ICNS 2012

The Eighth International Conference on Networking and Services

LMPCNA 2012

The Fourth International Workshop on Learning Methodologies and Platforms used in the Cisco Networking Academy

ISBN: 978-1-61208-186-1

March 25-20, 2012

St. Maarten, Netherlands Antilles

ICNS 2012 Editors

Toan Nguyen, INRIA - Grenoble - Rhone-Alpes, France

Reijo Savola, VTT Technical Research Centre of Finland, Finland

Vladimír Šulc, MICRORISC, Czech Republic

ICNS 2012

Foreword

The Eighth International Conference on Networking and Services (ICNS 2012), held between March 25-30, 2012 - St. Maarten, Netherlands Antilles, continued a series of events targeting general networking and services aspects in multi-technologies environments. The conference covered fundamentals on networking and services, and highlighted new challenging industrial and research topics. Ubiquitous services, next generation networks, inter-provider quality of service, GRID networks and services, and emergency services and disaster recovery were considered.

IPv6, the Next Generation of the Internet Protocol, has seen over the past three years tremendous activity related to its development, implementation and deployment. Its importance is unequivocally recognized by research organizations, businesses and governments worldwide. To maintain global competitiveness, governments are mandating, encouraging or actively supporting the adoption of IPv6 to prepare their respective economies for the future communication infrastructures. In the United States, government's plans to migrate to IPv6 has stimulated significant interest in the technology and accelerated the adoption process. Business organizations are also increasingly mindful of the IPv4 address space depletion and see within IPv6 a way to solve pressing technical problems. At the same time IPv6 technology continues to evolve beyond IPv4 capabilities. Communications equipment manufacturers and applications developers are actively integrating IPv6 in their products based on market demands.

IPv6 creates opportunities for new and more scalable IP based services while representing a fertile and growing area of research and technology innovation. The efforts of successful research projects, progressive service providers deploying IPv6 services and enterprises led to a significant body of knowledge and expertise.

With the growth of the Internet in size, speed and traffic volume, understanding the impact of underlying network resources and protocols on packet delivery and application performance has assumed a critical importance. Measurements and models explaining the variation and interdependence of delivery characteristics are crucial not only for efficient operation of networks and network diagnosis, but also for developing solutions for future networks.

Local and global scheduling and heavy resource sharing are main features carried by Grid networks. Grids offer a uniform interface to a distributed collection of heterogeneous computational, storage and network resources. Most current operational Grids are dedicated to a limited set of computationally and/or data intensive scientific problems.

Optical burst switching enables these features while offering the necessary network flexibility demanded by future Grid applications. Currently ongoing research and achievements refers to high performance and computability in Grid networks. However, the communication and computation mechanisms for Grid applications require further development, deployment and validation.

The conference has the following independents tracks:
ENCOT: Emerging Network Communications and Technologies

COMAN: Network Control and Management
SERVI: Multi-technology service deployment and assurance
NGNUS: Next Generation Networks and Ubiquitous Services
MPQSI: Multi Provider QoS/SLA Internetworking
GRIDNS: Grid Networks and Services
EDNA: Emergency Services and Disaster Recovery of Networks and Applications
IPv6DFI: Deploying the Future Infrastructure
IPDy: Internet Packet Dynamics
GOBS: GRID over Optical Burst Switching Networks

ICNS 2012 also included:

LMPCNA 2012: The Fourth International Workshop on Learning Methodologies and Platforms used in the Cisco Networking Academy

We welcomed technical papers presenting research and practical results, position papers addressing the pros and cons of specific proposals, such as those being discussed in the standard forums or in industry consortia, survey papers addressing the key problems and solutions on any of the above topics short papers on work in progress, and panel proposals.

We take here the opportunity to warmly thank all the members of the ICNS 2012 technical program committee as well as the numerous reviewers. The creation of such a broad and high quality conference program would not have been possible without their involvement. We also kindly thank all the authors that dedicated much of their time and efforts to contribute to ICNS 2012. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

We hope that ICNS 2012 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress in networking and services.

We are certain that the participants found the event useful and communications very open. The beautiful places of St. Maarten surely provided a pleasant environment during the conference and we hope you had a chance to visit the surroundings.

ICNS 2012 Chairs
Pedro Andrés Aranda Gutiérrez, Telefónica I+D - Madrid, Spain
Carlos Becker Westphall, Federal University of Santa Catarina, Brazil
Eugen Borcoci, University 'Politehnica' Bucharest, Romania
Jaime Lloret Mauri, Polytechnic University of Valencia, Spain
Sathiamoorthy Manoharan, University of Auckland, New Zealand
Yoshiaki Taniguchi, Osaka University, Japan
Toan Nguyen, INRIA - Grenoble - Rhone-Alpes, France
Abdulrahman Yarali, Murray State University, USA
Emmanuel Bertin, France Telecom R&D - Orange Labs, France
Steffen Fries, Siemens, Germany

LMPCNA 2012 Chairs
Petre Dini, Concordia University - Montreal, Canada / China Space Agency Center - Beijing, China
Pascal Lorenz, University of Haute Alsace, France

ICNS 2012

Committee

ICNS Advisory Chairs

Pedro Andrés Aranda Gutiérrez, Telefónica I+D - Madrid, Spain
Carlos Becker Westphall, Federal University of Santa Catarina, Brazil
Eugen Borcoci, University 'Politehnica' Bucharest, Romania
Jaime Lloret Mauri, Polytechnic University of Valencia, Spain
Sathiamoorthy Manoharan, University of Auckland, New Zealand
Yoshiaki Taniguchi, Osaka University, Japan
Toan Nguyen, INRIA - Grenoble - Rhone-Alpes, France
Abdulrahman Yarali, Murray State University, USA
Emmanuel Bertin, France Telecom R&D - Orange Labs, France
Steffen Fries, Siemens, Germany

ICNS 2012 Technical Program Committee

Ryma Abassi, Higher School of Communication of Tunis /Sup'Com, Tunisia
Ferran Adelantado i Freixer, Universitat Oberta de Catalunya, Spain
Javier M. Aguiar Pérez, Universidad de Valladolid, Spain
Rui L.A. Aguiar, University of Aveiro, Portugal
Basheer Al-Duwairi, Jordan University of Science and Technology, Jordan
Ali H. Al-Bayatti, De Montfort University - Leicester, UK
Mario Anzures-García, Benemérita Universidad Autónoma de Puebla, Mexico
Pedro Andrés Aranda Gutiérrez, Telefónica I+D - Madrid, Spain
Patrick Appiah-Kubi, Towson University, USA
Mohamad Badra, Dhofar University, Oman
Mohammad M. Banat, Jordan University of Science and Technology, Jordan
Javier Barria, Imperial College of London, UK
Mostafa Bassiouni, University of Central Florida, USA
Michael Bauer, The University of Western Ontario - London, Canada
Carlos Becker Westphall, Federal University of Santa Catarina, Brazil
Tarek Bejaoui, University of Carthage, Tunisia
Mehdi Bennis, University of Oulu, Finland
Luis Bernardo, Universidade Nova de Lisboa, Portugal
Emmanuel Bertin, France Telecom R&D - Orange Labs, France
Alex Bikfalvi, Madrid Institute for Advanced Studies in Networks - Madrid, Spain
Eugen Borcoci, University "Politehnica"of Bucharest (UPB), Romania
Fernando Boronat Seguí, Polytechnic University of Valencia, Spain
Kalinka Branco, University of São Paulo, Brazil
Ioannis Broustis, Alcatel-Lucent - Murray Hill, USA
Jens Buysse, Ghent University/IBBT, Belgium
Maria Calderon Pastor, Universidad Carlos III, Madrid, Spain
Maria Dolores Cano Baños, Polytechnic University of Cartagena - Campus Muralla del Mar, Spain
Tarik Caršimamovic, BHTelecom, Bosnia and Herzegovina

Patryk Chamuczynski, Technisat Digital R&D, Poland
Bruno Chatras, Orange Labs, France
Wei Cheng, University of California - Davis, USA
Jun Kyun Choi, KAIST, Korea
Hugo Coll Ferri, Universidad Politecnica de Valencia, Spain
Todor Cooklev, Indiana University - Purdue University Fort Wayne, USA
Alejandro Cordero, Amaranto Consultores, Spain
Noelia Correia, Universidade do Algarve, Portugal
Félix Cuadrado Latasa, UPM, Spain
Taiping Cui, Inha University - Incheon, Korea
Carlton Davis, École Polytechnique de Montréal, Canada
João Henrique de Souza Pereira, University of São Paulo, Brazil
Wei Ding, New York Institute of Technology, USA
Zbigniew Dziong, ETS - Montreal, Canada
Giuseppe Durisi, Chalmers University of Technology - Göteborg, Sweden
El-Sayed El-Alfy, King Fahd University of Petroleum and Minerals, Saudi Arabia
Fakher Eldin Mohamed Suliman, Sudan University of Science and Technology, Sudan
Juan Flores, University of Michoacan, Mexico
Steffen Fries, Siemens, Germany
Sebastian Fudickar, University of Potsdam, Germany
Michael Galetzka, Fraunhofer Institute for Integrated Circuits - Dresden, Germany
Alex Galis, University College London, UK
Ivan Ganchev, University of Limerick, Ireland
Elvis Eduardo Gaona G., Universidad Distrital Francisco José de Caldas, Colombia
Abdennour El Rhalibi, Liverpool John Moores University, UK
Stenio Fernandez, Federal University of Pernambuco, Brazil
Gianluigi Ferrari, University of Parma, Italy
Miguel Garcia Pineda, Universitat Politecnica de Valencia, Spain
Gordana Gardasevic, University of Banja Luka, Bosnia and Herzegovina
Rosario Garroppo, Università di Pisa, Italy
Sorin Georgescu, Ericsson Research, Canada
Marc Gilg, University of Haute Alsace, France
Debasis Giri , Haldia Institute of Technology, India
Ivan Glesk, University of Strathclyde - Glasgow, UK
Ann Gordon-Ross, University of Florida, USA
Victor Govindaswamy, Texas A&M University-Texarkana, USA
Dominic Greenwood, Whitestein, Switzerland
Jean-Charles Grégoire, INRS - Université du Québec - Montreal, Canada
Vic Grout, Glyndwr University - Wrexham, UK
Ibrahim Habib, City University of New York, USA
Go Hasegawa, Osaka University, Japan
Hermann Hellwagner, Klagenfurt University, Austria
Enrique Hernandez Orallo, Universidad Politécnica de Valencia, Spain
Zhihong Hong, Communications Research Centre, Canada
Per Hurtig, Karlstad University, Sweden
Naohiro Ishii, Aichi Institute of Technology, Japan
Arunita Jaekel, University of Windsor, Canada
Peter Janacik, University of Paderborn, Germany

Imad Jawhar, United Arab Emirates University, UAE
Ravi Jhawar, Universitàdegli Studi di Milano - Crema, Italy
Sudharman K. Jayaweera, University of New Mexico - Albuquerque, USA
Ying Jian, Google Inc, USA
Fan Jiang, Tuskegee University, USA
Eunjin (EJ) Jung, University of San Francisco, USA
Enio Kaljic, University of Sarajevo, Bosnia and Herzegovina
Georgios Kambourakis, University of the Aegean - Karlovassi, Greece
Hisao Kameda, University of Tsukuba, Japan
Nirav Kapadia, Fujitsu America, USA
Georgios Karagiannis, University of Twente, The Netherlands
Masoumeh Karimi, Technological University of America, USA
Aggelos K. Katsaggelos, Northwestern University - Evanston, USA
Sokratis K. Katsikas, University of Piraeus, Greece
Bithika Khargharia, Cisco Systems, Inc., USA
Dong Seong Kim, University of Canterbury, New Zealand
Kyungtae Kim, NEC Labs. America, USA
Younghan Kim, Soongsil University - Seoul, Republic of Korea
Mario Kolberg, University of Stirling - Scotland, UK
Lisimachos Kondi, University of Ioannina, Greece
Jerzy Konorski, Gdansk University of Technology, Poland
Elisavet Konstantinou, University of the Aegean, Greece
Kimon Kontovasilis, NCSR "Demokritos", Greece
Igor Kotenko, St. Petersburg Institute for Informatics and Automation, Russia
Evangelos Kranakis, Carleton University, - Ottawa, Canada
Suk Kyu Lee, Korea University at Seoul, Republic of Korea
DongJin Lee, Auckland University, New Zealand
Leo Lehmann, OFCOM, Switzerland
Ricardo Lent, Imperial College London, UK
Alessandro Leonardi, AGT Group (R&D) GmbH - Darmstadt, Germany
Qilian Liang, University of Texas at Arlington, USA
Wen-Hwa Liao, Tatung University - Taipei, Taiwan
Fidel Liberal Malaina, University of Basque Country, Spain
Thomas Little, Boston University, USA
Giovanni Livraga, Università degli Studi di Milano - Crema, Italy
Jaime Lloret Mauri, Polytechnic University of Valencia, Spain
Edmo Lopes Filho, Algar Telecom, Brazil
Albert Lysko, Meraka Institute/CSIR- Pretoria, South Africa
Zoubir Mammeri, ITIT - Toulouse, France
Sathiamoorthy Manoharan, University of Auckland, New Zealand
Moshe Masonta, Tshwane University of Technology - Pretoria, South Africa
Katerina Mitrokotsa, EPFL, Switzerland
Klaus Moessner, University of Surrey- Guildford, UK
Mohssen Mohammed, Cape Town University, South Africa
Carla Monteiro Marques, University of State of Rio Grande do Norte, Brazil
Lorenzo Mossucca, Istituto Superiore Mario Boella - Torino, Italy
Mary Luz Mouronte López, Telefónica I+D, Spain
Arslan Munir, University of Florida - Gainesville, USA

Nikolai Nefedov, Nokia, Finland
Petros Nicopolitidis , Aristotle University of Thessaloniki , Greece
Toan Nguyen, INRIA, France
Bruce Nordman , Lawrence Berkeley National Laboratory, USA
Serban Obreja, University Politehnica - Bucharest, Romania
Kazuya Odagiri, Yamaguchi University, Japan
Máirtín O'Droma, University of Limerick, Ireland
Tae (Tom) Oh, Rochester Institute of Technology, USA
Jinwoo Park, Korea University, Korea
Harry Perros, North Carolina State University, USA
Francisca Aparecida Prado Pinto, Federal University of Ceará, Brazil
Francesco Quaglia, Sapienza Università di Roma, Italy
Karim Mohammed Rezaul, Centre for Applied Internet Research (CAIR), NEWI, University of Wales, UK
Abdelmounaam Rezgui, George Mason University, USA
Oliviero Riganelli, University of Milano Bicocca, Italy
David Rincon Rivera, Technical University of Catalonia (UPC) - Barcelona, Spain
Diletta Romana Cacciagrano, Università di Camerino, Italia
Paolo Romano, Instituto Superior Tecnico/INESC-ID - Lisbon, Portugal
Sattar B. Sadkhan, University of Babylon, Iraq
Alessandra Sala, University of California - Santa Barbara, USA
Francisco Javier Sánchez Bolumar, Centro de Formación Tecnológica - Valencia, Spain
Luz A. Sánchez-Gálvez, Benemérita Universidad Autónoma de Puebla, México
Panagiotis Sarigiannidis, University of Western Macedonia - Kozani, Greece
Stefan Schmid, TU Berlin & Telekom Innovation Laboratories (T-Labs), Germany
René Serral Garcia, Universitat Politècnica de Catalunya, Spain
Xu Shao, Institute for Infocomm Research, Singapore
Fangyang Shen, City University of New York, USA
Jian Shen, Chosun University, Gwangju, Republic of Korea
Tsang-Ling Sheu, National Sun Yat-Sen University - Kaohsiung, Taiwan
Eunsoo Shim, Avaya Labs Research, USA
Simone Silvestri, University of Rome "La Sapienza", Italy
Navjot Singh, Avaya Labs Research, USA
Charalabos Skianis, University of Aegean - Karlovasi, Greece
Vasco Soares, Instituto de Telecomunicações / University of Beira Interior / Polytechnic Institute of Castelo Branco, Portugal
José Soler, Technical University of Denmark, Denmark
Gritzalis Stefanos, University of the Aegean, Greece
Yoshiaki Taniguchi, Osaka University, Japan
Olivier Terzo, Istituto Superiore Mario Boella - Torino, Italy
Christian Timmerer, Alpen-Adria-Universität Klagenfurt, Austria
Petia Todorova, Fraunhofer Institut FOKUS - Berlin, Germany
Binod Vaidya, University of Ottawa, Canada
Hans van den Berg, TNO / University of Twente, The Netherlands
Dario Vieira, EFREI, France
José Miguel Villalón Millan, Universidad de Castilla - La Mancha, Spain
Manuel Villén-Altamirano, Universidad Politécnica de Madrid, Spain
Demosthenes Vouyioukas, University of the Aegean - Karlovassi, Greece
Arno Wacker, University Duisburg-Essen, Germany

Bin Wang, Wright State University - Dayton, USA
Mea Wang, University of Calgary, Canada
Tingkai Wang, London Metropolitan University, UK
Michelle Wetterwald, EURECOM - Sophia Antipolis, France
Ouri Wolfson, University of Illinois - Chicago, USA
Feng Xia, Dalian University of Technology, China
Qin Xin, Université Catholique de Louvain - Louvain-la-Neuve, Belgium
Homayoun Yousefi'zadeh, University of California - Irvine, USA
Vladimir S. Zaborovsky, Polytechnic University/Robotics Institute - St.Petersburg, Russia
Sherali Zeadally, University of the District of Columbia, USA
Yifeng Zhou, Communications Research Centre, Canada
Ye Zhu, Cleveland State University, USA
Piotr Zuraniewski, University of Amsterdam (NL), The Netherlands /AGH University of Science and Technology, Poland

LMPCNA 2012

LMPCNA 2012 Technical Program Committee

Nalin Abeysekera, Open University of Sri Lanka, Sri Lanka
Giancarlo Bo, Technology and Innovation Consultant – Genova, Italy
Maiga Chang, Athabasca University, Canada
Pavel Cicak, Slovak University of Technology, Slovakia
Giuseppe Cinque, Consorzio ELIS - Rome, Italy
Dumitru Dan Burdescu, University of Craiova, Romania
Kristen DiCerbo, Cisco Systems, Inc., USA
Adam M. Gadomski, ECONA (Centro Interuniversitario Elaborazione Cognitiva Sistemi Naturali e Artificiali) - Rome, Italy
Ján Genci, Technical University of Kosice, Slovakia
Juraj Giertl, Technical University of Kosice, Slovakia
Ron J. Kovac, Ball State University, USA
Eugenijus Kurilovas, Vilnius Gediminas Technical University, Lithuania
Jaime Lloret Mauri, Universidad Politécnica de Valencia, Spain
Iain Murray, Curtin University of Technology – Perth, Australia
Elisabetta Parodi, Giunti Labs S.r.l., Italy
Josep Prieto Blázquez, Open University of Catalonia, Spain
Jelena Revzina, Transport and Telecommunication Institute, Latvia
Shahram S. Heydari, University of Ontario Institute of Technology - Oshawa, Canada
Richard Seaton, The Open University, UK
Andrew Smith, The Open University, UK

Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission or reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article is does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

Table of Contents

Airborne Surveillance Networks with Directional Antennas

William Huba, Nirmala Shenoy
Networking Security and Systems Administration Department,
Rochester Institute of Technology, Rochester NY 14623, USA
nxsvks@rit.edu

Abstract— **Surveillance using unmanned aerial vehicles (UAVs) is an important application in tactical networks. Such networks are challenged by the highly dynamic network topologies, which result in frequent link and route breaks. This requires robust routing algorithms and protocols. Depending on the coverage area, several UAVs may be deployed thus requiring solutions that are scalable. The use of directional antennas mitigates the challenges due to limited bandwidth, but requires a scheduling algorithm to provide conflict free schedules to transmitting nodes. In this article we introduce a new approach, which uses a single algorithm 1) that facilitates multi hop overlapped cluster formations to address scalability and data aggregation; 2) provides robust multiple routes from data generating nodes to data aggregation node and; 3) aids in performing distributed scheduling using a Time Division Multiple Access protocol. The integrated solution was modeled using Opnet and evaluated for success rate in packet delivery and average end to end packet delivery latency primarily. The notably high success rates important for surveillance purposes coupled with low latencies validate the use of the proposed solution in critical surveillance applications.**

Keywords-airborne surveillance; network of unmanned aerial vehicles; directional antennas; TDMA

I. INTRODUCTION

Surveillance networks comprising of airborne nodes such as unmanned aerial vehicles (UAVs) are a category of mobile ad hoc networks (MANETs), where the nodes are travelling at speeds of 300 to 400 Kmph. Surveillance requires aggregation of data captured by all nodes in the network at few nodes, from where the data is then sent to a center for further action. Due to high mobility of nodes and varying wireless environment, the topology in surveillance networks is subject to frequent and sporadic changes. Such MANETs thus face severe challenges when forwarding data from node to node, which is the task of the medium access control (MAC) protocol and also in discovering and maintaining routes between source and destination nodes, which is the task of the routing protocols. Another challenge faced is the scalability of the protocols to increasing number of nodes which can be addressed partly through clustering which can also aid in data aggregation

In this article, a unique solution for surveillance networks comprising of UAVs, equipped with directional antennas is investigated. The solution *uses a single algorithm* for several operations such as 1) multi-hop overlapped cluster formation, 2) routing of data from cluster clients to cluster head to aid in data aggregation, and 3) scheduling time slots to transmitting nodes using a Time Division Multiple Access (TDMA) based MAC protocol, which avails the directional antenna capabilities. To best leverage the strengths of this approach, the MAC, clustering and routing functions were implemented as processes operating using a single address that collaboratively address the challenges faced in surveillance networks. Due to the critical nature of the application such a unified or integrated approach is justified. This is vetted by the performance tests conducted in networks with twenty, fifty and seventy five UAVs. The performance metrics of primary interest were success in packet delivery and packet delivery latency which are very important in surveillance applications.

