manipulation of records.

A new authentication approach is proposed taking into account the security measures of services and their dynamic changes, provided by the service providers through the help of the Sec- SLAs. The proposed authentication interface is trusted third party trusted by the service provider and the enterprise. The access rights of an end user while accessing the cloud service are considered. The new approach tries to solve the following issues of previously proposed authentication approaches.

a. In Pippal et al. [9] proposed approach the service authentication credentials are given to those employees also who don't have the access rights for the particular requested service. The possibility of denial of service attack increases, as service authentication credential granting server will get overloaded with the unauthorized employees demand.

b. In Hota et al. [11] proposed model, an enterprise shares information regarding employee's credentials and access rights of employees with the service providers, proving as a source of leakage and misuse of employees information crucial to an enterprise.

c. Authentication token is used in approach proposed by the Tao et al. [10] to resolve the problem of the single sign on of the user to access a service. If the same user wants to access the same service very frequently, the employee validity for the service is checked again and again through the authentication token, thus increasing the overhead.

d. An employee has to check the trust level of a service provider before accessing the service. For this an employee should have all the information regarding the trust level and the features provided by a service provider increasing the load on an end user.

e. Previously proposed models only consider SLAs while registering with cloud providers, don't consider security measures of services in general and their dynamic variations with time.

The rest of the paper is organized as follows. The next section discusses various related work done in the context of authenticating a user to access a cloud service and the service level agreements so far. Section 3 proposes a new trusted authentication interface between the service provider and an end user. In section 4 the proposed approach results are evaluated. The final section concludes the work followed by the references.

2 Related Work

Cloud Computing is the challenging area of research in IT field, many authors are working on it and tackling the various issues regarding the cloud. Chaves et al [1], Bernsmed et al. [2] and Kandukuri et al. [3] discusses the work done on the service level agreements and pro-posed an approach for the management of the service level agreements in the con-text of hybrid clouds. Authors discusses the security issues in the cloud computing and how they can be solved by the help of the service level agreements. To fulfill the service requests in cloud, Ahuja et al. [4], Clark et al. [5], Daniel and Lovesum [6] and Kubert and Wesner [7] discusses the usage of SLAs and their monitoring. Authors propose a service selection algorithm which allows re-provisioning of resources on the cloud in the event of failures. Author main focus is to provide a fair deal to the consumers, enhance quality of service delivered to the consumer as well as generation of optimal revenue. Liu et al. [8] and Pippal et al. [9] discuss approaches for mutual authentication and authorization, addressing the issue of establishing trust across heterogeneous domains. Tao et al. [10] discusses a generic authentication interface to allow the user to access the diverse Clouds in a unified way by the help of the authentication tokens. The proposed interface also combines different clouds enabling inter-cloud communications. Hota et al. [11] and Zhang et al. [12] addresses the issues of data security in cloud storage system and access control by using the capability based access control techniques. Data storage safety issue is tried to address by dividing the various technologies to make the data storage safe roughly into the storage protect, transfer protect and authorize.

3 Proposed Approach

In the proposed model as shown in Fig. 2.an enterprise and the cloud service provider are present in different domains having different security policies. Assumptions made for the proposed model are:

- An employee is registered with his enterprise.
- A role is assigned to an employee and according to it the permissions to access a service that are registered with an enterprise.
- An enterprise is registered with the service providers which are providing the services of its need after negotiating the service level agreements.
- Employee uses a service and records the security measures of services (like integrity of the data, backup frequency, reliability, confidentiality of the data, data store laws according to the domain) provided by the service provider and give feedback to the enterprise about the service provider.

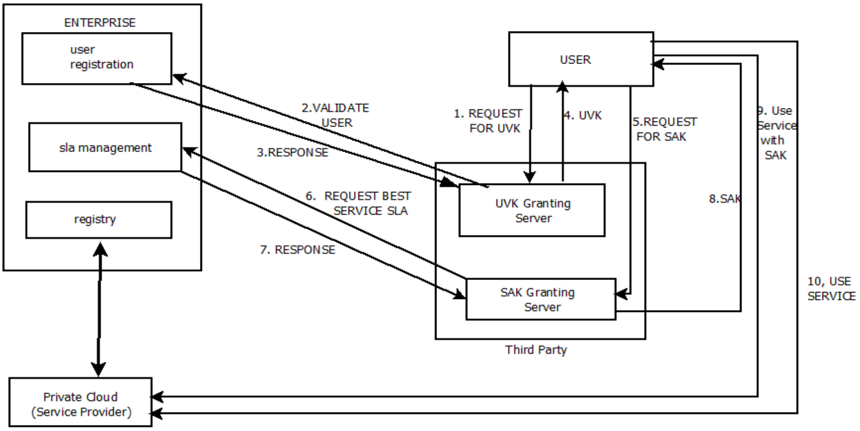


Fig. 2. Proposed Model

The enterprise as shown in Fig. 2. consists of the following units:

Registry: Responsible for registering a service provider with an enterprise after negotiating the service level agreements.

Employee Registration Unit (ERU): An employee registers with an enterprise through this unit and it provides an employee with a unique employee id and a unique password. Records of employee credential (Password) are maintained, to check whether the employee is the registered employee of an enterprise or not before a user authentication credential i.e. User Validation Key (UVK) is granted.

Sec-SLA Management Unit (SMU): It keeps record of the security measures of services provided by the various service providers. A Sec-SLA is maintained about the service reliability, availability, backup frequency, data confidentiality, data integrity, domain where the data get store as the security policies change for different domains. Records of employee's access rights are maintained, to check the validity of employee to access a service, before a service access credential i.e. Service Access Key (SAK) is granted.

Third party authentication interface as shown in Fig. 2. consists of the following units:

User Validation Key Granting Server (UGS): It is responsible for granting UVK to an employee which works as an authentication token. The same authentication token can be used for single sign on till the lifetime of the UVK. Before granting the UVK, it takes the help of the Employee Registration Unit to validate an employee credentials as they are not shared with the third party because of security reasons. UVK is encrypted with the symmetric key K_u, known to the third party i.e. the authentication/authorization interface.

UVK: E (K_u [empid||special_code||timestamp|||lifetime||random_no])

Service Access Key Granting Server (SGS): It is responsible for granting a SAK to an employee to access a service after validating the UVK. Employee can use the same SAK to access the same service up to a certain time limit, till the SAK don't expire. SAK is encrypted with symmetric key Ks known to the third party i.e. the authentication/authorization interface and the service provider, shared between them initially by a secure procedure.

SAK: E (K_s [empid||special_code||timestamp|||lifetime||random_no])

In the Proposed model the message flows as shown in the Fig. 3. total ten numbers of messages are required to access a service as described below.

1. An employee sends a request to UGS for the UVK.
   E→UGS: empid||password
2. UGS asks the ERU for the validity of the employee.
3. ERU returns a positive response if the employee pass correct credentials otherwise returns a negative response.
   ERU→UGS: True/False
4. UGS grants the UVK to the employee if the response of the ERU is true.
   UGS→E: UVK
5. Employee requests the SGS for SAK.
   E→SGS: empid||special_code||UVK||serviceid
   SGS verifies the validity of the UVK if its valid then goes to step 6.
6. SGS requests the appropriate service provider reference from the SMU.
   SGS→SMU: serviceid||empid
7. SMU gives the service provider reference to the SGS by checking out the security measures of services provided by the service providers and access rights of the employee.
8. SGS grants the SAK to the employee for a particular service.
   SGS→E: SAK
9. The employee requests the service from the service provider.
   E→SGS: SAK||empid||special_code||serviceid
10. The service provider checks the validity of the SAK and grants the service to the employee if SAK is valid.

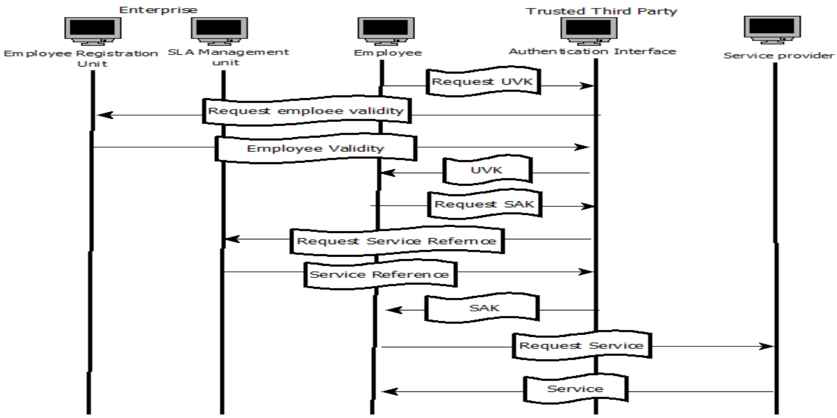


Fig. 3. Message flow sequence of the proposed model

A Rivest-Shamir-Adleman (RSA) based approach is used to make the UVK and SAK generation and validation algorithms more secure.

1. Select two prime numbers p and q.
2. n=p*q.
3. φ (n) = (p-1)(q-1), where φ is Euler's Totient Function.
4. Select an integer e such that 1 < e < φ (n) and greatest common divisor of (e, φ (n)) = 1, i.e. e and φ(n) are co prime, e will work as the public key.
5. d = e–1 mod φ(n); i.e. d is the multiplicative inverse of e mod φ(n), where d will work as the private key.

Different private and public keys are generated and used for the SAK and the UVK generation and validation process. Private key d and public key e used for UVK are known to UGS only. For SAK private key d_1 is known to the SGS and public key e_1 is known to the service provider providing the service for which the SAK is granted. A random no is included in the UVK and the SAK by the help of the d and e.

Algorithm to generate the UVK

1. Generate a random number r, through random number generator function.
2. K2= r^e mod n, by using standard cryptography Rivest-Shamir-Adleman (RSA) algorithm. Here e is the public key.
3. Add a constant c. K3=K2+c.
4. String=concatenation of the employee id and the special code passed by the user while requesting the UVK.
5. Compute the substring KEY1 of 10 bytes from the symmetric key K1 using random number r.
6. Encrypt the String with KEY1. enc = encrypt (KEY1, String).
7. Compute Ticket1 i.e. the concatenation of the K3, current time T, lifetime L for which the UVK has to be valid and enc.
8. Finally encrypt the Ticket1 with the symmetric key K_u to get the final authentication token UVK. UVK=encrypt (K_u , Ticket1).

Algorithm to check the validity of the UVK

1. Decrypt the UVK with the symmetric key K_u . Ticket1= decrypt (K_u , UVK)
2. Fetch the lifetime L and T from the Ticket1.
3. Check the expiration of the lifetime of UVK by using the current time.
4. Fetch K3 from the Ticket1.
5. K2=K3- c
6. r = $K2^d$ mod n; by using standard cryptography Rivest-Shamir-Adleman (RSA) algorithm. Here d is the private key.
7. Compute KEY1 of 10 bytes from the symmetric key K1 using random number r.
8. Fetch the enc from the Ticket1.
9. Decrypt the enc with the KEY1, get the String. String= decrypt (KEY1, enc).
10. Compare the String and the concatenation of the employee id and the special code provided by the user while using UVK. If both are same the UVK is valid otherwise not.

The algorithm to generate and validate the SAK is same as of UVK except the symmetric key K_s is used in place of K_u. K_s is shared between the service provider and the authentication interface initially by a secure means.

4 Results and Analysis

Message Flow: Proposed model uses total 10 numbers of messages for accessing a distributed cloud service. The comparison with other models based on the number of messages required to access a service is shown in Table 1. Based on the number of messages exchanged in different models, the load on the systems implementing the models due to message overhead varies as shown in the Fig. 4.

Table 1. Comparison based on the number of messages

		Kerberos	Pippal et al. Model [9]	Proposed Model
1.	Number of messages to get the User Authentication Key	2	4	4
2.	Number of messages to get the Service Access Key	4	6	8
3.	Number of messages to access the service	6	12	10
4.	Total number of messages	6	12	10

In the proposed model the number of messages increases while acquiring the following benefits:

To acquire a user authentication credential i.e. UVK, two extra messages are used to validate the employee credentials through ERU by the UGS. This overcomes the problem of misusing the employee credentials by third party authentication interface, which is stored at ERU and not shared with the third party.

To acquire a service access credential SAK, two extra messages are used to check the access rights of an employee and providing the reference of a service provider, considering security measures of services provided by the different service providers through SMU by the SGS. This overcomes the chances of employee access rights alterations and misuse by third party authentication interface as they are stored at SMU and not shared with the third party. All the information regarding the security measures of services provided by a service provider are handled by the SMU, thereby relieving the employee.

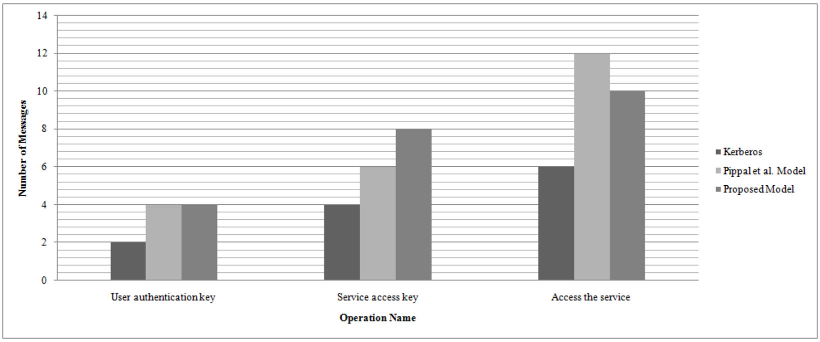


Fig. 4. Graph showing message overhead required in different models

Security Service Level Agreements (Sec-SLAs)

Security measures of services provided by a service provider are handled by SMU without the intervention of an end user through the help of the Sec-SLAs, thereby relieving the end user of service providers as shown in Fig.5. in comparison to the Hota et al. [11] approach where all the information regarding a service provider are handled at the user end only. Selection of service provider is done by considering the Sec-SLAs maintained of different service providers.

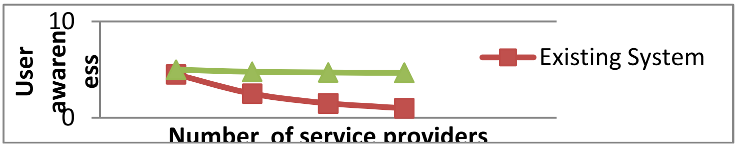


Fig. 5. Graph showing user awareness of service providers

Security Analysis

- *Denial of Service Attack (DOS):* Access rights of an employee for a particular service is checked before granting the SAK through SMU, the SAK granting server will not get overloaded by the requests of the unauthorized users, thus reducing DOS.
- *Access Control Rights:* Access rights of an employee are checked before providing him the SAK. Access rights are not shared with the authentication inter-face; there usage is done from the enterprise end only, increasing the trust level.

- *Man in the Middle Attack:* If a third person is able to capture SAK, even then this key cannot be used, as the special code is incorporated into SAK that is known only to the valid user that owns the SAK. A random number is used in SAK to make it more resistive from the security breaches.
- *Two Level Authentication:* Two level authentication is done at the authentication interface, first the enterprise authentication and after that the employee authentication.
- *Trust Level:* The user validation key (UVK) and service access key (SAK) are delivered by the use of a third party authentication interface trusted by both the service provider and the enterprise.
- *Encryption:* Encryption and decryption in the proposed model are taking place to encrypt and decrypt the UVK and the SAK with the help of the symmetric key. A RSA based asymmetric keys are also used to incorporate a encrypted random number into the UVK and SAK overcoming the issues of man-in-the middle attack.
- *Security Measures:* The Security measures of services provided by a service provider and their dynamic change are considered in the proposed model while accessing a service.
- *Single sign on:* In the new proposed model user authentication token UVK is used for user single sign on, service authentication token SAK is used for the single sign on for a particular service making the approach more efficient.

On the various features a comparison of the proposed model with the Pippal et al. [9] and Tao et al.[10] proposed approaches is shown in Table 2. The results show that the proposed model provides extra new features with the features already provided by the proposed models.

Table 2. Feature based comparison with previously proposed models

Particulars	Pippal et al. approach [9]	Tao et al. approach [10]	Proposed Model
User Single Sign On	Yes	Yes	Yes
Service Single Sign On	Yes	No	Yes
Checking of security Parameters	No	No	Yes
Adequate service registry maintenance	Yes	Yes	Yes
Use of symmetric keys	Yes	Yes	Yes
Checking the access rights	No	Yes	Yes
Handling of Denial of Service Attack	No	No	Yes
Total	4	4	7

5 Conclusion

The proposed model explains the number of messages involved in the process of authenticating employees of an enterprise and provides them access of the services.

Access control rights of a user for a particular service are considered before granting the service access to the user of that service at the enterprise only. To make the system more securely intact, the access rights are not shared with the authentication interface.

Denial of service attack is controlled by not granting authentication tokens to unauthorized users. Man in the middle attack is handled by encrypting the authentication tokens with the help of symmetric and asymmetric encryption. The trust is established between the end user and the service provider through the authentication interface. Security measures of services, maintained through Sec-SLAs are considered while selecting a service to fulfill the user request in order to provide more trusted cloud service. At enterprise end, based on the security measures of service provided by different service providers, service is selected. Thereby completely relieving the end user from the nitty-gritty of service providers as compared to the models proposed in the past that considers the handling of security measures of services by end users. Thus the proposed methodology overcomes the drawbacks of previously defined models.

References

1. Chaves, S.A.D., Westphall, C.B., Lamin, F.R.: SLA Perspective in Security Management for Cloud Computing. In: Sixth International Conference on Networking and Services, pp. 212–217 (2010)
2. Bernsmed, K., Jaatun, M.G., Meland, P.H., Undheim, A.: Security SLAs for Federated Cloud Services. In: Sixth International Conference on Availability, Reliability and Security, pp. 202–209 (2011)
3. Kandukuri, B.R., Paturi, V.R., Rakshit, A.: Cloud Security Issues. In: IEEE International Conference on Services Computing, pp. 517–520 (2009)
4. Ahuja, R., De, A., Gabrani, G.: SLA Based Scheduler for Cloud for Storage & Computational Services. In: International Conference on Computational Science and Its Applications, pp. 258–262 (2011)
5. Clark, K.P., Warnier, M.E., Brazier, F.M.T., Quillinan, T.B.: Secure Monitoring of Service Level Agreements. In: International Conference on Availability, Reliability and Security 2010, pp. 454–461 (2010)
6. Daniel, D., Lovesum, S.P.J.: A novel approach for scheduling service request in cloud with trust monitor. In: Proceedings of 2011 International Conference on Signal Processing, Communication, Computing and Networking Technologies. ICSCCN, pp. 509–513 (2011)
7. Kubert, R., Wesner, S.: Service level agreements for job control in high performance computting. In: Proceedings of the International Multiconference on Computer Science and Information Technology, pp. 655–661. IEEE (2010)
8. Liu, P., Zong, R., Liu, S.: A new model for Authentication and Authorization across Heterogeneous Trust-Domain. In: International Conference on Computer Science and Software Engineering, vol. 03, pp. 789–792. IEEE Computer Society (2008)
9. Pippal, S.K., Kumari, A., Kushwaha, D.K.: CTES based Secure approach for Authentication and Authorization of Resource and Service in Clouds. In: International Conference on Computer & Communication Technology (ICCCT), pp. 444–449 (2011)
10. Tao, J., Marten, H., Kramer, D., Karl, W.: An Intuitive Framework for Accessing Computing Clouds. In: International Conference on Computational Science. ICCS, pp. 2049–2057 (2011)
11. Hota, C., Sanka, S., Rajarajan, M., Nair, S.K.: Capability-based Cryptographic Data Access Control in Cloud Computing. Int. J. Advanced Networking and Applications 03, 1152–1161 (2011)
12. Zhang, X., Hong-Tao, D., Chen, J.Q., Lin, Y., Zeng, L.J.: Ensure Data Security in Cloud Storage. In: International Conference on Network Computing and Information Security, pp. 284–287 (2011)

Load Balancing in Cluster Using BLCR Checkpoint/Restart

Hemant Hariyale, Manu Vardhan, Ankit Pandey, Ankit Mishra,
and Dharmender Singh Kushwaha

Motilal Nehru National Institute of Technology Allahabad
Department of Computer Science and Engineering
Allahabad-211004, U. P., India
{cs1010,rcs1002,it088003,it088015,dsk}@mnnit.ac.in

Abstract. Modern computation is becoming complex in a way that the resource requirement is gradually increasing. High Throughput Computing is one technique to deal with such a complexity. After a significant amount of time, computing clusters gets highly overloaded resulting in degradation of performance. Since there is no central coordinator in Computer Supported Cooperative Working (CSCW) load-balancing is more complex. An overloaded node does not participate in a CSCW network as they are already overloaded. This paper proposes migration of computation intensive jobs from overloaded nodes, which will allow overloaded nodes to be able to participate in CSCW. The proposed solution improves the performance by making more nodes participating in CSCW by migrating compute intensive jobs from overloaded nodes to underloaded nodes. Evaluation of proposed approach shows that the availability and performance of the CSCW clusters is improved by 30%-40% with fault-tolerance based load balancing.

Keywords: Checkpoint/Restart, CSCW, Fault Tolerance, Job Migration, Load Balancing.

1 Introduction

Computer Assisted Cooperative Working is concept, in which a group of computer nodes participate to share services, computational workload and perform compute intensive jobs cooperatively. Cluster computing is becoming more popular these days and the prime reason for that is high availability, more reliability and more efficient computing. Checkpoint/Restart is a technique of storing the state of process at particular intervals. If the process fails at some point of time, previously stored process states can be used to restart the process from the last stable state. Storing the state of a running process i.e. process id, register values, stack pointer etc. in a file known as context or checkpoint file is known as creating a checkpoint. Checkpointing/Restart allows starting a job from the most recent stable running state saving a lot of computational time and reducing response time of crashed jobs in overloaded conditions. Berkley Lab Checkpoint/Restart (BLCR) is an open source

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 729–737.
springerlink.com

checkpointing and restart utility for Linux clusters that targets the space of typical High Performance Computing applications (HPC) developed at Computational Research Division, Ernest Orlando Lawrence Berkeley National Laboratory.

Cluster is a group of computers interconnected for providing efficient computation for High Throughput Computation (HPC) application. Balancing load for achieving optimal performance forces the job migration. Job migration is termination of job on the overloaded CPU and restarting it on a underloaded or idle CPU. Since a CSCW does not have a central controller or coordinator, balancing load in such situations is more complex and needs to be done in decentralized manner. Migration of job requires transfer of files required for restarting a job on new node. Due to increasing complexity, requirements of computational intensive jobs, the possibility that the process may fail during the course its execution will increase.

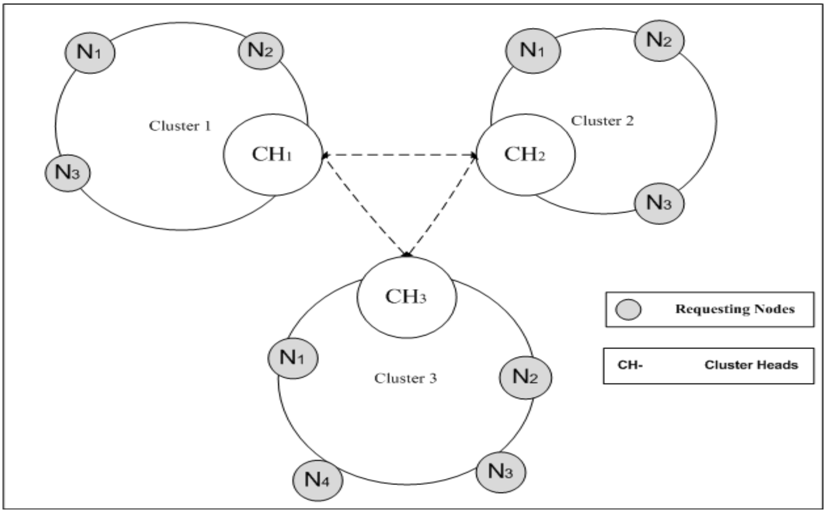


Fig. 1. Architecture

Figure 1 shows three clusters, each having a cluster head. Each cluster head manages some node locally. If one of the cluster head fails because of overloading, all the nodes in its cluster zone will not be able to participate in computer supported cooperative working. Job migration helps in handling issues related to fault tolerance for critical processes as well as for balancing the load among the nodes of cluster. In all job migration improves availability of cluster node in CSCW.

The rest of the paper is organized as follows. The next section discusses the related work done in the context of job migration, checkpointing and load balancing. Section 3 discusses the proposed work. In section 4 performance evaluation is performed which is followed by conclusion in section 5.

2 Related Work

Selikhov and Germain [1] propose a fault tolerant message passing environments that protects parallel applications against node failures. Very large scale computing systems, ranging from large clusters to worldwide Global Computing systems, require a high level of fault tolerance in order to efficiently run parallel applications. Khaled and Kassem [2] present an algorithm for migrating processes in a general purpose workstation environment. In such an environment, parallel applications are allowed to co-exist with other applications, using workstations when released by their owners, and off loading from workstations when they are reclaimed. The applications can dynamically create and terminate processes during their execution. Author presents a migration algorithm that conforms to a set of stated objectives. The presented algorithm enables to implement migration in an environment of general purpose homogeneous workstations. In such an environment, workstations are added and removed from the pool of available resources to the general users based on the release and reclaim of the workstation by its owner, respectively. Joshua and Richard [3] propose an Algorithm Based Fault Tolerance (ABFT) technique to improve the efficiency of application recovery beyond what traditional techniques alone can provide. Applications will depend on libraries to sustain failure-free performance across process failure to continue to use High Performance Computing (HPC) systems efficiently even in the presence of process failure. Optimized Message Passing Interface (MPI) collective operations are a critical component of many scalable HPC applications. However, most of the collective algorithms are not able to handle process failure. Chtepen et al. [4], introduces several heuristics that dynamically adapt the abovementioned parameters based on information on grid status to provide high job throughput in the presence of failure while reducing the system overhead. Furthermore, a novel fault-tolerant algorithm combining checkpointing and replication is presented. Fault tolerance forms an important problem in the scope of grid computing environments. To deal with this issue, several adaptive heuristics, based on job checkpointing, replication, and the combination of both techniques, were designed. Lapriore [5] proposes migration paradigms with reference to a memory environment implementing the notion of a single address space. The operations defined by a given type are the services that may be provided by an object of this type to a client process. When the process and the object are located in different nodes, migrations may represent valid alternatives to remote procedure calls. Migration of the server object causes the memory area storing the internal representation of this object to be copied into the node of the client process. Migration of the client process causes execution of this process to proceed in the node of the server object.