The proposed solution with its various components was modeled using Opnet. Surveillance applications require low packet loss and low information or packet delivery latencies. The unique approach introduced in this article achieves these performance goals. However due to lack of similar published work and the availability of models for such application scenarios, this presentation is limited to result from simulations of the proposed solution.

The rest of the paper is organized as follows. Section II describes related work in the area of TDMA MAC, routing in large MANETs and clustering techniques. The benefits of the integrated approach are highlighted in the light of these discussions. Section III describes the *meshed tree* algorithm - the single algorithm, which is able to support all three operations while allowing them to interact efficiently. Section IV describes the link assignment strategy. Section V provides the performance analysis conducted using Opnet. Conclusions are provided in Section VI.

II. RELATED WORK

The solution in this work targets an integrated approach, facilitated also by the use of a single algorithm. To the best of our knowledge there is no published work that integrates different operations such clustering, scheduling at MAC and routing using a single algorithm and a single address for large surveillance MANETs. In this section, we thus present some related work conducted separately in the areas of TDMA based MAC and scheduling, routing protocols for large MANETs and clustering techniques and conclude by highlighting the advantages of an integrated approach.

TDMA based MAC: To achieve higher capacity and better delay guarantees in networks that use directional antennas, *Spatial reuse Time Division Multiple Access* (STDMA) MAC can be employed. In STDMA, multiple transmissions can be scheduled in a way to avoid packet

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

interference [1]. STDMA thus takes advantage of the spatial separation between nodes to reuse time slots. Such schemes require strict time synchronization among participating nodes. In addition, if the nodes are mobile, periodic changes in the network topology require regular and timely updates to the schedules. The most challenging task is generating conflict free schedules to aid multiple nodes to transmit simultaneously in a time slot. Several algorithms [1-3] have been investigated for the purpose. Some adopt a centralized approach requiring information about all links at the centralized scheduler. Distributed scheduling eases this requirement but at the expense of higher complexity.

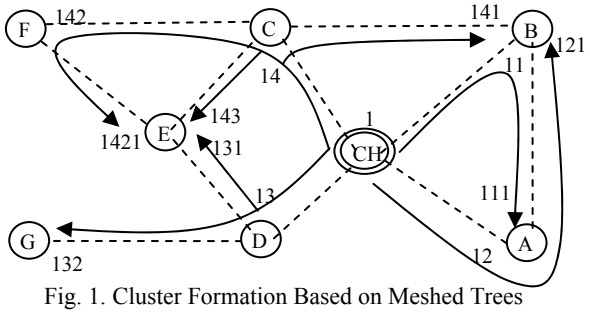


Fig. 1. Cluster Formation Based on Meshed Trees

Literature is rich with work conducted in the area of routing and clustering for MANETs in general. The related work presented here is limited to those routing protocols that are zone limited and multi hop clustering.

Clustering or zoning can be efficiently employed for the type of convergecast traffic encountered in surveillance networks, were the primary traffic flow is from cluster clients (CC) to cluster head (CH) [4, 5]. In such cases proactive routing approaches are recommended as the routing is limited to the cluster or zone and will also reduce stale routes. However proactive routing algorithms require the dissemination of link state information to all routers in the network or zone, which can introduce latency in realizing or breaking a route, and high overhead. In the *Zone Routing Protocol* (ZRP) [6] each node pre-defines a zone centered at itself. ZRP proposes a framework, where any proactive routing protocol can be adopted within the zone and any reactive routing protocol can be adopted to communicate outside of the zone. *Multi path distance vector zone* routing protocol [6] is an implementation of ZRP that uses multi path *Destination Sequence Distance Vector* [7] for proactive routing. *LANMAR* [6] routing protocol defines logical groups to address scalability where landmark nodes keep track of the groups. A local scope routing based on *Fisheye State Routing* is used in the group.

Multi Hop clustering techniques such as the d-hop or k-hop clustering [8] algorithms can offer flexibility in terms of controlling the cluster size and cluster diameter, but are often complex to implement.

Advantages of the Single Algorithm Approach: From the above discussions it would be clear that clustering, routing and scheduling are different operations and hence normally are based on different algorithms. When combining the different operations, it becomes essential to define an interworking mechanism for the different algorithms. This adds processing complexity. It also results in added overhead for the operation of the combined functions. If all these operations can be based off a single algorithm the complexity and overhead can be reduced.

Advantages of Interacting Modules: If the above approach were possible, and if the MAC, routing, clustering and scheduling can use a single address for their operation (unlike our current protocol stack, where MAC protocol uses 48 bit MAC addresses for its operation and routing protocols use 32 bit IP addresses (or 128 bits If IPv6)), we can achieve a solution, where the processes can closely interact and also avoid issues and overhead due to protocol layering, handling different headers and complex cross layered techniques. This would also make the solution compact and efficient and foster close interworking among the different operations. Such an approach would be ideal for critical tactical surveillance networks.

III. THE INTEGRATED APPROACH

The multi meshed tree (MMT) algorithm [10-12] is the one proposed to support the integrated approach and will be briefly explained first. Cluster formations, proactive routing and TDMA scheduling based on this algorithm will be explained subsequently.

A. The Multi Meshed Tree Algorithm

The formation of a single *meshed tree* based on the MMT algorithm is described with the aid of Fig. 1. The dotted lines connect nodes that are in communication range with one another at the physical layer. The node designated as CH is the root of the meshed tree. For ease in explanation, the meshed tree formation is kept simple and restricted to nodes that are connected to the CH by a maximum of 3 hops. At each node several values or IDs have been noted. These are the virtual IDs (VIDs) assigned to the node when it joins one of the tree branches in the meshed tree. Without loss of generality, assume that the CH has a VID '1'. All nodes connected to this CH will have '1' as the first digit in their VIDs. Extending the above logic, a node gets a VID, which will inherit as its prefix the VID of the node upstream in the tree branch (the parent node), followed by a single (or multiple) digit(s) which indicates the child number under that parent. In the presented work the child number is restricted to a single integer - validated by the fact that having more than nine children under a single parent node could cause bottleneck issues during traffic aggregation. In Fig. 1, each arrow from CH is a tree branch that connects the nodes to the root.

Flexible Multi-hop Cluster Formation: Except for the CH, each node in Fig. 1 is a CC that will send the captured surveillance data to the CH. The size of the tree branch can be limited by limiting the length of the VID, which in turn allows control of the diameter of the cluster. Each node that joins the cluster has to register with the CH, by forwarding a registration request (reg_req) along the branch of the VID.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

This confirms the path defined by the VID and also allows the CH to accept /reject a joining node to control the cluster size. The number of VIDs allowed for a node can control the amount of meshing in the tree branches of the cluster.

Multiple Dynamic Proactive Paths: The branches of the meshed tree provide the *route* to send and receive data and control packets between the CCs and the CH. The branch denoted by VIDs 14, 142 and 1421 connects nodes C (via VID 14), F (via VID 142) and E (via VID 1421), respectively, to the CH. Consider packet forwarding based on VIDs in which the CH has a packet to send to node E. If the CH decided to use E's VID 1421, it will include this as the destination address and broadcast the packet. En route nodes C and F will pick up the packet and forward to E. This is possible as the VIDs for nodes C and F are contained in E's VID. The VID of a node thus provides a virtual path vector from the CH to itself. Note that the CH could have also used VIDs 143 or 131 for node E, in which case the path taken by the packet would have been CH-C-E or CH-D-E respectively. Thus, between the CH and node E there are multiple routes as identified by the multiple VIDs. The support for multiple proactive routes through the multiple VIDs allows for robust and ***dynamic route adaptability*** to topology changes in the cluster, as the nodes request for new VIDs and joins different branches as their neighbors change. This keeps the routes non-stale.

B. Scheduling

The VIDs carry link information between a pair of nodes that share a parent-child relationship. Thus a link assignment strategy was adopted in this work. The structure of the VIDs, also allows each node in a cluster to be aware of its neighbors due to the parent-child relationship defined by the VIDs. This allows a node to schedule time slots with its neighbors (parent or child) taking into consideration its current committed time slots to its other neighbors.

C. Scalability

A surveillance network can comprise of several tens of nodes; hence the solutions for surveillance networks have to be scalable to that many nodes. We assume that several 'data aggregation nodes (i.e., CHs)' are uniformly distributed among the non-data aggregation nodes during deployment of the surveillance network. Meshed tree clusters can be formed around each of the data aggregation nodes by assuming them to be roots of the meshed trees. Nodes bordering two or more clusters are allowed to join the different meshed trees and thus reside in the branches originating from different CHs. Such border nodes will inform their CHs about their multiple VIDs under the different clusters. When a node moves away from one cluster, it can still be connected to other clusters, and thus the surveillance data collected by that node is not lost. Also, by allowing nodes to belong to multiple clusters, the single meshed tree cluster based data collection can be extended to *multiple* overlapping *meshed tree* (MMT) clusters that can collect data from several tens of nodes deployed over a wider area with a very low probability of losing any of the captured data. This addresses the *scalability* requirements in surveillance networks.

D. Interworking of Modules

It is important to understand the interworking of the modules and their interaction with the directional antenna system. Hence, the directional antenna system is first described followed by the interactions among the modules and their use of the directional antenna systems.

Directional Antenna System: All nodes in the surveillance network are assumed to be equipped with four phased array antennas capable of forming two beam widths. One beam width is focused with an angle of 10^o and the other is defocused with an angle of 90^o. The defocused beams are used for sending broadcast packets, while the focused beams are used for unicast or directed packets. Each antenna array covers a quadrant (90^o) and is independently steerable to focus in a particular direction within that quadrant in the focused beam mode.

We also assume that each node is equipped with a Global Positioning System (GPS) which is used for time synchronization and to provide node position. The latter information is used in a tracking algorithm to estimate the location of a receiver node, so transmitting nodes can direct

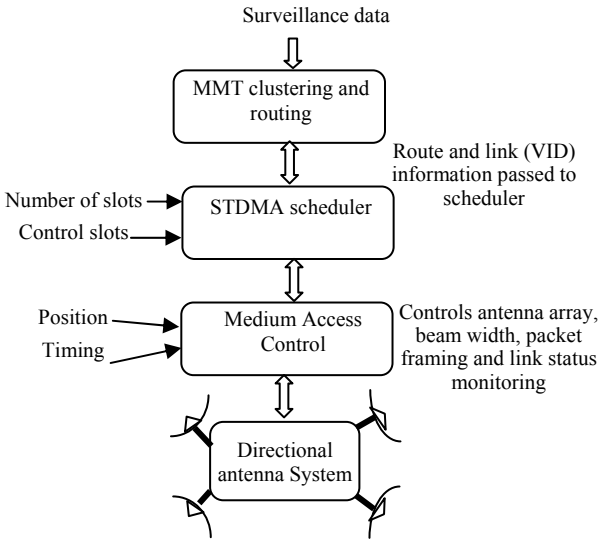


Fig. 2. Scheduler Operations with other Modules

their beams to the destination node.

Interworking Principles: The surveillance data collected by the nodes is passed to the MMT clustering and routing module, which decides the route or VID to use to forward the data to the CH. Once the route has been decided, the node knows the address of the next hop node which will forward the packet. This information will be used by the STDMA scheduler to schedule slots, taking as input the number of slots, slot time and control slots. This

information is then passed to the MAC to create the frame and forward to the next node. Before forwarding, the MAC, locates the destination node position and controls the antenna array to transmit the packet using a directed beam.

Scheduler Operations: The scheduling algorithm has to schedule time slots for (1) cluster formation after deployment of the UAV nodes, (2) subsequent cluster and route maintenance, and (3) data aggregation. It should also send updated schedules in a timely manner as network topology changes. For all of these operations different categories of time slots as described below were used.

- *Broadcast Slots:* Some slots are preselected as broadcast slots in which they announce their VIDs, location, and current schedule, in a *configuration* (conf) packet, so neighboring nodes can listen and decide to join the cluster.
- *Directed Slots:* All other slots are used in a directed mode, where one node is transmitting using the directed beam to its listening neighbor. Directed slots can be *assigned* slots or *temp* (unassigned) slots.
- *Temp Slots* are used by nodes to negotiate for a common time slot for data transfer.
- *Assigned Slots:* Temp slots become assigned slots after a mutual negotiation by a pair of nodes. In the assigned slots control information for cluster and route maintenance, link maintenance (*lnk_ mnt*) control packet generated by the MAC and data packets are sent and received. Assigned slots are unidirectional and are used either for transmitting (data-tx) or receiving (data-rx). If there are data packets to be sent in such slots slot, the control packets are sent first, followed by the data packets. When there are no data packets to send, the MAC sends lnk-mnt packets to monitor the link status.

IV. LINK ASSIGNMENT

The approval of a new node by the CH is an indication that the CC has both a physical and logical path towards the CH. Scheduling slots for the new node starts subsequent to its acceptance into a cluster by the CH. Nodes individually schedule data slots in a distributed manner with their one-hop neighbors making the scheme truly distributed. The end to end information is carried by the VIDs. Time slots are scheduled for as long as at least one VID remains between a node pair. The process of mutual scheduling is explained with the aid of Fig. 3 below.

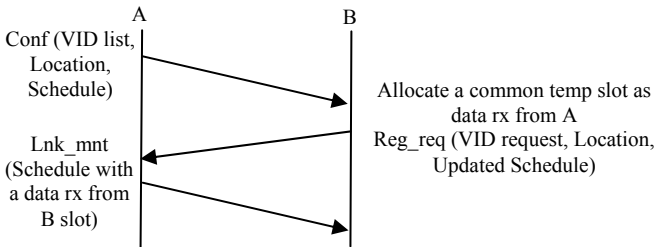


Fig. 3. Distributed Scheduling Across Neighbors

Details about cluster formation, VID acquisition and control packets are covered in [10-12]. When node A advertises its VIDs via a *conf* packet it attaches its current schedule and GPS coordinates. Node B receives the packet and decides to request a VID under one of the advertised VIDs. Node B will then reserve a data-rx slot from one of the temp slots advertised by the parent that matches with its own temp slot and respond with a registration request (reg_req), and the updated schedule to node A. Node A in turn assigns another temp slot that is common to the pair as a data-rx slot for receiving packets from node B. It then forwards the registration request from node B towards the CH. During the next frame, node A will send a *lnk_mnt* packet to node B with the updated schedule. Thus a set of slots for transmitting and receiving between nodes A and B are decided. No other node's schedule is taken into account unless it directly affects the current link between two negotiating nodes.

The process of allowing a new requesting node a VID to reserve a data_rx slot in which the parent node can transmit allows the parent node to resolve conflicts in case the suggested data_rx slot is not available. The parent node does this by sending a lnk_mnt packet on a data-rx slot reserved by the child, requesting that it change its data-rx slot.

A. On Demand Slot Allocation

The above negotiation can be tuned to traffic demands at a node. For example if node A's buffer indicates packet (to be sent to node B) accumulation beyond a threshold value, then in the next lnk-mnt packet, A can request node B to set aside *x* data-rx slots, where the value x is capped to avoid one node taking up all available slots. Node B will respond with the updated schedule by setting aside the *x* slots provided it has no such similar demands from its other neighbors. If there are similar demands, it will allocate slots proportional to the demands of tis neighbors. The on demand allocation can result in increased number of data-rx slots at B (to receive from node A) though the single data-tx towards node A will be maintained unless changed by a demand. The tuning of the on-demand slots is executed every frame.

V. SIMULATIONS

The performance evaluations of the surveillance network using the proposed solution was carried out using Opnet (version 14.5) simulation tool. All the processes explained above were modeled in Opnet. For surveillance data, each CC generated a 1 MByte file, which was then sent to the CH for aggregation. Normally UAVs travel in elliptical trajectories for surveillance purposes. In the models, we used circular orbits, to introduce more route breaks and thus stress test the solution. These circular orbits had a diameter of 20 Km (which defines the areas for each scenario), while the maximum transmission range was limited to 15 Km. the overlap between trajectories is seen in Fig. 4. A maximum of 5 UAVs were allowed in one circular trajectory, thus the UAVs were deployed over a wider area, which was covered with several trajectories. For example, in the 20 node scenario, there were four circular trajectories with slight overlap in their trajectories, to avoid physical network

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

segmentation as shown in Fig 4. In the trajectories, the speed of the UAVs varied between 300 to 400 Kmph; hence, the different colors for the trajectories.

The physical layer parameters were maintained invariant. Packets with 1 bit error rate were dropped and no *Forward Error Correction* was implemented. In the focused beam mode the data rate is 50 Mbps and in the defocused mode the data rate is 1.5 Mbps. A single frame had 50 timeslots each of 4 ms duration and 0.5 ms guard time. These values were optimized based on our prior work [4, 5].

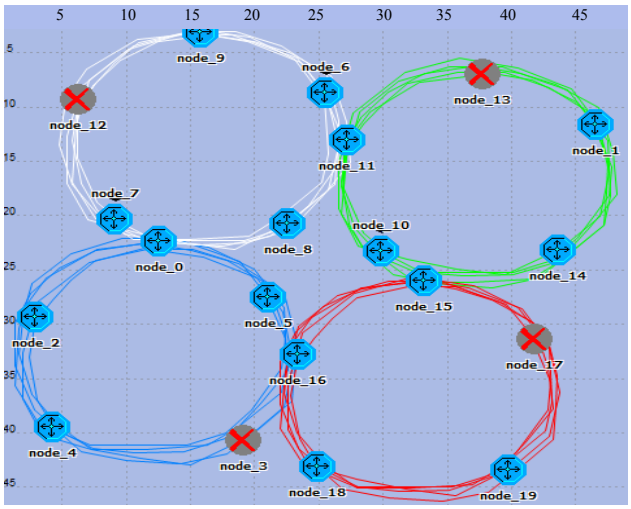


Fig. 4. Typical Deployment and UAVs

Due to the lack of similar published work and comparable models in Opnet the performance of the presented solution is analyzed with respect to the performance goals we had stated for surveillance networks earlier namely success in packet delivery, and latency in packet and file deliveries. Included in the performance graphs are the overhead incurred by the MAC and routing protocols, and the average hops encountered during packet delivery, which is useful in explaining some results

Overhead is the % of control traffic as a ratio of all traffic i.e., including data traffic in the network. Packet latency recorded was the end to end latency i.e. from the time the packet was sent by a sender node till it was received by the CH. File delivery latency was calculated similarly.