Dynamic Load Balancing (DLB) by Pyali et al. [6] provides application level load balancing for parallel jobs using system agents and DLB agent. The approach requires a copy of system agents on all the system so that DLB agent may collect load information from these systems and perform load balancing. The other contemporary work includes grid load balancing using Intelligent Agents by Cao et al. [7], proposes a combined approach using intelligent agents and multi-agents for effectively scheduling the local and global grid resources that also incorporate peer to peer advertisement and service discovery to balance the workload. The approach requires a copy of system agents on all system so that DLB agent may collect load information from these systems and perform load balancing. Yagoubi and Slimani [8] puts forward a dynamic tree based model to represent grid architecture and proposes Intra-site, Intracluster and Intra-grid load balancing. Nehra et al. [9] addresses issues to balance the load by splitting processes in to separate jobs and then distributing them to nodes using Mobile Agent (MA). The authors propose a pool of agents to perform this task.

3 Proposed Work

This paper proposes a solution in which, a process running on one of the machine of a cluster is migrated to other node if the load on the CPU on which the process originally scheduled, is more than threshold load of the cluster. Scripts are installed that runs on every node in the cluster which periodically calculate the CPU load on that machine, periodically checkpoints the process using BLCR utility and checks for any files that are currently in open state by the process. If at any point of time, CPU load gets over the threshold value the migration script is invoked. In migration procedure, to enabling checkpointing for the process, requires running the process with *cr_run* (blcr utility). The migration script finds the node that has the lowest CPU load in the cluster in past one minutes. Migration of a process starts by copying the exe files, checkpoint file and files that were in use by the job at the time of taking checkpoint, to the node having less CPU load. The migration script resume the job at the chosen node using the *cr_restart* (blcr utility) with the checkpoint file created at the source node.

The whole procedure of migrating job is automated. Hence migration process does not require manual support. This paper proposes a solution, which tries to balance the load by migrating jobs and enable fault tolerance for compute intensive job. Sometimes, a job executing on a node gets crashed during course of its execution, one possible reason for the crash is less resources which implies overloading. Hence approach implicitly does load balancing and also enable fault tolerance. Figure 2 is process flow diagram of the proposed approach for balancing load in CSCW clusters. The proposed approach balances load as well as provides load aware fault tolerance. Average CPU load in considered as the percentage utilization of the CPU during a particular interval. Load is calculated with help of UNIX utility “uptime”, which gives metric of 3 load average numbers. One can interrupt the values as average during past 1, 5 and 15 minutes. A node is considered to be overloaded if it is over utilized as compared to other nodes in the cluster. The threshold load value is decided as double of the average of CPU load of all the nodes in the cluster. For example: if there are 3 nodes in the cluster and their respective load is L1, L2, and L3, load threshold value will be described by eq. 1.

$$L_m = (2 / 3) * (L_1 + L_2 + L_3) \tag{1}$$

Hence for a cluster with n nodes, load threshold will be calculated as shown in eq. 2.

$$L_m = (2 / n) * (L_1 + L_2 + L_3 + \dots\dots\dots + L_n) \tag{2}$$

This threshold value is calculated periodically on every node that has a process running on it. There are two situations in this solution in which the migration procedure is invoked. Migration achieved by copying the checkpoint and I/O files to the destination node using a secure tunnel (Ssh). Results are copied back to the source node after successful termination. In this approach it is assumed that all the nodes in the cluster are trusted user on trusted host. Every node in the cluster should be able to login or copy data from one another without need of any password with secure shell login. A remote user should have the privilege execute process on remote nodes in the cluster.

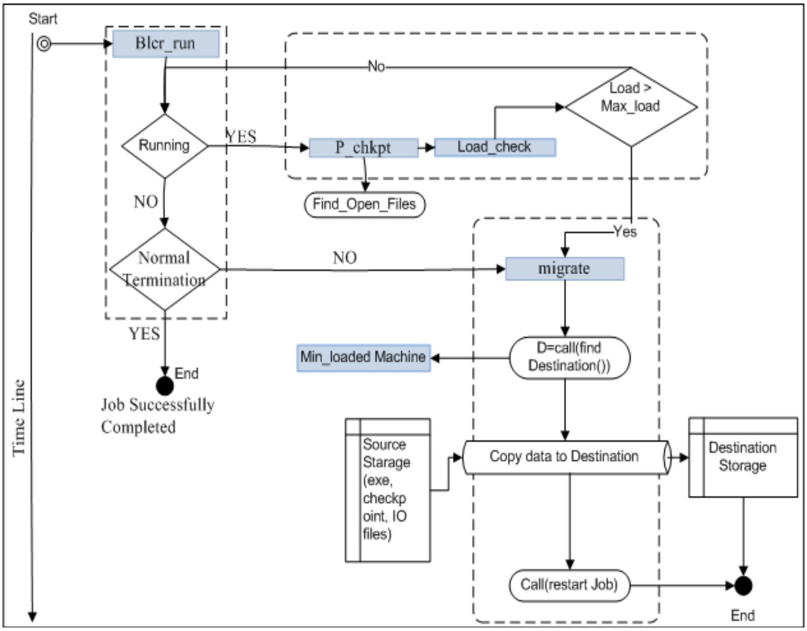
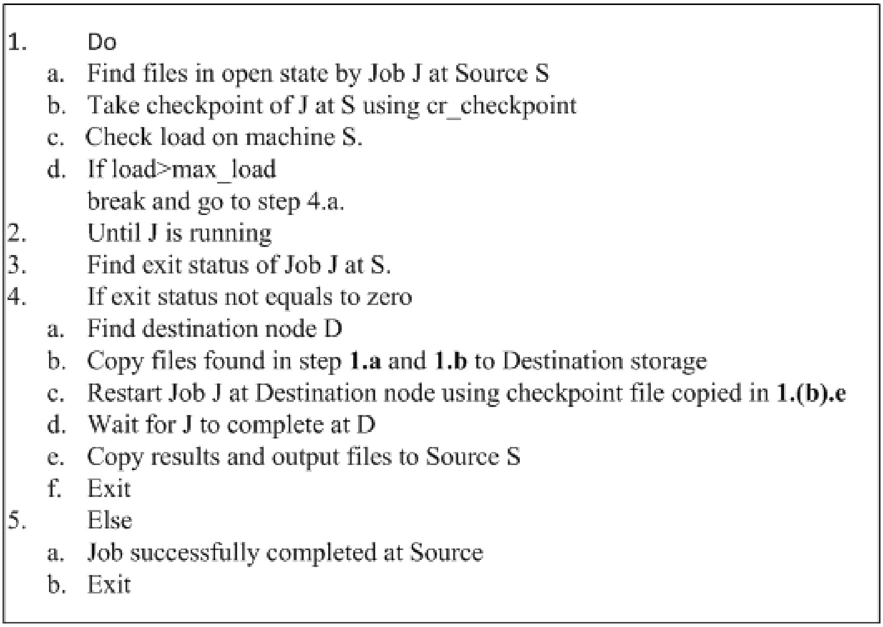


Fig. 2. Flow chart showing load balancing with fault tolerance

Figure 2 is process flow diagram of the proposed approach for balancing load in CSCW clusters. It describes the migration of a process upon abnormal termination, which may be due to resource requirement of the process or overloading of the execution node. In this approach the process is periodically checkpointed using BLCR checkpoint and restart tool. The load on the node is calculated periodically. The process is migrated to a less loaded remote machine using the latest checkpoint created if it terminates abnormally or the load on that machine goes over the calculated load threshold. The required files are copied to remote machine using a secure tunnel created using Ssh (Secure Shell). The process is restarted at destination node with checkpoint files copied previously. The flow chart show in figure 2 describes 4 services that needs to be deployed on every node in the cluster to implement the propose algorithm. Figure 3 shows the Algorithm used for load balancing in CSCW cluster.



```
1.      Do
    a.  Find files in open state by Job J at Source S
    b.  Take checkpoint of J at S using cr_checkpoint
    c.  Check load on machine S.
    d.  If load>max_load
        break and go to step 4.a.
2.      Until J is running
3.      Find exit status of Job J at S.
4.      If exit status not equals to zero
    a.  Find destination node D
    b.  Copy files found in step 1.a and 1.b to Destination storage
    c.  Restart Job J at Destination node using checkpoint file copied in 1.(b).e
    d.  Wait for J to complete at D
    e.  Copy results and output files to Source S
    f.  Exit
5.      Else
    a.  Job successfully completed at Source
    b.  Exit
```

Fig. 3. Algorithm of *Migrate_Job_terminate* (Job=J, Source=S, Destination=D, Pool=N)

4 Performance Evaluation

4.1 Experimental Setup

To evaluate the performance of proposed model, Three Personal Computers (Dual core, 1GB RAM, 100 Mbps Ethernet) in our lab at Computer Science and Engineering Department, MNNIT, Allahabad are configured. All machines has Linux version (ubuntu 10.04) of operating system and BLCR configured on every machine. The nodes communicate through secure shell protocol. While testing the performance the behavior for different number and variety of process is observed. The following scripts/processes were installed on every node in the cluster to automate the process of migration and load balancing.

- *Blcr_run.sh* script starts the execution of the job (process). It starts the execution of the process using *cr_run* utility provided by BLCR checkpointing and restart tool. It starts *p_chkpt.sh* for taking periodic checkpoints of the previously started process and waits for the process to terminate. When process terminates it checks the exit status of the terminated process and if found unsuccessful termination, then the *migrate.sh* will be executed.
- *P_chkpt.sh*, child process created by *blcr_run.sh* takes periodic checkpoints of the running process. It takes checkpoints periodically at particular intervals, which can be specified by the user using *cr_checkpoint* utility of BLCR tool. It also finds the list of files opened by the running process during

execution. It stores and updates the list periodically in a file name *<process name>.list.*

- *Migrate.sh* process is started by *blcr_run.sh* script if the process started terminates unsuccessfully. It starts with finding a destination node in the cluster pool having the lowest load among all nodes present in the cluster. The machine having minimum load is found by *load_check.sh*. The nodes in the cluster communicate using the Ssh (secure shell protocol). It finds the suitable node for the unsuccessfully terminated process and transfers the checkpoint file and files opened by the process before terminating abnormally by reading the file *<process name>.list* file created by *p_chkpt.sh*. It resumes the process execution by restarting the process using the *cr_restart* utility of BLCR which takes the checkpoint file as an argument. When the process successfully terminates at remote node the migrate.sh script transfers the output files to source node and removes the temporary files from destination node.
- Load_check.sh:- returns the node having minimum CPU load within last minute in the cluster pool and load of the host machine. Information about nodes available in the pool is present in a file named nodes, which is updated periodically.

4.2 Results

The threshold value of CPU load is calculated on the fly. For evaluating the results, 6 CPU intensive jobs started on each machine in the cluster using the proposed approach as well as Mosix (Multicomputer Operating System for UNIX). Average CPU load is the percentage utilization of CPU during a respective amount of time. All UNIX system generates a metric of three load averages for 1, 5 and 15 minutes upon querying current results by running uptime command. For example:

Table 1. CPU load on different nodes of cluster

	Proposed Approach	Default Mosix Behavior
Node 1	4.1	3.3
Node 2	3.5	6.13
Node 3	4.3	2.5

A CPU load of 1.73 means CPU is overloaded by 73% (1 CPU with 1.73 ready queue process so that 0.73 process has to wait). During performance evaluation load from other process/demons is kept minimal on all machines. During Idle conditions (No cluster or other Job running or other user process are running) average CPU load varies from 0%-15%. Load is observed by running 5-6 process on each node in Mosix and proposed model. Load threshold is kept at 4.5 for overloading condition i.e. only three waiting process are allowed at a node. The behavior of the node in terms of CPU loads is observed. The Comparison of both the approaches is shown in Table 1 and Figure 4.

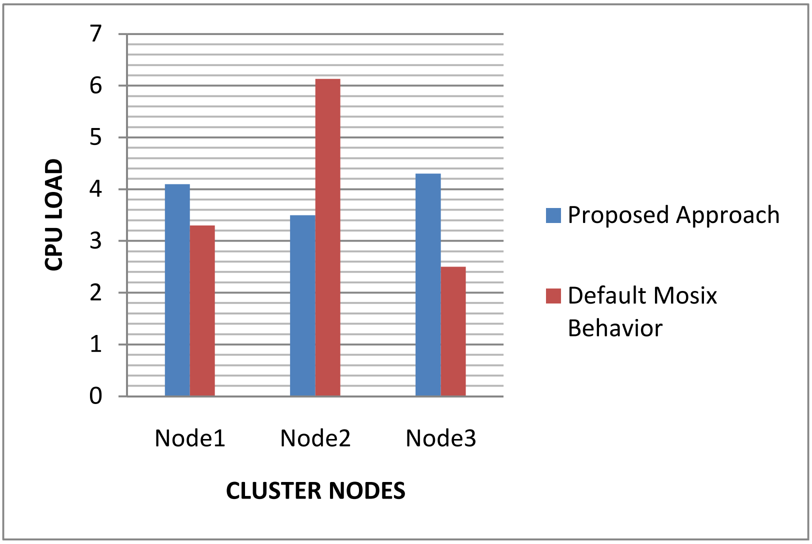


Fig. 4. Comparison of CPU loads on nodes of cluster

The Data in table 1 and figure 4 is reflection of the state of the cluster nodes at a particular instant of time during the execution of process in both the approaches. In case of MOSIX, even if it does load balancing, the difference between Node1 and Node2 load is high. MOSIX only considers peer nodes load for load balancing. Proposed approach results show better performance in terms of load balancing. The proposed solution improves the performance by 30%-40%, avoids overloading of nodes and improves their availability.

5 Conclusion

The proposed solution for load balancing with fault tolerance of Computer Supported Cooperative Working (CSCW) has been described and evaluated. The proposed model provides new services that can be deployed on any CSCW based cluster, extending existing services. Recent researches in checkpoint and restart has made this more attractive and feasible option. Discussed approach indicates that distributed fault tolerant load-balancing is not only feasible, but also produces reasonable results. Proposed model discuss two approaches, one proposes only load balancing considering load factors on every node in cluster. Second approach provides fault tolerance for compute intensive jobs, at the same time monitors the load in cluster nodes, and migrate the job in case of overloading. Overall performance evaluation shows that, performance of nodes in terms of availability and load in CSCW cluster is improved by 30%-40%.

References

1. Selikhov, A., Germain, C.: A Channel Memory based fault tolerance for MPI applications. Future Generation Computer Systems 21(5), 709–715 (2005)
2. Al-Saqabi, K.H., Saleh, K.A.: An efficient process migration algorithm for homogeneous clusters. Information and Software Technology 38(9), 569–580 (1996)
3. Hursey, J., Graham, R.L.: Analyzing fault aware collective performance in a process fault tolerant MPI. Parallel Computing 38(1-2), 15–25 (2012)
4. Chtepen, M., Claeys, F.H.A., Dhoedt, B., De Turck, F., Demeester, P., Vanrolleghem, P.A.: Adaptive Task Checkpointing and Replication: Toward Efficient Fault-Tolerant Grids. IEEE Transactions on Parallel and Distributed Systems 20(2), 180–190 (2009)
5. Lopriore, L.: Object and process migration in a single-address-space distributed system. Microprocessors and Microsystems 23(10), 587–595 (2000)
6. Payli, R.U., et al.: DLB—a dynamic load balancing tool for grid computing. Scientific International Journal for Parallel and Distributed Computing 07(02) (2004)
7. Cao, J., et al.: Grid load balancing using intelligent agents. Future Generation Computer Systems 21(1), 135–149 (2005)
8. Yagoubi, Slimani, Y.: Task load balancing for grid computing. Journal of Computer Science 3(3), 186–194 (2007)
9. Nehra, N., Patel, R.B., Bhatt, V.K.: A framework for distributed dynamic load balancing in heterogeneous cluster. Journal of Computer Science (2007)
10. Hargrove, P.H., Duell, J.C.: Berkeley lab checkpoint/restart (BLCR) for Linux clusters, https://ftg.lbl.gov/assets/projects/CheckpointRestart/Pubs/LBNL-60520.pdf
11. Rodríguez, G., Pardo, X.C., Martín, M.J., González, P.: Performance evaluation of an application-level checkpointing solution on grids. Future Generation Computer Systems 26, 1012–1023 (2010), doi:10.1016/j.future.2010.04.016

Adaptive Region Based Huffman Compression Technique with Selective Code Interchanging

Utpal Nandi[1] and Jyotsna Kumar Mandal[2]

[1] Dept. of Computer Sc. & Engg., Academy of Technology,
Hooghly-712121, West Bengal, India
[2] Dept. of Computer Sc. & Engg., University of Kalyani
Nadia –741235, West Bengal, India
{nandi.3utpal,jkm.cse}@gmail.com

Abstract. Adaptive version of loss-less Region Based Huffman compression techniques are proposed where a proposed region formation algorithm is used to divide the input file into a number of regions that adapts region size depending on the ASCII value difference of symbols. Huffman codes are obtained for entire file after formation of regions. Code interchanging between the maximum frequency element of a region and maximum frequency element of entire file is done before symbols of that region are compressed. Another variation of the technique where region wise interchanging of code is done based on an additional condition. Comparisons are made among these two compression techniques with Region Based Huffman compression technique, Size Adaptive Region Based Huffman compression technique and classical Huffman technique. The proposed techniques offer better results for most of the files.

Keywords: Data compression, Huffman tree, Frequency Table (FT), Symbol Code Table (SCT), compression ratio, Region Based Huffman (RBH), Size Adaptive Region Based Huffman (SARBH).

1 Introduction

To reduce the time of data transmission over network and the storage space requirement, the data compression techniques are used. Among the two major class of data compression techniques, the loss-less techniques generate exact duplicate of the original data after compress/expand cycle. But, the lossy techniques concede a certain loss of accuracy. One of the well established loss-less technique is Huffman Coding [4,6] which is based on the frequency of elements of entire file . If an element has maximum frequency, it gets shortest code. But if we divide a file into a number of regions, it is obvious that in each region the maximum frequency element may not the maximum frequency element of entire file and has large code length. If the large codes produced by Huffman coding are used for the elements which have maximum frequency for each region, the size of compressed file increases. In light of this Region Based Huffman (RBH) [2] coding has been introduced. The RBH coding technique divides the total input file/stream into a number of regions N. The maximum frequency elements for each region are calculated. Huffman codes are obtained based on frequency of elements for entire file/stream. Now for first region, if the code length of maximum frequency element

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 739–748.
springerlink.com

of that region is larger than the code length of maximum frequency element of entire file/stream, the code between maximum frequency element of that region and maximum frequency element of entire file/stream is interchanged. This interchanged information is attached with the compressed file/stream. The elements of that region are compressed with the changed codes and interchanged codes are reset. Otherwise, same symbol code table is used. Similarly, all other regions are compressed repeatedly. The main problem of RBH coding is that the performance depends on the number of regions of the file. Modified Region Based Huffman (MRBH) [2] coding also suffers from the same problem if the optimum value of number of region does not lie in the specified range. Fixed size regions are not able to adapt its size based on symbols that offers better compression. Section 2.0 discusses the weakness of Region Based Huffman technique. The Size Adaptive Region Based Huffman Compression (SARBH) [1] can be used to overcome this limitation. But, the performance of the same is not so well. To enhance the performance of SARBH, two variants of this technique are proposed where a proposed region formation algorithm is used and termed as Adaptive Region Formation (ARF) Algorithm as discussed in section 3.0. The proposed compression techniques based on ARF are discussed in section 4.0. Results have been given in section 5.0 and conclusions are drawn in section 6.0.

2 Limitation of RBH Coding

The main problem of Region Based Huffman (RBH) coding is that the performance depends on the number of regions of the file and therefore also on the size of region of the file. Compression ratios of same file with different region size are not same. It is very difficult to determine the optimum region size that offers maximum compression of a file as it depends on the symbols of the file. As different region contains different frequency of symbols, the compression ratios of different region are also not same. For example let us consider a file/stream containing the message as–CACBABCBCCABACBABABACBBADBDBEB (say MSG). Huffman tree is build based on the frequency of symbols from Table 1 and given in Fig. 1. Code of each symbols are obtained from the tree and placed in same Table 1. Now input message stream MSG is grouped into regions of size 10 as given in Fig. 2. In region 1, the maximum frequency symbol is C. The same under entire message stream is B. Therefore, code of B and C are interchanged and the symbols of region 1 are compressed. Symbol codes are reset as obtained from Huffman tree. In region 2, code of B and A are interchanged and the symbols are compressed. Symbol codes are reset as obtained from Huffman tree. In Region3, the maximum frequency symbol is B and is same with the maximum frequency symbol of entire message stream. Therefore, no interchange of symbol codes is occurred and the symbols of region3 are compressed using the same symbol codes. The region wise compressed message for entire message stream given in Table 2. The effective compression ratio may be obtained as follows: Original message size = 30x8 bits = 240 bits, Frequency Table size =7x8 bits = 56 bits, Code interchange information size = (3+3 + 2) bits = 8 bits, Value of number of region takes 5 bits, Only Compressed message size= 56 bits, Compressed message size including Frequency Table , code interchange information and value of number of region = (56+56+8+5) bits = 125 bits, Compression ratio = {(240–125)/240}X100% = 47.9 %. Now the same message is compressed again considering region size = 6. Then the input message stream is grouped into 5 regions as given in Fig.3. The 5 regions are compressed similarly as before. The compressed message stream will be 0100111 1011101110010111011110010001011100101101011010000 and the compression ratio is 45%.

Table 1. Frequency and code of Symbols

Symbol	Frequency	Code
A	8	10
B	12	0
C	7	111
D	2	1101
E	1	1100

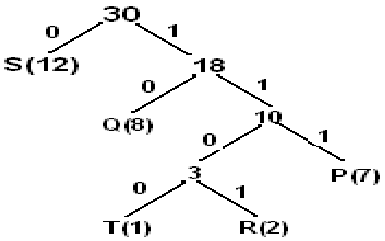


Fig. 1. Huffman tree based on frequency of Table 1

CACBABCBCC	**ABACBABABA**	**CBBADBDBEB**
← Region 1 →	← Region 2 →	← Region 3 →

Fig. 2. Region wise symbols

Table 2. Compressed symbols of region 1, 2 and 3

Region 1		Region 2		Region 3	
Symbol	Code	Symbol	Code	Symbol	Code
C	0	A	0	C	111
A	10	B	10	B	0
C	0	A	0	B	0
B	111	C	111	A	10
A	10	B	10	D	1101
B	111	A	0	B	0
C	0	B	10	D	1101
B	111	A	0	B	0
C	0	B	10	E	1100
C	0	A	0	B	0

CACBAB	**CBCCAB**	**ACBABA**	**BACBBA**	**DBDBEB**
← Region 1 →	← Region 2 →	← Region 3 →	← Region 4 →	← Region5 →

Fig. 3. Region wise symbols

For the same file/stream (MSG), compression ratio is 47.9% for region size 10 and compression ratio is 45% for region size 6. Therefore, the proper region size (or number of region) must be chosen for better compression of file/stream. Modified Region Based Huffman (MRBH) coding also suffers from the same problem if the optimum value of number of region does not lie in the specified range. Fixed size regions are not able to adapt its size based on symbols that offers better compression. The Size Adaptive Region Based Huffman Compression overcome this limitation. But, the performance of the same is not so well. To enhance the performance of SARBH, two compression techniques are proposed where a proposed region formation algorithm is used that has the ability to adapt its region size based on symbols and termed as Adaptive Region Formation (ARF) Algorithm and discussed in section 3.0.

3 The Proposed Region Formation Algorithm

It is found that ASCII value differences among group of adjacent characters are not so high most of the time. The proposed Adaptive Region Formation (ARF) algorithm uses this concept. The aim is to group sequence of characters into regions such that the differences among the ASCII values of characters in a region do not exceed a specified value (r). Therefore, after grouping into regions, the information of each region can be preserved by storing the number of symbols, minimum ASCII value and the differences among other ASCII values of symbols in the region with the minimum ASCII value. After that each region contains ASCII values not exceeding the specified value except first two (the number of symbols and the minimum ASCII value). The proposed Adaptive Region Formation (ARF) algorithm is given in Fig. 4.

4 The Proposed Compression Techniques

Two compression techniques are proposed in this section to enhance the performance of SARBH, all of which use ARF algorithm to group file/message stream into variable size regions as described below.

4.1 Size Adaptive Region Based Huffman Compression with Code Interchanging (S*ARBHI)* Technique

The schematic diagram SARBHI coding is shown in Fig. 5. Initially, variable length regions (R1 , R2 , R3 . . Rn) of input file / Stream are formed using ARF algorithm with a specified value(r). The frequencies of all the numbers of all regions are obtained whose value lie in the range 0 to r-1. Huffman Codes of the same are also constructed to obtain the code of each numbers. Code between maximum frequency element of entire file/stream and the same of that region are interchanged. During compression, for each region first two numbers (number of element and minimum value symbol) are kept unchanged and all other numbers (whose values lie in the range 0 to r-1) are coded by corresponding Huffman code. But, there is a limitation of the technique. The code interchanging may increase the compressed message size sometime. To overcome such limitation, another technique is proposed in the following section 4.2.