In each of the test scenarios, a certain number of nodes were randomly selected to send a 1 MByte file to the CH. These selected nodes sent the files simultaneously, thus stress testing the solution. Furthermore the number of sending nodes was increased to include all of the nodes except the data aggregation nodes, which is a highly stressful test scenario. Each test scenario was repeated with 20 different seeds (high prime numbers) and the results averaged over these seeds. The simulations were limited 20 runs in each case due to the stable outcomes noticed with different seeds.

A. 20 Nodes Scenario

Figures 5A to 5C are the plots for the twenty UAV scenario with 4 clusters. The x axis in all plots shows the number of nodes that are simultaneously sending aggregation traffic, i.e., 1 MByte file to the 4 CHs. The number of sending nodes was varied from 4 to 16. In the last case all 16 CCs were sending a 1 MByte file simultaneously to the CHs.

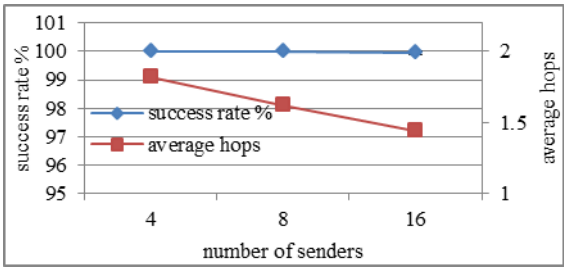


Fig. 5A. Success rate % and Avg hops vs senders

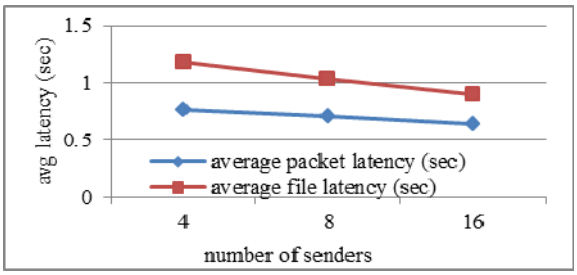


Fig. 5B. Average packet and file latency vs senders

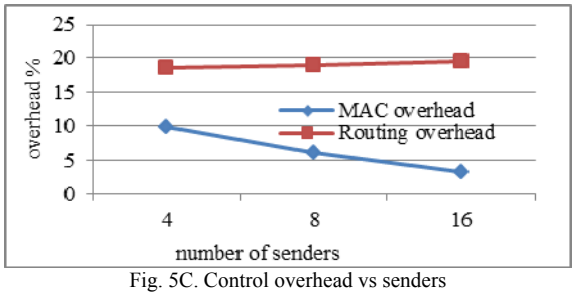


Fig. 5C. Control overhead vs senders

With increasing number of senders, the success rate hardly dropped below 100%. This shows the efficiency of the scheduler to successfully schedule all the packets that are arriving simultaneously. The average hops recorded in graph 1 however shows a decrease when the number of sending nodes was increased. When 20 nodes were selected to send traffic they encountered an average hop distance of 1.8 hops; which dropped to 1.4 hops when all 16 nodes were sending traffic. This is because of the random way in which the sending nodes were selected. The average hops graph can be interpreted thus – the first four nodes that were selected were farther away from the CHs, but as more nodes were randomly picked they were closer to the CH. The impact of this is noticeable in the packet and file latencies recorded in graph B, which shows a decrease with increasing number of senders.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

In Fig. 5B, the average packet latency recorded was less than 0.8 seconds. Acceptability of packets arriving at this latency depends on the criticality of the surveillance application. If an upper limit was specified then that could be used as a cut off to drop packets arriving late. The file delivery latency is only slightly higher at around 1.2 seconds, which shows that all packets in the 1 MByte file were transported from the data collection node to the aggregation nodes, i.e., the CH within the time.

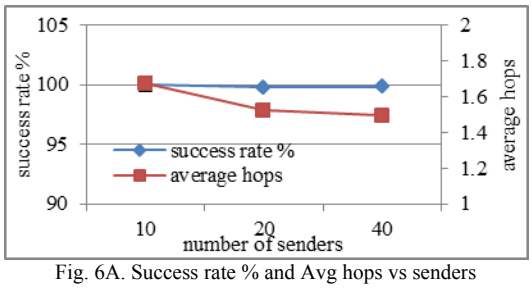


Fig. 6A. Success rate % and Avg hops vs senders

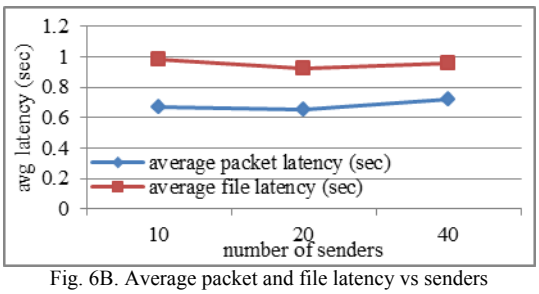


Fig. 6B. Average packet and file latency vs senders

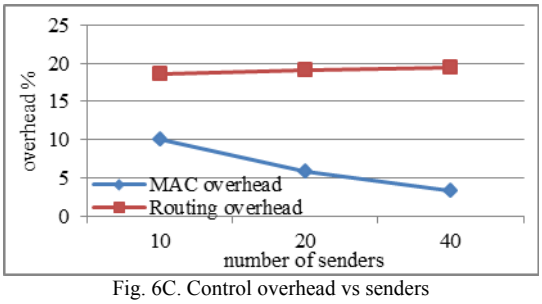


Fig. 6C. Control overhead vs senders

Fig. 5C is the plot of a very important parameter as it shows the channel bandwidth used by the control traffic both by the MMT based routing protocol as well as the MAC protocol. The MAC and routing overhead were recorded to show the ratio of messages used for control purposes by two operations.

The MMT routing overhead was below 20% while the MAC overhead reduced from 10% when there were 4 sending nodes to less than 5% when there were 16 sending nodes. It should be noted that the MMT routing traffic also includes the cluster formation control traffic.

The MAC overhead shows a decrease with increasing number of senders, because when there are fewer data packets to send (with less senders) the MAC still sends maintenance packets, thus the ratio of control bits to the total bits that travelled the network, shows a decrease when there are more data packets in the network. The routing overhead records a very slight increase (around 1%) with increasing senders, which can be attributed to more route maintenance which will be triggered to correctly route the high amount traffic generated.

B. 50 Nodes Scenario

Figures 6A to 6C are the plots for the 50 UAV scenario with 10 clusters. The number of UAVs sending 1 MByte file simultaneously was varied from 10, 20 to 40. Thus in the case of the 40 senders, all CCs were sending 1 MByte files to the CHs simultaneously.

The success rate in graph A shows a slight drop to around 99.7 % as the senders increased, which shows the reliability in data transfer of the proposed solution and its scalability as the number of surveillance nodes and data sending nodes increased. The average hops which is plotted along with success rate graph does not show a linear decrease as in Fig 5 graph A. This is again attributed to the random selection in sending nodes. The first 10 senders were on an average of 1.7 hops from the CH, the added 10 senders for the 20 node case reduced the average hops to slightly above 1.5, and the last 20 senders brought the average hops to 1.5.

Figure 6B reflects the impact of the average hops in the packet and file delivery latency. There is drop when the senders increase from 10 to 20, this is because the average hops has a steep decrease from 1.7 to 1.5. However the average hops drops very slightly when senders are increased from 20 to 40 nodes, this and the fact that there is more traffic and more buffering by the nodes, the packet and file latency increase with increase in senders from 20 to 40.

The MAC and routing overhead in Fig 6C show a similar trend as observed in Fig 5. Though the number of nodes has increases, control traffic is calculated as a ratio of control traffic to total traffic in the network during the time that the files are being delivered.

C. 75 Nodes Scenario

Figures 7A, 7B and 7C are the performance plots for the test scenario with a total of 75 UAVs and 15 clusters, the number of sending nodes was varied from 15, 30 to 60. Hence again when 60 nodes are sending 1 Mbyte file it is the case of all CCs sending traffic to the CHs. The success rate dropped to around 98.7% with increasing number of senders – reflecting the robustness of the proposed solution and its scalability to increasing UAVs and increasing number of senders. The plot of the average hops again shows a decrease from 1.55 to 1.47 as the number of senders selected randomly to send the traffic to the CH was increased.

Figure 7B is the plot for the packet and file latency. The plot shows an increase because the change in the average hops was 0.06 as the number of senders was increased. The latency trends reflect the average hops trend. Figure 7C

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

which is the plot of the MAC and routing overhead has a

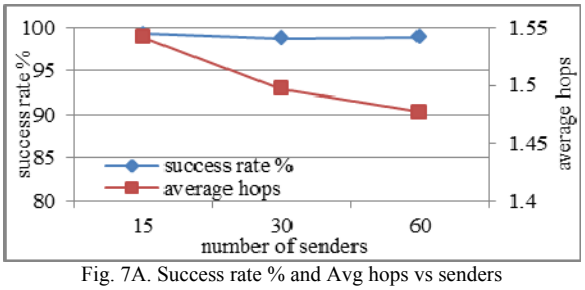


Fig. 7A. Success rate % and Avg hops vs senders

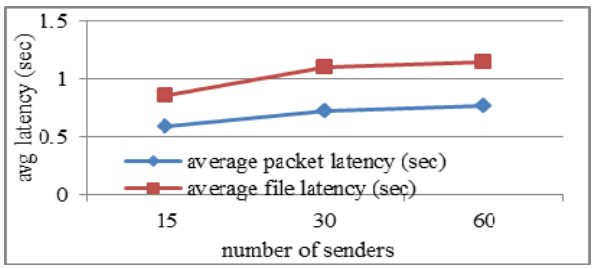


Fig. 7B. Average packet and file latency vs senders

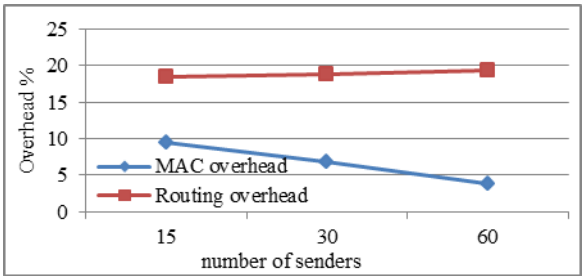


Fig. 7C. Control overhead vs senders

similar trend as noted for the 20 and 50 node scenarios.

Summarizing, the performance graphs indicate the high robustness of the proposed solutions to highly mobile and stressful MANET conditions. The continually high value of success rate despite the increase in the network size and the increase in the number of sending nodes indicate the reliability of the proposed solutions and its scalability. The packet and file latencies never exceeded 0.8 seconds and 1.2 seconds respectively in the three network setups. This indicates the robustness of the scheduling algorithm.

The overheads noted have similar trends and show very little difference as they were calculated as a ratio of the traffic in the network. The senders in each case were a quarter of the CCs, half of the CCS and the rest of the CCs. The control traffic increases with the increase in the number of nodes in a scenario, but as it is expressed as a ratio of all the traffic in the network including the data traffic, and due to the ratio of senders being consistent in all scenarios, this value can be noticed to be very close in all scenarios.

VI. CONLUSION

Surveillance networks are critical tactical applications, and hence require special consideration during solution design. The primary goal in surveillance networks of UAVs is to collect the captured data reliably at few nodes, and with low latencies. In this work we presented a solution that uses an integrated approach where MAC, routing and scheduling are based off a single algorithm and use a single address - the VIDs for their operations. This results in a low complexity yet robust and scalable solution.

The solution was evaluated in a UAV surveillance network of varying sizes of 20, 50 and 75 nodes. In each case the numbers of simultaneous 1 MByte file senders were increased from one quarter to one half to all of the remaining nodes besides the aggregation nodes. This was a highly stressful test case. The results achieved under such stress situations were very good. The drop in reliable and timely delivery was very low as the numbers of senders were increased. These results thus validate the use of the solution to such critical tactical applications.

ACKNOWLEDGMENT

This work was partly by funded by AFRL, Rome NY under contract no. 30822, and partly from ONR.

REFERENCES

[1] Fernandez S. G., “On the performance of STDMA Link Scheduling and Switched Beamforming Antennas in Wireless Mesh Networks”, Master's thesis, King's College London United Kingdom, 2009.

[2] Grönkvist J. and Hansson A., “Comparison between graph-based and interference-based STDMA scheduling,” in *MobiHoc*, 2001 pp 255-258

[3] Grönkvist J., “Assignment methods for spatial reuse TDMA, in”, First Annual Workshop on Mobile and Ad Hoc Networking and Computing, 2000, pp. 119–124.

[4] Gerla M. and Tsai J., “Multicluster, mobile, multimedia radio network,” *Wirel. Netw.*, vol. 1, pp. 255-265, 1995

[5] Lian, J.; Agnew, G.B. and Naik, S., “A variable degree based clustering algorithm for networks,” *Computer Communications and Networks, 2003. ICCCN 2003. Proceedings. The 12th International Conference on* , vol., no., pp. 465-470, 20-22 Oct. 2003

[6] Hong X., Xu K., and Gerla M., “Scalable Routing Protocols for Mobile Ad Hoc Networks”, *IEEE Network Journal,* July/Aug 2002, Vol 16, issue 4, pp 11-2.

[7] Abolhasan M., Wysocki T. and Dutkiewicz E., “A review of routing protocols for mobile ad hoc networks”, *Journal of ad hoc networks,* Elsevier publications, 2004

[8] Orakwue C., Al-Mousa Y., Martin N. and Shenoy N., “Cluster Based Time Division Multiple Access Scheme for Surveillance Networks using Directional Antennas” *ICSPCS, Brisbane Australia* Dec 2010.

[9] Amis, A. D., Prakash, R., Vuong, T.H.P. and Huynh, D.T., “Max-min d-cluster formation in wireless ad hoc networks,” INFOCOM 2000. vol.1, no., pp.32-41 vol.1.

[10] Huba W., Martin N., Al-Mousa Y., Orakwue C. and Shenoy N., “A Distributed Scheduler for Airborne Backbone Networks with Directional Antennas”, *ComsNets,* Bangalore India 2011

[11] Martin N., Al-Mousa Y. and Shenoy N., “An Integrated Routing and Medium Access Control Framework for Surveillance Networks of Mobile Devices” *12th ICDCN 2011*, Springer Verlag, pp 315-323.

[12] Daniel L., “A comprehensive overview about selected Ad hoc networking routing protocols”, *Technical Report, Department of Computer Science,* Munich Technical University, Germany

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

Merging Parallel Swarms for BitTorrent Performance Improving and Localization

Yang Wang, Jessie Hui Wang, Donghong Qin, and Jiahai Yang
The Network Research Center, Tsinghua University
Tsinghua National Laboratory for Information Science and Technology (TNList)
Beijing, China
{*wayang84, donghong.qin* }*@gmail.com,* {*hwang, yang*}*@cernet.edu.cn*

***Abstract*—With the prevalence of file distribution systems, P2P traffic has caused great challenges to Internet service providers or network operators. The corresponding problems thereby have been a concern for industrial and academic fields. In this paper, we first present a measurement study of torrents and swarms in BitTorrent systems. We find that most swarms suffer from lack of peers and many resource files are shared by multiple trackers. Based on these two observations, we propose a new architecture, named trackers' tracker (T-Tracker) to provide a new way of traffic control for network providers, bringing a little change to current BitTorrent protocol. With T-Tracker, our architecture can provide more peers by merging parallel swarms for the performance improvement of BitTorrent system and provide more choices for biased peer selection of traffic localization schemes. Therefore, quality of BitTorrent service and network utilization can be optimized at the same time.**

Keywords-BitTorrent; merging swarms; performance improvement; P2P traffic localization.

I. INTRODUCTION

Peer-to-peer content sharing on the Internet has become one of the most popular applications in recent years. BitTorrent, almost the most successful P2P file sharing system, has been widely used for the distribution of large files. Downloading files among BitTorrent peers generates a huge amount of traffic over inter-ISP links. It is a challenge to deal with the crowded network for Internet service providers (ISPs), especially at the links between ISPs.

There are many trackers deployed by different organizations in the Internet, and they are working independently. The set of active peers maintained by a same tracker and sharing the same content is referred to as a *swarm*. In BitTorrent system, one resource file might be shared by independent multiple trackers, which means that there are several swarms in these trackers distributing the same resource file and working in parallel. Peers in these different swarms cannot find each other although they are sharing the same file. Let us define these different swarms as a group of parallel swarms, and the resource file distributed in these swarms is referred to as trans-swarm resource file.

In this paper, we first conduct a measurement study on 2258909 swarms from 41 trackers. From the measurement results, we have two observations. First, most swarms have too few peers based on the snapshot. For example, 46.9% swarms have less than 2 active peers. In these small swarms, downloader cannot connect with enough peers to finish downloading in a reasonable time. Moreover, P2P locality schemes [1] [5] cannot work well since there are no enough local peers to be selected even in some large swarms. The second observation is that, there are lots of parallel swarms and trans-swarm resource files in BitTorrent system. This phenomenon may be caused by several reasons: the content provider tries to share his file among more clients and for longer time, and he makes metadata files with different trackers in order to avoid single point failure of the trackers; tracker's operator wants to reduce its work load, and these trackers trade peer information to balance the load between each other; different content providers share the same file without knowing each other.

If one can get together peers in a group of parallel swarms, so that the group of multiple small swarms are merged into a single bigger one, then one can not only improve the file sharing performance, but also can provide more potential peers for localization, which suffers greatly from the lack of peers. Motivated by these observations, we propose a Tracker's tracker system to merge parallel swarms groups to improve availability and performance of BitTorrent and provide feasibility for P2P localization schemes. The administrator can adopt many kinds of peer recommendation in this system, as required.

The remainder of this paper is organized as follows. Related works and background information of BitTorrent protocol are introduced in Section II. We describe our measurement methodology and the data set we collect in Section III. In Section IV, we present the two important observations from our analysis of the data set. In Section V, we further quantify potential benefit of merging swarms based on two metrics we define. Based on these observations, we propose a system to implement the parallel swarms merging in Section VI. Finally, this work is summarized in Section VII.

II. BITTORRENT AND RELATED WORK

A. BitTorrent protocol

In BitTorrent system, client's downloading a shared content starts from a metadata file (with the .torrent suffix name). The metadata file contains information about the

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

resource file, including its length, name, piece length, hashing information, and the URLs of one or multiple trackers, etc.A tracker is the central component, storing and managing shared contents information, and is responsible for helping peers sharing the same resource file to connect with each other by providing a list of peers randomly. The shared resource file maintained by a tracker is called *torrent*. A set of peers sharing the same resource file with the help of a tracker is called a *swarm*. A peer uploading and downloading the shared content at the same time is called a *leecher*. When it holds the whole shared content and uploads only, the peer is called a *seeder*.

Although multiple trackers' URLs may exist in a metadata file, the BitTorrent protocol only allows a peer to associate with one tracker. So, peers sharing a trans-swarm resource file but in different parallel swarms cannot collaborate because they cannot find and communicate with each other.