4.2 Size Adaptive Region Based Huffman Compression with Selective Code Interchanging (S*ARBHS)* Technique

Similar with S*ARBHI* technique, codes between maximum frequency element of a region and the same of entire file/stream are interchanged. But, one additional condition is checked before interchange of code. Code interchange is not allowed if the overhead (code interchange information size) is more than the benefit (reduction of size for code interchange).

```
Input: Input file, specified value(r).
Output: Collection of variable length regions ($R_1,R_2,R_3,\ldots R_n$) in a array B.

1.  Initially, set no_of_region=0, i=1.
2.  Read one symbol from input file and store it in a one dimensional
    array A[i] and increment i by one.
3.  Find minimum and maximum ASCII value of symbols of already
    read symbols ($x_1,x_2,x_3,\ldots\ldots x_{m+1}$) in array A as min and max
    respectively.
4.  Find diff = max - min.
5.  If diff > r , then
5.1 A region is formed with symbols m, min, ($x_1$-min), ($x_2$-min), ($x_3$-
    min),…. , ($x_m$-min) where m is the number of elements of the
    region excluding first two number m and min.
5.2 Increment no_of_region by one and save the value of m and min
    in two dimensional array B as B[no_of_region] [1]=m and
    B[no_of_region][2]=min.
5.3 Set B[no_of_region] [i]=($x_{i-2}$-min) , for i=3 to m+2.
5.4 Release all symbols from the array A except ($x_{m+1}$-m) from
    A[m+3] and shift $x_{m+1}$ to the first position of the array as A[1]=
    $x_{m+1}$ and set i=2.
    End If.
6.  Repeat step 2 to step 5 until end of file.
7.  Stop.
```

Fig. 4. Adaptive Region Formation (ARF) algorithm

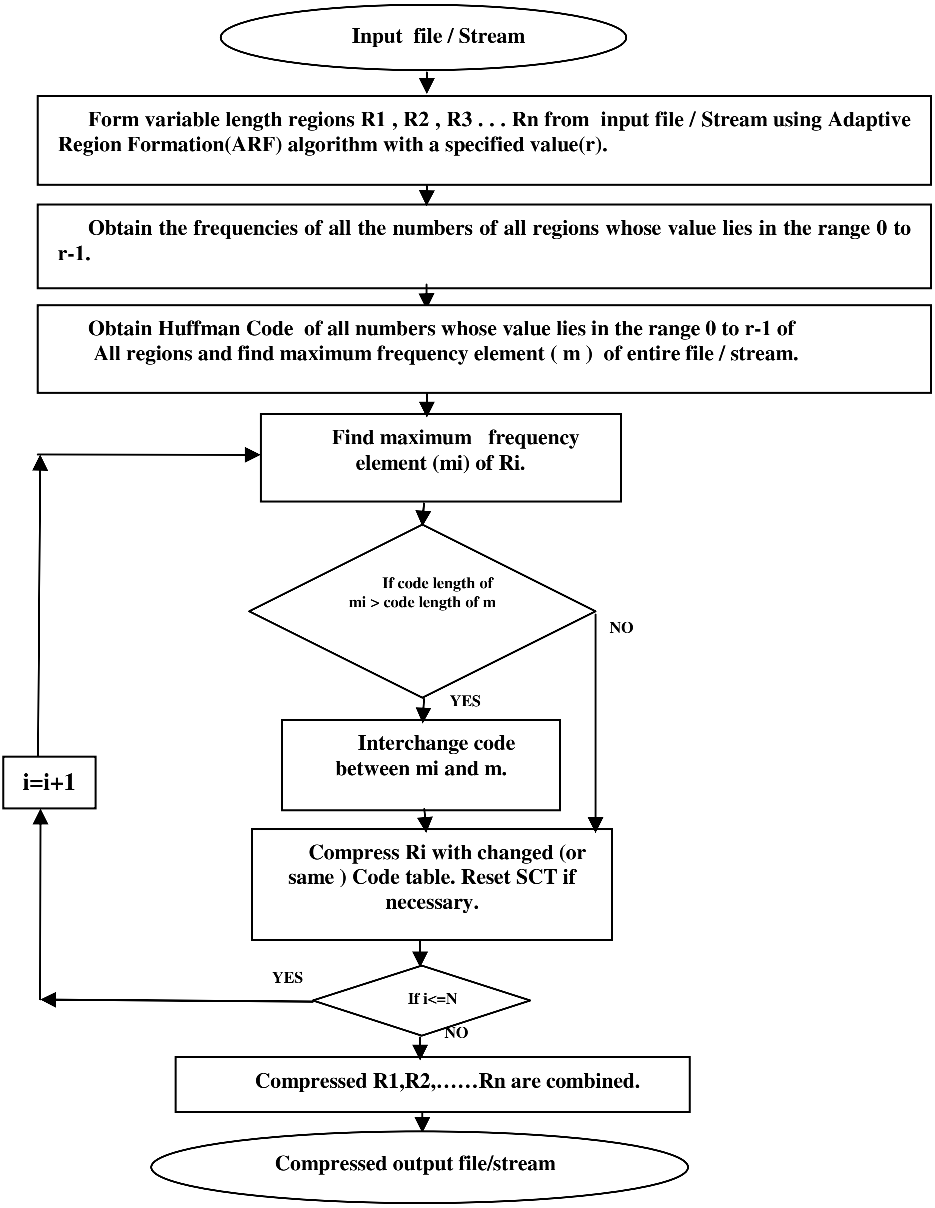

Input file / Stream
Form variable length regions R1 , R2 , R3 . . . Rn from input file / Stream using Adaptive Region Formation(ARF) algorithm with a specified value(r).
Obtain the frequencies of all the numbers of all regions whose value lies in the range 0 to r-1.
Obtain Huffman Code of all numbers whose value lies in the range 0 to r-1 of All regions and find maximum frequency element (m) of entire file / stream.
Find maximum frequency element (mi) of Ri.
If code length of mi > code length of m
NO
YES
Interchange code between mi and m.
i=i+1
Compress Ri with changed (or same) Code table. Reset SCT if necessary.
YES
If i<=N
NO
Compressed R1,R2,……Rn are combined.
Compressed output file/stream

Fig. 5. Schematic diagram of SARBHI coding

4.3 Example

Let us consider a same file/stream – **ABAABDADAAWXXZXWXYZXXYPQP SQ SPR** (say MSG1). ARF algorithm is used with specified value(r) =16 to form regions of MSG1. In ARF algorithm, MSG1 is grouped into a number of regions in such a way that each region does not have two symbols with ASCII value difference gather than or equal to 16 as shown in Fig. 6. Information of each regions are kept by storing the number of symbol of each region, minimum ASCII value of all the symbols and ASCII value difference of all the symbols with minimum ASCII value of symbol in the corresponding region as shown in Fig. 7. After formation of regions, compression techniques are applied.

ABAABDADAA	**WXXZXWXYZXXY**	**PQPSQSPR**
← Region 1 →	← Region 2 →	← Region 3 →

Fig. 6. Symbols of variable length regions

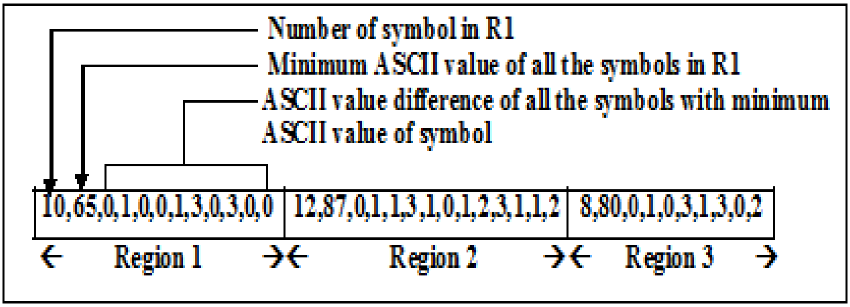


Fig. 7. Variable length regions of MSG1

<i> *SARBH* Technique: Frequencies of all the numbers in the range 0 to 15 are found as given in Table 3 and Huffman tree based on the frequency of numbers is constructed as shown in Fig. 8 and codes of each number are obtained and placed in same table 3. Maximum frequency number of the entire file /stream (m) is 0. Same of R1 (m1) is 0. As m=m1, no interchange of code is occurred for R1 before compression. Maximum frequency number of R 2 (m2) is 1. As code length of m2 is larger than code length of m**,** interchange of code between m and m2 is occurred and the numbers of R2 are compressed. Maximum frequency number of R 3 (m3) is 0. As m=m3, no interchange of code is occurred for R3 during compression. Compressed numbers of R1, R2 and R3 are given below in Table 4. Compressed size of R1, R2 and R3 (excluding first two numbers of each region) are 16, 22, 16 bits respectively.

Therefore, the compressed message will be-'10','65',0,10,0,0,10,111,0,111, 0,0, '12', '87',10,0,0, 111,0,10,0,110,111,0,0,110,'8','80',0,10,0,111,10,111,0,110. Size of all regions (excluding first two numbers of each region) is 16+22+16 bits =54 bits. Size of interchange information is 3+2 bits =5 bits. Frequency Table size =6x8 bits =48 bits, size of first two numbers of three region=3x2x8 bits =48 bits. Total Compressed message size=(54+48+48+5) bits=155 bits, Compression ratio={(240–155)/240}X 100 % = 35.41 %.

Table 3. Frequency and code of symbols

Number	Frequency	Code
0	11	0
1	10	10
2	3	110
3	6	111

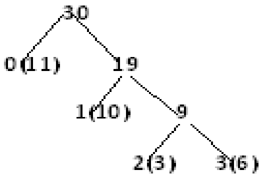


Fig. 8. Huffman tree based on Table1

Table 4. Compressed numbers of region 1, 2 and 3 except first two numbers of each region

Region 1		Region 2		Region 3	
Number	Code	Number	Code	Number	Code
0	0	0	10	0	0
1	10	1	0	1	10
0	0	1	0	0	0
0	0	3	111	3	111
1	10	1	0	1	10
3	111	0	10	3	111
0	0	1	0	0	0
3	111	2	110	2	110
0	0	3	111	-	-
0	0	1	0	-	-
-	-	1	0	-	-
-	-	2	110	-	-

<ii> ***SARBHS*** **Technique:** Frequencies of all the numbers , Huffman tree and codes of each number are obtained like previous technique. Maximum frequency number of the entire file /stream (m) is 0. Maximum frequency number of R 1 (m1) is 0. As m=m1, no interchange of code is occurred for R1. Maximum frequency number of R 2 (m2) is 1. As code length of m2 is gather than code length of m and overhead (interchange information size = 2bits) is less than benefit (reduction of number of bits=4bits), interchange of code between m and m2 is occurred. The resultant symbol code table is shown in Table 14. Maximum frequency number of R3 (m3) is 0. As m=m3, no interchange of code is occurred for R3. Compressed size of R1, R2 and R3 (excluding first two numbers) are 16, 22 and 16 bits respectively as shown in Table 11.Therefore, the compressed file/message stream will be- ‘10’,’65’,0,10,0,0,10, 111, 0,111,0,0,‘12’,’87’,10,0,0,111,0,10,0,110,111,0,0,110,‘8’,’80’,0,10,0,111,10,111,0, 110. Similarly, total compressed message size and the compression ratio are calculated as 155 bits and 35.41 % respectively.

5 Results

Comparison of compression ratios of Huffman technique, RBH coding, MRBH coding and SARBH coding with proposed SARBHI , SARBHS have been made using different type of files as shown in Table 5. Here the specified value(r) of ARF algorithm is taken as 128. The graphical representation of the same is shown in Fig. 9. The results show that both the proposed techniques offer significant compression ratio with respect to its counterpart for almost all type of files.

Table 5. Comparison of compression ratios in different techniques

File Name	Huffman	RBH With N=5	RBH With N=10	RBH With N=20	RBH With N=30	MRBH With Range 10 - 25	SARBH	SARBHI	SARBHS
Sample.txt	26.84	27.13	28.12	28.61	28.47	28.71	29.01	29.37	29.37
Task.txt	30.31	30.29	30.44	31.13	31.64	31.41	31.34	31.12	31.21
DDA.exe	18.41	18.57	18.92	19.10	19.25	19.37	19.59	19.65	19.71
TRY1.exe	21.09	21.38	21.67	21.81	21.92	21.86	21.84	21.98	21.92
ChipsetCHS.dll	51.25	51.22	51.20	51.25	51.29	51.39	50.09	51.17	51.25
ChipsetARA.dll	24.42	24.68	24.81	24.99	25.30	25.25	25.21	25.37	25.32
Dolly.doc	35.35	35.34	35.37	35.36	35.37	35.38	35.34	35.36	35.36
Resume.doc	35.16	35.17	35.19	35.20	35.20	35.20	35.31	35.37	35.39
Complex.java	35.48	35.48	35.50	35.49	35.52	35.51	35.71	35.79	35.81
Inharit.java	35.24	35.24	35.24	35.25	35.26	35.27	35.21	35.24	35.24

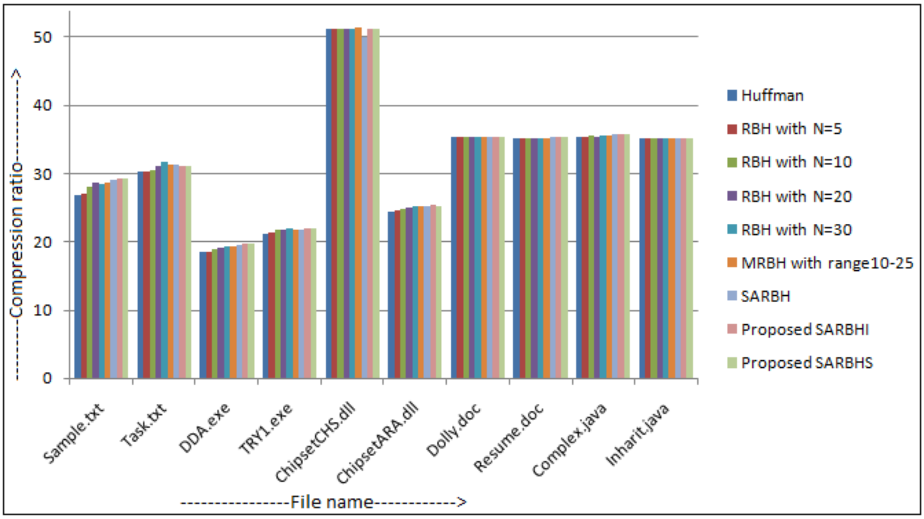


Fig. 9. The graphical representation of Comparison of compression ratios in different techniques

6 Conclusion

The proposed SARBHI and SARBHS coding techniques enhance the performance of SARBH, RBH and MRBH coding by introducing the concept of ARF algorithm which adapts region size based on symbol's ASCII value difference and region wise interchanging codes or selective interchanging of codes. The performances of proposed techniques are better than Huffman coding most of the time. For some files, the proposed techniques offer better results than RBH and MRBH coding also. Among the proposed two techniques, SARBHS is more effective for all most all types of files. The presented scheme has also a better scope of modification. Region wise multiple code interchanging can be done instead of region wise single interchanging of code. The techniques can also be applied for image compression.

Acknowledgement. The authors extend sincere thanks to the department of Computer Science and Engineering and IIPC Cell, University of Kalyani, Nadia, West Bengal, India and Academy Of Technology, Hooghly, West Bengal, India for using the infrastructure facilities for developing the technique.

References

1. Nandi, U., Mandal, J.K.: Size Adaptive Region based Huffman Compression. In: National Symposium on Emerging Trends in Computer Science (ETCS), Barrackpore, India (2012)
2. Nandi, U., Mandal, J.K.: Region based Huffman(RB H) Compression Technique with Code Interchange. Malayasian Journal of Computer Science(MJCS), Malayasia 23(2), 111–120 (2010)
3. Mandal, J.K., Kumar, A.: A Compression Technique Based on Optimality of Huffman Tree (OHT). In: 12th International Conference of IEEE on Advanced Computing and Communications, Ahmedabad, India, pp. 589–595 (2004)
4. Ziv, J., Lempel, A.: Compression of individual sequences via variable-rate coding. IEEE Transactions on Information Theory 24(5), 530–536 (1978)
5. Mandal, J.K., Gangopadhayay, R.: Implementation of Two Data Compression Schemes. In: First International Workshop on Telematics, pp. 154–162. NERIST, India (1995)
6. Reglebati, H.K.: An Overview of Data Compression Techniques. IEEE Computer, 71–75 (1981)
7. Ziv, J., Lempel, A.: Compression of individual sequences via variable-rate coding. IEEE Transactions on Information Theory 24(5), 530–536 (1978)
8. Welch, T.: A Technique for High-Performance Data Compression. IEEE Computer 17(6), 8–19 (1984)
9. Witten, I.H., Neal, R.M., Cleary, J.G.: Arithmetic Coding for Data Compression. Communications of the ACM 30(6), 520–540 (1987)
10. Ziv, J., Lempel, A.: A universal algorithm for sequential data compression. IEEE Transactions on Information Theory 23(3), 337–343 (1977)
11. Nelson, M.: The Data Compression Book, 2nd edn. BPB Publications, India (2008)
12. Kanitkar, Y.: C project, 2nd edn. BPB Publications, India (2002)

Human Tracking in Video Surveillance

Helly Patel[1] and Mahesh P. Wankhade[2]

[1] Department of Computer Engineering,
Zeal Education Society's Dnyanganga College of Engineering and Research, Pune,
Maharastra, India
[2] Department of Computer Engineering, Sinhgad College of Engineering, Pune,
Maharashtra, India
{hellypatel_786,mwankhade}@yahoo.com

Abstract. Video Surveillance has become area of research and development due to increase in terrorist activities, thefts and other activities that can cause harm to human property and lives. Due to this researchers have been trying to add more and more intelligence in the video monitoring systems, so that it could automatically detect the malicious activities in the area under surveillance and raise alarm. The goal of visual surveillance is not only to put cameras in place of human eyes, but also to accomplish the entire surveillance task as automatically as possible. We can say that video surveillance is nothing but taking the video, identifying unwanted entities, tracking their actions, understanding their actions and raising an alarm. In this paper, we will be study the phases of the video surveillance system. We will study how a video is divided into frames, those frames are then sent to detect any change detection (human detection), further this information about detected humans in one frame is correlated with the information of the other frame and thus the detected human is tracked in the subsequent frame and this is called human tracking. We will see most salient region method for tracking and in this paper we propose a method of handling occlusion using velocity and direction information.

1 Introduction

Video surveillance has been with us since a long time. In the traditional surveillance system, the camera captured video would be directly displayed on a monitor in a control room where human resources will be continuously monitoring the video for any abnormal activities. But, due to increase in thefts, terrorist activities and other criminal activities the demand for the sensitivity and accuracy of the surveillance system has increased and any small human error or delay in monitoring abnormal activity may lead to drastic damage to society. We all know that humans have their limitations. Due to this, an effort was made to automate the video surveillance system by giving extra information about the objects captured along with the captured video, by

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 749–756.
springerlink.com

triggering alarm in many cases or by generating reports. This extra information will help in accurate monitoring of the sensitive areas with less human resources occupied in the whole process.

Automatic video surveillance is becoming increasingly important in many applications, including traffic control, urban surveillance, home security and healthcare. In this paper we will study the process of automatic video surveillance. Also, we will see a proposed method for human tracking.

There are basically three conventional approaches for moving object detection: temporal differencing, optical flow and background estimation methods. Temporal differencing [2] is very adaptive to dynamic environments, but generally does a poor job of extracting all relevant feature pixels [5]. Optical flow [2, 3, 10] can be used to detect independently moving targets in the presence of camera motion, however most optical flow computation methods are very complex and are inapplicable to real-time algorithms without specialized hardware. Background subtraction is a particularly popular method for human detection especially under those situations with a relatively static background. It attempts to detect moving regions in an image by differencing between current image and a reference background image in a pixel-by-pixel fashion. However, it is extremely sensitive to changes of dynamic scenes due to lighting and extraneous events.

Human detection is one of the critical phases of video surveillance system as it is not only responsible for the extraction of moving objects but all the remaining phases process based on the output of this phase. A large number of people detection and tracking algorithms rely on the process of background subtracting, a technique which detects changes from a model of the background scene. Let study different techniques that are used to detect human and let us see why its output is importance for the next phase (human tracking).

2 Video Surveillance System

Video surveillance process that takes video as an input, processes the video and performs actions or outputs the captured video along with analyzed information. The process of video surveillance consists of many phases as shown in the figure 1.

In section 3 we will discuss about the image generation phase. Section 4 discusses about human detection. Following which sections 5, 6 and 7 discuss about Noise removal, human tracking along with occlusion handling and the activity analysis and triggering alarm respectively.

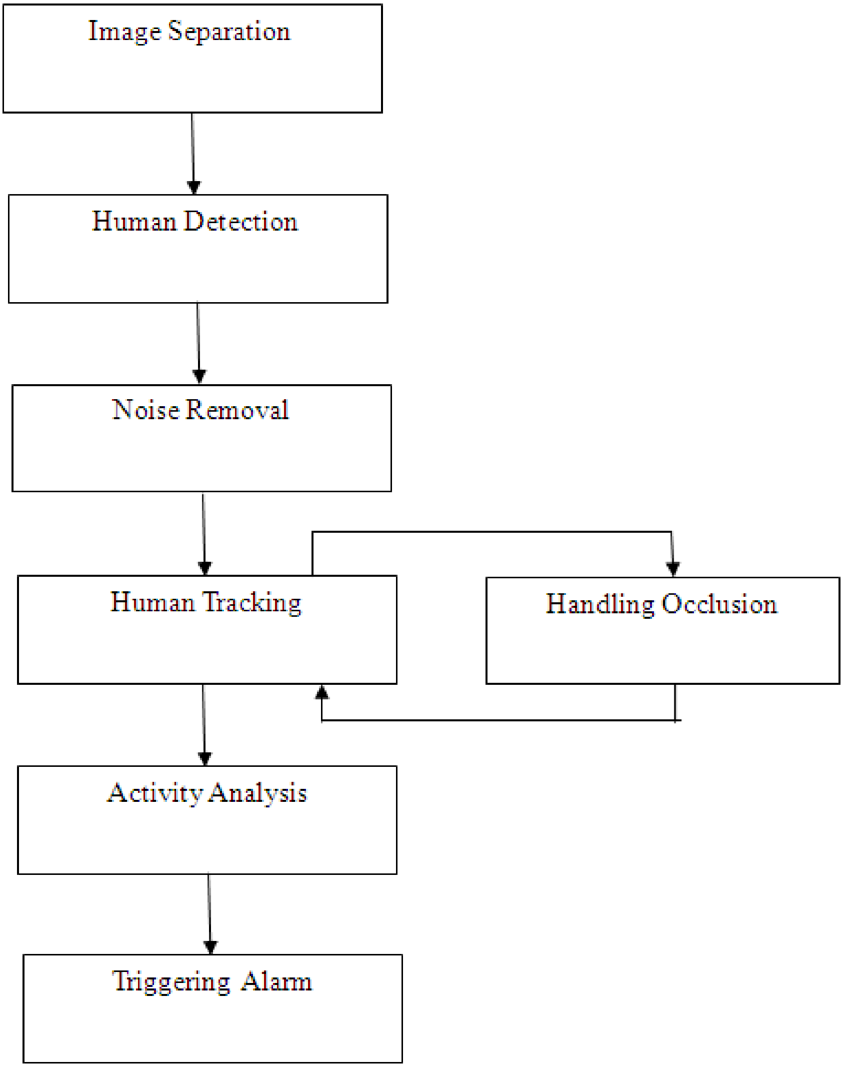


Fig. 1. Phases of Video Surveillance System

3 Image Separation

When we talk about processing a video, it is actually not a video that we are dealing with but the video must be divided into sequence of images which are further processed. The speed at which a video must be divided into images depends on the implementation of individuals. From [1] we can say that, mostly 20-30 images (or frames) are taken per second which are sent to the next phases for further processing.

4 Human Detection

In this phase, an attempt is made to detect human entities in the area under surveillance. There are mainly three ways in which this task can be accomplished 1) Simple Background Subtraction 2) Optical Flow and 3) Temporal Difference.

4.1 Simple Background Subtraction

The Background subtraction approach is mostly used when the background is static [4]. The principle of this method is to use a model of the background and compare the current image with a reference. In this way the foreground objects present in the scene are detected. It attempts to detect moving regions in an image by differencing between current image and a reference background image in a pixel-by-pixel fashion as shown by equation below:

$$| P_c - P_{bk} | > T \tag{4.1}$$

where, P_c – Current image pixel,

P_{bk} – Background image pixel, T – Threshold value.

This method is very useful for static background. Here the threshold is fixed such that the foreground pixels are extracted from the background image. When the pixel difference is above the threshold value, it is considered as foreground image.

4.2 Optical Flow

Optical flow can be used to segment a moving object from its background provided the velocity of the object is distinguishable from that of the background, and has expected characteristics. Optical flow is the amount of image movement within a given time period [2, 3, 10].

Let us study an optical flow method entitled Lucas-Kanade Method [2, 10]. This method calculated the motion between two image frames which are taken at time *t* and *t* +δ*t* for every pixel position. As a pixel at location (*x,y,t*) with intensity *I* (*x,y,t*) will have moved by δ*x*, δ*y* and δ*t* between the two frames, the following image constraint equation can be given:

$$I (x, y, t) = I (x + dx , y + dy , t + dt) \tag{4.2}$$

Assuming that the movement is small enough, the image constraint at *I*(*x*, *y*, *t*) with Taylor series can be derived to give:

$$I(x + dx, y + dy, t + dt) = I(x, y, t) + \partial I/\partial x * dx + \partial I/\partial y * dy + \partial I/\partial t * dt \tag{4.3}$$

From (4.2) and (4.3) we get,

$$\partial I/\partial x * dx + \partial I/\partial y * dy + \partial I/\partial t * dt = 0 \tag{4.4}$$

Or

$$\partial I/\partial x * dx/dt + \partial I/\partial y * dy/dt + \partial I/\partial t * dt/dt = 0 \tag{4.5}$$

Equation (4.5) can be further written as:-

$$\partial I/\partial x * Vx + \partial I/\partial y * Vy + \partial I/\partial t = 0 \tag{4.6}$$

Where, *Vx* and *Vy* are the *x* and *y* components of the velocity or optical flow of I(x, y, t) and ∂I/∂t at a given pixel is just how fast the intensity is changing with time.

4.3 Temporal Differencing

Unlike, the background subtraction method where the base image was the background image, here in temporal differencing [5] the reference image is the previous images. Hence the previous frame is subtracted from the current image and the subtraction value must be greater than a threshold value in order to give a difference image.

$$\begin{aligned} I_{current}(x, y) &= 1, \text{diff}_{\ I(x, y)} > \text{threshold} \\ &= 0, \text{diff}_{\ I(x, y)} <= \text{threshold} \end{aligned} \tag{4.7}$$

From the above three human detection methods, the background subtraction is the simplest and mostly used when the background is stationary. In our implementation we will be keeping static background and hence will use the simple background subtraction method of human detection.

5 Noise Removal

The image is expected to contain noises. These noises might be included in the image due to environmental factors (for example, humidity or fog in the area under surveillance), due to illumination changes, during transmission of video from the camera to the processing unit, and many more. Due to noise it is possible that there might be either added or erased portions in the obtained images. Hence, we might get any extra or lesser portion in the human detection phase.

Such noise has to be handled before the detected object is sent for further processing. This can be done by performing morphological operations [6] like opening and closing on the subtracted image. Opening is a combination of erosion and dilation operations with erosion followed by dilation whereas closing is dilation followed by erosion [15].

6 Human Tracking and Handling Occlusion

Human tracking means deriving a correspondence between the object detected in one frame with the object detected in the next frame. If a correspondence is found than we can say that the object found in the previous and the current frame is the same and the objects in both the frames are marked with same colored rectangle. In order to find the correspondence a simple strategy is followed. Few features [11] of the object detected in the previous frame are stored and the features are then matched with the object detected in the current frame. If they match, then the object detected in the current and the previous frame are said to be the same. Features can be color, orientation, speed, posture, speed, intensity or any other information that can be obtained from a pixel. Hence, selection of the features plays an important role in tracking an object.

The tracking methodologies [14] can be mainly divided into 2 types:-

1. Region Based Tracking: Here, the features of the blob, detected in one image frame are matched to the blob detected in the other frame. If there is a match then the detected image is linked with the image in the previous frame.
2. Contour Based Tracking [7]: Here, the energy of the boundary/contour of the blob detected in the previous frame is matched with the energy of the boundary of the blob detected in the current image.

Region tracking is very efficient with stationary background and hence we are taking region tracking. Region tracking stores the features of whole object and matches the features with the features of the object in the next frame. This wastes time, instead most salient region method only matches the most salient region of the previous frame with the most salient region of the current frame. Hence, reducing the amount of time required to match the whole image. The most salient region tracking [8, 12] works as follows:

- The initial features for color, orientation and intensity are fetched from the image [12].
- From the above fetched features, the feature vector is calculated for color, orientation and intensity using center-surround method [13].
- Once the feature map is obtained, the 3 features are weighted in order to find out which feature more uniquely identifies the object. Once weight is obtained, using their weight we get the saliency map for the detected object.
- Finally feature weight vector is calculated for this most salient region. This feature weight vector is matched with the subsequent frame's feature weight vector, if the match is above the threshold value then there is a match otherwise, the search area is doubled and the above procedure is repeated again.

Even after doubling the search area, a match is not found then object is expected to be occluded by some other object or any stationary background object.

As seen in figure1, occlusion handling is not a separate process rather it is a parallel process to human tracking process. When an object is detected its centroid can easily be obtained and when a match is found between the object in the previous and the current frame object, we will get the centroid of the current image. From the centroids of these two images we will obtain following:

- The speed (velocity) at which object is moving using the simple distance upon time formula as shown below
- The direction of the motion by the tanθ formula.

In case, the object is not found even after doubling the search area, the object can be said to be occluded and its next position can be predicated by the previously calculated velocity and direction of motion. In order to calculate the velocity and the direction of motion, minimum 5 frames should be processed.

7 Activity Analysis and Alarm Triggering

Activity analysis is like adding intelligence to the whole process. The information about the detected human is sent to a program which will study its position, pose and motion. After analysis, in case any abnormal pose, motion or position is found then an alarm is triggered.

Abnormal activity can be any action like moving into any highly secure area, moving with speed more than a limit in a secure place, any typical pose that is not normal, and many other actions can trigger the alarm. The list of abnormal activities varies from customer to customer.

Also, alarm triggering [9] may include actually ringing alarm/bell, sending notification to any department through e-mail or SMS, generating a report, etc. Thus, this phase totally deals with how you want to utilize the information that has been obtained from the previous phases and react on or handle abnormal activities.

8 Conclusion

Our proposed method of finding the speed and direction while performing most salient region tracking will help in fast and efficient occlusion handling. Also, as discussed above, in case the most salient region method is not able to find the human in initial search window, we just double the search window and if this fails too then the position of the human is predicted. This reduces the processing time in case the object is hidden as we are not searching whole frame instead just doubling the search area.

9 Future Enhancement

In this paper, we have considered static background; in future it can be enhanced for changing/non-static background. Here, we are making an assumption that the detected entity is a Human. But, we can further enhance it to classifying the detected entities into human and non-human entities.

References

[1] Ekinci, M., Gedikli, E.: Silhouette Based Human Motion Detection and Analysis for Real-Time Automated Video Surveillance. c T"UBITAK 13(2), 199–229 (2005)

[2] Lu, N., Wang, J., Wu, Q.H., Yang, L.: An improved Motion Detection Method for Real-Time Surveillance. IAENG International Journal of Computer Science 35, 1, IJCS_35_1_16

[3] Denman, S.P., Chandran, V., Sridharan, S.: An adaptive optical flow technique for person tracking systems. Pattern Recognition Letters 28(10), 1232–1239 (2007)

[4] Srinivasan, K., Porkumaran, K., Sainarayanan, G.: Intelligent human body tracking, modeling, and activity analysis of video surveillance system: A survey. Journal of Convergence in Engineering, Technology and Science 1, 1–8 (2009)

[5] Zhang, P., Cao, T.-Y., Zhu, T.: A Novel Hybrid Motion Detection Algorithm Based On Dynamic Thresholding Segmentation. In: 2010 12th IEEE International Conference on Communication Technology (ICCT), pp. 853–856 (2010)

[6] Li, L., XU, J.: Moving Human Detection Algorithm based on Gaussian Mixture Model. In: Proceeding of 29th Chinese Conference, pp. 2853–2856 (2010)

[7] Cai, L., He, L., Takayoshi, Y., Xu, Y., Zhao, Y., Yang, X.: Robust Contour Tracking by Combining Region and Boundary Information. IEEE Transactions on Circuit and Systems for Video Technology, 1-10 (2011)

[8] Frintrop, S., Kessel, M.: Most Salient Region Tracking. In: Proceedings of the ICRA 2009, pp. 1869–1874. IEEE (2009)

[9] Ramli, S.B., Ghazali, K.H.B., Ali, M.F.B.M., Hisahuddin, Z.H.B.: Human Motion Detection Framework, pp. 158–161. IEEE (2011)

[10] Yokoyama, M., Poggio, T.: A contour-based moving object detection and tracking, pp. 271–276. IEEE (2005)

[11] Collins, R.T., Liu, Y., Leordeanu, M.: Online Selection of Discriminative Tracking Features. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(10), 1631–1643 (2005)

[12] May, S., Klodt, M., Rome, E., Breithaupt, R.: GPU-accelerated Affordance Cueing based on Visual Attention. In: Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 3385–3390 (October 2007)

[13] Parkhurst, D., Law, K., Niebur, E.: Modeling the role of salience in the allocation of overt visual attention. Vision Research 42, 107–123 (2002)

[14] Yilmaz, A., Javed, O., Shah, M.: Object Tracking: A Survey. ACM Computing Surveys 38(4), Article 13, 1–4 (2006)

[15] Fujiyoshi, H., Lipton, A.J., Kanade, T.: Real-time Human Motion Analysis by Image Skeletonization. IEICE Trans. Inf. & Syst. E87-D(1), 113–120 (2004)

Performance Evaluation of V2VCommunication by Implementing Security Algorithm in VANET

Manpreet Kaur[1], Rajni[2], and Parminder Singh[3]

[1] Research Scholar
SBSCET, Ferozepur
[2] Assistant Professor
SBSCET, Ferozepur
[3] Senior Lecturer
CGC, Landran
er_rimpi@rediffmail.com, rajni_c123@yahoo.co.in,
Singh.parminder06@gmail.com

Abstract. Adhoc networks are the networks, deployed in the areas where the needed infrastructure is not feasible to install. Mobile Adhoc Networks (MANETs) use wireless medium in which each node acts as both data terminal and router without the need of any centralized control. Vehicular Adhoc Networks (VANETs) are subset of MANETs. The successful deployment of vehicular communication requires Vehicle-to-Vehicle (V2V) & Vehicle–to-Infrastructure (V2I) communication with security which requires confidentiality, integrity, availability and authenticity to improve road safety and optimize road traffic. The technique used for secure communication in the presence of unauthorized parties (adversaries) is known as cryptography. Cryptography refers to encryption in which a plaintext message is converted into a ciphertext. Encryption can be done with private-key or public-key. In this paper, an algorithm based on private key encryption is used to make communication possible between two people with QualNet simulator.

Keywords: Mobile Adhoc Networks (MANETs), Vehicular Adhoc Networks (VANETs), Intelligent Transport System (ITS), Security, Vehicle-to-Vehicle (V2V), Vehicle –to-Infrastructure (V2I).

1 Introduction

The process of exchange of information via some transmission media is known as network. VANETs are used for inter-vehicular communication considering all the security aspects required by the process of cryptography. In Safety applications, there is a requirement of V2V & V2I communication.

B.Karp in 2000 [1] discussed Greedy Perimeter Stateless Routing (GPSR) scheme in which the default packet forwarding strategy is the greedy strategy, where the sender selects the closest vehicle to the destination as the next hop. If the sender cannot find a forwarder based on the greedy strategy, it forwards the packet around the perimeter of the region containing itself and the destination.

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 757–763.
springerlink.com

Giuseppe Ateniese in 2005 [2] presented a new re-encryption schemes that realizes a stronger notion of security and demonstrated the usefulness of proxy re-encryption as a method of adding access control to a secure file system.

Maxim Raya in 2006 [3] described the security problems of the emerging vehicular networks and it outlined the solution architecture and several of its components. Tim Leinmuller in 2006 [4] aimed to define a consistent & future-proof solution to the problem of V2V/V2I security by focusing on SEVECOM (Secure Vehicle Communication). Pandurang Kamat in 2006 [5] proposed a security framework for vehicular networks using Identity-Based Cryptography (IBC), that provides authentication, confidentiality, message integrity, non repudiation and pseudonymity.

Xiaonan Liu in 2007 [6] described an Intelligent Transport System (ITS) which can be used under the security pattern to provide the appropriate solving measures in concern with the security issues of VANETs from some aspects.

Department for Transport in 2008 [7] concluded that it had been seen from various studies that the number of lives lost in motor vehicle crashes worldwide every year is by far the highest among all the categories of accidental deaths.

P.Caballero-Gil in 2009 [8] analyzed the features of inter-vehicle and vehicle-to-roadside communications to propose differentiated services for node authentication, according to privacy and efficiency needs.

Zuowen in 2010 [9] proposed an improved privacy-preserving mutual authentication protocol for vehicle ad hoc networks by using secure identity-based group blind signature, the private encryption system and the public encryption system. Surabhi Mahajan in 2010 [10] discussed a comparison between the two schemes that are used to reduce the overhead in authentication, when roaming – proxy re-encryption scheme and new proxy re encryption scheme. Hatem Hamad in 2010 [11] proposed a new method of message security by using the coordinates in GPS (Global Positioning System) service.

Umar Farooq Minhas in 2011 [12] described an important methodology required to enable effective V2V communication via intelligent agents. Nizar Alsharif in 2011 [13] explained that the reliability of position-based routing (PBR) in VANETs is ensured by proposing a set of plausibility checks that can mitigate the impact of PBR attacks without extra hardware requirement.

2 Security Challenges

The necessities required to provide security in VANETs are:

1. Authentication: An authentication framework is necessary to identify and ensure that the participants are valid and they are whom they claim to be to operate securely in VANETs.
2. Integrity: The data sent between two communicating nodes should be accurate that is to protect data accuracy is security issue desirable in VANETs.
3. Confidentiality: The major challenge is to protect data content/access from the third parties.
4. Non-Repudiation: Non-Repudiation refers to somebody who possesses the private key corresponding to the signing certificate with reasonable certainity

but if the key is not properly safeguarded by the original owner, a major concern can be digital forgery.

5. Pseudonymity: The state of describing a disguised identity is Pseudonomity which is the major issue of concern in the security of VANETs. A holder that is one or more human beings are identified but don't disclose their true names.
6. Privacy: The protection of personal information of drivers within the network from other nodes but extracted by authorities in case of accidents is a major privacy issue which is desirable for VANETs.
7. Scalability: The ability of a network to handle growing amount of work in a capable manner securely is Scalability, which is the main challenge in VANETs.
8. Mobility: The nodes communicating in VANETs constantly change their locations with different directions and speeds making the network dynamic in nature. So, in order to make communication successful, it is challenging to establish security protocols.
9. Key-Management: The key is used to encrypt and decrypt information during communication process. When designing security protocols for networks like VANET, the issue of key management must be resolved.
10. Location-verification: This is necessary to prevent many attacks and is helpful in data validation process. Thus to improve the security of VANETs, a solid method is required to verify the nodes positions.

3 Scenario Used

When implementing encryption, the choice of algorithm should be dictated by the purpose of the encryption. Private Key encryption is faster than public key encryption. However, private key encryption does not provide for digital signatures or the signing of information. It is also important to choose well-known and well-reviewed algorithms. Such algorithms are less likely to include back doors that may compromise the information being protected.

The security policy should define acceptable encryption algorithms for use within the VANET System. The security policy should also specify the required procedures for key management. In order to successfully gain access to the information transmitted over the VANET, an attacker must

- Capture the entire session, which means that a sniffer must be placed between the two end points at a location where all the VANET traffic must pass.
- Use a substantial amount of computer power and time to brute-force the key and decrypt the traffic.
- It would be much easier for an attacker to exploit vulnerability on the user's computer or to steal a portable computer in an airport. Unless the information is extremely valuable, any well-known, strong algorithm is appropriate for use in the VANET System.

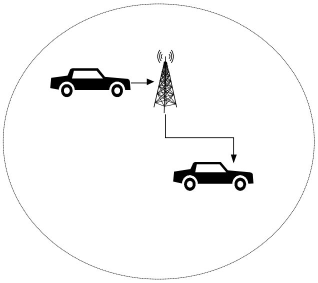


Fig. 1. V2V Communication in VANET

4 Algorithm Description

The Algorithm used in VANET system by implementing Private Key Encryption (In this Algorithm *i* and *j* remain secret) in a straight road. Assume there were two Vehicles V1 and V2 that need to communicate securely and thus need to agree on an encryption key.

1. V1 and V2 agree on two large integers *a* and *b* such that $1 < a < b$.
2. V1 then chooses a random number *i* and computes $I = a_{\mathrm{i}} \bmod b$. V1 sends *I* to V2.
3. V2 then chooses a random number *j* and computes $J = a_{\mathrm{j}} \bmod b$. V2 sends *J* to V1.
4. V1 computes $k1 = J_{\mathrm{i}} \bmod b$.
5. V2 computes $k2 = I_{\mathrm{j}} \bmod b$.
6. We have $k1 = k2 = a_{\mathrm{ij}} \bmod b$ and thus *k*1 and *k*2 are the secret keys to use for the other transmission.

5 Implementation Tool

The performance results have been evaluated using QualNet Simulator. QualNet is network evaluation software, which is entirely modeled as a Finite State Machine (FSM) and is written purely in C++. It can run on a variety of operating systems like UNIX, Windows, MAC and Linux and is equipped with an extensive range of libraries for simulating a variety of networks. The layered architecture of QualNet comprises of Application, Transport, Medium Access Control (MAC), Physical layer. A unique capability for accurate, efficient simulation of large-scale, heterogeneous networks can be provided by the following features of QualNet:

- QualNet can simulate a robust set of wired and wireless networks.
- QualNet has support for ITM (Irregular Terrain Model).
- QualNet executes scenarios 5-10x times faster than commercial alternatives.
- Processors addition makes execution of simulation multiples faster with QualNet.
- QualNet can simulate fading with Rayleigh and Ricean methods.

6 Results and Discussions

In this paper, the different results have been evaluated using QualNet simulator. The results shown indicate that the packet is sent between sender and receiver by using private key encryption algorithm. This means the same key is used for encryption and decryption. This leads to increase in throughput with reduce delay.

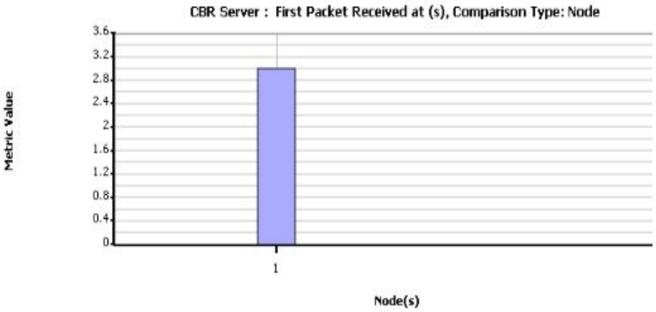


Fig. 2. First packet received during transmission

From the figure 2, we have measured the throughput of moving Vehicles communicating with each other via Access Point (A.P) already shown in figure 1. V1 Communicates with V2 via Access Point which leads to the calculation of the throughput of V1. From the Figure1, when the communication starts between AP and V2 then the throughput of first packet is calculated at its arrival. The first packet arrives from V1 having Node ID 1=2.8 seconds. The simulation carried from the help of QualNet Simulator and the length of Packet = 1024 Bytes.

From the Figure 3, the throughput of Last Packet received by the AP from V1 (called Source Node) =100 Seconds approximately is calculated. After the Packet is received, the Vehicle Node sends the termination signal to do end of communication between two Vehicles with the Base Station (AP) Node.

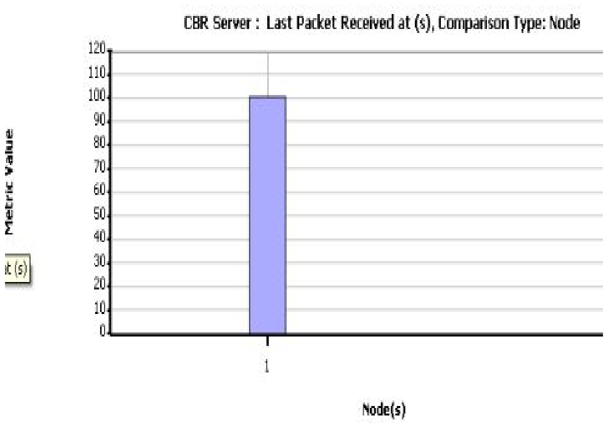


Fig. 3. Last packet received during transmission

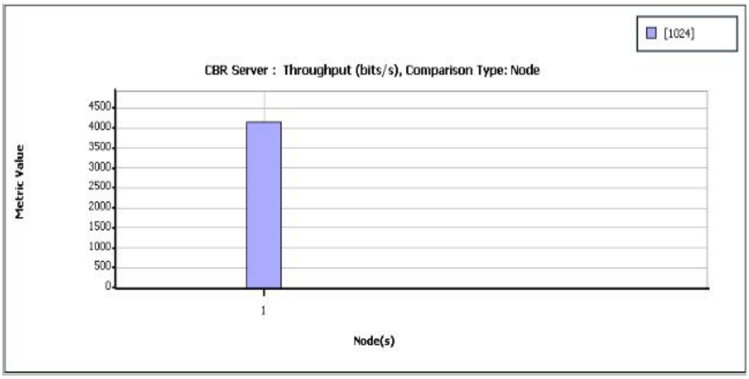


Fig. 4. Throughput observed during transmission (with private key)

The figure 4 represents the Throughput of Two Nodes with AP by applying Security Algorithm and also calculates the speed at which the two vehicles move when sharing the data. The speed at which the Vehicles move was 60m/s with a packet size of 1024 Bytes and approximately calculated value was 4200bits/s.

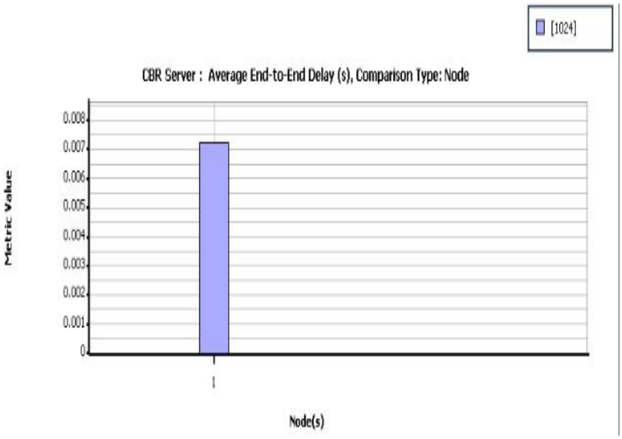


Fig. 5. Average End-to-End Delay during transmission (with private key)

The average end-to-end delay in a VANET means that the source node (V1) sends packet to destination Node (V2) and total average time to reach the packet to destination Node. In this Figure 5 the total average time was 0.007 seconds approximately. This delay was due to retransmission and applying security Algorithm at 802.11 networks.

7 Conclusion

Security measures protect data during transmission and guarantees that the transmissions of data are authentic that is data is accessible only by authorized parties. Various security techniques can be applied to vehicular users in Vehicular Adhoc Networks. The algorithm used in the paper is private key encryption in which sender and receiver communicate securely with the help of encryption key. The algorithm reduces delay, increases throughput, provides authentication and higher security level in VANETs.

8 Future Work

The algorithm discussed here leads to solution of the security problems that are encountered in VANET. The system is costly, so an effective cost management analysis of the system can be a great future research issue. Various performances of data transmission in VANETs can be tested by applying more encryption algorithms to provide more security with increased throughput and reduced delay.

References

[1] Karp, B., Kung, H.T.: GPRS: Greedy Perimeter Stateless Routing for wireless networks. In: Proc. 6th Annual Int. Conf. on Mob. Computing and Netw., pp. 243–254 (2000)

[2] Ateniese, G., Fu, K., Green, M., Hohenberger, S.: Improved Proxy Re-Encryption Schemes with Applications to Secure Distributed Storage. In: Proc. of the 12th Annual Netw. and Distributed Syst. Security Symp., pp. 29–44 (2005)

[3] Raya, M., Papadimitratos, P., Hubaux, J.P.: Securing Vehicular Networks. J. IEEE Wirel. Commun. (2006)

[4] Leinmuller, T., Buttyan, L., et al.: SEVECOM-Secure Vehicle Communication. In: Proc of IST Mob. Summit (2006)

[5] Kamat, P., Baliga, A., Trappe, W.: An Identity-Based Security Framework for VANETs. In: Proc of the 3rd Int. Workshop on Vehicular Adhoc Netw., pp. 94–95 (2006)

[6] Liu, X., Fang, Z., Shi, L.: Securing Vehicular Ad Hoc netw. IEEE (2007)

[7] Department for Transport, Reported road casualties Great Britain: 2008 Annual Report. Road Casualties G. B. UK (2008)

[8] Caballero-Gil, P., Hernandez-Goya, C., Fuster-Sabater, A.: Differentiated Services to Provide Efficient Node Authentication in VANETs (2009)

[9] Zuowen, T.: A Privacy-Preserving Mutual Authentication Protocol for Vehicle Ad Hoc Networks. J. of Convergence Inf. Technology (2010)

[10] Mahajan, S., Jindal, A.: Security and Privacy in VANET to reduce Authentication Overhead for Rapid Roaming Networks. Int. J. of Comput. Applications (2010)

[11] Hamad, H., Elkourd, S.: Data Encryption using the dynamic location and speed of mobile node. J. Media and Commun. Stud., pp. 067-075 (2010)

[12] Minhas, U.F., et al.: A Multifaceted Approach to Modeling Agent Trust for Effective Communication in the Application of Mobile Ad Hoc Vehicular Networks, pp. 407–420. IEEE (2011)

[13] Wasef, A.N., Shen, X.: Mitigating the effects of Position-Based Routing Attacks in Vehicular Ad Hoc Networks. In: IEEE ICC Proc. (2011)

An Efficient Approach to Secure Routing in MANET

Dipayan Bose[1], Arnab Banerjee[1], Aniruddha Bhattacharyya[1], Himadri Nath Saha[1],
Debika Bhattacharyya[1], and P.K. Banerjee[2]

[1] Department of Computer Science & Engineering
Institute of Engineering & Management, Salt Lake Kolkata, India
connect2dipayan@gmail.com,
arnab.saheb@gmail.com,
aniruddha.aot@gmail.com,
him_shree_2004@yahoo.com,
bdebika@yahoo.com
[2] Department of Electronics and Telecommunication Engineering
Jadavpur University, West Bengal, India

Abstract. In MANET secure routing is an important issue because of its self organizing and cooperative nature, capable of autonomous operation, rapid changing topologies, limited physical security and limited energy resource. So our proposal is a new scheme which significantly differs from other available schemes dealing with security attacks against mobile ad hoc networks. In this paper, our proposed scheme, Efficient Secure Routing Protocol in MANET (ESRP) provides a new routing scheme based on trust, which is an integer value that helps to select administrator inside the network for routing. The comparison between our proposed protocol and parameters of ad hoc network shows the performance according to secure protocol. We have also implemented the message confidentiality and integrity in our proposed scheme. Our simulation result shows the robustness, reliability and trustworthiness of our scheme.

Keywords: manet, ESRP, trust, administrator, digital signature, secure routing protocol, willingness function, olsr, secure routing.