B. Related studies on BitTorrent

Related studies of recent years can be classified into two categories. One is measurement to understand BT systems [2] [8] [9]. The other is to propose schemes to improve performance [1] [5].

In [2], the authors provide new finding regarding the limitations of BitTorrent systems, e.g., the existing BitTorrent system provides poor service availability, fluctuating downloading performance, and unfair services to peers. The recent measurements [8] show that 82 percent of the active swarms have no more than 10 peers. We find a similar problem on BitTorrent in our work.

Several implementations of traffic localization have been proposed, such as iPlane [5], Ono [11]. Localization requires that peers should preferentially select neighbors in multi-scales (city, AS, ISP) or choose the peers according to the network's status and the preferences regarding the application traffic [1]. Choffnes and Bustance [11] propose a method for localizing BitTorrent traffic without need for any additional infrastructure such as iTrackers [1] or cooperation between applications and ISPs. They use a plugin called Ono for a BitTorrent client which does the peer selection. The challenge is that what BT users concern is to maximize the speed of replication, while ISP wants to make the best use of the network and avoids congestion, as well as too much inter-ISP traffic. Furthermore, many solutions proposed by ISPs to control or localize P2P traffic ignore downgrade of availability caused by localization and the fact that all the torrents could not be localized.

Some recent works [4] [6] on the BitTorrent locality provide a characterization of swarms and the distribution of peers to autonomous systems (ASes). The results suggest that most ASes do not have enough potential for locality. This is a real problem faced by every localization scheme in reconciling the interest between ISPs and BT users. So it is more favorable for the applications of locality enhancements which can provide more peers and peer recommendation.

III. Measurement methodology

The BitTorrent trackers can response to two kinds of HTTP GET requests. The first kind is the most common and called as Announce request. It is used by clients to participate in the torrents. It includes the resource file's identification (info-hash, 20-byte SHA1 hash) and metrics from clients that help the tracker keep overall statistics of this peer. A resource file can be identified exactly by a 20-byte SHA1 hash (info-hash), which is calculated from the data including resource file name, file length, piece length, hash of every piece, etc. If a resource file has different names or is blocked in various ways in different torrents, the resource files will be considered as distinct resource files because they have different info-hash.

A lot of trackers support Scrape request which is the second kind of HTTP GET requests. If the Scrape request includes info-hash of a particular resource file, trackers will return statistics information of the swarm distributing this resource. Otherwise, if the scrape request is without info-hash of any files, trackers will return statistics of all swarms they host including info-hash of the file, number of downloaders (downloaded the complete file), seeders and leechers. For example, by sending HTTP GET request "http://tracker.prq.to:80/scrape", a client can get stats of all swarms in this tracker. In this paper, we develop a client using java to send Scrape requests without info-hash to collect stats of all torrents from as many trackers as we can. We conduct following steps to collect our data set for our analysis:

1. Starting from metadata files. We try to find as many metadata files as possible from websites and ftp sites, etc.

2. Extracting Trackers' URLs. Then we can extract trackers' URLs out of each metadata file. In our measurement, we find about 720 trackers' URLs.

3. Getting Scrape pages. We send Scrape requests to trackers in the trackers' URL set and get all torrents' information.

4. Identifying independent trackers. We remove the redundant trackers whose content is of high similarity. We find that many trackers' URLs seem to be totally different, but pointing to a same IP; and the swarm information of these trackers is identical. In addition, some trackers having different URLs and IP addresses maintain the same content. We remove these two kinds of redundant trackers and identify 41 independent trackers from many countries and areas, e.g., China, America, Malaysia, Holland, Sweden, Germany, France, Hong Kong, Taiwan, etc. We captured the stats of 2258909 swarms in the 41 independent trackers which support Scrape convention.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

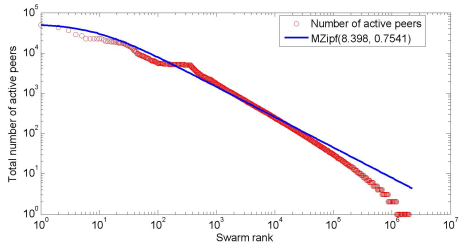


Figure 1. The number of active peers in different swarms (log-log)

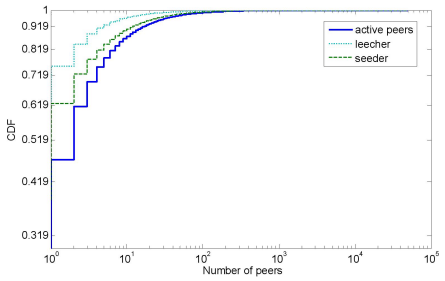


Figure 2. The CDF of the number of active peers of each swarm (log-log)

IV. Observations on Swarms

In this section, we begin to focus on the characteristic of swarms according to our measurement.

A. A Snapshot of swarms' size

Figure 1 shows the number of active peers in different swarms, ordered from the largest to smallest swarm. We can find the curve of the figure can be fitted as a Zpif-Mandelbrot distribution (q=8.398, s=0.7541). Not as shown, a little weak correlation coefficient between the number of seeders in a swarm and that of leechers is 0.6113. The number of seeders is more than that of leechers in about 51.5% swarms. Seeders and leechers have the same number in about 16.2% swarms.

Typically, when a tracker receives Announce requests for a peer list, the tracker chooses at most 50 peers randomly from all active peers in the swarm and returns this peer list to clients. A client seeks to maintain connections to a number of peers, e.g., 30-55 in the official BitTorrent client version 3. When the client maintains fewer connections, it re-contacts the tracker and tries to obtain additional peers.

Figure 2 shows the CDF of the number of active peers of each swarm. We find, unfortunately, most torrents have a small swarm according to our measurement: more than 80% swarms have less than 10 active peers; over 90% swarms have less than 10 leechers or seeders. The downloading performance of peers in these small swarms may be very poor, and it is difficult for peers in these swarms to complete resource file downloading.

B. Parallel swarms

In our measurement, we find 1409659 unique info-hash from 2258909 torrents. Let us define the number of parallel

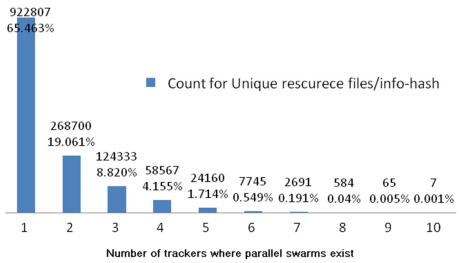


Figure 3. The number of trans-swarm resource files across different amount of trackers

swarms distributing the same resource file F_i as N_i. In Figure 3, we plot the distribution of N_i for all resource files in our data set. We can see that MAX(N_i)=10, and 65.46% of unique resource files are not trans-swarm resource files, which means they only appear at one trackers. The left 34.54% of unique resource files are shared by 40.85% torrents holding trans-swarm resource files. These resource files can benefit from merging parallel swarms potentially.

The total number of requests of all resource files on a tracker reflects how many times BitTorrent users use this tracker to download contents they need since the tracker has been built. The total number of active peers of all swarms maintained by a tracker reflects how many peers the tracker is serving at that time. The number of resource files in the tracker shows how many shared contents for which the tracker can provide service. All these three factors can evaluate a tracker's influence and scale in BitTorrent system.

In our study, we further analyze the 41 trackers and plot the number of trans-swarm resource files that each tracker hosts in Figure 4. In order to study the relationship between the total of trans-swarm resource files of a tracker and other factors, we also plot total of torrents, the number of active peers and BT download amount of all trackers. In this figure, x-axis are trackers sorted in ascending order of the number of trans-swarm resource files in each tracker, and y-axis denotes the percentage of each data item in different trackers.

The tracker size is the total of torrents in one tracker which could have many trans-swarm resource files. We can see that the tracker size follow the similar trend as the number of trans-swarm resource files in each tracker in Figure 4. The correlation coefficient between them is 0.9551. However, it is hard to establish the relationship between the number of active peers and the number of trans-swarm resource files of the tracker.

In Figure 5, all trans-swarm files are divided into 10 parts according to their N_i. We plot the average number of leechers, active peers and BitTorrent download amount of trans-swarm resource files in each part. All three curves show that resource files with higher N_i tend to be downloaded by more peers. We can conclude that the resources files that gain more popularity are more likely to appear at more swarms or trackers.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

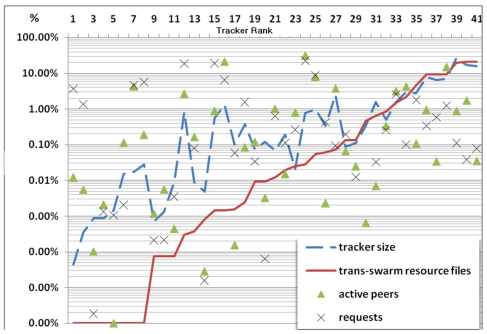


Figure 4. The percentage of trans-swarm resource files and other metrics of 41 trackers under study (semi-log)

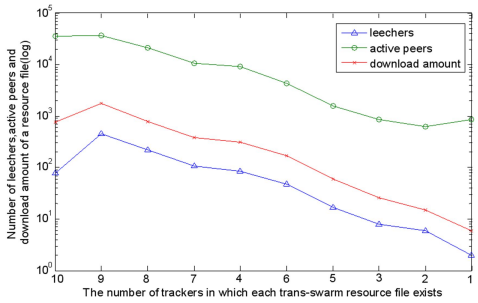


Figure 5. The average number of leechers, active peers and BT download amount of trans-swarm resource files (semi-log)

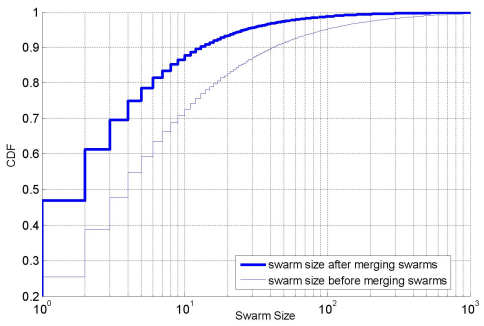


Figure 6. The CDF of swarm size before and after merging (semi-log)

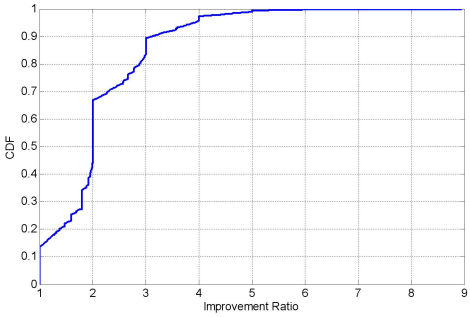


Figure 7. The CDF of improvement ratio of trans-swarm resource files

V. Potential Benefit of Merging Parallel Swarms

The number of active peers is particularly important for BitTorrent where service availability relies purely on the participation of leechers and the volunteer seeders. In general, it is hard to define whether a swarm should be considered to be small or large. But, the result of our measurement reveals there are too few peers in most swarms. From Figure 2, we can find 46.9% swarms have less than two peers. Additionally, it is important to provide more active peers not only for smaller swarms but also for larger ones, because localization in some of larger swarms is still poor in nature [4]. We have found a lot of groups of parallel swarms in our measurement, as illustrated in Figure 3. Also, it would be helpful for implementing all kinds of peer recommendation as well. In this section, we would like to study the benefit of merging parallel swarms.

Based on the snapshot of all swarms, we analyze potential benefit of merging swarms from two aspects: increase in quantity of active peers and the value of swarms.

A. More active peers

Figure 6 shows the CDF of swarm size before and after merging parallel swarms. The numbers of torrents with less active peers are expected to be significantly reduced after merging swarms. The number of swarms with no active peers has been reduced from as many as 268760 to 175125 in this snapshot. It means about 35% of resource files in the swarms with no peers before can be downloaded now. 86.62% swarms had less than 10 peers before merging, and now the percentage becomes 70.96% after merging parallel swarms.

B. Enhancing the value of swarms

As mentioned before, the number of active peers is particularly important for the performance of a swarm. The meaning of performance includes many things, such as the usability, the network efficiency [9], the stability and the accessibility. So, it is not easy to evaluate the importance of swarm size. We use Metcalfe's Law to evaluate and analyze quantitatively the value of swarms. A BitTorrent swarm is an overlay network actually. Metcalfe's Law states that the total value of a network of a number of nodes (n) is proportional to n (n-1)/2, proportional to n^2 asymptotically, if we regard all active peers and connections present the same value.

We define the value of a swarm as F(N_i)= αN_i^2. N_i is the number of active peers in swarm i, α is a constant. Then we use this model to calculate the value of merged swarms. If a trans-swarm resource file shared by a group of m parallel swarms existing in independent trackers, and the percentage of peers in the swarm of tracker i is x_i (i=1...m). So the improvement ratio is defined as $1/\sum_{i=1}^{m} x_i^2$, which is in fact the ratio of the whole swarm value to the sum of each small swarm value.

Figure 7 shows the CDF of improvement ratio of trans-swarm resource files. We observe a sharp increase in the

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

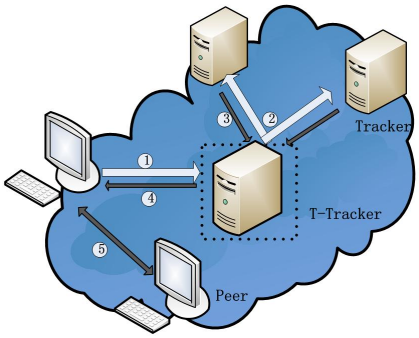


Figure 8. BitTorrent System with T-Tracker

number between 1 and 3. We see that value enhancement takes place in nearly 86% trans-swarm resource files, not in all of them. There are two reasons: 1) all the swarms in some groups of parallels warms have no peers at all. 2) Only one swarm has peers while other swarms do not. In these conditions, the improvement ratio is 1, which means there is no improvement.

VI. TRACKERS' TRACKER ARCHITECTURE

From the analysis in Section V, we can see that merging swarms would be beneficial for BitTorrent system. In this section, we will propose a new architecture to merge parallel swarms for performance improvement of BitTorrent system.

A. T-Tracker

We introduce a proxy tracker in traditional BitTorrent system, called T-Tracker (trackers' tracker) as shown in Figure 8. T-Tracker is deployed by ISPs or some third-party organizations. It is responsible for two tasks: 1) Retrieve peer lists from multiple standard trackers on behalf of clients; 2) After T-Tracker gets peer lists, it can choose which peers to be returned to clients. This is called as peer selection strategy of T-Tracker which can be very flexible, such as localization.

Obviously, both clients and ISPs can benefit from T-Tracker. For clients, there might be more active peers sharing the same resource file in multiple swarms, which can improve the downloading performance. Another advantage can be gained by clients when all the trackers specified in the metadata file are unavailable. Because T-Tracker may seek out other trackers sharing the same file; for ISPs, T-Tracker provides an approach to "tame the torrent" generated by BitTorrent, e.g., localization, which is beneficial for both ISPs and peers [1] [5].

Client software needs to be modified to exploit the benefits brought by T-Tracker. A peer can still use current client software to contact with standard trackers as before, if it doesn't want to rely on T-Tracker. But, we believe peers have incentives to use new software for the advantages mentioned above.

B. Sharing procedure

The file sharing steps in BitTorrent system with T-Tracker are listed as follows:

0. T-Tracker maintains a database of resource files, containing the data of standard trackers. It should maintain not only the information of the trackers but also the addressed and the state of peers. We will talk more about the database below.

1. When a client wants to download a resource file, the client constructs a query for a peer list to T-Tracker through Announce request. The client needs to report to the T-Tracker with standard trackers specified in metadata file when it first connects with T-Tracker. The information of trackers comes from the metadata file which the client gets.

2. T-Tracker searches its database and find which trackers the file may exists in. And then it asks standard trackers which hold the file for the active peers and it acts like an ordinary client when doing so. These trackers may be not the ones which the client reports to T-Tracker.

3. The standard trackers response to the T-Tracker with their peer lists.

4. T-Tracker collects peer lists and recommends a selected peer set according to T-Tracker's peer selection algorithm for the client. The peer set also contains which trackers these peers come from.

5. The client receives the peer set and begins to share the file with these peers.

If a client fails to maintain a fixed number of connections, it re-contacts the T-Tracker to obtain additional peers. Then T-Tracker re-contacts standard trackers to get more active peers and then forwards the selected peers to the client.

C. Implementation of T-Tracker System

On an implement level, we only need to do a little extension to BitTorrent protocol by adding a process of the client connecting with T-Tracker instead of standard trackers specified in metadata files. And T-Tracker serves BitTorrent users like a standard tracker.

In BitTorrent protocol, some information about upload and download rates, joining, completing or leaving event is sent to the tracker periodically for statistics gathering. In our system, clients need to report their state to the T-Tracker as well as standard trackers whose peers the clients are connecting with.

In step 0, we assume that T-Tracker has the knowledge of from which trackers it can get peer lists of the resource files that the clients are interested in. In fact, T-Tracker gets this knowledge in the following two ways: 1) A client needs to notify T-Tracker of the information of standard trackers from a metadata file when it begins to download a resource file. T-Tracker holds an active tracker set, and adds the new trackers to the tracker set if it can connect with these standard trackers. T-Tracker updates the resource

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

file database as well. 2) T-Tracker sends Scrape requests to the trackers at intervals in their spare time.

When T-Tracker receives a request for a content that it doesn't maintain before, T-Tracker only sends requests to the trackers which the client reports instead of all trackers in its active tracker set. This is because sending requests to all trackers will bring a great burden to standard trackers if they don't host that shared content.

For the trackers that do not support Scrape requests, T-Tracker can only find the related trackers through the first way. When T-Tracker is used by more BitTorrent users and works for a longer time, it will find more trackers and cover more parallel swarms.

D. Discussion on working load

T-Tracker hosts files on lots of trackers it can connect with. But the load of T-Tracker is not unbearable due to following reasons:

1. Bandwidth requirement of T-Tracker: the bandwidth requirements of the tracker are very low. The tracker's responsibilities are strictly limited to helping peers find each other. The network bandwidth is consumed mainly between peers, and the cost of uploading pieces of the file to downloaders is carried by peers, not by T-Tracker, due to the advantage of P2P.

2. The number of torrents and users: there would be not too many users or torrents for T-Tracker. In our measurement, the total number of torrents we find is only 2.77 times more than that of the tracker with the most torrents, and the number of active peers which need to report their stats to T-Tracker is 2.24 times more than that of the tracker connecting with the most peers.

3. T-Tracker will send Announce requests to related standard trackers, only when users ask for active peers to download a resource file. T-Tracker can cache results from different users to improve T-Tracker query performance.

4. For file sharing system, time-delay is considered to be tolerated. So tolerate delay can decrease pressure for T-Tracker.

VII. Conclusion

We performed measurement for 41 independent BitTorrent trackers, which collectively maintain state information of a combined total of over 2.2 million unique torrents. The measurement data we present in this paper shows that peers of most swarms suffer from the lack of active peers. And a lot of peers in parallel swarms can be connected with each other to improve performance.

Based on our measurement of BitTorrent System, we propose a Trackers' tracker (T-Tracker) as an extension to BitTorrent system, aiming to look for a solution to enable inter-tracker collaboration. Our scheme can enhance availability of small swarms. Furthermore, all kinds of localization/traffic improvement solutions always suffer from the lack of active peers to recommend. Our T-Tracker can draw out the potential of BitTorrent system, and provide more active peers sharing the same file by merging multiple smaller swarms into a single one. Therefore, T-Tracker provides useful means for ISPs to carry out P2P localization. It can also be used easily with other peer selection algorithms by ISPs for other traffic engineering goals.