1 Introduction

Mobile Ad Hoc Network (MANET) is a network consisting of a collection of nodes capable of communicating with each other in a self-organized and non predefined infrastructure. Ad Hoc networks are new paradigm of wireless communication for mobile hosts where node mobility causes frequent changes in topology. It has been used in a wide range of applications ranging from a battlefield to the user's living room. Many efficient routing protocols have better network performance however they are more vulnerable to security threats. Ad hoc network has faced even more serious security problems as compared to traditional wireless networks. Several security solutions require a centralized server for key distribution or a secret understanding between communicating entities. This lack of infrastructure has posed serious threats as far as routing security is concerned. Secondly, the vulnerability of the nodes towards

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 765–776.
springerlink.com

physical compromise gives rise to serious internal threats within the network which make the issues of authentication, integrity and confidentiality even more challenging than conventional wireless networks. Thirdly, without support from fixed infrastructure it is undoubtedly arduous for people to distinguish the insider and outsider of the wireless network thereby difficult to tell apart the legal and illegal participants in wireless network. This assumption is also coupled with pre-configuration of nodes with encryption keys prior to joining the network. Public key cryptography with digital signature makes the network stronger to stand up against the attackers and secure communication is assured. However, due to the limitation of battery energy of mobile nodes, methods of prolonging the lifetime of nodes as well as the network become the key challenge in MANET. The performance of MANET depends on the routing scheme employed and the traditional routing protocols do not work efficiently in MANET. Developing routing protocols for has been an extensive research area in recent years, and many proactive, reactive and hybrid protocols have been proposed from a variety of perspectives[1]. Section I introduces our research on the security of MANET. Section II describes the Working Methodology of ESRP, while section III explores related works in this domain. Section IV describes our proposed algorithm. In section V, we present the proposed packet format while section VI gives us the picture of the performance evaluation. Lastly, section VII deals with all future work and section VIII express the conclusion in relation to this domain.

2 Working Methodology of ESRP

Our proposed routing algorithm, ESRP (Efficient Secure Routing Protocol) is a proactive routing protocol inspired by OLSR [2]. In this algorithm trust has been established using signed acknowledgement based on asymmetric key cryptography. Key distribution is out of the scope of this paper and any popular key distribution methodology can be followed. This protocol concentrates in dispersal of packets from source to network through administrator. We have selected Admin node as a minimal subset of all nodes that can form a fully connected network. It consists of all the administrators which can reach out to all the neighbor nodes. This administrator node selection depends on symmetric link, node coverage, willingness of that node and TRUST.

3 Related Works

Till date many secure routing protocols have been developed like SOLSR, TAODV, SAODV, etc. SOLSR [3] (based on OLSR) has used symmetric key for encrypting all data and control packets but Trust concept has not been implemented yet. While TAODV [4] (based on AODV [5]) does not use any encryption technique but it uses the trust factor. Again when considering the case of SAODV [6], it uses public key cryptography and digital signature to protect RREQ & RREP messages [7]. It also uses hash-chain to authenticate hop-count of each message. Few secured routing protocols like SRP [8], FTAODV [9], Ariadne [10] and others [11], [12], [13] have similar kind of approaches and so are not included in this paper. So in our protocol we have blended the concepts of both cryptography and trust factor to enhance the security of the protocol. We are using digital signature in each acknowledgement packet to prevent generation of forged packet.

4 Proposed Algorithms

This paper outlines the mechanism for selection of administrator, based on willingness & trust value of a node considering more exhaustive parameters so as to keep the admin node count to the minimum (as per basic OLSR) and make routing more secure. Our algorithm forces the same path return as traversed while sending. Only when absolutely necessary, admin node can switch from one node to another, offloading their job to the other node, increasing network runtime. The algorithms to maintain a secure and reliable network run in each individual node.

First we will discuss about Admin node selection algorithm. The value of willingness will be derived from next algorithm and trust from algorithm given in section 4.

4.1 Dynamic Willingness Function

Our algorithm takes a weighted sum of battery power of a node, coverage area and reliability of the node while calculating willingness value. The weighted values are experimentally tested and optimized. The power factor in MANET is crucial so it has been assigned the highest weighted value and so on. All weights are experimentally tested and optimized value for the scheme.

$$\text{Willingness}(P, C, R) = (0.75 * P) + (0.15 * C) + (0.1 * R) \quad (1)$$

Where, P: power available for that node (in %)

$$P = (\text{current node power/rated capacity of the node})*100 \quad (2)$$

C: coverage (in %)

C = (no of 1-hop neighbors of that node / no of 2-hop neighbors of nodes that want to select this node as its ADMIN)*100 (3)

R: reliability of the node (in %)

Reliability (R) is calculated from various sensor inputs regarding outside environment condition where R ranges from 0% to 100% depending upon the node's position. R = {0% … 100%} (4)

4.2 Admin Node Selections

This algorithm selects the administrator node which can cover most of the 2-hop neighbor of its selector. Selection also takes care of willingness and trust value of node. In case of tie, node with higher trust/power will be selected.

Few Definitions

- ADMIN(x): Admin set of node x which is running this algorithm.
- N1(x): One hop neighbor set of node x (symmetric neighbors)
- N2(x): Two hop neighbor set of node x [symmetric neighbors of nodes in N(x)].The two hop neighbor set N2(x) of node x does not contain any one hop neighbor N(x) of node x.

- D(x,y) : Degree of one hop neighbor node y (where y is a member of N1(x) -- means y belongs to N1(x)), is defined as the number of symmetric one hop neighbors of node y EXCLUDING the node x and all the symmetric one hop neighbors of node x, i.e.,

$$D(x, y) = N(y) - \{x\} - N1(x) \tag{5}$$

- W = Current willingness value of the node. [can range from 0 to 7]
- T = Current trust value of the node. [can range from 0 to 10]
- Trust_Threshold = Implementation dependent [we choose 2]

Initialization

1. *Initialize* ***Node_Trust*** *table with default trust value 3 for each node.*
2. *Initialize PATHLIST = [].*

Algorithm

```
Step 1: Start with an empty ADMIN(x) set.
Step 2: Calculate D(x, y), where y is a member of N1(x), for all nodes in N1(x) (put
        for all +ve sign)
Step 3: First select as ADMINs those nodes in N1(x) which provides the "only path"
        to reach some of the nodes in N2(x). [Trivial case]
Step 4: For each node in N1(x)
        {
          4.1 SELECT current node as a ADMIN as per table 1.
          4.2 While if some nodes still exists in N2(x) that is not covered by ADMIN(x):
          {
            For each node in N1(x), calculate the no. of nodes in N2(x) which are not
            yet covered by ADMIN(x) and are reachable through this one hop neighbor
            of x.
          }
          4.3 Select as a ADMIN that node of N1(x) which reaches the maximum
              number of uncovered nodes in N2(x) & refer table 1.
          4.4 If a tie occurs, select that node as ADMIN whose D(x,y) is greater & refer
              table 1.
        }
Step 5: To optimize, process each node y in ADMIN(x), one at a time, if ADMIN(x) -
        {y} still covers all nodes in N2(x) then remove y from ADMIN(x).
Step 6: After that Convert the link between node x and ADMIN as SYM_LINK to
        ADMIN_LINK
Step 7: Exit
```

Table

Table 1. Admin Selection in case of tie

NODE 1:		NODE 2:		
TRUST (T1) %	POWER (P1)%	TRUST (T2)%	POWER (P2)%	SELECTION
L	L	L	L	WHEN BOTH THE NODES HAVE THE SAME VALUES THEN SOURCE NODE CAN BROADCAST THE MESSAGE TO THE NETWORK THROUGH EITHER OF THE NODES. EITHER NODE1 OR NODE2
L	L	L	H	NODE2
L	L	H	L	NODE2
L	L	H	H	NODE2
L	H	L	L	NODE1
L	H	L	H	(IF P1>P2 THEN NODE1 ELSE NODE2) ELSE (IF P1==P2 THEN IF T1>T2 THEN NODE1 ELSE NODE2)
L	H	H	L	(IF P1-TH_PWR>T2-TH_TR & T1-TH_TR > P2-TH_PWR THEN NODE1) ELSE (IF T2-TH_TR>P1-TH_PWR & P2-TH_PWR > T1-TH_TR THEN NODE2)
L	H	H	H	NODE2
H	L	L	L	NODE1
H	L	L	H	(IF P1>P2 THEN NODE1 ELSE NODE2)
H	L	H	L	(IF T1>T2 & P1-TH_PWR>P2_TH_PWR THEN NODE1) ELSE (IF T2>T1 & P2-TH_PWR>P1_TH_PWR THEN NODE2) ELSE
H	L	H	H	NODE2
H	H	L	L	NODE1
H	H	L	H	NODE1
H	H	H	L	NODE1
H	H	H	H	(IF P1>P2 THEN NODE1 ELSE NODE2)

4.3 Digital Signature and Trust Value Calculation

Sender Node's Job

Step 1: Encrypt the message with Public Key of destination
ENC_MSG←ENCRYPT (PlainText_MSG)

```
Step 2: Calculate HASH VALUE for ENC_MSG
        HASH_VAL ← HASH (ENC_MSG)
Step 3: Create a entry for PATHLIST table with following data:
        < HASH_VAL, DEST_NODE_ID >
Step 4: Set a TIMER for this entry with timeout value T.
        [Value of T is implementation dependent]
```

Original Message

If the Node is Intermediate Node

```
Step 1:   Receive the encrypted message.
Step 2:   Append next Hop ID to the variable Path.
Step 3:   Update the packet size to reflect the modified Path.
Step 4.1: Calculate: HASHVAL←HASH (MSG.ENC_MSG)
Step 4.2: Store the following entry in PATHLIST table:
          <HASHVAL, DEST_NODE_ID +MSG.PATH>
Step 5:   Set TIMER with T sec Timeout for this entry.
Step 6:   Forward the updated encrypted message.
```

If the Node is Intended Receiver Node (DEST)

```
Step 1: Extract PATH from received message:
        PATH ← MSG.PATH
Step 2: Extract message:
        MSG←DECRYPT (MSG.ENC_MSG)
Step 3: Create a HASH value for ACK message generation:
        HASHVAL_C←HASH (MSG.ENC_MSG)
Step 4: Sign the ACK message:
        SIGN← ENCRYPT (HASHVAL_C, PVT_KEY_DEST)
Step 5: Transmit the ACK message with SIGN to Previous
        Node found in PATH.
```

For Acknowledge Message.

```
Step 1: Receive the ACK packet.
Step 2: Extract the encrypted hash value.
        HASHVAL_R←ACK.ENC_HASH
        [Where ACK.ENC_HASH = ENC (HASH (ENC_MSG), PRK_DEST))]
Step 3: Find entry in PATHLIST with HASHVAL_R
Step 4.1: If entry found
        i)    Extract stored path
              E_PATH ←Entry.PATH
        ii)   If
              the last node in E_PATH is Sender Node of this ACK packet
              then
               increase TRUST of Sender Node by 1.
```

```
                Else
                decrease TRUST of Sender Node by 1 and discard the packet.
                Remove this entry from PATHLIST.
                Round off TRUST to within 0 to 10.
                GOTO Step 5
      iii)      Update the E_PATH of ACK packet by removing the Sender Node
                ID.
                Remove this entry from PATHLIST.
      iv)       Forward the ACK message to the previous hop in E_PATH.

Step 4.2: If entry not found decrease TRUST of Sender node
          by 1(round off within 0 to 10) and discard the
          packet.
          Remove this entry from PATHLIST.
          GOTO Step 5

Step 5:   Done.
```

On expiration of time out for particular entry in path list

```
Step 1: Extract path from Time out entry:
        PATH_T ← Timeout_Entry.PATH
Step 2: Decrease TRUST value for last node in PATH_T by 1unit
Step 3: Remove the entry from PATHLIST table.
```

5 Packet Format

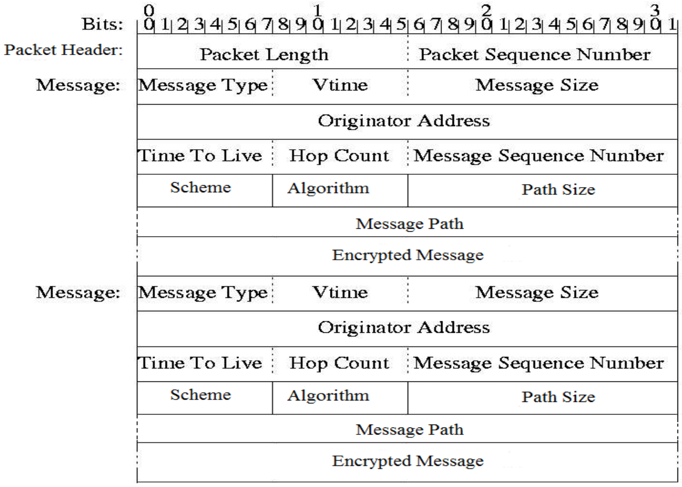


Fig. 1. Message Packet Format

As multiple packet are piggybacked (as in OLSR) into a single packet, each message part will contain its own path and separate encrypted message content. All message type except HELLO_MESSAGE will be encrypted with destination node's public key. Scheme & Algorithm field is used to send ATSR specific data. In each hop, Message Path field is updated to add the current hop address. Accordingly, Message Size & Packet Length is updated. For ACK packet (Message Type = ACK), Message Path is omitted for ACK packet. Instead of Encrypted Message part, following is send:

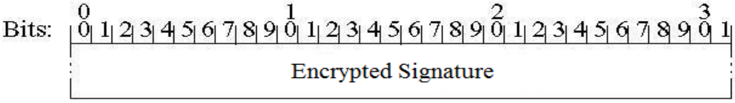


Fig. 2. ACK packet format

6 Performance Evaluation

6.1 Admin Node Selections

We used OLSR protocol implementation from Niigata University for Glomosim. [14], [15].

Table 2. Simulation parameters

Parameter	Value
Terrain Dimension	(600x500) sq. meter
Simulation Time	500 minutes
Channel	Noisy
Noise Figure	10 dB
Radio Frequency	2.4 Ghz
Radio Receive Threshold	-65.046 dBm
Radio Transmit Power	22.5 dBm
Node Placement	Random
Mobility Speed	0-10 m/s
MAC Protocol	802.11
MAC Propagation Delay	1000 ns
Bandwidth	11 Mbps
Routing Protocol	OLSR, ESRP, SAODV
Number of Interface per node	2
Rated Battery Power (each node)	1500 mAh
Data Packet Type	FTP, CBR
Data Packet Size	2044 byte
Cryptographic algorithm	RSA (512 bit)

To simulate the proposed algorithm we used Glomosim 2.03 network simulator [14]. Glomosim can simulate both wired and wireless network with layered TCP/IP stack with model based on noisy & noiseless channel with MAC protocol 802.11/CSMA/MACA/TSMA and various network, transport & application layer protocols. Glomosim is written using PARSEC language [16], a C derivative for large scale parallel simulation.

6.2 Energy Consumption Model

We are using IEEE 802.11b (DSSS modulation) as MAC protocol. The transceiver uses energy both to transmit and to listen for incoming packet. It also consumes energy in idle state. Let, the energy needed to transmit a packet E_t for duration t_t and to receive a packet E_r for duration t_r .Also assume it waits for t_i consuming energy E_i . Then total energy consumed by that node will be approximately:

$$E_c = E_t * t_t + E_r * t_r + E_i * t_i \qquad (6)$$

We assumed each node will use 5V DC battery with rated capacity of 1500 mAh. Transmission energy consumed will depend on radio signal strength of transmission; here we assumed 22.5 dBm; which approximately translate into 177.83 mW.

$$A = \frac{W}{V} \qquad (7)$$

From equation (7) we get, A= 35.57 mA for V=5V DC. If we draw the same amount of current, using 1500 mAh battery, we'll get approximately 42 hour of runtime before the battery dies. Adding Idle and receiver power we'll get less than that.

6.3 Simulation Results

We have made a comparative study between OLSR, SAODV and our protocol ESRP. We carried out the result is based on the simulated data, the ACK being sent and frequency of data transfer. First we evaluate number of admin in the network by both protocols variant as a function of number of nodes. Maximum numbers of nodes were set to 50. Also to simulate attack vector, we configured Glomosim in such a way that 20% of those nodes will randomly drop packet or delay the delivery to next hop.

Simulation results are illustrated in following figures:

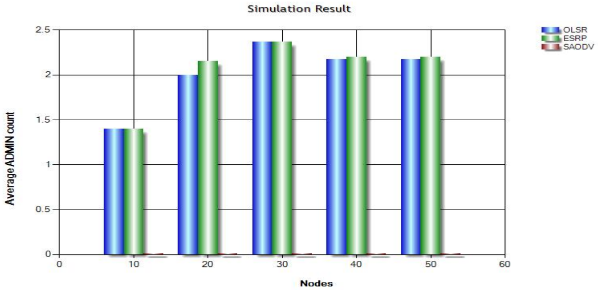


Fig. 3. ADMIN Count

Here we can’t see much difference in average ADMIN count over basic OLSR protocol. We can also see that number of ADMIN count has increased slightly when numbers of nodes were 20, 40 & 50. This increase in ADMIN count is due to shift in responsibility as the node’s willingness & trust changes with time. Significant increase in ADMIN count adversely decreases network performance. But here the count has increased only slightly. SAODV does not use ADMIN concept.

But increase in ADMIN count will affect radio layer packet collision, as depicted in figure 4:

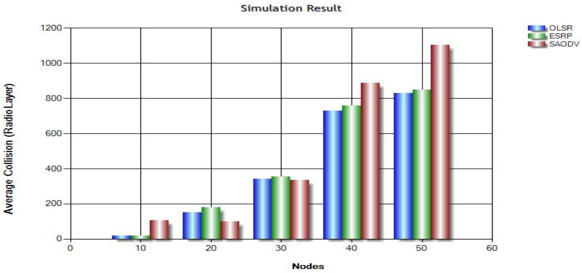


Fig. 4. Average Collision

We can see the collision in fact has increased, but only slightly as the increase of ADMIN count was not so drastic. This increase was due to reselection of ADMIN and subsequent topology message being broadcasted internally. It also increases due to sending and receiving of acknowledgement packets. For SAODV, the increase in collision is due to frequent route request-reply in each transmission. Collision increases with network density as more and more nodes are trying to compete for radio frequency. Using 802.11b reduced collision due to deliberate use of collision avoidance scheme (such as RTS/CTS) built into radio layer protocol itself. Also we saw a slight change in throughput in the protocol. SAODV’s performance was poor as compared to OLSR & ESRP. [Depicted in following figure]:

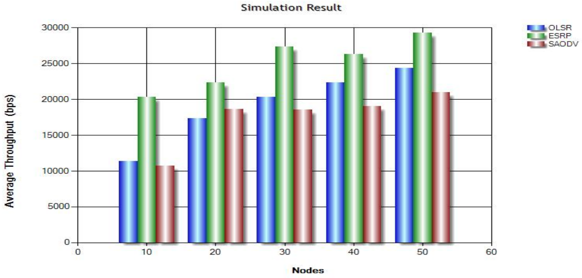


Fig. 5. Average Throughput

With 11 Mbps network bandwidth and multiple FTP and CBR data transfer, we saw average throughput stayed around 27 kbps. Actually the average throughput increases in the case of successful data transfer. Implementation of security helps us to avoid retransmission of packets as well as data packet flooding. We also found that end-to-end delay also increased with our proposed protocol compared to stock OLSR:

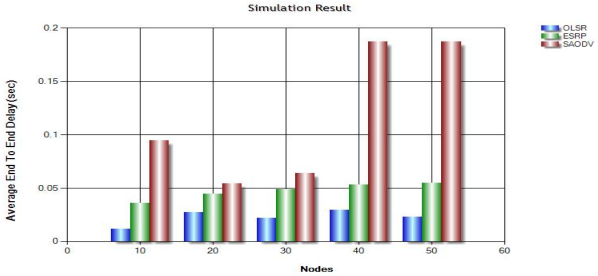


Fig. 6. Average End-To-End Delay

Compared to ESRP, SAODV has increased end-to-end delay; we suspect it is due to transmission through suboptimal path. End-To-End delay increases for encrypting each message though the transmission of every packet is secured. Then ACK transmission and encryption of messages also increases the end to end delay considerably.

We are trying to demonstrate that our protocol does not adversely affect the network performance compared to the existent solutions. Our protocol is quite robust as it protects from data and control traffic attacks.

7 Future Work

We have already implemented trust factor in ESRP using signed acknowledgement which has enhanced the security of the routing protocol. We have also been able to mitigate black hole, gray hole, forged ACK, snooping attacks using this protocol. We have also implemented parameterized willingness function in ESRP. Now our next future goal is to mitigate as many routing attacks as possible by simulating each of those attacks individually.

8 Conclusion

Our secure ESRP which is inspired from OLSR may not be energy efficient but is quite secure for end to end communication as compared to other routing protocols. In this paper Administrator and trust based routing has been proposed. This novel feature allows us to forward the data packets to the destination and by receiving the acknowledgement it verifies the validity of the nodes in the route. The performance of this routing algorithm in comparison to OLSR has improved. The security implementation has also protected the network from internal and external threats.

References

1. Mahfoudh, S., Minet, P.: An energy efficient routing based on OLSR in wireless adhoc and sensor networks. In: 22nd International Conference on Advanced Information Networking and Applications (2008)
2. Clausen, T., Jacquet, P.: Optimized Link State Routing Protocol (OLSR), RFC 3626, Workshops (2008), `http://tools.ietf.org/html/rfc3626`
3. Hong, F., Hong, L., Fu, C.: Secure OLSR. In: 19th International Conference on Advanced Information Networking and Applications, AINA 2005, vol. 1, pp. 713–718 (2005)
4. Li, X., Lyu, M.R., Liu, J.: A trust model based routing protocol for secure ad hoc networks. In: Proceedings of the IEEE Aerospace Conference 2004, vol. 2, pp. 1286–1295 (2004)
5. Perkins, C., Belding-Royer, E., Das, S.: Ad hoc On-Demand Distance Vector (AODV) Routing. IETF. RFC 3561 (July 2003)
6. Lu, S., Li, L., Lam, K.-Y., Jia, L.: SAODV: A MANET Routing Protocol that can Withstand Black Hole Attack. In: Conference on Computational Intelligence and Security, CIS 2009, pp. 421–425 (2009)
7. Juwad, M.F., Al-Raweshidy, H.S.: Experimental Performance Comparisons between SAODV & AODV. In: Second Asia International Conference on Modeling & Simulation, AICMS 2008, pp. 247–252 (2008)
8. Papadimitratos, P., Haas, Z.J., Samar, P.: The Secure Routing Protocol (SRP) for Ad Hoc Networks, draftsecure-routing-protocol-srp-00.txt (September 2002)
9. Martin Leo Manickam, J., Shanmugavel, S.: Fuzzy based Trusted Ad hoc On-demand Distance Vector Routing Protocol for MANET. In: 15th International Conference on Advanced Computing and Communications (2007)
10. Hu, Y.C., Perrig, A., Johnson, D.B.: Ariadne: A Secure On- Demand Routing Protocol for Ad Hoc Networks. In: Proceedings on Eighth Annual Int'l Conf. Mobile Computing and Networking (MobiCom), pp. 12–23 (2002)
11. Yang, Y.-T., Yuan, Z., Fang, Y., Zeng, P.: A Novel Authentication Scheme Based on Trust-value Updated Model in Adhoc Network. In: 31st Annual International Computer Software and Applications Conference, COMPSAC 2007 (2007)
12. Gilaberte, R.L., Herrero, L.P.: A secure routing protocol for ad hoc networks based on trust. In: Third International Conference on Networking and Services, ICNS 2007 (2007)
13. Gonzalez, J.M., Anwar, M., Joshi, J.B.D.: Trust-based Approaches to Solve Routing Issues in Ad-hoc Wireless Networks: A Survey. In: International Joint Conference of IEEE, TrustCom-11/IEEE CESS-11/FCST-11 (2011)
14. Zeng, X., Bagrodia, R., Gerla, M.: GloMoSim: A Library for Parallel Simulation of Large-scale Wireless Networks. In: Workshop on Parallel and Distributed Simulation (May 1998)
15. Niigata University, Information & Communication Networks laboratory, OLSR_Niigata, `http://www2.net.ie.niigata-u.ac.jp/olsr-e.php`
16. Bagrodia, R., Meyer, R., Takai, M., Chen, Y.-A., Zeng, X., Martin, J., Yoon, H.: Parsec: A Parallel Simulation Environment for Complex Systems (October 1998)

Trust Oriented Secured AODV Routing Protocol against Rushing Attack

Swarnali Hazra and S.K. Setua

[1] Computer Science and Engineering, University of Calcutta, India
{swarnali.hazra,sksetua}@gmail.com

Abstract. This research aims to identify the security threats in on-demand routing protocol, AODV for Ad-hoc network. In this context we have extended the AODV protocol with trust and recommendation to secure the network. In AODV, establishment of routing path depends on the faster route request and route reply packets. Rushing attacker exploits such faster traversal activities to attack the network. Firstly, rushing attackers are identified based on their misbehavior in comparison to other nodes of the network. Furthermore, trust value is assigned to the misbehaving node and the same is augmented with other aspects of trust like dependency pattern, context, previous history and dynamicity. Finally, based on threshold value of trust, trust evaluating node takes the decision to include or not to include the trustee node in routing path depending on the final trust value. To facilitate the trust computation and decision of our trust model, AODV is enhanced with different functional modules: Node Manager, Trust Module and Decision Manager. Lastly, AODV secures the routing path by isolating the rushing attacker, based on their trust value. Our analysis and simulation results show the effectiveness of our proposal against rushing attack.

Keywords: AODV, rushing attack, trustor, trustee, direct trust, indirect trust.

1 Introduction

In existing on-demand routing protocol AODV [9], a source node requiring communication to a destination node, broadcasts route request packet (RREQ), to find a communication route to that destination. Network is flooded with RREQ. To limit the overhead of such flooding, each node typically forwards only one RREQ that arrives first from each route discovery. Rushing attack [10] exploits this mechanism of the route discovery process. A rushing attacker forwards RREQ more quickly than legitimate nodes to increase the probability of discovering routes including itself. When a node receives rushed RREQ from the attacker, it broadcasts this first received RREQ and discards late arriving legitimate RREQs. As a result, AODV is unable to discover data delivery route without attacker.

To prevent this rushing attack, we incorporated trust concept in AODV, and proposed Context Aware Trusted AODV against Rushing attack (CAT-AODV-R). In our work, trust is represented as a level of one node's expectations about responsiveness of

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 777–788.
springerlink.com

another node and trust value varies in-between zero and one. Depending on trust, a binary relationship [2], which denotes belief-disbelief decision, is established between two interacting nodes. One node's trust evaluation is not restricted to direct expectations. It is extended to the combined effect of direct and indirect trusts under timing constraints and relevant contexts for more precision. Direct trust is evaluated by direct monitoring and indirect trust is considered as recommendations and notifications. Considering all current and old values of direct and indirect trusts in accordance to their contributing factors, final trust is evaluated for trustee at current time instant. Trust evaluating node is called current trustor (CT) and the node whose trust is evaluated by CT is called current trustee (TE). Using proposed TOR (Trusted On-demand Routing) model, final trust value is computed by CT, based on the context of MAC and/or routing layer delay, or on the context of using grater transmission range of communication. If TE is detected as rushing attacker by its CT, on the basis of computed trust value, CT does not consider malicious TE forwarded RREQ and discards it. As a consequence, rushing attacker is avoided and data delivery path become secured.

In the section 2, the status of the considered domain is presented and the rushing attack is discussed in section 3. In section 4, TOR model architecture is explained. In section 5, functionalities of CAT-AODV-R are discussed on the basis of trust computation of underlying TOR model. Simulation results of our experiments are presented in section 6. Section 7 includes the conclusion part.

2 Related Work

Trust becomes a popular approach to provide security in a distributed way for ad-hoc networks. Trust concept is proposed in different ways and in different aspects. In [7] , a trust monitoring architecture called TrAM (Trust Architecture for Monitoring), monitors trustworthiness of service users at run-time depending on trust rules and calculation mechanism for preventing occurrences of unwanted events. A system [4], based on path reputation and trust value, is proposed to enhances network throughput and reliability of discovered route. Here, trust value is incremented and decremented according to positive and negative observations, respectively. An integrated trust management model [8] is introduced to select trustworthy service by quantification of trust on platform of context-aware service. This model, addresses a basic set of trust aspects related to identity provisioning, privacy enforcement, and context provisioning activities.

Using trust concept AODV is secured by many proposed approaches. Trust-based SAODV [3] is proposed on the basis of intrusion detection mechanism (IDM) and trust-based mechanism (TBM) to penalize selfish nodes in AODV. In [1], a modified version of AODV is proposed on the basis of node trust and route trust to secure AODV. This work supports continuous node performance evaluation and neighbor node's recommendations collection. In [6], implicit trust relations between nodes are used to differentiate between trustworthy nodes from malicious nodes in AODV. Here, implicit trust relations of AODV are formalized. Based on these relations, each node is able to reason out the actions performed by its neighbors and deduces information about their knowledge. Finally, deduced information is used to supervise the behavior of neighbor and to detect malicious nodes. In [5], a trust-based framework is proposed for improving security and robustness of AODV. This mechanism is based on incentives and penalties depending on the behavior of network nodes.

3 Rushing Attack in AODV

Rushing attacker sends RREQ quickly to target node in comparison to other legitimate sender node. AODV protocol considers this first arriving rushed RREQ and discards late arriving legitimate RREQs. Rushing attackers send RREQ more quickly by ignoring MAC and/or routing layer delay, or by using higher transmission range.

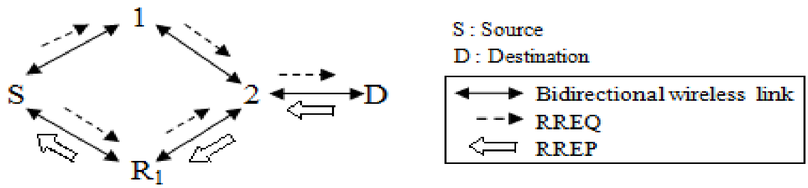


Fig. 1. Rushing attacker ignores routing or MAC layer delays

On-demand routing protocols are imposed by MAC and/or routing layer delays for collision avoidance among the routing packets. Rushing attacker ignores these delays to achieve faster transmission of RREQ. In Fig. 1, node R_1 is rushing attacker which sends RREQ more quickly to target node 2, in comparison to node 1, by ignoring delays. As a result, node 2 considers first received R_1 forwarded rushed RREQ for route discovery and discards node 1 forwarded and late arriving legitimate RREQ.

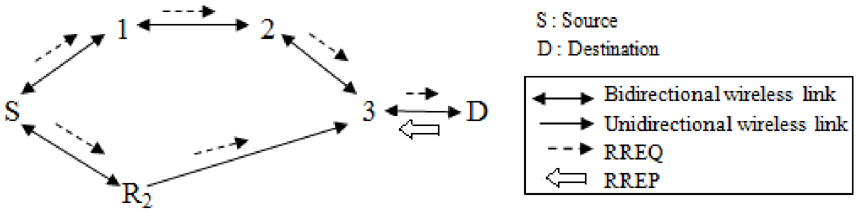


Fig. 2. Rushing attacker using higher transmission range

Another type of rushing attacker sends rushed RREQ to target node using higher transmission range. In Fig. 2, node R_2 (rushing attacker) is using higher transmission range in comparison to node 1, for sending RREQ more quickly to target node 3. Node 3 considers R_2 forwarded rushed RREQ for route discovery and discards lately received legitimate RREQ that reached form node 1 via node 2. But to establish the routing path, when RREP packet from node D will reach node 3, it can't be forwarded to R_2 since of shorter transmission range of node 3. As a consequence, node S will not get RREP packet and no route will be discovered between S and D.

4 Trusted On-Demand Routing (TOR) Model

In this work, AODV is extended with TOR model to establish secure routing path between source and destination by avoiding malicious nodes. The following section describes the structural and functional components of the model. TOR model (Fig. 3) consists of three functional modules (Node Manager, Trust Module and Decision Manager) along with the on-demand routing protocol, AODV.

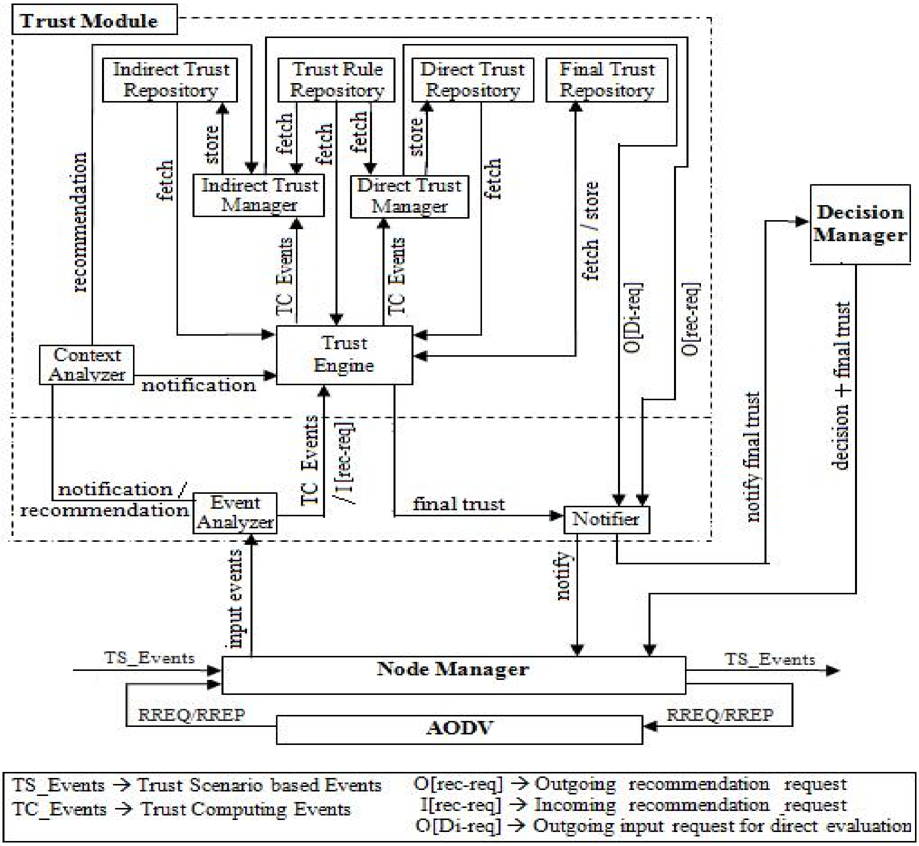


Fig. 3. TOR Model

4.1 Node Manager

On receipt of AODV specified RREQ, RREP and TSB_Events, CT's Node Manager sends TC_Events to Trust Module for computing trust value for TE. On the other hand, based on received decision (based on trust value) from Decision Manager, Node Manager either considers TE in the route discovery (case of belief decision) or avoids TE (case of disbelief decision). It also broadcasts the computed final trust value as notification to other nodes. When a node receives this notification, it sends the value to Trust Module for storing in Final Trust Repository. When CT requires recommendation about TE, Node Manager of recommender sends recommendation to CT's Node Manager.

4.2 Trust Module

Trust Module of TOR model is responsible for trust value computation of TE. Trust Module consists of Trust Engine, Direct Trust Manager and Indirect Trust Manager for computing different levels of trust values which are stored in respective repositories. Trust Module sends the computed final trust to decision Manager for taking belief-disbelief decision. Context Analyzer of Trust Module analyzes the contexts of incoming recommendations and trust notifications. Trust Module also has Event Analyzer for analyzing input events and Notifier for notifying output events. On the other hand, Trust Module sends recommendation from the Final Trust Repository and also stores the incoming notified final trust value for a TE.

4.3 Decision Manager

On the basis of received trust value from Trust Module, Decision Manager takes either belief or disbelief decision for TE. If the computed trust value is greater than 0.5, it takes belief decision otherwise it takes disbelief decision (Fig. 4). It sends this decision to Node Manager for considering or not considering TE in route discovery. It also forwards the received final trust value to Node Manager for notifying other nodes in the network.

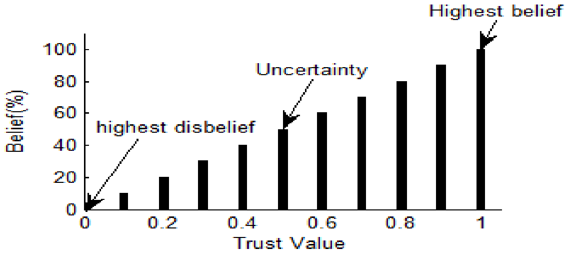


Fig. 4. Belief-disbelief graph

5 CAT-AODV-R: Context-Aware Trusted AODV against Rushing Attack

In CAT-AODV-R, context (C) is classified into context-1 and context-2, and on these considered contexts two types of misbehaviors (misbehavior-1, misbehavior-2) of rushing attacker is defined.

Context-1 (C1): C1 is defined with respect to ignorance of routing layer delay and/or MAC layer delay.

Context-2 (C2): C2 is defined with respect to usage of higher transmission range of a node in comparison to other nodes of the network.

Misbehavior-1 (M1): Based on C1, if a particular node sends rushed RREQ packets by ignoring MAC and/or routing layer delays, this behavior is considered as M1.

Misbehavior-2 (M2): Based on C2, if a particular node sends rushed RREQ using higher transmission range, this behavior is considered as M2.

In CAT-AODV-R, every node broadcasts RQres packet (response packet of RREQ). In Fig. 5, when TE receives RREQ, it broadcasts RQres and after receiving it, CT considers RQres receiving time for direct evaluation of TE. After necessary processing TE broadcasts RREQ and after receiving it, CT considers RREQ receiving time for direct evaluation. Next, CT broadcasts RQres in response of received RREQ. Against RQres, TS_Events are collected for direct trust evaluation, and also recommendations are collected for indirect trust evaluation.

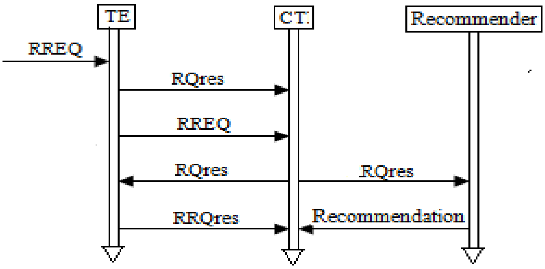


Fig. 5. Packet transfer sequence diagram

CAT-AODV-R deals with following set of symbols:

- T_{Packet}: (time taken for packet transmission and reception) + (packet travel time) + (MAC and routing layer delays) + (queuing time at receiver node). Here, Packet stands for RREQ / RQres / RRQres / recommendation packet.
- T_{Pi} : Processing time, where i = 1, 2,…. different levels of processing.
- $T_{Const\text{-}j}$: Different constant times for network. Where j = 1,2,… .

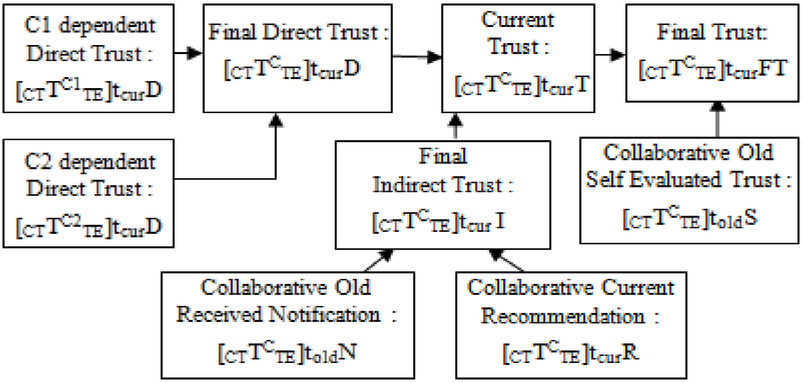

C1 dependent Direct Trust : $[{}_{CT}T^{C1}{}_{TE}]t_{cur}D$
Final Direct Trust : $[{}_{CT}T^{C}{}_{TE}]t_{cur}D$
Current Trust : $[{}_{CT}T^{C}{}_{TE}]t_{cur}T$
Final Trust: $[{}_{CT}T^{C}{}_{TE}]t_{cur}FT$
C2 dependent Direct Trust : $[{}_{CT}T^{C2}{}_{TE}]t_{cur}D$
Final Indirect Trust : $[{}_{CT}T^{C}{}_{TE}]t_{cur}I$
Collaborative Old Self Evaluated Trust : $[{}_{CT}T^{C}{}_{TE}]t_{old}S$
Collaborative Old Received Notification : $[{}_{CT}T^{C}{}_{TE}]t_{old}N$
Collaborative Current Recommendation : $[{}_{CT}T^{C}{}_{TE}]t_{cur}R$

Fig. 6. Trust Chain

In this work, CT evaluates final trust of TE depending on different level of trusts. Inter-dependencies among trust levels are shown in Fig. 6. In symbolic representation $[_{X}T_{Y}]$, T denotes trust of X on Y. Here C1, C2 and C are the contexts of computation; and t_{cur} or t_{old} are time instants at which respective trust values are computed.

CAT-AODV-R computes trust in phases. These are concerned with initiation, computation, decision and reaction phases. Phase-1 is for initiating Direct and Indirect Trust Manager to compute final direct trust and final indirect trust respectively. Considering phase-2, phase-3 and phase-4, final direct and indirect trust is computed in phase-5. In phase-6, current trust is computed by Trust Engine depending on final direct and indirect trust. Based on current trust and collaborative old self evaluated final trust, Trust Engine computes the final trust in this phase. Finally, on the basis of computed final trust, belief or disbelief decision is taken for TE in phase-7. Based on belief-disbelief decision, CAT-AODV-R avoids rushing attacker, and secures the route discovery.

5.1 Phase-1: Initiating Direct and Indirect Trust Manager

After receiving RREQ, TE broadcasts RQres and CT receives this RQres at time T_1(TE). Next, after necessary processing, TE broadcast the received RREQ and CT receives it at time T_2(TE). CT's Node Manager sends T_1(TE) and T_2(TE) as TC_Events to Trust Engine via Event Analyzer. Then Trust Engine initiates Direct Trust Manager by sending T_1(TE) and T_2(TE), for computing $[_{CT}T^{C1}{}_{TE}]t_{cur}D$. On the other hand, Trust Engine fetches stored notified trust values from Final Trust Repository and initiates Indirect Trust Manager by sending these fetched values for computing $[_{CT}T^{C}{}_{TE}]t_{old}N$.

5.2 Phase-2: $[_{CT}T^{C1}{}_{TE}]t_{cur}D$ and $[_{CT}T^{C}{}_{TE}]t_{old}N$ Computation

After receiving initial TC_Events (T_1(TE) and T_2(TE)) from Trust Engine, CT's Direct Trust Manager calculates T(TE). Here, TE = (T_2(TE) – T_1(TE)). Direct Trust Manager compares T(TE) with specific threshold time α. Here, $\alpha = (T_{RREQ} + T_{P1} + T_{Const\text{-}1})$ where, T_{P1} and $T_{Const\text{-}1}$ are constant for the network. For α, the time T_{RREQ} is the specified standard time, concerned with legitimate TE forwarded legitimate RREQ to CT. If T(TE) < α, Misbehavior M1 of TE is identified based on context C1. On the basis of M1 identification, CT's Direct Trust Manager assigns $[_{CT}T^{C1}{}_{TE}]t_{cur}D$ for TE as per equation (1).

$$f_1\{T(TE)\} = \begin{cases} For\ T(TE) < \alpha,\ [\ _{CT}T^{C1}_{TE}\ ]t_{cur}D = 0.1; & (disbelief) \\ ForT(TE) \geq \alpha, [\ _{CT}T^{C1}_{TE}\ ]t_{cur}D = 0.9; & (belief) \end{cases} \tag{1}$$

Lemma 1. For T(TE) < α, $[_{CT}T^{C1}{}_{TE}]t_{cur}D = 0.1$.

Proof: Here, T(TE) = (T_2(TE) – T_1(TE)) = $(T_{RREQ} + T_{P1}+ T_{Const\text{-}1})$. For T(TE), the time T_{RREQ} is concerned with TE forwarded RREQ to CT. If T(TE) < α, it implies that

T_{RREQ} is taking less time than standard time, specified for legitimate TE forwarded legitimate RREQ to CT, since T_{P1} and $T_{Const-1}$ are constant for the considered network. It implies that, TE is forwarding rushed RREQ to CT ignoring MAC and/or routing layer delays. As a consequence, CT's Direct Trust Manager identifies misbehavior-1 of TE on the basis of context-1 and therefore $[_{CT}T^{C1}{}_{TE}]t_{cur}D = 0.1$ as the case of disbelief. □

On the other hand, after receiving stored notified trust values ($[Ti^{C}{}_{TE}]t_{old}$, where i=1,2…n; n=total numbers of old notified trust values) from Trust Engine, CT's Indirect Trust Manager computes $[_{CT}T^{C}{}_{TE}]t_{old}N$ for TE as per equation (2). Here $e^{-(t_{old}-t_{0})}$ is time decaying function, where t_0 is initial time and t_{old} is the old specified time at which received notified trust value is computed.

$$[\,_{CT}T^{C}_{TE}\,]t_{old}N = \frac{1}{n}\times\{\sum_{i=1}^{n}[Ti^{C}_{TE}\,]t_{old}\times e^{-(t_{old}-t_{o})}\} \tag{2}$$

Next, Indirect Trust Manager and Direct Trust Manager send recommendations request (O[rec-req]) and request of input for direct evaluation (O[Di-req]) respectively to Node Manager via Notifier. Against O[rec-req] and O[Di-req], CT's Node Manager broadcasts RQres packet after initializing time to zero. Against RQres, CT not only collects recommendations from recommenders for indirect evaluation, but also collects response packet of RQres i.e. RRQres from TE for direct evaluation.

5.3 Phase-3: 2nd Time Initiation of Direct and Indirect Trust Manager

After broadcasting RQres at time instant zero, CT waits for RRQres till time (T_{RQres} + T_{P2} + T_{RRQres} + $T_{Const-2}$). Here, T_{RQres} is concerned with CT forwarded RQres to TE, and TR_{RRQres} is concerned with TE forwarded RRQres to CT. T_{P2} is the time taken for processing at TE after receiving RQres from CT and before sending RRQres to CT. In response to RQres, if CT does not get back RRQres from TE within specified time, CT's Node Manager sends Nack[RRQres] as TC_Events to Direct Trust Manager via Event Analyzer and Trust Engine, otherwise sends Ack[RRQres], for computing $[_{CT}T^{C2}{}_{TE}]t_{cur}D$.

As well as, in response to RQres, CT gets recommendations about TE from Recommenders. CT's Node Manager sends these recommendations as TC_Events to Context Analyzer via Event Analyzer. If the contexts of incoming recommendations are in valid context set C (set of C1 and C2), Context Analyzer sends these recommendations to Indirect Trust Manager, for computing $[_{CT}T^{C}{}_{TE}]t_{cur}R$.

5.4 Phase-4: $[_{CT}T^{C2}{}_{TE}]t_{cur}D$ and $[_{CT}T^{C}{}_{TE}]t_{cur}R$ Computation

If CT's Direct Trust Manager receives Nack[RRQres], it understands that CT did not receive RRQres in response of RQres, within specified time, since of CT's smaller transmission range. In this case, CT's Direct Trust Manager identifies Misbehavior M2 of TE on the basis of context C2, and it disbelieves TE. Based on M2 identification, CT's Direct Trust Manager assigns the value of $[_{CT}T^{C2}{}_{TE}]t_{cur}D$ for TE as per equation (3).

$$f_2\{\ Nack[RRQres],Ack[RRQres]\ \} =$$
$$\begin{cases} For\ Nack[RRQres], & [\ {}_{CT}T_{TE}^{C2}\ ]t_{cur}D = 0.1\,; \quad (disbelief\,) \\ ForAck[RRQres], & [\ {}_{CT}T_{TE}^{C2}\ ]t_{cur}D = 0.9\,; \quad (belief\,) \end{cases} \tag{3}$$

Lemma 2. For Nack[RRQres], $[_{CT}T^{C2}{}_{TE}]t_{cur}D = 0.1$.

Proof: If, Direct Trust Manager receives Nack[RRQres], it implies that CT did not receive RRQres from TE within specified time ($T_{RQres} + T_{P2} + T_{RRQres} + T_{Const-2}$). In this case, TE forwarded packet reaches CT (CT receives TE forwarded RREQ) but CT forwarded packet could not reach TE because of CT's smaller transmission range than TE. That means TE is forwarding rushed RREQ packet to CT using higher transmission range. As a consequence, CT's Direct Trust Manager identifies misbehavior-2 of TE on the basis of context-2 and therefore $[_{CT}T^{C2}{}_{TE}]t_{cur}D = 0.1$ as the case of disbelief. □

On the other hand, based on received recommendations ($R[_{TRi}T^{C}{}_{TE}]t$, where TRi is ith recommender and i=1,2,…n; n=total number of recommendations), Indirect Trust Manager computes $[_{CT}T^{C}{}_{TE}]t_{cur}R$ for TE as per equation (4). Here, $[_{CT}T^{C}{}_{TRi}]t$ is CT's trust for TRi , and $e^{-(t-t_0)}$ is time decaying function. The time: t is computing time instant.

$$[\ {}_{CT}T_{TE}^{C}\ ]t_{cur}R = \frac{\sum_{i=1}^{n}\{R[_{TRi}\,T_{TE}^{C}]t\times e^{-(t-t_o)}\}\times\{[\ {}_{CT}T_{TRi}^{C}]t\times e^{-(t-t_0)}\}}{\sum_{i=1}^{n}\{[\ {}_{CT}T_{TRi}^{C}]t\times e^{-(t-t_0)}\}} \tag{4}$$

5.5 Phase-5: $[_{CT}T^{C}{}_{TE}]t_{cur}D$ and $[_{CT}T^{C}{}_{TE}]t_{cur}I$ Computation

CT's Direct Trust Manager computes $[_{CT}T^{C}{}_{TE}]t_{cur}D$ for TE as per equation (5), and stores it in Direct Trust Repository. Then it sends storing acknowledgement of $[_{CT}T^{C}{}_{TE}]t_{cur}D$ to Trust Engine.

$$[\ {}_{CT}T_{TE}^{C}\ ]t_{cur}D = \{\ W_1 \times [\ {}_{CT}T_{TE}^{C1}\ ]t_{cur}D\ \} + \{\ (1-W_1)\times[\ {}_{CT}T_{TE}^{C2}\ ]t_{cur}D\ \} \tag{5}$$

If misbehavior M1 is identified in context C1 and misbehavior M2 in context C2 is not identified, value of W1 is 0.9. If misbehavior M2 in context C2 is identified and misbehavior M1 in context of C1 is not identified, value of W_1 is 0.1. On the other hand, if both M1 and M2 are identified or both are not identified in contexts C1 and C2 respectively, W1 is of value 0.5.

On the other hand, Indirect Trust Manager computes $[_{CT}T^{C}{}_{TE}]t_{cur}I$ for TE as per equation (6), and store it in Indirect Trust Repository. Then it sends the storing acknowledgement of $[_{CT}T^{C}{}_{TE}]t_{cur}I$ to Trust Engine.

$$[\,_{CT}T^{C}_{TE}\,]t_{cur}I = \{0.5\times[\,_{CN}T^{C}_{TE}\,]t_{cur}R\} + \{0.5\times[\,_{CN}T^{C}_{TE}\,]t_{old}N\} \tag{6}$$

5.6 Phase-6: $[_{CT}T^{C}{}_{TE}]t_{cur}T$, $[_{CT}T^{C}{}_{TE}]t_{old}S$ and $[_{CT}T^{C}{}_{TE}]t_{cur}FT$ Computation

After getting storage acknowledgement of Direct trust and Indirect trust into their respective repository, CT's Trust Engine fetches $[_{CT}T^{C}{}_{TE}]t_{cur}D$ and $[_{CT}T^{C}{}_{TE}]t_{cur}I$ from Direct Trust Repository and Indirect Trust Repository respectively. Then Trust Engine computes $[_{CT}T^{C}{}_{TE}]t_{cur}T$ for TE with the help of retrieved $[_{CT}T^{C}{}_{TE}]t_{cur}D$ and $[_{CT}T^{C}{}_{TE}]t_{cur}I$ as per equation (7).

$$[\,_{CT}T^{C}_{TE}\,]t_{cur}T = 0.8\times[\,_{CT}T^{C}_{TE}\,]t_{cur}D\} + \{0.2\times[\,_{CT}T^{C}_{TE}\,]t_{cur}I\} \tag{7}$$

Then Trust Engine retrieves old self evaluated final trust values($[_{CT}Ti^{C}{}_{TE}]t_{old}FT$, where i=1,2,…n ; n=total numbers of old self computed values) from Final Trust Repository. Next, it computes $[_{CT}T^{C}{}_{TE}]t_{old}S$ for TE as per equation (8).

$$[\,_{CT}T^{C}_{TE}\,]t_{old}S = \frac{1}{n}\times\{\sum_{i=1}^{n}[\,_{CT}Ti^{C}_{TE}\,]t_{old}FT\times e^{-(t_{old}-t_{o})}\} \tag{8}$$

Finally, Trust Engine computes $[_{CT}T^{C}{}_{TE}]t_{cur}FT$ for TE, with the help of $[_{CT}T^{C}{}_{TE}]t_{cur}T$ and $[_{CT}T^{C}{}_{TE}]t_{old}S$ for TE as per equation (9).

$$[\,_{CT}T^{C}_{TE}\,]t_{cur}FT = \{\ 0.7\times[\,_{CT}T^{C}_{TE}\,]t_{cur}T\ \} + \{\ 0.3\times[\,_{CT}T^{C}_{TE}\,]t_{old}S\ \} \tag{9}$$

Trust Engine stores this Final $[_{CT}T^{C}{}_{TE}]t_{cur}FT$ in Final Trust Repository, and sends it to Decision Manager for taking belief-disbelief decision.