Acknowledgment

Our work is supported by "973 Program" of China under Grant No. 2009CB320505, the National Science and Technology Supporting Plan of China under Grant No. 2008BAH37B05, the Natural Science Foundation of China (NSFC) under Grant No. 61170211, Specialized Research Fund for the Doctoral Program of Higher Education (S-RFDP) under Grant No. 20110002110056, and "863 Program" of China under Grant No. 2008AA01A303 and 2009AA01Z251.

References

[1] Haiyong Xie, Y. Richard Yang, Arvind Krishnamurthy, Yanbin Liu, and Abraham Silberschatz, *P4P: Provider Portal for Applications*, SIGCOMM Comput. Commun. Rev., vol. 38, no. 4, 2008, pp. 351-362.

[2] L. Guo, S. Chen, Z. Xiao, E. Tan, X. Ding, and X. Zhang, *Measurements, Analysis, and Modeling of BitTorrent-like Systems*, Proc. of the 5th ACM SIGCOMM conference on Internet measurement, CA, USA, 2005, pp. 35-48.

[3] M. Izal, G. Urvoy-Keller, E. W. Biersack, P. A. Felber, A. Al Hamra, and L. Garces-Erice, *Dissecting BitTorrent: Five Months in a Torrent's Lifetime*, Passive and Active Network Measurement, 2005, pp. 1-11.

[4] H. Wang, J. Liu, and K. Xu, *On the locality of BitTorrent-based video file swarming*, Proceedings of the 8th USENIX International Conference on Peer-to-Peer Systems, Boston, 21 April 2009, pp. 1-6.

[5] Harsha V. Madhyastha, Tomas Isdal, Michael Piatek, Colin Dixon, Thomas Anderson, Arvind Krishnamurthy, and Arun Venkataramani, *iPlane: An Information Plane for Distributed Services*, Proc. Symp. Operating Systems Design and Implementation, 2006, pp. 367-380.

[6] Tobias H., F. Lehrieder, D. Hock, S. Oechsner, Z. Despotovic, Kellerer W., and Michel M., *Characterization of BitTorrent Swarms and their Distribution in the Internet*, Computer Networks, 2011, pp. 1197-1215.

[7] S. Le Blond, A. Legout, and W. Dabbous, *Pushing BitTorrent locality to the limit*, Computer Networks, 2011, pp. 541-557.

[8] Chao Zhang, P. Dhungel, Di Wu, and K. W. Ross, *Unraveling the BitTorrent Ecosystem*, Parallel and Distributed Systems, IEEE Transactions, vol. 22, no. 7, 2011, pp. 1164-1177.

[9] A. Al Hamra, A. Legout, and C. Barakat, *Understanding the properties of the bittorrent overlay*, INRIA, Sophia Antipolis, July 2007, pp. 01-18.

[10] A. Bellissimo, B. N. Levine, and P. Shenoy, *Exploring the use of BitTorrent as the basis for a large trace repository*, Tech. Rep., University of Massachusetts, Amherst, 2004, pp. 04-41.

[11] D. R. Choffnes and F. E. Bustamante, *Taming the Torrent: A Practical Approach to Reducing Cross-ISP Traffic in P2P Systems*, Proc. ACM SIGCOMM, Seattle, WA, USA, Aug. 2008, pp. 363-374.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

A Framework for Classifying IPFIX Flow Data, Case KNN Classifier

Jussi Nieminen, Jorma Ylinen, Timo Seppälä, Teemu Alapaholuoma, Pekka Loula
Telecommunication Research Center
Tampere University of Technology, Pori Unit
Pori, Finland
jussi.nieminen@tut.fi, jorma.ylinen@tut.fi, timo.a.seppala@tut.fi, teemu.alapaholuoma@tut.fi, pekka.loula@tut.fi

***Abstract* — Flow-level measurement applications and analysis in IP networks are inevitably gaining popularity, due to the unstoppable increase in the amount of transmitted data on the Internet. It is not reasonable or even possible to examine each and every packet traversing through a network. Our research focuses on passive flow level data classification and characteristic identification. To be more exact, our goal is to design a framework for extracting certain classes, feature(s) and behavior from IP flow data. One of the goals is to achieve this without examining the payload of any of the IP packets and without compromising the anonymity of the flow counterparts. Traditionally, Deep Packet Inspection or port mapping techniques have been applied for this purpose. In this paper, we present an alternative framework for classifying the IP traffic, which we aim to utilize in the future for separating classes from the IP traffic for information security purposes.**

Keywords-Flow; IP; IPFIX; KNN; Classification

I. INTRODUCTION

In this paper, we study the possibility of identifying traffic characteristics from IP traffic, and more precisely from the IP/TCP/UDP/ICMP header data. We utilize the KNN Classifier method (K Nearest Neighbors) through passive data analysis on IPFIX [1] [2] flow data. The motivation for our research comes from the area of information security. We are keen on finding methods for separating classes from the data in order to be able to identify a measurable unit (IPFIX flow in this case) for example as normal or malicious in future analysis work. In this paper, we present a framework, which can be utilized for that purpose. Our research relies on total anonymity. The IP-addresses are either anonymized or cut off prior to analysis execution. The payload of each IP packet is cut off in the data capture phase, so all the details compromising the user privacy of the connection counterparts are discarded.

The KNN Classifier method determines the class of a new data point based on its K-nearest neighbors in a selected feature space. The class that exists the most among the K-nearest neighbors is given to the test data point. The KNN Classifier is based simply on the distance metric of data points. The Euclidean distance metric is the most common one, while also other metric methods are available. This obviously means that a variety of different KNN implementations have been introduced.

Our data for the analysis was captured from a large-scale local area network. The selected network is known to have a large amount of hosts and good set of services active. It is also known that the information security policy doesn't restrict the usage of any service in the network. This is a clear advantage from the analysis point of view, because the captured data is as pure as it can be without any restrictions or filtering in any way at any point.

The data was captured from the network and stored to disk in IPFIX format. In the analysis phase the data was first divided into two classes. We use a class distribution of WWW-type traffic versus other traffic in this paper. WWW as a service provides interesting viewpoints for future analysis, as it is commonly used, uses standard port numbers, and therefore also has a lot of information security aspects. The following step was to select the parameters for the classification execution. K-fold cross-validation was used as the classification framework to determine the best value for the constant 'K' in KNN-Classifier. Another important factor was to select suitable input parameters (features) for the classification. We came up with a set of three parameters. Once the parameters were selected, the actual classification was executed. As a result, the details were obtained about how the classification succeeded. The results were studied and written down, along with conclusions and observations about the functionality of the analysis framework and the methods used. Based on the analysis, we present our framework for classifying IP Flow data. In addition, some thoughts on how the results could be utilized in practice are provided.

This paper consists of seven sections. In the next section, the related work in the field of IP-traffic data classification is presented and analyzed briefly. In Section three, the data is presented in terms of how the data is obtained, how it is pre-processed, what is the total amount of data and how it is connected to real life time-wise. The theory behind the analysis is presented in Section four. Section five presents the analysis framework and the execution of each step during the analysis. The observations and results of the analysis are presented in Section six. Finally, conclusions and future plans are given in Section seven.

II. RELATED WORK

The quest for finding solutions for extracting IP-traffic characteristics from IP traffic has been a challenge for

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

researchers since the early years of the Internet. Words like generic, dynamic, effective, intelligent and self-learning are all features of a desirable solution. DPI (Deep Packet Inspection) techniques have been found effective to a certain degree by several studies. The drawback of DPI techniques is that you have to examine the payload of each and every IP packet, which is very expensive from the resource usage point of view in large-scale IP networks. Payload inspection might also compromise the anonymity of the connection counterparts, which might be unacceptable in some cases. Bendrath has examined the effects of DPI from the Internet governance point of view very carefully in his research [3]. Along with DPI techniques a variety of transport port-based methods have been introduced. These methods rely on a static port number mapping, where a certain port number is linked to a certain service in the network. Direct mapping is obviously effective but somewhat unreliable due to the possibility of port number faking or misuse. For example, Karagiannis et al. have made similar observations in their research work [4].

Various classification and clustering methods for grouping IP traffic have been introduced over the years. The focus is typically similar to this paper. By defining an analysis framework and utilizing a chosen method, a solution to a given problem is presented. Kumpulainen et al. have successfully utilized multi-level K-means clustering for separating traffic classes and behavioral patterns from IP-traffic [5]. Karagiannis et al. have used their own approach by classifying IP traffic in a three-layer classification setup. Their framework classifies the data in social, functional and application levels [4].

Moore et al. have successfully utilized a Naïve-Bayes classifier for identifying application details from network traffic. They achieved a significant improvement to the classification result by training the classifier with several simple operations [6].

In their paper, Yarifard et al. study unsupervised learning methods for identifying application specific behavior patterns from IPFIX flow data. They studied three different clustering algorithms and got good results from K-means and SNN-clustering [7].

Nguyen et al. examined a vast variety of machine learning techniques for classifying internet traffic in their paper [8]. This is an informational study rather than a survey focusing on mining the data with different methods. It provides a good overview of the methods studied and their benefits and drawbacks.

Countless studies with different goals and problem settings are available. Anyhow, there are not many papers focusing on IPFIX flow data classification. Furthermore, the use of K-Nearest-Neighbor Classification algorithm is rare in the area of IPFIX flow data classification. Based on the work of other researchers, and by our previous experience on IP traffic analysis, we decided to present our framework for IP traffic classification purposes.

III. THE DATA

The analyzed dataset was generated from a three-day trace taken in April 2011. The tracing was executed over a period of three weekdays from Tuesday to Thursday. The monitored network can be considered as a Wide Area Network (WAN) or a large-scale local area network. We use the latter term in this paper. The target network is ideal for capturing IP traffic for analysis purposes, because the information security policy of the administrating organization allows the use of any service as long as it is not illegal, does not violate the user privacy or disturb other users of the network.

The IPFIX format was used for the flow data. The IPFIX format was selected because it is the leading flow standard at the moment in terms of the level of standardization. IPFIX is based on Netflow [9], a trademark of Cisco Company. The IPFIX flow data was generated with the Maji program, provided by the WAND research team from the University of Waikato in New Zealand [10]. Maji relies heavily on the libtrace data capture library [11], which clearly also played a very important role during the data capture phase. Libtrace was also provided by the WAND -research team. Maji supports a variety of IPFIX-compliant parameters. From these parameters, we gathered a compact set of variables suitable for our purposes in order to avoid unnecessary load during the capture phase and in order to optimize the usage of storage space. The IPFIX flow data was first stored to hard disk in SQLite database format [12], from which we were able to post-process the data to CSV format for the analysis execution.

We further reduced the dataset to include only needed parameters. The dataset ended up holding in a 123 million rows, i.e., IPFIX flow records. For each flow record there is a set of parameters as follows:

1. Feature identifier (WWW-type or Other)
2. Source Transport Port
3. Destination Transport Port
4. Number of transmitted packets within the flow
5. Number of transmitted octets within the flow
6. Maximum Time To Live value within the flow

Parameters one, two, and three are related to the class definition. We separated the traffic that looks WWW-related from the rest of the data. We call this phase the basic profiling phase. The purpose of basic profiling is to highlight the desired feature or traffic class. After basic profiling we should be able to trust that the profiled traffic is what it looks like with acceptable probability. Parameters four, five and six were chosen as input parameters for the actual classification phase. They were selected after some preliminary testing and visual data mining of the flow parameters. As a guideline for parameter selection, we used the characteristics of the KNN Classifier, meaning that we tried to find parameters whose value distribution was somehow clustered or packed into clear groups within the value range. The better these conditions are met, the better is the probability of finding the correct class for a test

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

observation. The parameter selection also involved a behavioral factor. We went through the flow parameters and wrote down characteristics typical for traffic that looks like WWW traffic and then used those facts in the selection.

IV. THE THEORY

KNN classification belongs to the supervised learning methods in the field of machine learning techniques. Furthermore, KNN classification is a non-parametric learning method, meaning that it does not assume any known prior distribution. Naïve-Bayes for example assumes that the data follows normal distribution. Non-parametric methods are sometimes referred to as instance-based or memory-based methods.

KNN-Classifier is a simple, yet computationally expensive classification method. It is based on the distance metric of the classification features. The classifier algorithm is given a feature vector as an input and it places it in the feature space of the training dataset for comparison. Based on the constant 'K' and the selected distance metric, the algorithm computes the class for the new data point based on which class exists the most within the K nearest neighbors of the test feature vector.

KNN requires the whole training dataset to be available whenever a new test data point is set under classification. The classifier computes the distance of each test data point to each and every data point in the training dataset. This limits the use of KNN Classifier to being suitable mainly for passive data analysis rather than real-time applications.

The mathematics behind KNN Classifier is very simple. We have to compute each feature vector in the test data in order to define its location in the test feature space. Then each feature vector in the training data space is computed to define its location. Only after theses operations can we compare the locations of the test feature vector against the feature vectors inside the K neighbors in the training feature space. On the basis of that comparison we obtain the class for the test feature vector. Details about the mathematics are available in references [13] and [14].

There are four major questions one must ask himself/herself when designing a KNN-Classifier:

1. What is the characteristic in our dataset that defines the class distribution, and how should it be obtained, if not natively present?
2. What is the optimal value for the neighbor constant 'K', and how should it be obtained?
3. What distance metric should we use with this particular dataset?
4. What are the features in our dataset we need in order to be able to classify each test sample with the best possible accuracy and without redundancy?

Once these questions are answered, the rest is a straightforward matter of executing of the classification. Our framework binds together the workflow from the data capture and pre-processing to the result analysis.

V. THE EXECUTION

The execution stage defines the analysis framework and the workflow of the analysis process. The framework consists of six phases.

First, the data is captured from the target network. We are focusing mainly on the analysis methods, so this phase is not described in detail here.

The second phase involves parameter reduction, which means the removal of unnecessary flow parameters from the data. Parameters such as IP-addresses (anonymized) and timestamps are not needed in the classification phase, but are essential when the flow record is generated in the capture phase.

In third phase, the desired class distribution for the dataset is generated, if not natively present in the data. This phase is called the basic profiling phase. The purpose of this step is to ensure that each flow record belongs explicitly to one and only one class. We generated two classes: 'WWW-type' and 'Other' by using the known transport port numbers 80, 8000 and 8080.

The fourth phase deals with the classifier training, i.e., configuring the classifier. The first step of the classifier training consists of selecting the classification features. In the second step of training the distance metric for the classification was selected. We decided to use the Euclidean distance metric as it is by far the most common metric method used in data analysis in general. As the third step of the training, the KNN algorithm requires the neighbor constant 'K'. To determine the best value for 'K' we executed KNN Classifier with K-values 1-10 in a 10-fold cross-validation setup. We took a sample data of IPFIX flow data and divided it into 10 subsets of equal size. Each subset in turn acts as a test data and the other 9 subsets are combined to act as training data. All in all 100 separate classification executions are obtained, one for each combination of tested values of 'K' versus each possible cross-validation setup. The K-value with the best average classification success ratio should be selected for the actual analysis phase.

The fifth phase is the execution of the actual classification with the full dataset, and the final phase of the analysis is to analyze the results and make observations and conclusions. The analysis framework and workflow is described in Figure 1.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

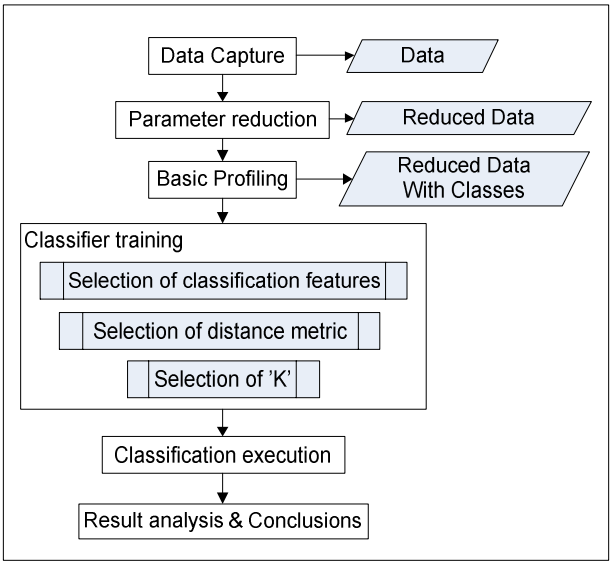


Figure 1: Analysis framework & workflow

In the actual classification phase, the three-day dataset was split into three separate datasets, as presented in Figure 2.

As we knew the week-day of each sample, we decided to split the dataset day-wise (N=3) instead of splitting the dataset into subsets of equal size. This meant that we could compare possible similarities and differences in the classification ratio between the days. A test data to Training data ratio of 1 to 9 (n=10) was used in each execution.

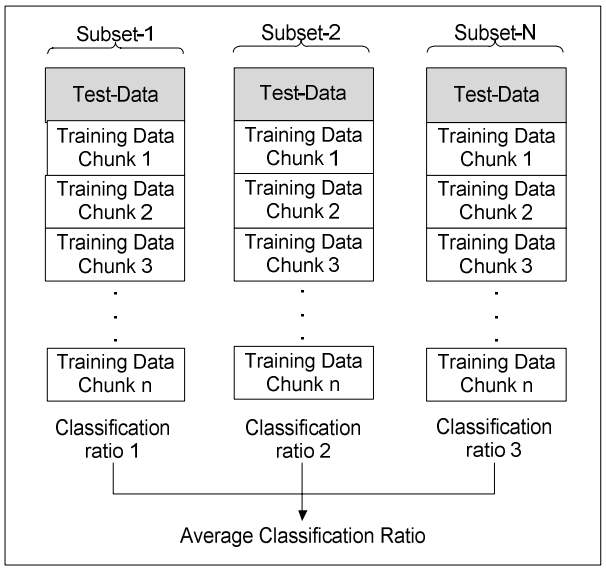


Figure 2: Test Data vs. Training Data setup

The arrangement in Figure 2 was used for two major reasons: the dataset size and to minimize the possible behavioral factor related to a certain day in the dataset. The total average classification ratio over the three-day daytime datasets was calculated in order to lighten the load and resource consumption of the classification execution. The behavioral-based division of the data derives from the fact that the amount of traffic and variety of services used in the network might be dependent on the day of the week.

KNN Classification was executed using the Euclidean distance metric. It is a straightforward and fair method for ranking observations. Moreover, the data in hand does not have any special characteristics that would require the use of more complex distance metrics. Distance-based weighting was not applied in this paper as a classifier training method.

VI. THE RESULTS

The results are handled in four parts. First we discuss the pre-processing of the dataset and the results obtained from the basic profiling phase. Then we handle the classification input parameter selection process. Subsequently, we go through the results of the actual classification. Finally, we discuss the functionality of the framework as a whole. In conclusion, we should have a view of how the applied classification mechanism and the framework in general suit the classifying of IPFIX flow data and distinguishing the feature vs. parameter relations in IPFIX flow data.

The basic profiling phase gave us a dataset with a class distribution of two classes: WWW-type and Other. We have not used the class name WWW, because we believe we can never achieve 100% success ratio in the basic profiling phase. There is always a room for error, such as measurement errors for example. However, in the case of behavior like WWW type behavior, we can be sure with an acceptable probability that the majority of the traffic traversing through ports 80, 8000 and 8080 is WWW-related. For comparison we could take DNS traffic for example. DNS is a service, which is tightly associated with port number 53. Furthermore the DNS query is static in terms of packet and flow record structure. These types of services are easier to identify in basic profiling and also easier to classify with the aid of flow features because the basis for class distribution is sufficiently solid. In this paper, the data to be classified was distributed to the aforementioned classes as follows:

TABLE I. DAILY FLOW COUNT AND CLASS DISTRIBUTION DETAILS

	Day 1	Day 2	Day 3
Flow count	41 751 116	40 350 846	40 946 094
WWW-type	12.01 %	12.27 %	11.39 %
Other	88.99 %	87.73 %	88.61 %

The data distribution was surprisingly even between the daily datasets. The traffic profile of the monitored network is very constant, at least where WWW-type traffic is concerned. The amount of WWW-type traffic was around 12% over the whole three-day dataset.

Classification input parameter selection was done by executing the classifier under several different setups and within several iterations. The goal was to train the

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

framework to be as generally applicable as possible for the analysis of IPFIX flow data.

In the selection of classification features, we aimed to find flow parameters that were descriptive from the client-server type of services aspect such as WWW. Another aim was to restrict the number of parameters. Our goal was to have 2 or 3 parameters to continue with. It is a clear benefit if the classification feature space has no more than three dimensions. For example, illustrating the feature space and the classification results is much easier that way. The third objective is a general goal for the classification parameter selection. The parameters should have as little redundancy as possible. Redundant parameters do not bring any distinguishable or useful information to the feature space. It is not sensible to use for example three redundant parameters if one parameter provides the same information for the classification execution. We observed that the data profiled as WWW-type consisted either of flows with a very small amount of transmitted packets or flows with a large amount of sent packets. It seemed to be a typical behavioral pattern for this type of traffic, so we decided to select the count of sent packets in the flow as one parameter. One would assume that the amount of sent packets within a flow would follow the same pattern, but in this case it did not. We decided to add it to our classification feature space as another parameter characterizing the flow without redundancy. The maximum TTL parameter was selected because it seemed to have a clear distribution into separate groups within its value range and because it was not redundant regarding the other two selected parameters.

The distance metric selection did not involve any data mining or any other characteristic examination of the data. No weighting algorithms were used either when determining the distances of the data points. The eEclidean distance metric is a clear and simple method for calculating distances between data points. Furthermore, it adds extra value to the illustration of the data since the data points can be presented and compared as a vector in a three-dimensional space.

The cross validation for examining the best value for the 'K' gave us surprisingly good results with all the tested K-values. The average classification success ratio was over 97 % and within 0.5 percent with K-values of 3-10. As a guideline, low odd values should be preferred. The best classification success ratio was achieved with a K-value of 9, both in scaled and unscaled feature space. As the difference in success ratio was very small we faced the problem of whether to go on with value 9 or to select a smaller odd value like 3 or 5 for the actual classification, which had also had a very good classification ratio throughout the cross-validation execution. Higher K-values lead to a more noise-tolerant system, but on the other hand it makes the class distribution less distinct within the k data points. Figure 3 illustrates the average classification success ratio both in un-scaled and logarithmic-scaled feature space with K-values from 1 to 10. The effect of scaling on the classification success ratio was very low. On average the success ratio increased by only about 0.2 % compared to the unscaled feature space. Our interpretation of this phenomena is that although the value ranges of Packet Count and Byte Count features are higher compared to the Maximum TTL feature, most of their data is located in the lower part of the value range, as are the values of Maximum TTL values. Therefore the effect of the logarithmic scaling has only a minor effect on the classification results.

Figure 3: Average classification ratio values from the Cross-Validation, Logarithmic scaling versus unscaled feature space, K = [1,10]

We decided to use the K-value 5 for the actual classification. None of the tested K-values in the cross-validation provided significantly better results than the others, and low odd values are typically recommended.

Several interesting pieces of information were obtained from the actual classification phase:

1. The average classification success ratio over the three classifications executed for the daily subsets.
2. The classification success ratio from each of the daily classification executions
3. The ratio of unsuccessful classifications per original class, i.e., how many 'WWW-type' flows were classified as 'Other' and vice versa.

The average classification ratio tells us the overall performance of the classification framework. The average classification ratio was 93,02 %, as shown in Table 2. This result is very good and the daily classification ratios are also very close to each other. This means that the similarity level of the daily subsets is high and the level of activity and WWW-type traffic behavior in this particular network is close to the same on different days of the week. Here we have one example of how this framework can be utilized, as a method for finding out the overall behavior of the dataset.

TABLE II. DAILY AND AVERAGE CLASSIFICATION SUCCES RATIOS

	Success ratio
Day 1	92.91 %
Day 2	91.76 %
Day 3	94.39 %
Average	93.02 %

The characteristics of the unsuccessfully classified flows were examined. The unsuccessfully classified flows are

interesting because they somehow differ from the typical behavior of the root class. We back-traced the unsuccessfully classified flow records back to the original data. The results were once again surprising. A clear majority of the unsuccessfully classified flows belonged to the WWW-type traffic profile.

TABLE III. COUNT AND DISTRIBUTION OF UNSUCCESSFULLY CLASSIFED FLOW RECORDS

	Day 1	Day 2	Day 3
Flow count	295884	332649	229839
Original Class WWW	90.86 %	80.16 %	90.33 %
Original Class Other	9.14 %	19.84 %	9.67 %

This might mean that these flows belong to some WWW-based service, which is clearly different from the mass, or even more interestingly, they might be somehow malicious. A clear benefit is also the fact that the amount of data for further analysis is significantly smaller than the amount we started with. It has come down from over 40 million flows to a few hundred thousand rows of interesting data. Clearly, the successfully classified data cannot be totally ignored in the belief that it does not hold in any false positives. However, the first step is to try to identify the true negatives or false negatives from the unsuccessfully classified data. Table 3 illustrates the total number of unsuccessfully classified rows per daily datasets, together with the proportions of the class in the original data.

As a whole the framework performed very well, and therefore can be recommended for feature identification and characteristics examination purposes of IPFIX flow data. The main benefit of the framework in general is that it provides a solid and defined workflow for classification analysis. In many cases the actual workflow is lost behind the results, and the repeatability of the results is therefore compromised. We designed the framework to be solid, yet flexible enough so it wouldn't cause too many restrictions to the analysis work. The framework does not restrict the classification algorithm selection in any way. If some other classifier is used, the top-level framework is applicable as it is. The classification algorithm selection affects to training and classification execution phases. The cornerstones of the framework are the basic profiling phase and the classification training phase. These are clearly also the places for fine-tuning the framework.

VII. CONCLUSION AND THE FUTURE

The goal was to define a framework for classifying IPFIX flow data with KNN Classifier and prove its functionality. The overall classification success ratios were at a very good and promising level throughout the research. Over 90% classification accuracy with a considerably large amount of flow data is an indication of a very good performance. We used cross-validation in the classifier training phase and a three-way classification setup in the actual classification phase in order to prove the classification framework to be solid and robust. In conclusion, we can state that the framework performed well and the results were very promising. They have certainly given us a boost to continue our research.

There are several interesting starting points for the further analysis. Our research is related to information security and the analysis of the WWW-type traffic has many interesting information security aspects, as it is a commonly used and therefore misused service. It utilizes mainly standard port numbering, which means that those ports are typically left open in firewall configurations, thus leaving some space in which the misusers and attackers can operate. In future research we aim to detect the misuse inside WWW-type traffic by trying to point out what is normal and what is not. We aim to do this with total anonymity so that misuse identification and the results analysis is not illegal or harmful to anyone.

Our intention is also to examine the limits of the KNN classifier by further training the classifier. Utilizing the framework with other types of classification methods is also in the scope of interest in the future research. There are public datasets available that can be used as reference datasets and for further validation of the framework.

REFERENCES

[1] IPFIX Working Group, http://datatracker.ietf.org/wg/ipfix/charter/, retrieved: January, 2012

[2] IPFIX Specifications, http://datatracker.ietf.org/wg/ipfix/, retrieved: January, 2012

[3] R. Bendrath. "Global technology trends and national regulation: Explaining Variation in the Governance of Deep Packet Inspection", ISA's 50th Annual Convention "Exploring The Past, Anticipating The Future", New Ypork city, NY, USA, Feb 15, 2009

[4] T. Karagiannis, K. Papagiannaki and M. Faloustos. "BLINC: Multilevel Traffic Classification in the Dark", SIGCOMM'05, Philadelphia, Pennsylvania, USA, August, 2005, pp. 22–26

[5] P. Kumpulainen, K. Hätönen, O. Knuuti and T. Alapaholuoma, "Internet Traffic Clustering Using Packet header Information", Joint International IMEKO Symposium Jena, 2011

[6] A. W. Moore and D. Zuev. "Internet Traffic Classification Using Bayesian Analysis Techniques". SIGMETRICS'05, Banff, Alberta, Canada, June 6-10. 2005

[7] A. A. Yarifard and M. H. Yaghmaee. "The Monitoring System Based on Traffic Classification", World Applied Sciences Journal 5:, 2008, pp. 150-160

[8] Thuy T.T. Nguyen and G. Armitage. "A Survey of Techniques for Internet Traffic Classification using Machine Learning", 2008, ISSN: 1553-877X, pp. 56-76

[9] RFC3954 – Cisco Netflow 9, http://www.ietf.org/rfc/rfc3954.txt, retrieved: January, 2012

[10] MAJI, http://research.wand.net.nz/software/maji.php, retrieved: January, 2012

[11] Libtrace, http://research.wand.net.nz/software/libtrace.php, retrieved: January, 2012

[12] SQLite, http://www.sqlite.org/, retrieved: January, 2012

[13] R. Herbrich. "Learning Kernel Classifiers. Theory And Algorithms". The MIT Press. Cambridge, Massachussets, London, England. ISBN 0-262-08306-X, 2002

[14] X. Wu, V. Kumar, J.R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. Mclachlan, A. Ng, B. Liu, P. S. Yu, Z-H. Zhou, M. Steinbach, D. J. Hand and D Steingerg. "Top 10 Algorithms in Data Mining", Survey Paper, DOI 10.1007/s10115-007-0114-2, pp. 14:1–37

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

A Formal Data Flow-Oriented Model For Distributed Network Security Conflicts Detection

Hicham El Khoury, Romain Laborde,
François Barrère, Abdelmalek Benzekri
IRIT – University Paul Sabatier
Toulouse, France
hkhoury@ul.edu.lb, Romain.Laborde@irit.fr,
Barrere.Francois@irit.fr, Abdelmalek.Benzekri@irit.fr

Maroun Chamoun
Saint Joseph University
Beirut, Lebanon
maroun.chamoun@usj.edu.lb

Abstract— **Network security is inherently a distributed function that involves the coordination of a set of devices, each device affording its specific security features. The complexity of this task resides in the number, the nature, and the interdependence of the mechanisms. Any security service can interfere with others creating a breach in the whole network security. We propose a formal data flow oriented model to detect network security conflicts. Network security services are represented by specific abstract functions that can modify the data flow. We have specified our model in hierarchical Colored Petri Nets to automate the conflicts detection analysis. This approach has been tested on various NAPT/IPsec scenarios to prove that without any a priori knowledge these conflicts can be detected.**

Keywords - network security; security conflict detection; data flow modeling; Colored Petri Nets.

I. INTRODUCTION

Network security is inherently a distributed function that involves the coordination of a set of devices, each device affording its specific security features. Any equipment (end and core devices) involved in a security solution requires a precise configuration. This configuration, determining its behavior, has a local impact on the security service provided by the equipment but also can affects the global network security. If a rule on equipment is poorly defined, the global security might be compromised (principle of the weakest link in the security chain).

The network security conflicts detection is therefore a major problem. Conflicts can be local to one security service, i.e. two rules for the same security mechanism on the same equipment may be incompatible (e.g. one filtering rule permits a data flows whereas another one on the same firewall blocks it). However, conflicts can be distributed too. In this case, the incompatibility can occur between different mechanisms on different equipment playing a role at different levels in the OSI layers (IPsec tunnels blocked by firewalls, HTTP proxy unable to filter HTTP traffic because it has been encrypted by an IPsec gateway, etc). These distributed conflicts are much harder to cope with because they expect to consider the dependencies between different equipments and/or different mechanisms.

It is therefore essential to develop tools to express and validate network security policies. One difficulty resides in the nature of network security information. How to express management information while taking into consideration constraints such as the heterogeneity of solutions, interdependencies between security mechanisms, and the sustainability of expression languages confronted to the fast evolution of technologies? The right level of abstraction should be produced, both independent of the security mechanisms and at the same time faithfully representative of the reality.

To address this problem, we have proposed a formal data flow oriented model. The first version of our model has been presented in [16] and was enhanced in [17]. In this approach, a data flow is represented as a sequence of logical elements to match physical data flow which is a sequence of bytes grouped according to the specifications of the network protocols. The security mechanisms are represented as functions handling these flows. Constraints applied to data flows and security mechanisms point out conflicts that can occur. In this article, we validate our model by specifying it in hierarchical Colored Petri Nets. This formal language is suitable for representing data flows and associated tools such as CPNtools can automate the validation task. In addition, we present how to detect distributed conflicts using our approach thought various NAPT/IPsec scenarios.

The rest of this article is organized as follows. Section 2 is dedicated to related works. Section 3 describes our modeling. Section 4 presents example of modeling. Section 5 illustrates our approach to conflict detection, on implementing IPSec and NAPT, using Colored Petri Networks. Finally, Section 6 concludes and presents our working tracks.

II. RELATED WORKS

Several works have focused on detecting misconfigurations. Their approach consists in modeling the configurations of devices and the network topology. Al-Shaer et al. [1] [5] proposed a classification of the anomalies that can appear in the configuration of one or more firewalls. Alfaro et al. [2] have improved this classification and introduced IDS. Fu et al. [3] endeavored to address the problem of inconsistency of IPSec tunnels and firewalls. Preda [7] considers firewalls, IPsec and IDS. These models describe correctly the reality. However, they are closely attached to limited set of technologies and it is difficult to adapt them to other technologies.

The other approach followed by distributed conflicts detection research considers data flow or IP datagram as the primary entity. Security technologies are then represented has functions applied on data flows. The advantage of this approach is the processing on data flow is related to the

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

abstract data flow model, not to the underlying technology. However, the strength of these solutions depends on the quality of the data flow modeling approach: not limited to one protocol or layer, but also not de-correlated from reality.

Guttman and Herzog [4] proposed an abstract model for IP datagrams by a 3-tuple $<l,k,\theta>$ where l represents the current location of the datagram, k represents the current state of treatment of the datagram and θ is the following IP header representing the history of the actions performed on the datagram. An IP header is also abstracted by a 3-tuple $<s,d,p>$ where s and d are respectively the source and destination and p represents other data. Nevertheless, the formalization is limited to some information contained in the IP header.

Laborde et al. [6] proposed a formal solution based on Colored Petri Nets for the specification coupled with the CTL logic for the analysis. Like Guttman and Herzog, this approach focuses on analyzing the data flow. A network is represented as an interconnection of generic functionality on the data flow: endings flow (terminal devices such as workstations, or servers), filtering functionality (such as firewalls or application gateways), transformation functionality (such as IPsec, NAT, etc.) and channels functionality (to represent the communication media such as WiFi, abstraction of a network). A data flow is represented by a 4-tuple < *efs, efd, r, t* > where *efs* and *efd* are source and destination end flows functionalities, *r* represents the permission used to generate this flow, and *t* represents the list of transformation applied to the data flow. This formalism allows the approach to be independent from technologies. However, the level of abstraction being too high, it does not represent explicitly what has been changed by a transformation. This level of abstraction issue is highlighted by "feasibility analysis" that was introduced in the refinement process, The goal of the feasibility analysis is to validate that something specified at the abstracted level can actually be implemented on the real device.

In a more recent article, Al Shaer et al. [15] have proposed a similar approach. They model the network as a finite state machine where each state depends on the location of IP packets $(ip_s, port_s, ip_d, port_d, location)$. However, they do not consider IP payload in their modeling. As a consequence, they had to add an extra valid bit to address the problem of IPsec encapsulation modeling. Security involving different mechanisms at different layers (TCP/IP stack or OSI model), this modeling approach is limited for describing the entire encapsulation stack.

III. A FORMAL DATA FLOW-ORIENTED MODEL

Our goal is to define a technology independent formalism to detect distributed conflicts (multi-mechanisms, multi-OSI layers). Our modeling approach is data flow oriented; the treatments of different network mechanisms are then seen as functions on flows.

It is important to consider that network security mechanisms are not applied to one single network layer only (firewalls and IPSec are both mechanisms at the IP level). Network security is multi-level (or cross-layer). For example, a VPN can be an IPsec but also L2TP, SSL or SSH. Filtering can be done at the IP level through a firewall and at the data link layer via a switch, or at the application level by a dedicated gateway.

It is also necessary to consider the influence of non "pure security labeled" devices. For example, the installation of a router into an Ethernet network composed only of switches may require the changing of MAC addresses filtering rules. Another example, NAT may have an adverse effect on the enforcement of IPSec VPNs.

A. Analysis of the problem

In concrete terms, a data flow is a contiguous set of bytes with variable size conveyed over a network. This sequence of bytes is divided into logical blocks according to encapsulation protocols. For example, a data flow corresponding to a HTTP request can be seen as < *Ethernet protocol block, IP protocol block, TCP protocol block, HTTP protocol block* >. The bytes in a logical block are not necessarily contiguous. Then, each protocol divides the block of bytes associated with fields in accordance to its description. For example, the control information of the Ethernet protocol are distributed into 14 bytes (destination MAC address, source MAC address, identifier of the encapsulated protocol) at the beginning of the frame and 4 bytes for the control field at the end.

Also, data flow is not static. A data flow evolves during its journey through the network according to the mechanisms implemented on network devices. Some mechanisms can:

- Add new blocks of bytes. For example, an IPsec gateway adds the AH header between the IP block and the UDP/TCP block when this protocol is used in transport mode,
- Remove blocks of bytes. e.g., an HTTP proxy removes the IP block when it receives a data stream,
- Modify fields. e.g., NAT changes the value of the source IP address field in the IP block or a router changes the time-to-live field,
- Authenticate fields. e.g., the AH protocol authenticates certain fields of IP and all other fields of the encapsulated protocols,
- Encrypt fields. For example, the ESP in transport mode encrypts all the fields of the encapsulated protocols,
- Etc…

In addition, the network mechanisms perform these treatments according to a subset of data flow fields they can perceive. E.g., a stateless firewall analyses only the protocol id, IP source and IP destination fields in the IP block, and port source and destination fields in the UDP/TCP block.

B. Modeling data flows

We propose a data flow model that is independent from the underlying protocols. We want this model to be able to anticipate future protocols. The difficulty is to reach the good level of abstraction between security mechanisms independence and reality closeness.

Foremost, we define our core entities:

- $\mathcal{A}$ is the set of possible attributes. An attribute $a \in \mathcal{A}$, represents a couple <*name, value*> where *name* is a field that can be found in a protocol, and *value* is its content,
- $\mathcal{P}$ is the set of protocols, i.e., the set of logical blocks. An instance of protocol $p \in \mathcal{P}$ is a couple <*protoid, attributes*> where *protoid*=<*name, id*> is the name of the protocol and an unique identifier and *attributes* is

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

defined on the Power-set of $\mathcal{A}$, i. e., $attributes \in \mathbb{P}(\mathcal{A})$,

- $\mathcal{E} = \mathcal{P}^{\mathbb{N}}$, is the set of finite sequences over $\mathcal{P}$. This set represents all the possible encapsulation chain of protocols, i.e. sequences of logical blocks. For reasons of notation simplicity, we use: $\text{next}(p_i) = p_{i+1}$ and $\text{rest}(p_i) = < p_{i+1}, \dots, p_n >$ for a given sequence $e = < p_1, p_2, \dots, p_i, \dots, p_n >$,
- $\mathcal{S}$ is the set of security algorithms addressing the encapsulation chain of protocols (for instance, DES, 3DES, HMAC-SHA1, etc.).