5.7 Phase-7: Decision and Reaction

If CT's Decision Manager receives $[_{CT}T^{C}{}_{TE}]t_{cur}FT$, having value greater than 0.5, it takes belief decision for TE, otherwise it takes disbelief decision. Decision Manager sends the final trust value and decision to Node Manager. If CT's Node Manager receives belief decision, it broadcasts TE forwarded RREQ, and if it receives disbelief decision, it avoids TE by discarding the TE forwarded RREQ. Finally, CT's Node Manager notifies final trust value.

If CT believes TE and broadcasts TE forwarded RREQ, CT appends the IP address of TE in RREQ. If CT discards a TE forwarded RREQ, CT stores a tag which indicates TE as malicious RREQ sender. Next, when a RREQ of same route discovery reaches to that CT with appended IP address of CT evaluated malicious TE, CT discard that RREQ immediately.

When destination node receives RREQ, it evaluates the trust of the RREQ sender node (destination is CT and RREQ sender node is TE) by the same process.

6 Simulation Result

We conducted simulation experiments to evaluate the performance of the proposed routing algorithm CAT-AODV-R, in presence of rushing attacker which behaves according to Misbehavior-1 (M1) or Misbehavior-2 (M2). Our CAT-AODV-R is also compared with existing routing algorithm, AODV. The traffic type is CBR. Here considered network is over a 1000m×1000m terrain. Fig. 7 shows that CAT-AODV-R detects rushing attacker efficiently as the detection rate is efficient, with respect to network of 100 nodes. Detection rate denotes the rate of detection of rushing attacker among all rushing attackers present in network. Fig. 8 shows that legitimate RREQ success rate in CAT-AODV-R is much higher than AODV, with respect to network of 50 nodes. Legitimate RREQ success rate denotes the win of legitimate RREQ against rushed RREQ. Presented results are evaluated with 100 simulation runs.

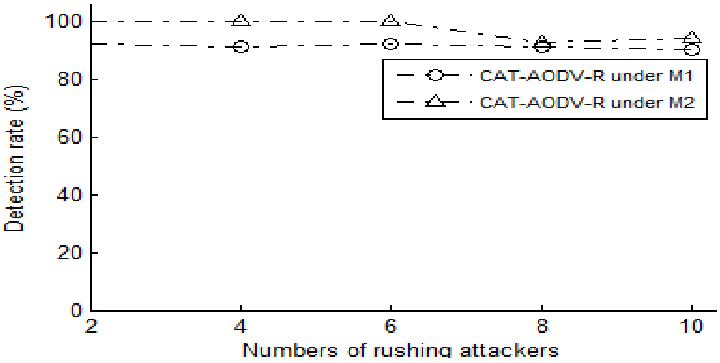


Fig. 7. Detection rate vs. numbers of rushing attackers

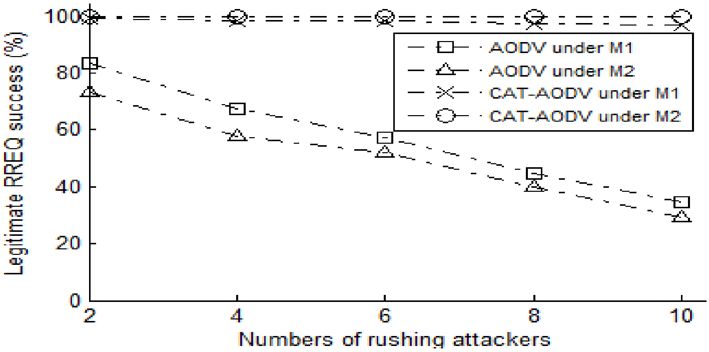


Fig. 8. Legitimate RREQ success vs. numbers of rushing attackers

7 Conclusion

CAT-AODV-R protocol for ad hoc network has been presented. We proposed a new solution using trust concept, against the rushing attack problem for existing on demand routing protocol AODV, in ad hoc networks. With the help of proposed trust model, all CAT-AODV-R supported nodes cooperate together to detect and avoid misbehavior-1 (M1) and/or misbehavior-2 (M2) behaving rushing attacker nodes in a more reliable fashion. Our detection-avoidance scheme detects the misbehaving rushing attacker nodes and isolates them from the active data forwarding and routing on the basis of belief-disbelief decision which comes from evaluated trust value. More research into this novel mechanism for secure routing is necessary. For further research, we are working on improving the proposed trust model, which may provide solutions against other attacks in ad-hoc network.

References

1. Menaka, A., Pushpa, M.E.: Menaka Pushpa, A.: Trust Based Secure Routing in AODV Routing Protocol. In: IMSAA, Bangalore, pp. 1–6 (2009)
2. Xiu, D., Liu, Z.: A Formal Definition for Trust in Distributed Systems. In: Zhou, J., López, J., Deng, R.H., Bao, F. (eds.) ISC 2005. LNCS, vol. 3650, pp. 482–489. Springer, Heidelberg (2005)
3. De Rango, F., Marano, S.: Trust-based SAODV Protocol with Intrusion Detection and Incentive Cooperation in MANET. In: IWCMC 2009, Leipzig, Germany (2009)
4. Li, J., Moh, T.-S., Moh, M.: Path-Based Reputation System for MANET Routing. In: van den Berg, H., Heijenk, G., Osipov, E., Staehle, D. (eds.) WWIC 2009. LNCS, vol. 5546, pp. 48–60. Springer, Heidelberg (2009)
5. Meka, K.D., Virendra, M., Upadhyaya, S.: Trust Based Routing Decisions in Mobile Adhoc Networks. CiteSeerX (2009)
6. Ayachi, M.A., Bidan, C., Abbes, T., Bouhoula, A.: Misbehavior Detection using Implicit Trust Relations in the AODV Routing Protocol. In: International Conference on Computational Science and Engineering (2009)
7. Uddin, M.G., Zulkernine, M.: A Trust Monitoring Architecture for Service Based Software. Springer Science+Business Media, LLC (2009)
8. Neisse, R., Wegdam, M., van Sinderen, M., Lenzini, G.: Trust Management Model and Architecture for Context-Aware Service Platforms. Springer, Heidelberg (2007)
9. Basagni, S., Conti, M., Giordano, S., Stojmenovic, I.: Mobile Ad Hoc Networking (ch. 10). A John Wiley & Sons, Inc.
10. Hu, Y., Perrig, A., Johnson, D.B.: Rushing Attacks and Defense in Wireless Ad Hoc Network Routing Protocol. In: WISE 2003, San Diego, California, USA, September 19 (2003)

SEPastry: Security Enhanced Pastry

Madhumita Mishra, Somanath Tripathy, and Sathya Peri

Indian Institute of Technology Patna, India
{madhumita,som,sathya}@iitp.ac.in

Abstract. Pastry is one of the most popular DHT overlay used in various distributed applications, because of its scalability, efficiency and reliability. On the other hand, Pastry is not resistant against the more generous attacks include Sybil attack, Eclipse attack etc. In this paper, we propose SEPastry (security enhanced pastry) to heighten the security features of Pastry without using any computational cryptographic primitives. SEPastry is found to be resistant against various forms of node-id attacks like Sybil attack, Eclipse attack, etc..

Keywords: Structured p2p, Pastry, Security, node-id attack.

1 Introduction

Structured Peer to Peer (P2P) systems are the distributed hash table (DHT), which provides an efficient decentralized look up facilities. P2P network services are characterized by features like high scalability and efficiency, well capable of handling random node failures. Mostly structured P2P is used in various distributed network applications. Sharing of file, audio, video, mails, document and electronic commerce are the widely used distributed application. Security concern rises with the increase of the connectivity and sharing.

In structured systems peers and keys of content objects are identified by using a set of IDs. DHT provides a self organizing substrate for large scale P2P applications. Structured overlays guarantee that the number of hops required to reach any node in the network is upper-bounded by O(logN) where N is the number of participating nodes [9]. Additionally there is the guarantee that a document if present in the network will definitely be reached. P2P network uses distributed hash table (DHT) to establish an association among peers and resources. Structured overlays allow applications to locate any object in a probabilistically bounded, small number of network hops, while requiring per-node routing tables with only a small number of entries. There are various structured peer to peer overlay architectures such as Chord, Pastry, CAN, etc. impose a specific linkage structure between nodes, Pastry posses huge potential to build self organizing applications to today′ s programmers. Pastry provides scalability with a low management overhead, reliability and are theoretically able to find data in O(log n) steps [1]. Applications like SCRIBE, PAST have been built on top of pastry because of its inherent advantages. SCRIBE [2] is used for group communication and event notification while PAST [11] is a peer to peer archival

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 789–795.
springerlink.com

storage utility implemented using Pastry. The popularity of pastry, in real world distributed application is rapidly increasing. However, to make it more acceptable, the security of Pastry needs to be heightened.

In a Pastry network a node is randomly assigned a nodeId, given a message and numeric key, the node routes the message to a nodeId numerically closest to the key. While routing a given message the node first checks the leafset. The leafset of node i contains the L/2 nodes numerically greater closest to node i and the L/2 nodes numerically smaller closest to node i. If the given key falls within it then the message reaches its destination in one hop else refers to the routing table. The nodeId of the network can be harnessed by attackers to induce malicious behaviour in the network. Tampering the nodeId in the Pastry networks gives rise to several security issues in the network. In this paper we propose a mechanism named Security Enhanced Pastry (SEPastry), to secure Pastry from the most generous attack called nodeId attack. At any instant of time a node attached to the internet may wish to join the existing the overlay network for obtaining services or can even leave the P2P network. The protocol design aims to mitigate the threat of joining of bogus node. The attractive feature of this scheme is that it disallows the non-registered nodes to join into the network without involving any computational complex cryptographic mechanisms.

The paper is organized as follows. The related work is elucidated in Section 2. The SEPastry protocol to secure the node Id and node joining process has been revealed in Section 3. Section 4 illustrates the security analysis of the protocol. A brief comparison with other protocols is presented in Section 5 and concluded in Section 6.

2 Related Work

Exploitation of nodeId in structured P2P network is possible in various ways.By taking advantage of this fact attackers induce hazardous impact on the working of network protocol. Effort to secure structured P2P network is the area of focus of several researchers. In [4], NodeId attacks are categorised into two types ID mapping attack and Sybil attack. ID mapping attack[3] is utilized to obtain a set of particular identifiers. User can choose its own identifier and can obtain a desired position in the overlay network. This eventually allows a malicious user to gain control over certain resources. In Sybil attack [6], a single malicious user creates multiple fake peer identities and pretends to be multiple, distinct physical nodes in a system.

Approach to detect and recover the structured overlays from identity attacks is attempted in [7]. Mechanism proposed is based on the reliable performance of nodes in the presence of malicious peers. Periodically nodes construct and disseminate existence proofs for each name space regions to the set of proof managers for that region. A node queries the proof manager , successful replies provides indisputable evidence of an attempted identity attack.The mechanism proposed by them is based on the reliable performance of nodes in the presence of malicious peers. After detecting and identifying the malicious node , the set of other nodes constantly avoid them when routing the KBR requests.

An admission control system(ACS) for structured P2P systems is given in [4]. The system constructs a tree-like hierarchy of cooperative admission control nodes, from which a joining node has to gain admission via client puzzles. ACS defends against Sybil attacks by adaptively constructing a hierarchy of cooperative admission control nodes. A node wishing to join the network is serially challenged by the nodes from a leaf to the root of the hierarchy. Nodes completing the puzzles of all nodes in the chain are provided a cryptographic proof of the examined identity. Borisov proposes to add computational challenges to Chord in order to defend against Sybil attacks. Castro et al. suggest using a set of trusted certification authorities to produce signed certificates that bind a random node identifier to a public key and node's IP address. According to them inclusion of IP address in the certificate makes it difficult for an attacker to swap certificates between nodes it controls and also allows optimization based on minimizing communication delays.This mechanism works well with DHTs such as Chord, Pastry and Tapestry, where the identifiers are fixed. Distributed registration procedure is proposed in [5] for Chord. According to this system, each virtual node registers r registration nodes in the Chord ring. The r registration nodes are computed using the hash of the IP address and an integer j $(1 < j < r)$ .Registration nodes maintain a list of registered virtual nodes for each IP address and reject registration if the number of registered nodes for each IP address exceeds a system wide constant a. This approach provides a reasonable level of protection by regulating the number of identities that a malicious IP address can get. Wang et al. proposed a concept called net-print to build a secure DHTs.Net-print of a node is built using a node's default router IP address, its MAC address and a vector of RTT measurements between the node and a set of designated landmarks. According to them physical network characteristics can be used to identify nodes. The proposed mechanism attempts to make identity theft difficult. Bazzi and Konjevod proposed a defence mechanism based on physical network characteristics. The proposed mechanism aims to identify individual nodes and guarantee that identities in different groups are not controlled by the same entity. Informant protocol proposed in [10] based on game theory principles to detect rather than prevent Sybil attack. SEPastry has been designed with an approach to completely delimit the fact of existence or joining of malicious node with two simple operational phases.

3 SEPastry: The Proposed Mechanism

The Security Enhanced Pastry (SEPastry) comprises of the following two operational phases: Registration Phase and Joining Phase. A node willing to (access/) provide the services to (/from) p2p network, needs to register with a centralized server called registration server (RS), before it joins. Thus the non-registered nodes can be prevented easily, to participate in the networking operation during the joining phase. Note that if a node leaves from the network, it informs to RS, which makes the RS to exclude that node from the potential leafset in future.

Registration Phase: Each node X executes the registration phase before joining/ accessing to the network services. X sends the registration request comprising of IP address to the Registration Server (RS). RS generates a random number and assigned it to nodeId of X (ID_X). RS replies with the node-id (ID_X) to X in a secure way. Further, RS sends the (ID_X) and corresponding IP address (IP_X), to the potential leafset through a secure channel. Figure 1 illustrates the phase of registration.

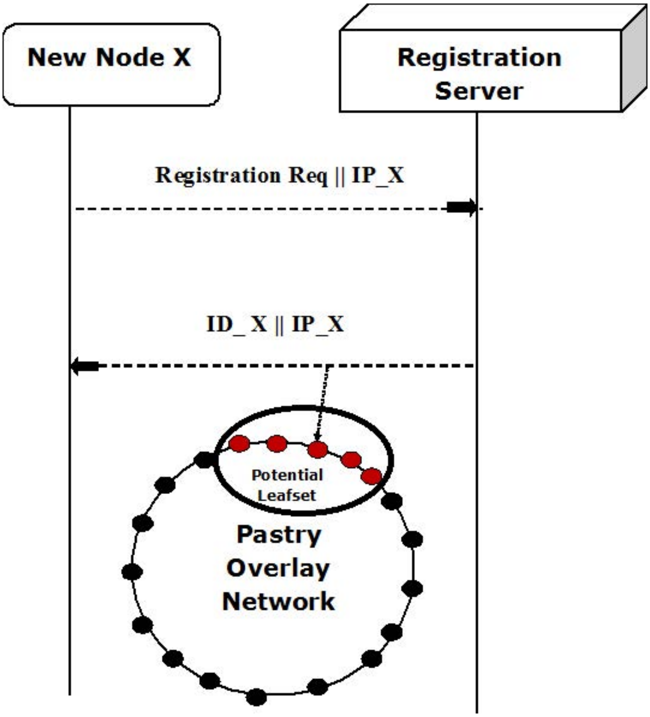


Fig. 1. Registration Phase

Joining Phase: The new node X willing to join, is assumed to know about node A on the basis of proximity metric.Node X then asks A to route the joining request message which comprises of the nodeId (ID_X) and IP address (IP_X).

This request needs to be routed to the existing node Z whose Id is numerically equivalent to ID_X. Now the leafset NodeIds of node Z verifies the validity of the given parameter in the request and responds with a positive or negative acknowledgement message accordingly.The new node receives positive responses greater than or equal to threshold value(β) from the potential leafset Ids.In such a situation the nodes A, Z and all nodes encountered on the path from A to Z send their state tables information to node X. Otherwise the new node is refrained from initializing its state table.The joining phase is as shown in figure 2.

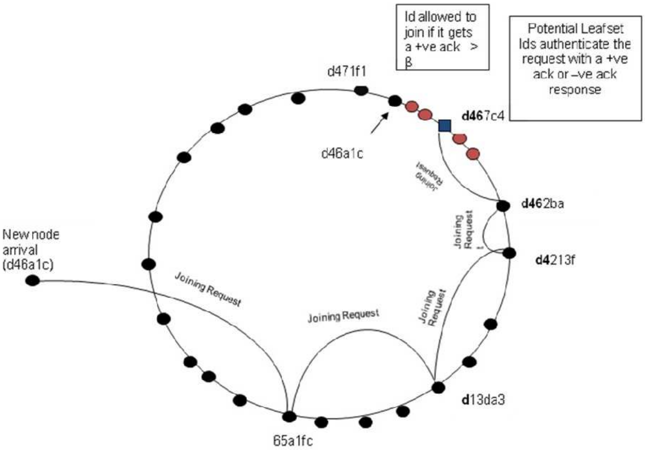

Id allowed to join if it gets a +ve ack > β
Potential Leafset Ids authenticate the request with a +ve ack or –ve ack response
d471f1
d467c4
d46a1c
Joining Request
d462ba
Joining Request
d4213f
New node arrival (d46a1c)
Joining Request
Joining Request
Joining Request
d13da3
65a1fc

Fig. 2. Joining Phase

4 Security Analysis of SEPastry

The level of defence offered by SEPastry against the various possible attacks in the structured P2P network is illustrated as follows:

Sybil Attack: In Pastry, nodeId is obtained as the hash digest of the node's IP address. A malicious user can simultaneously spoof many IP addresses to quickly obtain a multitude of identities.The registration phase in SEPastry restricts the possibility of creating multiple identities. As a result the proposed protocol provides high level of security against Sybil attack.

Eclipse Attack: Successful restriction of Sybil attack in a way reduces the possibility of Eclipse attack. However Eclipse attack is also possible in presence of defence against Sybil attack such as nodeId certificate solution [9].

A small set of malicious node with legitimate identities is sufficient to carry out Eclipse attack . The two phases of SEPastry provides strong defence against

Eclipse attack as no malicious node is allowed to place itself in between the nodes and reroute the message.

Message Forwarding Attack: In case of an honest node where all the entries are valid, the message is delivered to the root node for the key after an average of number of hops. There are two cases in this category of attack, presence of faulty node in the path or the root node may be faulty. Routing may fail in the presence of a faulty node along the path. Presence of faulty node along the path may simply drop the message, route the message to the wrong place of the node. As SEPastry restricts to join the illegitimate node into the network, the message forwarding like attacks are reduced. However, if an inside node becomes malicious, it requires detection mechanism to exclude the said node. This is beyond the scope of this work.

5 Discussion

In comparison to Bootstrap server mechanism this provides additional security as the joining node contacts a set of leafset Ids and waits for threshold response. Failure of one or two leafset Ids does not have an impact in the joining process. In contrast to identity based cryptography protocol, SEPastry does not involve the computational complexity of cryptographic technique. During the joining phase the cost incurred in sending request messages and getting response from Leafset nodeId is negligible. The process of securing the node joining procedure improves the overall routing performance of the network as it mitigates the threat of routing table poisoning. Overall scalability, reliability and efficiency of the network improves. SEPastry boost up the process of safe and sound node joining as compared to any other proposed protocol. The mechanism guarantees the process of assigning each peer with a unique nodeId. Strongly forbids attacks like Sybil attack and eclipse attack. SEPastry provides optimized flow control, load balancing and QoS routing. In this proposed mechanism, there is no issue of assigning certificate. Issuing certificate mechanism to authenticate the joining process is time consuming. SEPastry protocol secures the whole structure relatively in shorter time span. In Eigen trust algorithm to establish trust in the system it requires a large number peers to cooperate. In contrast SEPastry protocol requires only a few set of Ids to secure the joining procedure.

6 Conclusion

NodeId attacks have the potential of completely paralysing the whole network structure. This paper proposed SEPastry to enhance the security features of the existing Pastry without using any computational intensive cryptographic operations. SEPastry is found to be robust against Sybil attack, Eclispse attack, etc.

References

1. Rowstron, A., Druschel, P.: Pastry: Scalable, Decentralized Object Location, and Routing for Large-Scale Peer-to-Peer Systems. In: Guerraoui, R. (ed.) Middleware 2001. LNCS, vol. 2218, pp. 329–350. Springer, Heidelberg (2001)
2. Rowstron, A., Kermarrec, A.-M., Druschel, P., Castro, M.: Scribe: The design of a large-scale event notification infrastructure. In: Intl. Workshop on Networked Group Communication (NGC 2001) (June 2001)
3. Cerri, D., Ghioni, A., Paraboschi, S., Tiraboschi, S.: ID mapping attacks in P2P networks. In: IEEE Global Telecommunications Conference, GLOBECOM 2005, December 3 (2005)
4. Rowaihy, H., William, E., Patrick, M., Porta, T.L.: Limiting sybil attacks in structured peer-to-peer networks. Technical Report NAS-TR-0017-2005, Network and Security Research Center, Department of Computer Science and Engineering, Pennsylvania State University, University Park, PA, USA (2005)
5. Dinger, J., Hartentstein: Defending the Sybil Attack in P2P Networks: Taxonomy, Challenges,and a Proposal for Self-Registration. In: Proc. 1st International Conference on Availability, Reliability and Security, Vienna, Austria, pp. 756–763. IEEE Computer Society Press, Los Alamitos (2006)
6. Douceur, J.R.: The Sybil Attack. In: Druschel, P., Kaashoek, M.F., Rowstron, A. (eds.) IPTPS 2002. LNCS, vol. 2429, pp. 251–260. Springer, Heidelberg (2002)
7. Puttaswamy, K., Zheng, H., Zhao, B.: Securing structured overlays against identity attacks. IEEE Transactions on Paralleland Distributed Systems 2010, 1487–1498 (2009)
8. Aiello, L.M., Milanesio, M., Ruffo, G., Schifanella, R.: Tampering Kadmelia with a Robust Identity Based System. Computer science Department - Universit'a degli Studi di Torino, Italy
9. Castro, M., Druschel, P., Ganesh, A., Rowstron, A., Wallach, D.S.: Secure routing for structured peer-to-peer overlay networks. In: Proc. of the 5th Usenix Symposium on Operating Systems Design and Implementation, Boston, MA (December 2002)
10. Margolin, N.B., Levine, B.N.: Informant: Detecting Sybils Using Incentives. In: Dietrich, S., Dhamija, R. (eds.) FC 2007 and USEC 2007. LNCS, vol. 4886, pp. 192–207. Springer, Heidelberg (2007)
11. Druschel, P., Rowstron, A.: PAST: A large-scale, persistent peer-to-peer storage utility. In: Proc. HotOS VIII, Schloss Elmau, Germany (May 2001)

Hybrid Scenario Based Analysis of the Effect of Variable Node Speed on the Performance of DSDV and DSR

Koushik Majumder[1], Sudhabindu Ray[2], and Subir Kumar Sarkar[2]

[1] Department of Computer Science & Engineering, West Bengal University of Technology, Kolkata, India
koushik@ieee.org

[2] Department of Electronics and Telecommunication Engineering, Jadavpur University, Kolkata, India

Abstract. Routing in mobile ad hoc network is considered as a challenging task due to the drastic and unpredictable changes in the network topology resulting from the random and frequent movement of the nodes and due to the absence of any centralized control. Routing becomes even more complex in hybrid networking scenario where the MANET is combined with the fixed network for covering wider network area with less fixed infrastructure. Although, several routing protocols have been developed and tested under various network environments, but, the simulations of such routing protocols have not taken into account the hybrid networking environments. In this work we have carried out a systematic simulation based performance study of the two prominent routing protocols: Destination Sequenced Distance Vector Routing (DSDV) and Dynamic Source Routing (DSR) protocols in the hybrid networking environment under varying node speed. We have analyzed the performance differentials on the basis of three metrics – packet delivery fraction, average end-to-end delay and normalized routing load using NS2 based simulation.

Keywords: Mobile ad hoc network, hybrid network scenario, varying node speed, performance analysis, packet delivery fraction, average end-to-end delay, normalized routing load.

1 Introduction

A group of mobile devices can form a self-organized and self-controlled network called a mobile ad hoc network (MANET) [1-6]. The main advantage of these networks is that they do not rely on any established infrastructure or centralized server. These networks are autonomous where a number of mobile nodes equipped with wireless interfaces communicate with each other either directly or through other nodes. The communication is multi-hop and each node has to play the role of both the host as well as the router. But due to the limited transmission range of the MANET

N. Meghanathan et al. (Eds.): Advances in Computing & Inform. Technology, AISC 176, pp. 797–806.
springerlink.com

nodes, the total area of coverage is often limited. Also due to the lack of connectivity to the fixed network, the users in the MANET work as an isolated group. However, many applications require connection to the external network such as Internet or LAN to provide the users with external resources.