Definition 1: *Formal definition of data flows*

Based on above definitions, we present the set of data flows as: $\mathcal{F} \subseteq \mathcal{E} \times \text{AUTHN} \times \text{CONF}$ such that:

- $\mathcal{E}$ is the encapsulation chain of protocols,
- $\text{AUTHN} \subseteq (\mathcal{A} \times \mathcal{P} \times \mathcal{A} \times \mathcal{P} \times \mathcal{S})$ represents the attributes of the data flow that have been authenticated such that $(a_1, p_1, a_2, p_2, s) \in \text{AUTHN}$ indicates that attribute a_1 of protocol p_1 guarantees the integrity of attribute a_2 of protocol p_2 via the security algorithm s,
- $\text{CONF} \subseteq \text{BAG}(\mathcal{A} \times \mathcal{P} \times \mathcal{S})$ represents the attributes of the data flow that have been encrypted, such that $(a, p, s) \in \text{CONF}$ indicates that attribute a of protocol p is encrypted via the security algorithm s. We used a multi-set because an attribute can be encrypted several times by the same algorithm. A multi-set is a set of elements where an element may appear several times.

A treatment on a data flow is then seen as a particular function of $\mathcal{F}$ to $\mathcal{F}$, called *transform function* as in [6]. This function represents the capability to modify the data flows. It can symbolize encryption protocols such as IPsec where one transform function adds some security services (e.g. confidentiality) and another removes it, or the NAT where only one transform function is concerned. According to the underlying technology, each treatment considers a subset of attributes from data flow as input (e.g., for IPSec, attributes are source address, destination address and transport protocol field of the IP header as well as source port and destination port of the TCP/UDP header) and modifies the data flow. For example, when using ESP, treatment adds new protocols in the encapsulation chain of protocols and new instances in the relationship AUTHN and new attributes in multi-set CONF.

Examples:

Given flow $f = (e, \text{AUTHN}, \text{CONF}) \in \mathcal{F}, e = < p_1, p_2 >$, *where* $p_1 = ((p, 1), \{(\alpha, v1), (\beta, v2), (\gamma, v3)\})$ and $p_2 = ((p, 2), \{(\kappa, v4), (\mu, v5)\})$:

1. If $\text{AUTHN} = \{\ \}$, and $\text{CONF} = \{\ \}$, then this data flow includes two encapsulated protocols for which no field is protected,
2. If $\text{AUTHN} = \{(\kappa, p_2, \gamma, p_1, s_1), (\kappa, p_2, \mu, p_1, s_1)\}$ and $\text{CONF} = \{(\mu, p_2, s_2)\}$, then field κ in protocol p_2 authenticates fields γ and μ in protocol p_1 by algorithm s_1 and field μ in protocol p_1 is encrypted by algorithm s_2.

In the rest of the article, we use the following notation to simplify the readability:

- we designate by p_1 the protocol element whose *protoid* equals to (p,1),
- attributes(p) for the set of attributes of a protocol p. E.g. attributes(p_1)= {(α, v1), (β, v2), (γ, v3)},
- when there is no ambiguity, we use the name of the attribute instead of the couple (name,value).

Definition 2: *Data flow integrity*

Data flow integrity indicates that no authenticated attribute has been changed. This is described in our model as follows: Let $f = (e, \text{AUTHN}, \text{CONF}) \in \mathcal{F}$, integrity of f is satisfied iff $\forall (a, a') \in \mathcal{A}$ x $\mathcal{A}, \forall (p_1, p_2) \in \mathcal{P}$ x $\mathcal{P}, \forall s \in \mathcal{S}, (a, p_1, a', p_2, s) \in \text{AUTHN} \Longrightarrow a \in \text{attributes}(p_1) \wedge a' \in \text{attributes}(p_2)$.

We do not provide any definition of data flow confidentiality in this article. Confidentiality refers to non-disclosure of sensitive information. Sensitive information in the context of data flow is a subset of the protocols fields/payload that are required to be confidential. In our attribute-based modeling, attributes within the multi-set CONF represent the encrypted information in the data flow. It can be noticed that our model faithfully represents reality. A real data flow cannot be completely encrypted (if the IP destination field is encrypted, routers won't be able to route the packet). However, particular values of specific attributes might be considered as confidential (e.g., if the knowledge that two devices are communicating over the Internet is confidential, it is important to hide their IP addresses values for example in an IPsec tunnel where only the IP addresses of the two IPsec gateways will be revealed). Thus, the set attributes required to confidential depends on external security requirements that are out of scope of this article.

IV. CASE STUDY

In this section, we present the modeling of security mechanisms to validate the expression capability of our approach. We describe examples related to IPsec [10] (namely the AH and ESP protocols) and NAPT. Although, conflicts between these technologies are well know, our intent is to explain how our approach can be used.

.

In the following examples, we use IP, TCP, UDP, AH, ESP protocols and a data flow f = (< …,ip_1,… >, AUTHN,CONF). In our formalism, they can be defined as follows:

- attributes(ip_i)={version,hlength,tos,tlength,id,flags,offset, ttl,proto,checksum,ips,ipd,options},
- attributes(tcp_i)={ports,portd,seq,ack,hlength,reserved, tcpflags,win,options,checksum},
- attributes(udp_i)={ports,portd,len,checksum},
- attributes(ah_i)={nexthdr,payloadlength,reserved,spi, seq,ad},
- attributes(esp_i)={ spi,seq,padlength,nextheader,ad}.

A. Specification of AH

AH (Authentication Header) [11] is designed to ensure integrity and authenticity of IP datagrams without data encryption. The Authentication Data (AD) field guarantees the integrity of the datagram. The AH protocol has two modes: transport and tunnel.

1) In the **transport mode**, AH is inserted after the IP header and before next layer (Fig. 1).and protects the entire IP packet except mutable fields (i.e. the fields DSCP, ECN, Flags, Offset, TTL, Header Checksum).

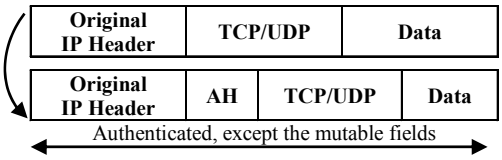


Figure 1. IP datagram before and after applying AH in transport mode

Definition 3: *Specification of AH in the transport mode*

The application of AH in transport mode on the flow *f* (as defined above) is the transformation function $tf_{AH}^{transport} : \mathcal{F} \rightarrow \mathcal{F}$, generating the flow $tf_{AH}^{transport}$ =*f'*=(<…,ip_1,ah, p…>, AUTHN', CONF') where:

- $attributes(ip_1') = attributes(ip_1) \backslash \{(proto, x)\} \cup \{(proto, 51)\}$ where 51 is the value of AH protocol,
- $AUTHN' =$ $AUTHN \cup$ $\cup_{\forall x \in attributes(ip_1') \backslash \{DSCP,ECN,\ldots,checksum\}} \{(ad, ah, x, ip_1', s)\} \cup$ $\cup_{\forall y \in attributes(ah) \backslash \{ad\}} \{(ad, ah, y, ah, s)\} \cup$ $\cup_{\forall p \in rest(ah) | \forall z \in attributes(p)} \{(ad, ah, z, p, s)\}$ It indicates that integrity of all immutable fields of the IP protocol and every fields of all protocols, which are encapsulated by AH, is guaranteed by using security algorithm *s*.

2) In the **tunnel mode**, the inner IP header carries the ultimate IP source and destination addresses, while an outer IP header contains the addresses of the IPsec peers (Fig. 2) and protects the entire inner IP packet, including the entire inner IP header. The position of AH in mode tunnel, relative to the outer IP header, is the same as for AH in the transport mode. In fact, in AH Tunnel mode the entire original IP header and data becomes the "payload" for the new packet. The new IP header is protected exactly the same as the IP header in Transport mode.

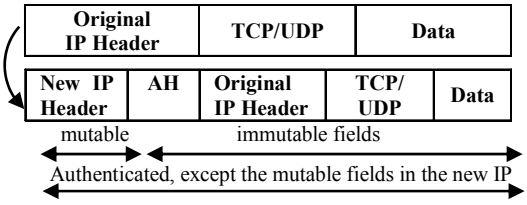


Figure 2. IP datagram before and after applying AH in tunnel mode

Definition 4: *Specification of AH in the tunnel mode*

The application of AH in the tunnel mode on the flow *f* (as defined above) is the transformation function: $tf_{AH}^{tunnel} : \mathcal{F} \rightarrow \mathcal{F}$, generating flow $tf_{AH}^{tunnel}(f) = f' = (< \ldots, ip_2, ah, ip_1, p, \ldots >, AUTHN', CONF)$ where:

- $attributes(ip_2) = \{(ips, sourcegateway), (ipd, destinationgateway), (proto, 51), \ldots\}$,
- $AUTHN' =$ $AUTHN \cup$ $\cup_{\forall x \in attributes(ip_2) \backslash \{DSCP,ECN,\ldots,checksum\}} \{(ad, ah, x, ip_2, s)\} \cup$ $\cup_{\forall y \in attributes(ah) \backslash \{ad\}} \{(ad, ah, y, ah, s)\} \cup$ $\cup_{\forall p \in rest(ah) | \forall z \in attributes(p)} \{(ad, ah, z, p, s)\}$ This indicates that integrity of all immutable fields of the new IP header, and every fields of all protocols, which are encapsulated by AH (including the original IP header), is guaranteed by using security algorithm *s*.

B. Specification of ESP

ESP protocol (Encapsulating Security Payload) [12] provides authentication and encryption of data carried in IP datagram. Like AH protocol, the Authentication Data (AD) field guarantees the integrity of the datagram and ESP has two modes: transport and tunnel.

1) In the **transport mode**, the ESP bounds the transport and data layers (Fig. 3). ESP authenticates the data transported in the IP datagram but not the IP header. In addition, it encrypts the data protocol transport layer.

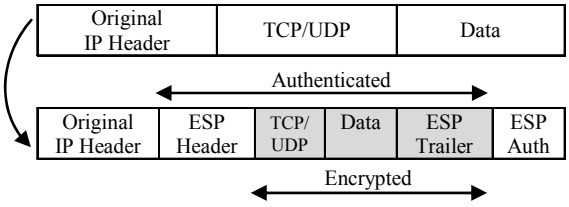


Figure 3. IP datagram before and after applying ESP in transport mode

Definition 5: *Modeling of ESP in transport mode*

Our model considers protocols as logical blocks. So, we do not differentiate between the header and the tail of ESP.

The application of ESP in the transport mode on the flow *f* (as defined above) is the transformation function $tf_{ESP}^{transport} : \mathcal{F} \rightarrow \mathcal{F}$, generating flow $tf_{ESP}^{transport} = f' = (< \ldots, ip_1', esp, p, \ldots >, AUTHN', CONF')$ where:

- $attributes(ip_1') = attributes(ip_1) \backslash \{(proto, x)\} \cup \{(proto, 50)\}$ where 50 is the value of ESP protocol.
- $AUTHN' =$ $AUTHN \cup$ $\cup_{\forall x \in attributes(esp)\{ad\}} \{(ad, esp, x, esp, s_1)\} \cup$ $\cup_{\forall p \in rest(esp) | \forall y \in attributes(p)} \{(ad, esp, y, p, s_1)\}$ indicating integrity of every fields of all protocols, which are encapsulated by ESP, is guaranteed by using security algorithm s_1,
- $CONF' =$ $CONF \cup \cup_{\forall x \in attributes(esp) \backslash \{spi,seq,ad\}} \{(x, esp, s_2)\} \cup$ $\cup_{\forall p \in rest(esp) | \forall y \in attributes(p)} \{(y, p, s_2)\}$. This indicates that every fields of all protocols, which are encapsulated by ESP, are encrypted using security algorithm s_2,

2) In the **tunnel mode**, the inner IP header carries the ultimate IP source and destination addresses, while an outer IP header contains the addresses of the IPsec

gateways (Fig.4) and protects the entire inner IP packet, including the entire inner IP header. The position of ESP in the tunnel mode, relative to the outer IP header, is the same as for ESP in transport mode. The integrity of the datagram is checked against the field Authentication Data (AD). In fact, in ESP Tunnel mode the entire original IP header and data becomes the "payload" for the new packet. The new IP header is not protected.

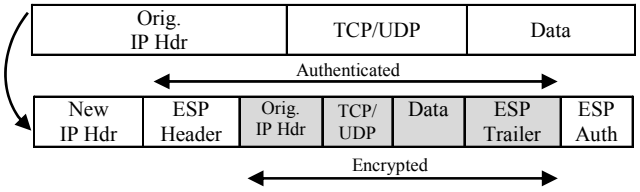


Figure 4. IP datagram before and after applying ESP in tunnel mode

Definition 6: *Modeling of ESP in the tunnel mode*

The application of ESP in tunnel mode on a flow f (as defined above) is the transformation function $tf_{ESP}^{tunnel}: \mathcal{F} \rightarrow \mathcal{F}$, generating flow $tf_{ESP}^{tunnel}(f) = f' = (< \ldots, \mathrm{ip}_2, \mathrm{esp}, \mathrm{ip}_1, \mathrm{p}, \ldots >, \mathrm{AUTHN}', \mathrm{CONF}')$ where:

- $\mathrm{attributes}(\mathrm{ip}_2) = \{(\mathrm{ips}, \mathrm{sourcegateway}), (\mathrm{ipd}, \mathrm{destinationgateway}), (\mathrm{proto}, 50), \ldots\}$,
- $\mathrm{AUTHN}' = \mathrm{AUTHN} \cup \bigcup_{\forall x \in \mathrm{attributes(esp)} \setminus \{\mathrm{ad}\}} \{(\mathrm{ad}, \mathrm{esp}, \mathrm{x}, \mathrm{esp}, \mathrm{s}_1)\} \cup \bigcup_{\forall p \in \mathrm{rest(esp)} \mid \forall y \in \mathrm{attributes(p)}} \{(\mathrm{ad}, \mathrm{esp}, \mathrm{y}, \mathrm{p}, \mathrm{s}_1)\}$ indicating that integrity of every fields of all protocols, which are encapsulated by ESP, is guaranteed by using security algorithm s_1,
- $\mathrm{CONF}' = \mathrm{CONF} \cup \bigcup_{\forall x \in \mathrm{attributes(esp)} \setminus \{\mathrm{spi,seq,ad}\}} \{(\mathrm{x}, \mathrm{esp}, \mathrm{s}_2)\} \cup \bigcup_{\forall p \in \mathrm{rest(esp)} \mid \forall y \in \mathrm{attributes(p)}} \{(\mathrm{y}, \mathrm{p}, \mathrm{s}_2)\}$. This indicates that every fields of all protocols encapsulated by ESP are encrypted using security algorithm s_2.

C. Modeling of NA(P)T

NA(P)T (Network address and port translation) [8] transforms the IP source address of an IP datagram to allow the communication between an equipment with private IP address with another one connected to the Internet.

The operation of this system can be summarized as follows:

1. NAPT generates dynamically a source port,
2. NAPT records the association (old IP source address, old source port, new IP address, new source port),
3. NAPT modifies the fields source port and checksum fields of the UDP / TCP,
4. NAPT modifies the source IP address and the checksum in the IP header.

Definition 7: *Basic NAPT*

Consequently, we can represent the NAPT processing system by the transformation function tf_{NAPT}^{B} as follows:

- Pre-condition 1: the protocol following IP header should be either tcp or udp.
  $\forall f = (< \ldots \mathrm{ip}, \mathrm{next(ip)} \ldots >, \mathrm{AUTHN}, \mathrm{CONF})$, with: $\{\mathrm{ports}, \mathrm{portd}, \mathrm{checksum}\} \in \mathrm{attributes(next(ip))}$.
- Pre-condition 2: NAPT must be able to read the source IP address and the source port:
  $\forall f = (< \ldots, \mathrm{ip}, \mathrm{next(ip)} \ldots >, \mathrm{AUTHN}, \mathrm{CONF})$,
  $\nexists s \mid (\mathrm{ips}, \mathrm{ip}, s) \in \mathrm{CONF} \vee (\mathrm{ports}, \mathrm{next(ip)}, s) \in \mathrm{CONF}$.

NAPT transforms the following fields: source IP address, checksum of IP, the source port, and the checksum of transport protocol. Therefore, tf_{NAPT}^{B} transforms a flow $f = (< \ldots, \mathrm{ip}, \mathrm{next(ip)} \ldots >, \mathrm{AUTHN}, \mathrm{CONF})$ into a data flow $f' = (< \ldots, \mathrm{ip}', \mathrm{next(ip)}' \ldots >, \mathrm{AUTHN}, \mathrm{CONF})$ such that:

- $\mathrm{attributes}(\mathrm{ip}') = \mathrm{attributes}(\mathrm{ip}) \setminus \{(\mathrm{ips}, \mathrm{value}), (\mathrm{checksum}, \mathrm{value})\} \cup \{(\mathrm{ips}', \mathrm{new\_value}), (\mathrm{checksum}', \mathrm{new\_value})\}$,
- $\mathrm{attributes}(\mathrm{next(ip)}') = \mathrm{attributes}(\mathrm{next(ip)}) \setminus \{(\mathrm{ports}, \mathrm{value}), (\mathrm{checksum}, \mathrm{value})\} \cup \{(\mathrm{ports}', \mathrm{new\_value}), (\mathrm{checksum}', \mathrm{new\_value})\}$

This treatment considers that protocol TCP or UDP follows immediately the IP header, which leads to the problem of the evolution of the arrangement of protocols encapsulation. We therefore propose an advanced version of NAPT processing that is not limited by this assumption. This version, called advanced NAPT, is able to search the TCP or UDP protocol deeper in the data flow.

Definition 8: *Advanced NAPT*

The transformation function tf_{NAPT}^{adv} represents an advanced NAPT which is defined as:

- Pre-condition 1: the IP protocol encapsulates directly or indirectly the TCP or UDP protocol: e.g.,
  $\forall f = (< \ldots, \mathrm{ip}, \ldots, \mathrm{p}, \ldots >, \mathrm{AUTHN}, \mathrm{CONF})$,
  $\exists p \in \mathrm{rest(ip)} \mid \{\mathrm{ports}, \mathrm{portd}, \mathrm{checksum}\} \in \mathrm{attributes(p)}$. thus p is a transport protocol.
- Pre-condition 2: NAPT must be able to read the IP source address and the source port: e.g.,
  $\forall f = (< \ldots, \mathrm{ip}, \ldots, \mathrm{p}, \ldots >, \mathrm{AUTHN}, \mathrm{CONF})$,
  $\nexists s$ such that $(\mathrm{ips}, \mathrm{ip}, s) \in \mathrm{CONF} \vee (\mathrm{ports}, \mathrm{p}, s) \in \mathrm{CONF}$

NAPT transforms the fields IP source address, IP checksum and the source port and checksum of the transport protocol. Therefore, tf_{NAPT}^{adv} transforms a flow $f = (< \ldots, \mathrm{ip}, \ldots \mathrm{p} \ldots >, \mathrm{AUTHN}, \mathrm{CONF})$ into a flow $f' = (< \ldots, \mathrm{ip}', \ldots \mathrm{p}' \ldots >, \mathrm{AUTHN}, \mathrm{CONF})$ such that:

- $\mathrm{attributes}(\mathrm{ip}') = \mathrm{attributes}(\mathrm{ip}) \setminus \{(\mathrm{ips}, \mathrm{value}), (\mathrm{checksum}, \mathrm{value})\} \cup \{(\mathrm{ips}', \mathrm{new\_value}), (\mathrm{checksum}', \mathrm{new\_value})\}$,
- $\mathrm{attributes}(\mathrm{p}') = \mathrm{attributes}(\mathrm{p}') \setminus \{(\mathrm{ports}, \mathrm{value}), (\mathrm{checksum}, \mathrm{value})\} \cup \{(\mathrm{ports}', \mathrm{new\_value}), (\mathrm{checksum}', \mathrm{new\_value})\}$

V. CONFLICT ANALYSIS IN COLORED PETRI NETS (CPN)

In this section, we demonstrate that it is possible to detect conflicts between security mechanisms without a priori knowledge. The problems between IPsec and NAPT are well known [9] but they are not trivial without using the human expertise. Our goal here is to detect these conflicts by applying our formalization only. Our approach, being independent of the underlying technologies, can handle conflicts between other technologies.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

In order to automate the conflict detection task, we have specified our formalism in colored Petri nets; this formal language being adapted to data flows oriented approach [6] and featured with tools (CPN tools [14]) to validate our formal methodology.