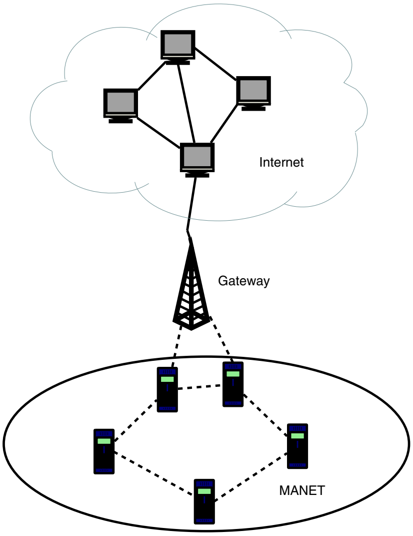


Fig. 1. Hybrid Network

Sometimes a hybrid network can be formed by combining the ad hoc network with the wired network. By using this combination we can cover a larger area with less fixed infrastructure, less number of fixed antennas and base station and can reduce the overall power consumption. Due to the hybrid nature of these networks, routing is considered a challenging task. Several routing protocols have been proposed and tested under various traffic conditions. However, the simulations of such routing protocols have not taken into account the hybrid network scenario. In this work we have carried out a systematic performance study of the two prominent routing protocols: Destination Sequenced Distance Vector Routing (DSDV) and Dynamic Source Routing (DSR) protocols in the hybrid networking environment under different node speed.

The rest of the paper is organized as follows. Section 2 describes the related work. Section 3 and section 4 details the simulation model and the key performance metrics respectively. The simulation results are presented and analyzed in section 5. Finally the conclusion has been summarized in the section 6. The last section gives the references.

2 Related Work

Several simulation based experiments have been made to compare the performance of the routing protocols for mobile ad hoc network.

Das et al. [7] made performance comparison of routing protocols for MANET based on the number of conversations per mobile node for a given traffic and mobility model. Small networks consisting of 30 nodes and medium networks consisting of 60 nodes were used. Simulation was done using the Maryland Routing Simulator (MARS).

Performance comparison results of two on demand routing protocols – AODV and DSR is presented in the work of Das, Perkins and Royer [8]. They used NS2 based simulation. CBR sources were used with packet size of 512 bytes. Two different simulation set ups were used. One with 50 nodes and 1500m x 300m simulation area and the other with 100 nodes and 2200m x 600m simulation area. The performance metrics studied were: packet delivery fraction, average end-to-end delay and normalized routing load.

Johansson, Larsssson, Hedman and Mielczarek [9] in their work incorporated new mobility models. A new mobility metric was introduced to characterize these models. Using this metric, mobility was measured in terms of relative speeds of the nodes instead of absolute speeds and pause times. The network consisted of 50 nodes. There were 15 sources and the data packets transmitted were of 64 bytes. Performance analysis was made in terms of throughput, delay and routing load.

Park and Corson [10] made a performance comparison between TORA and an “idealized’ link state routing protocol. Many simplifications were made in the simulation environment. For example, in the simulation scenario packets were transmitted at the rate of only 4, 1.5, or 0.6 packets per minute per node for avoiding congestion. Total duration of the simulation run was 2 hours. The network was connected in a “honeycomb” pattern. The node density was kept constant artificially. The notion of true node mobility was missing. Every node was connected to a fixed set of neighboring nodes through separate links. Each link switched between active and inactive states irrespective of other links. Immediate feedback was available when a link went up or down which is not the case in reality.

These works, however, do not take into consideration the influence of hybrid network scenario over the performance of the routing protocols. In this work we have studied the effect of varying node speed on the performance of two prominent routing protocols for mobile ad hoc network – Destination Sequenced Distance Vector Routing (DSDV) and Dynamic Source Routing (DSR) protocol in the hybrid networking environment.

3 Simulation Model

We have done our simulation based on ns-2.34 [11-14]. NS is a discrete event simulator. It was developed by the University of California at Berkeley and the

VINT project [11]. Our main goal was to measure the performance of the protocols under a range of varying network conditions. We have used the Distributed Coordination Function (DCF) of IEEE 802.11[15] for wireless LANs as the MAC layer protocol. Data packets were transmitted using an unslotted carrier sense multiple access (CSMA) technique with collision avoidance (CSMA/CA) [15].

The protocols have a send buffer of 64 packets. In order to prevent indefinite waiting for these data packets, the packets are dropped from the buffers when the waiting time exceeds 40 seconds. The interface queue has the capacity to hold 80 packets and it is maintained as a priority queue. We have generated the movement scenario files using the setdest program which comes with the NS-2 distribution. The total duration of our each simulation run is 900 seconds. We have varied our simulation with movement patterns for six different node speed: 5m/s, 10m/s, 15m/s, 20m/s, 25m/s, 30m/s. In our simulation environment the MANET nodes use constant bit rate (CBR) traffic sources when they send data to the wired domain. We have used two different communication patterns corresponding to 30 and 40 sources. The complete list of simulation parameters is shown in Table 1.

Table 1. Simulation Parameters

Parameter	Value
Protocols	DSDV, DSR
Number of mobile nodes	70
Number of fixed nodes	10
Number of sources	30,40
Transmission range	250 m
Simulation time	900 s
Topology size	900 m X 600 m
Source type	Constant bit rate
Packet rate	5 packets/sec
Packet size	512 bytes
Pause time	100 seconds
Node speed	5m/s, 10m/s, 15m/s, 20m/s, 25m/s, 30m/s
Mobility model	Random way point

3.1 Hybrid Scenario

We have used a rectangular simulation area of 900 m x 600 m. In our simulation we have used two ray ground propagation model. Our mixed scenario consists of a wireless and a wired domain. The simulation was performed with 70 wireless nodes and

10 wired nodes. For our hybrid networking environment we have a base station located at the centre (450,300) of the simulation area. The base station acts as a gateway between the wireless and wired domains. For our mixed simulation scenario we have turned on hierarchical routing in order to route packets between the wired and the wireless domains. The domains and clusters are defined by using the hierarchical topology structure. As the base station nodes act as gateways between the wired and wireless domains, they need to have their wired routing on. In the simulation setup we have done this by setting the node-config option–wiredRouting on. After the configuration of the base station, the wireless nodes are reconfigured by turning their wiredRouting off.

4 Performance Metrics

We have primarily selected the following three performance metrics in order to study the performance comparison of DSDV and DSR.

Packet delivery fraction: This is defined as the ratio between the number of delivered packets and those generated by the constant bit rate (CBR) traffic sources.

Average end-to-end delay: This is basically defined as the ratio between the summation of the time difference between the packet received time and the packet sent time and the summation of data packets received by all nodes.

Normalized routing load: This is defined as the number of routing packets transmitted per data packet delivered at the destination. Each hop-wise transmission of a routing packet is counted as one transmission.

5 Simulation Results and Analysis

In this section we have analyzed the effect of varying node speed on the performance of DSDV and DSR in the hybrid simulation scenario.

5.1 Packet Delivery Fraction (PDF) Comparison

From Fig. 2 we observe the difference in the packet delivery performances of DSDV and DSR from our simulation experiments. We have measured the packet delivery fraction of these two protocols by varying the node speed with respect to 30 and 40 numbers of sources. From the graphs we see that DSDV shows better packet delivery performance than DSR at lower node speed. This happens due to the fact that, at lower node speed, the network remains relatively stable and once a route is established, it continues to be available for a longer period of time. Due to the proactive nature of DSDV, routing information exchanges take place regularly between the nodes and each node maintains routing information to every destination all the time. Consequently, most of the packets can be delivered smoothly without having to wait for the path setup time. This results in better packet delivery performance of DSDV. On the contrary, DSR, being a source routing protocol, a significant time is required for initial path setup. During this time, no packets can be delivered to the destination due to unavailability of routes. This results in lower packet delivery fraction of DSR in comparison to DSDV.

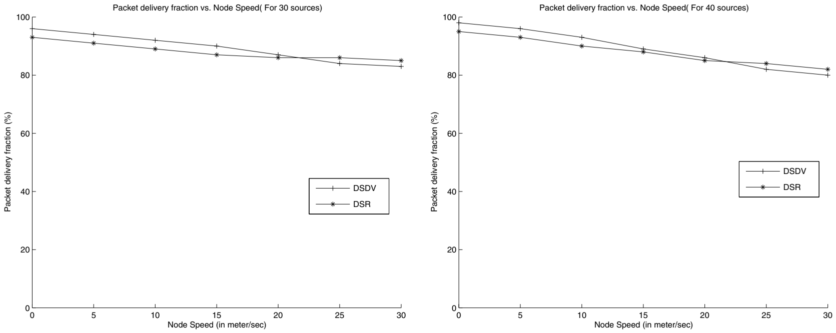


Fig. 2. Packet Delivery Fraction vs. Node Speed for 30 and 40 sources

With higher node speed, the network topology becomes highly dynamic and link breaks become more frequent. The unavailability of routes causes the nodes to show deterioration in the packet delivery performance for both DSDV and DSR. The periodic nature of operation of DSDV makes it less adaptive to these frequent changes. It requires greater number of full dumps to be exchanged between the nodes in order to maintain up-to-date routing information at the nodes. This huge volume of control traffic occupies a significant part of the channel bandwidth and lesser channel capacity remains available for the data traffic which results in reduced packet delivery fraction of DSDV at higher node speed.

DSR on the contrary, is more adaptive to the frequently changing scenario due to its on-demand nature of functioning. DSR maintains multiple routes in the cache. Thus, even if a link is broken due to higher node speed, alternative routes can be obtained from the cache. This reduces the number of dropped packets and results in better packet delivery performance of DSR.

5.2 Average End-to-End Delay Comparison

From Fig. 3 we can observe the fact that DSDV has less average end to end delay in comparison to DSR. DSDV is a proactive routing protocol. In DSDV, nodes periodically exchange routing tables between them in order to maintain up-to-date routing information to all destinations. Due to this regular route optimization, nodes have access to fresher and shorter routes to the destinations all the time. Hence, whenever a source node wants to send a packet to a destination node, with the already available routing information it can do so without wasting any time for path setup. This instant availability of fresher and shorter routes thus results in less average end-to-end delay in the delivery of data packets in case of DSDV.

DSR, on the contrary, is a reactive source routing protocol and routing information exchanges do not take place regularly. Instead, if a node in DSR wants to send a packet to a destination node, it has to first find the route to the destination in an on demand fashion. This route discovery latency is a part of the total delay. DSR being a source routing protocol, the initial path set up time is significantly higher as during the route discovery process, every intermediate node needs to extract the information before forwarding the data packet. Moreover in DSR, the source needs to wait for all the replies sent against every request reaching the destination. This increases the delay.

From the figures it is evident that the average end-to-end delay becomes more with higher node speed and greater number of sources for both the protocols. Frequent changes in the network topology due to increasing node speed results in greater number of link breaks. This together with the greater number of sources requires DSR to invoke the route discovery process more frequently in order to find new routes. The frequent invocation of the route discovery creates huge amount of control traffic. The data traffic to be delivered also becomes more with greater number of sources. This results in more collisions, further retransmissions and higher congestion in the network. Consequently, the route discovery latency increases due to the constrained channel. This in turn increases the average end-to-end delay. In addition to that, due to the higher priority of the control packets, the data packets need to spend more time in the queue waiting for the huge volume of control packets to be delivered. This also increases the end-to-end delay in delivering the data packets. In case of DSDV, due to higher speed of the nodes and frequent link breaks, routes become unavailable and nodes need to wait till the next routing information exchanges for new routes. Thus the delay increases depending upon the duration of the interval between the successive routing information exchanges.

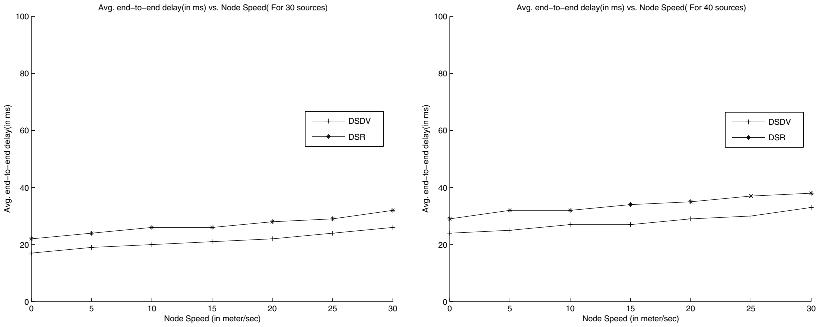


Fig. 3. Average End to End Delay vs. Node Speed for 30 and 40 Sources.

5.3 Normalized Routing Load Comparison

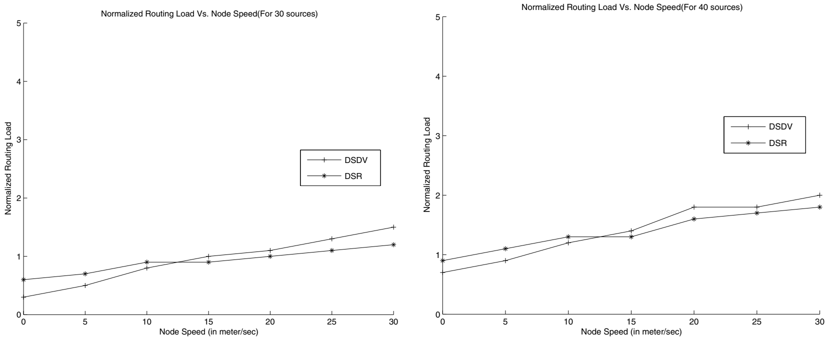


Fig. 4. Normalized Routing Load Vs. Node Speed for 30 and 40 Sources

From Fig. 4 we note that initially at lower node speed, DSR has greater normalized routing load. This is attributed to the fact that DSR being a source routing protocol, with every packet the entire routing information is embedded. In addition to that, in response to a route discovery, replies come from many intermediate nodes. This increases the total control traffic. In case of DSDV, initially, at lower node speed, the network topology remains relatively stable. Hence, nodes need to exchange only incremental dumps rather than full dumps. This results in lesser overhead of DSDV.

Both DSDV and DSR suffer from increased normalized routing load with higher node speed and greater number of sources. In case of DSR, with increasing node speed, the route discoveries need to be invoked more often due to increase in the number of broken links. Furthermore, as DSR does not use route optimization until the route is broken and continues using longer and older routes, the chances of link breaks also increase. This further adds to the number of route discoveries which ultimately results in huge control traffic and subsequently higher normalized routing load. Greater number of sources also causes frequent invocation of the route discovery which significantly increases the volume of control overhead. Higher volume of data and control traffic creates congestion in the network. This results in further collisions, more retransmissions and newer route discoveries and further adds up to the already increased control overhead which ultimately results in higher normalized routing load.

With higher node speed, the network topology experiences frequent and high volume of changes. DSDV, due to its proactive nature of operation, is less adaptive to this highly dynamic scenario. Therefore, nodes need to exchange full dumps in order to maintain up-to-date routing information. This causes greater routing overhead for DSDV. In comparison, DSR uses aggressive caching strategy and the hit ratio is quite high. As a consequence, in highly dynamic scenario, even if a link breaks, DSR can resort to an alternate link already available in the cache. Thus the route discovery process can be postponed until all the routes in the cache fail. This reduces the frequency of route discovery, which ultimately results in less routing overhead of DSR.

6 Conclusion

In this paper we have carried out a detailed ns2 based simulation to study and analyze the performance differentials of DSDV and DSR in the hybrid scenario under varying node speed with different number of sources. Our work is the first in an attempt to compare these protocols in hybrid networking environment. From the simulation results we see that at lower node speed, DSDV shows better packet delivery performance than DSR mainly due to the instant availability of fresher and newer routes all the time. On the other hand, with higher node speed, DSDV shows more deterioration in the packet delivery performance than DSR mainly due to its less adaptability to the highly dynamic network topology. DSR's better performance is attributed to its ability to maintain multiple routes per destination and its use of aggressive caching strategy. In terms of the average end-to-end delay, DSDV outperforms DSR. The poor performance of DSR in terms of average end-to-end delay is primarily due to its source routing nature and its inability to expire the stale routes. Both the approaches suffer form greater average end-to-end delay when we increase the speed of the nodes and the numbers of sources. At higher node speed we observe that DSR shows lower routing load in comparison to DSDV. DSR applies aggressive caching technique and maintains multiple routes to the same destination. Hence, in highly dynamic scenario, even if a link is unavailable due to link break, DSR can resort to an alternate link already available in the cache. This results in reduced frequency of route discovery which ultimately reduces the routing overhead of DSR. On the other hand, at lower node speed, the network topology remains relatively stable. Hence, in DSDV, nodes need to exchange only incremental dumps rather than full dumps. This results in lesser overhead of DSDV. Thus we can conclude that if routing delay is of little concern, then DSR shows better performance at higher mobility in terms of packet delivery fraction and normalized routing load in hybrid networking scenario. Under less stressful scenario, however, DSDV outperforms DSR in terms of all the three metrics.

References

1. Dow, C.R.: A Study of Recent Research Trends and Experimental Guidelines in Mobile Ad-Hoc Networks. In: Proceedings of 19th International Conference on Advanced Information Networking and Applications, vol. 1, pp. 72–77. IEEE (March 2005)
2. Freisleben, B., Jansen, R.: Analysis of Routing Protocols for Ad hoc Networks of Mobile Computers. In: Proceedings of the 15th IASTED International Conference on Applied Informatics, pp. 133–136. IASTED-ACTA Press, Innsbruck, Austria (1997)
3. Royer, E.M., Toh, C.K.: A Review of Current Routing Protocols for Ad hoc Mobile Wireless Networks. IEEE Personal Communications Magazine, 46–55 (April 1999)

4. Anastasi, G., Borgia, E., Conti, M., Gregori, E.: IEEE 802.11 Ad-hoc Networks: Protocols, Performance and Open Issues. Ad hoc Networking. IEEE Press Wiley, New York (2003)
5. Arun Kumar, B.R., Reddy, L.C., Hiremath, P.S.: A Survey of Mobile Ad hoc Network Routing Protocols. Journal of Intelligent System Research (January- (June 2008)
6. Rappaport, T.S.: Wireless Communications, Principles & Practices. Prentice-Hall (1996)
7. Das, S.R., Castaeda, R., Yan, J.: Simulation-based Performance Evaluation of Routing Protocols for Mobile Ad hoc Networks. Mobile Networks and Applications 5, 179–189 (2000)
8. Das, S.R., Perkins, C.E., Royer, E.M.: Performance Comparison of two On-demand Routing Protocols for Ad hoc Networks. In: Proceedings of the IEEE Conference on Computer Communications (INFOCOM), Tel Aviv, Israel, pp. 3–12 (March 2000)
9. Johansson, P., Larsson, T., Hedman, N., Mielczarek, B.: Routing Protocols for Mobile Adhoc Networks - A Comparative Performance Analysis. In: Proceedings of the 5th International Conference on Mobile Computing and Networking (ACM MOBICOM 1999) (August 1999)
10. Park, V.D., Corson, M.S.: A Performance Comparison of TORA and Ideal Link State Routing. In: Proceedings of IEEE Symposium on Computers and Communication 1998 (June 1998)
11. Fall, K., Vardhan, K. (eds.): Ns notes and documentation (1999), `http://www.mash.cd.berkeley.edu/ns/`
12. Network Simulator-2 (NS2), `http://www.isi.edu/nsnam/ns`
13. The CMU Monarch Project: The CMU Monarch Projects Wireless and Mobility Extensions to ns (1998), `http://www.monarch.cs.cmu.edu`
14. Altman, E., Jimenez, T.: NS Simulator for Beginners, Lecture notes. Univ. de Los Andes, Merida, Venezuela and ESSI. Sophia-Antipolis, France (2003)
15. IEEE Computer Society LAN MAN Standards Committee. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, IEEE Std 802.11-1997. The Institute of Electrical and Electronics Engineers, New York (1997)

Author Index

Abirami, K.V. 141
Abulaish, Muhammad 131
Aggarwal, Alok 151
Ali Khan, Md. Hashem 551
Anitha Kumari, K. 245
Atiq, Shaikh Mohammed 477

Bajpai, Durgesh 719
Balaji, N. 81
Bandyopadhyay, Soma 395
Banerjee, Arnab 765
Banerjee, Indrajit 597, 621
Banerjee, P.K. 765
Barnwal, Rajesh P. 641
Basha, M.S.S. 525
Bassi, Paras 299
Bhattacharyya, Abhijan 395
Bhattacharyya, Aniruddha 765
Bhattacharyya, Debika 765
Bhumireddy, Swathi 189
Bhutkar, Ganesh 317
Biswas, Suparna 103, 447
Biswas, Utpal 209
Boddukolu, Sarath Chandra 279
Borade, Deepali R. 575
Bose, Dipayan 765

Chaki, Rituparna 263, 273, 691
Chanak, Prasenjit 597, 621
Chandramohan, D. 81, 525
Chandrasekhar, C. 361
Chaudhari, Payal 651
Chauhan, Tarulata 239
Chetna, Bhandari 245
Chouksey, Ramanuj 71

Das, Ayan Kumar 691
Deshmukh, Achala 375
Dey, Priyanka 103
Dhavachelvan, P. 279, 525
Dhwaj, Kirti 543
Diwanji, Hiteishi 651
Dogmus, Zeynep 507
Doshi, Nishant 239
Dutta, Chaitali Biswas 209
Dutta, Kamlesh 11
Dutta, Subrata 671

Garg, Rachit 543
Geetha, G. 457
GnanaPrakasi, O.S. 405
Goel, Akhil 709
Goswami, Nishant 239
Goyal, Mayank Kumar 151, 299
Gupta, Ashish 385
Gupta, Paras 709
Gupta, Sachin 719

Haldar, Nur Al Hasan 131
Halder, Srimanta 597, 621
Hambir, Nitin 219
Hariyale, Hemant 729
Hazra, Swarnali 777

Indu Vadhani, S. 435
Ingle, Dayanand 477

Jabbehdari, Sam 61
Jayakumar, J. 497
Jaya Nirmala, S. 255

Jegatha, M. 661
Joshi, Ramesh Chandra 31
Joshila Grace, L.K. 307
Joy, Preetha Theresa 609
Julia, J. 489

Kadambi, Govind R. 575
Kakarla, Jagadeesh 289
Kant, Krishna 171
Karkhanis, Manjusha 375
Karunakar Reddy, D. 633
Kaur, Manpreet 757
Khapre, Shailesh 81
Khare, Nilay 339
Kumar, Amish 95
Kumar, Ravinder 719
Kushwaha, Dharmender Singh 709, 719, 729

Lakshmi, K. 141
Lee, Moon Ho 551

Mahajan, Premraj 43
Mahalingam, P.R. 1
Maheswari, V. 307
Majumder, Koushik 797
Mandal, Jyotsna Kumar 199, 739
Mandal, Shanta 263
Manvi, S.S. 515
Mary Saira Bhanu, S. 255
Matam, Rakesh 189, 587
Mazumdar, Monomita 597, 621
Meenakshi, M. 457, 489
Menaga, R.S. 141
Meshram, B.B. 477
Mishra, Ankit 729
Mishra, Gaurav 543
Mishra, Madhumita 789
Mishra, Shivani 171
Mishra, Vineet 159
Misra, Arun Kumar 227
Misra, Paurush Kumar 299
Mondal, Uttam Kr. 199
Mukherjee, Nandini 533, 671
Mukhopadhyay, Debajyoti 425

Nagamalai, Dhinaharan 307
Nagaradjane, Prabagarane 563
Nagaraj, S.V. 23
Nagesha 515
Naik, Bukke Devendra 279
Nandi, Subrata 159
Nandi, Utpal 739
Neogy, Sarmistha 447, 671
Nezakatolhoseini, Majid 61

Ojha, Gaurav 699

Pandey, Aakanksha 339
Pandey, Ankit 729
Park, Daechul 551
Pasha, Syed Asim 131
Patel, Ahtesham Akhtar 255
Patel, Helly 749
Patil, Shailaja 385, 467
Patil, Tulsidas 317
Peri, Sathya 789
Polouse Jacob, K. 609
Pourmina, Mohammad Ali 61
Prabhakar, S. 633
Priya, Lakshmi 457
Priya Zach, M. 497
Purohit, Neetesh 543

Rahimunnisa, K. 497
Rajni 685, 757
Ramesh Babu, S. 633
Ravichandran, Shruthi 563
Ravichandran, Sriram 563
Ray, Sudhabindu 797
Rhymend Uthariaraj, V. 141
Roy, Monideepa 671
Roy, Sarbani 533
Roy, Tumpa 11
Rubika, S. 245

Saha, Himadri Nath 765
Saha, Sujoy 159
Saira Banu, J. 55
Sakthi Ganesh, M. 369
Sambasivam, G. 81
Sampath, Amritha 417
Sandhya, C.H. 575
Sardana, Anjali 31, 43
Sarkar, Pratima 273
Sarkar, Subir Kumar 797
Sasirekha, M. 55
Sen, Sevil 507
Sengupta, Satadal 159
Setua, S.K. 777
Shanmugam, M. 81

Sharma, Rinki 575
Shyamala, R. 349
Sikri, Gaurav 685
Singh, Parminder 757
Singhai, Piyush 71
Sisodia, Poonam 11
Siva Kumar Reddy, B. 369
Siva Sathya, S. 289
Sivashankari, R. 71
Soniya, B. 87
Sreenivas, Banda 633
Srinivasan, Nithin 563
Srivastav, Ambrish 219
Srivastava, Amit Kumar 227
Sudha Sadasivam, G. 245
Sujatha, Pothula 279
Sumaiya Thaseen, I. 55
Suresh Kumar, S. 497

Tak, Gaurav Kumar 699
Tarapore, Noshir 317
Tasare, Suyog 425
Thakur, Arnab 641
Thapa, Chandra 361
Tiwari, Namita 95
Tripathy, Somanath 189, 587, 789
Tripti, C. 417

Upadhyay, Divye 11

Vaidyanathan, Sundarapandian 329
Valarmathi, J. 141
Valli, S. 349
Varalakshmi, P. 405
Vardhan, Manu 709, 719, 729
Varshney, Gaurav 31
Veeraiah, D. 81
Vengattaraman, T. 525
Venkadeshan, R. 661
Venkata Krishna, P. 369
Verma, Abhinav 709
Verma, Rohit 159
Verma, Yatendra Kumar 299
Vij, Sheetal 425
Vinayagasundaram, B. 435
Vithya, G. 435

Wankhade, Mahesh P. 749
Wilscy, M. 87

Yadav, R.S. 171
Youssef, Amr M. 113

Zahno, Roger 113
Zaveri, Mukesh 385, 467