A. Introduction to CPN

Colored Petri Net (CPN) is a formal specification language consisting of a set of tokens whose type is represented by a color, a set of transitions, and a set of places with a domain (which defines the types of tokens that can be stored in that place) and a set of arcs connecting places and transitions. It allows creating formal models of systems.

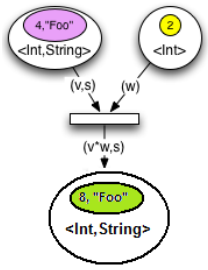


Figure 5. Example of CPN specification

The state of a system (Fig. 5) is represented by the distribution of tokens in places. It changes when a transition is fired. A Boolean expression called guard may be associated with a transition to set the conditions required to fire the transition. If the tokens contained in places connected by incoming arcs in the transition satisfy the guard, then they are removed from these places and new tokens are created in the places connected to outgoing arcs of the transition.

Our choice of Colored Petri Nets formalism [13] to address the design of modeling data flow is motivated by the following reasons: Colored Petri Nets are well-known for their graphical and analytical capabilities for the specification and verification of concurrent, asynchronous, distributed, parallel and nondeterministic systems. Various features contribute to such a success include graphical nature, the simplicity of the model and the firm mathematical foundation. It also provides modularity in design.

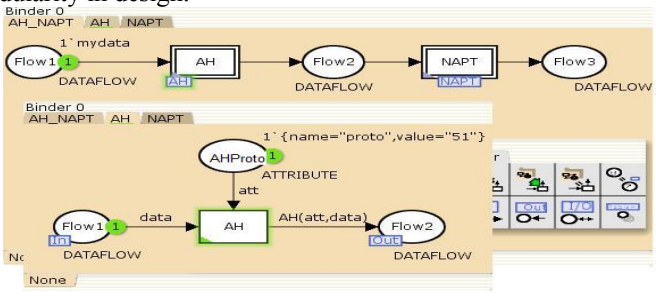


Figure 6. Navigating through marking menus

In addition to colors, it is possible to create hierarchical descriptions, i.e., structure a large description as a set of smaller pieces by using the facilities within CPN Tools through well-defined interfaces and relationships to each other. This is similar to the use of modules in a programming language. Conceptually, CPNs with substitution transitions are CPNs with multiple layers of detail. It enhances the readability of the CPN specification. Figure 6 shows an example of navigation from a super-page to a subpage. The substitution transition AH (with double line in the CPN at the top of figure 6) is actually a black box view of a more detailed CPN (at the bottom) that specifies AH in transport mode.

B. Net structure and declaration

We simulate and validate our CPN model with "CPN Tools" [14]. The CPN development environment uses an extension of ML language to formally specify colors of tokens, guards at transitions, and functions on arcs. Fig. 7 presents the definition data flow in CPN-ML.

```
// Definition of attributes
color ATTRIBUTE = record name:STRING * value:STRING;
// Definition of protocol identification
color PROTOCOLID = record name :STRING * id :INT;
// Definition of the list of attributes
Color ATTList = list ATTRIBUTE;
// Definition of protocols
color PROTOCOL = record protoid:PROTOCOLID*value:ATTList;
// Definition of encapsulation chain of protocols
Color ENCAPSULATION = list PROTOCOL;
// Definition of security algorithm
color SECALGO = with DES | 3DES | HMAC | …;
// Definition of authentication elements
color AUTHN = product ATTRIBUTE*PROTOCOL*ATTRIBUTE
*PROTOCOL*SECALGO;
// Definition of the list of authentication elements
color AUTHNLIST = list AUTHN;
// Definition of confidentiality elements
color CONF = product ATTRIBUTE*PROTOCOL*SECALGO ;
// Definition of the list of confidentiality elements
color CONFLIST = list CONF;
// Definition of data flows
Color DATAFLOW = product ENCAPSULATION*AUTHNLIST
*CONFLIST;
```

Figure 7. Definition in CPN-ML of data flows

C. Scenarios

We choose three scenarios to validate our model. The first one is AH in the transport mode transformed data flow through NAPT to show the use of the authentication “AUTHN”. The second one is ESP flow in the transport mode transformed data flow through NAPT to show the use of both authentication “AUTHN” and confidentiality “CONF”. Finally, AH in the tunnel mode data flow through NAPT presents tunneling.

1) Scenario 1: AH flow in transport mode with NAPT

In this first scenario, we study the interaction between a mechanism that implements AH in the transport mode and the NAPT mechanism (Fig. 8). Our study consists in analyzing, for a given data flow $f = (< ip_1, tcp >, \{\}, \{\})$, the transformation chain $tf_{NAPT} \circ tf_{AH}^{transport}$.

Based on definition 3, data flow f is transformed to $f' = tf_{AH}^{transport}(f) = (< ip'_1, ah, tcp >, AUTHN, \{\})$ where:

- $attributes(ip'_1) = attributes(ip_1) \backslash \{(proto, 4)\} \cup \{(proto, 51)\}$,
- AUTHN = $\cup_{\forall x \in attributes(ip'_1) \backslash \{DSCP, ECN, \ldots, \text{checksum}\}} \{ (ad, ah, x, ip'_1, s) \} \cup$

$\cup_{\forall y \in attributes(ah)\backslash\{ad\}}\{(ad, ah, y, ah, s)\} \cup$
$\cup_{\forall z \in attributes(tcp)}\{(ad, ah, z, tcp, s)\}.$

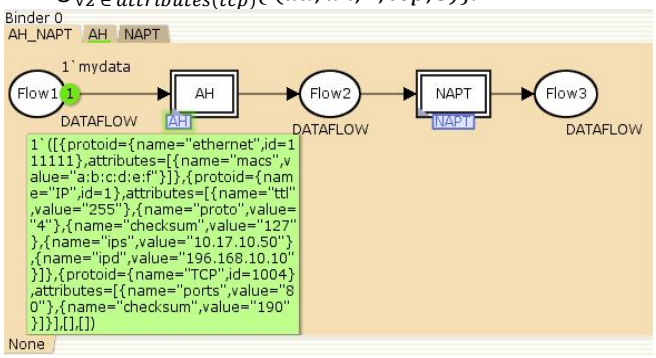


Figure 8. AH in the transport mode through NAPT

Then, data flow f' can't be transformed by *basic NAPT* (definition 7), because f' does not verify pre-condition 1; i.e. the protocol encapsulated directly by IP is AH. As a result, the token representing the data flow if blocked in place Flow2.

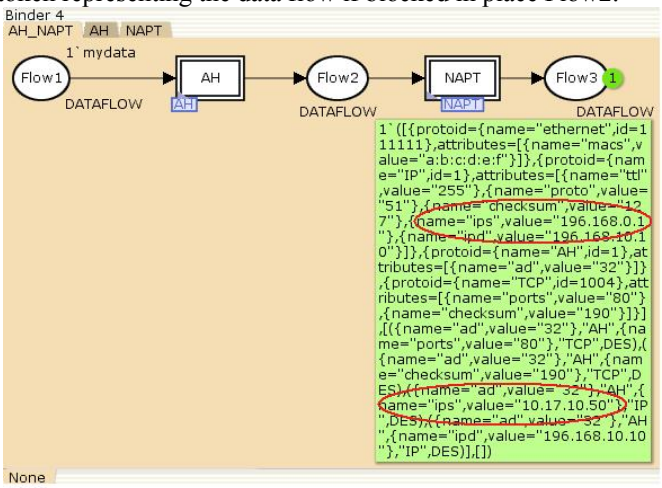


Figure 9. Conflict detection: AH in the transport mode through Advanced NAPT

While using the *advanced NAPT* (definition 8), we get the following data flow:

$f'' = tf_{NAPT}^{adv} \circ tf_{AH}(f) = (< ip_1', ah, tcp' >, AUTHN, \{\})$, where:

- $attributes(ip_1') = attributes(ip_1)\backslash\{(ips, old), (checksum, old)\} \cup \{(ips, new), (checksum, new)\}$,
- $attributes(tcp') = attributes(tcp)\{(ports, old), (checksum, old)\} \cup \{(ports, new), (checksum, new)\}$.

In this case, integrity of f'' (definition 2) is violated (Fig. 9) because:

- $(ad, ah, (ips, old), ip_1', s) \in$ AUTHN and $(ips, old) \notin ip_1'$,
- $(ad, ah, (checksum, old), ip_1', s) \in$ AUTHN and $(checksum, old) \notin ip_1'$,
- $(ad, ah, (ports, old), tcp', s) \in$ AUTHN and $(ports, old) \notin tcp'$,
- $(ad, ah, (checksum, old), tcp', s) \in$ AUTHN and $(checksum, old) \notin tcp'$.

2) Scenario 2: a flow ESP in the transport mode with NAPT

In our second scenario, we study the interaction between a mechanism that implements ESP in the transport mode and the NAPT mechanism. Our study consists in analyzing, for a given data flow $f = (< ip_1, tcp >, \{\}, \{\})$, the transformation chain $tf_{NAPT} \circ tf_{ESP}(f)$.

Based on definition 5, data flow f is transformed to $f' = tf_{ESP}^{transport}(f) = (< ip_1', esp, tcp >, \text{AUTHN}, \text{CONF})$ where:

- $attributes(ip_1') = attributes(ip_1)\backslash\{(proto, 4)\} \cup \{(proto, 50)\}$,
- AUTHN = $\cup_{\forall x \in attributes(esp)\backslash\{ad\}}\{(ad, esp, x, esp, s_1)\} \cup \cup_{\forall y \in attributes(tcp)}\{(ad, esp, y, tcp, s_1)\}$,
- CONF = $\cup_{\forall x \in attributes(esp)\backslash\{spi,seq,ad\}}\{(x, esp, s_2)\} \cup \cup_{\forall y \in attributes(tcp)}\{(y, tcp, s_2)\}$.

Using transformation function *basic NAPT* (definition 7), there is a conflict because f' does not verify pre-condition 1; i.e. the protocol encapsulated directly after IP is ESP. As a consequence, the token representing the data flow is blocked in place Flow2. On the other hand, using the *advanced NAPT* (definition 8), pre-condition 2 is not satisfied because $(ports, tcp, s_2) \in$ CONF, i.e. the source port of TCP is encrypted and therefore incomprehensible for NAPT (Fig. 10).

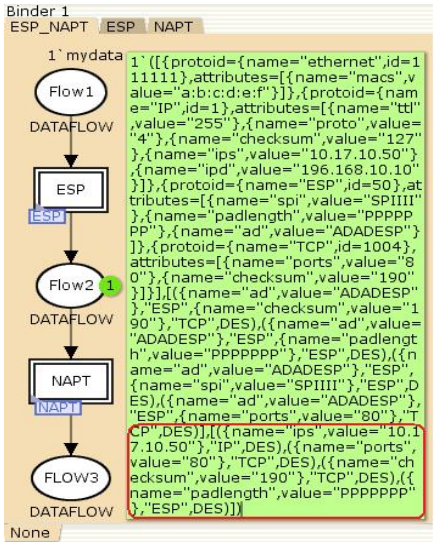


Figure 10. Conflict detection: ESP in the transport mode through NAPT

3) Scenario 3: a flow AH in tunnel mode with NAPT

In our third scenario, we study the interaction between a mechanism that implements AH in the tunnel mode and the NAPT mechanism. Our study analyzes, for a given data flow $f = (< ip_1, tcp >, \{\}, \{\})$, the transformation chain $tf_{NAPT} \circ tf_{AH}^{tunnel}(f)$.

Basing on definition 4, data flow $f' = tf_{AH}^{tunnel}(f) = (< ip_2, ah, ip_1, tcp >, AUTHN, \{\})$ where:

- $attributes(ip_2) =$

$\{(ips, sourcegateway), (ipd, destinationgateway), (proto, 51), \dots\}$,

- AUTHN =
  $\bigcup_{\forall x \in \text{attributes}(ip_2)\setminus\{\text{ttl,tos,flags,...}\}}\{(ad, ah, x, ip_2, s)\} \cup$
  $\bigcup_{\forall y \in attributes(ah)\setminus\{ad\}}\{(ad, ah, y, ah, s)\} \cup$
  $\bigcup_{\forall z \in attributes(ip_1)}\{(ad, ah, z, ip_1, s)\} \cup$
  $\bigcup_{\forall t \in attributes(tcp)}\{(ad, ah, t, tcp, s)\}$.

Data flow f' cannot be transformed by *basic NAPT* (definition 7), because f' does not verify pre-condition 1; i.e. the protocol encapsulated directly after IP is AH.

While using *advanced NAPT* (definition 8), we get the following data flow: $f'' = tf_{NAPT}^{adv} \circ tf_{AH}^{tunnel}(f) = (< ip_2', ah, ip_1, tcp' >, AUTHN, \{\})$, where:

- $attributes(ip_2') = attributes(ip_2)\setminus \{(ips, old), (checksum, old)\} \cup \{(ips, new), (checksum, new)\}$,
- $attributes(tcp') = attributes(tcp)\setminus \{(ports, old), (checksum, old)\} \cup \{(ports, new), (checksum, new)\}$.

In this case, integrity of f'' (definition 2) is violated (Fig. 11) because:

- $\left(ad, ah, (ips, old), ip_2^{'}, s\right) \in$ AUTHN and $(ips, old) \notin ip_2^{'}$,
- $\left(ad, ah, (checksum, old), ip_2^{'}, s\right) \in$ AUTHN and $(checksum, old) \notin ip_2^{'}$,
- $(ad, ah, (ports, old), tcp', s) \in$ AUTHN and $(ports, old) \notin tcp'$,
- $(ad, ah, (checksum, old), tcp', s) \in$ AUTHN and $(checksum, old) \notin tcp'$.

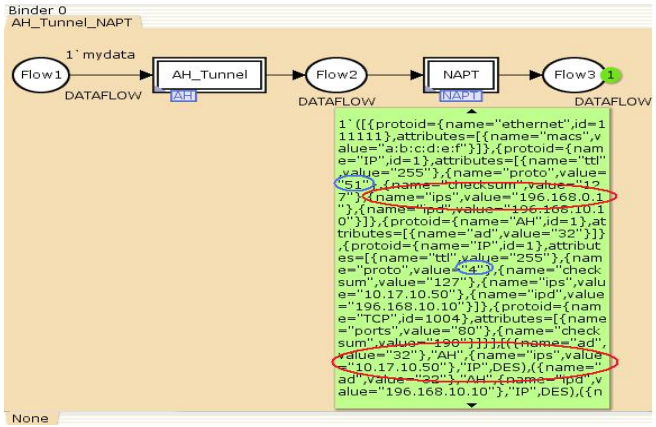


Figure 11. Conflict detection: AH in the tunnel mode through Advanced NAPT

VI. CONCLUSION

Network security requires the coordination of various heterogeneous and interdependent devices. As a consequence, multi-mechanisms and multi-levels conflicts may occur. The fast evolution of technologies imposes network security analysis to be independent from current technology. We proposed in this article a formal data flow oriented approach to analyze conflicts related to security policy deployment. This formal representation is based on an abstraction of a physical data flow consisting of blocks and the relation between each block and the underlying protocol.

By using CPN Tools, we have validated our approach based on well-known scenarios (IPsec and NAPT). This work allows us to verify the capacity of our model to express and analyze real distributed inconsistency. The results are encouraging because conflicts have been detected without requiring any a priori knowledge or experience. Indeed, we do not have to specify the semantics of the elements of a data flow. This leads us to believe that our approach can be used to detect unknown conflicts involving new security mechanisms that may occur at different levels.

Our future work will, therefore, focus on generalizing the model, especially on providing a generic model of devices' configurations built on our attribute-based approach. This will improve the reusability of hierarchical CPN specifications. Then, we will consider more realistic scenarios by looking at the future protocols available as draft specifications to prove our approach is adapted to new technologies too.

REFERENCES

[1] E. Al-Shaer and H. Hamed, "Discovery of Policy Anomalies in Distributed Firewalls", in IEEE INFOCOM, vol.4, pp. 2605-2616, 2004.

[2] J. G. Alfaro, N. B. Cuppens, F. Cuppens, "Complete analysis of configuration rules to guarantee reliable network security policies", in International Journal of Information Security,7(2), pp. 103-122, 2008.

[3] Z. Fu, F. Wu, H. Huang, K. Loh, F. Gong, I. Baldine, C. Xu, "IPSec/VPN Security Policy: Correctness, Conflict Detection and Resolution", in IEEE POLICY, pp. 39-56, 2001.

[4] J. D. Guttman and A. M. Herzog, "Rigorous automated network security management", in International Journal of Information Security, 4(3), pp. 29-48, 2005.

[5] H. Hamed and E. Al-Shaer, "Taxonomy of Conflicts in Network Security Policies", in IEEE Communications Magazine, 44(3), pp. 134-141, 2006.

[6] R. Laborde, M. Kamel, F. Barrère, and A. Benzekri, "Implementation of a Formal Security Policy Refinement Process in WBEM Architecture", in Journal of Network and Systems Management, 15(2), pp. 241-266, 2007.

[7] S. Preda, "Reliable context aware security policy deployment with applications to IPv6 environments", PhD thesis, Télécom Bretagne, 2010.

[8] P. Srisuresh and K. Egevang, "Traditional IP Network Address Translator (Traditional NAT)", IETF RFC 3022, 2001.

[9] B. Aboba and W. Dixon, "IPsec-Network Address Translation (NAT) Compatibility Requirements", IETF RFC 3715, 2004.

[10] S. Kent and K. Seo, "Security Architecture for the Internet Protocol", RFC 4301, IETF, 2005.

[11] S. Kent, "IP Authentication Header (AH)", IETF RFC 4302, 2005.

[12] S. Kent , "IP Encapsulating Security Payload (ESP)", IETF RFC 4303, 2005.

[13] K. Jensen and L. M. Kristensen, "Coloured Perti Nets: Modelling and Validation of Concurrent Systems", Springer, 2009.

[14] http://cpntools.org/, <retreived: 03, 2012>

[15] Al-Shaer, E., Marrero, W., El-Atawy, A., ElBadawi, K., "Network configuration in a box: towards end-to-end verification of network reachability and security", in IEEE ICNP, pp. 123-132, 2009.

[16] I. El Khoury, R. Laborde, F. Barrère, A. Benzekri, "Towards a Formal Data Flow Oriented Model for Network Security Policies Analysis", in SAR-SSI, pp. 1-7, 2011.

[17] I. El Khoury, R. Laborde, F. Barrère, A. Benzekri, C. Maroun, "A generic data flow security model", in SAFECONFIG, pp. 1-2, 2011.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

Mobile Ad hoc Networks for Ground Surveillance

Mathew McGee, Nirmala Shenoy
Rochester Institute of Technology
Rochester, NY, USA
mxm9106@rit.edu, nxsvks@rit.edu

***Abstract*— Ground surveillance networks are an important application of mobile ad hoc networks. The mobile nodes used in such applications benefit by a compact set of protocols that focus on reliable and timely data delivery. A solution that allows for closely integrated operation of routing, medium access control (MAC) and clustering is presented in this article, where clustering is used to improve data aggregation. The solution is evaluated for its performance and compared with two schemes using Optimized Link State Routing (OLSR) and Ad hoc On demand Distance Vector (AODV) routing over wireless LAN 802.11 MAC. Significant performance improvements indicate the potential for integrated approaches.**

Keywords - ground surveillance applications, MANET architectures, clustering, routing.

I. INTRODUCTION

Data aggregation for surveillance is an important application, which requires normally mobile *data collection nodes* to be deployed over the area of interest. From the collection nodes, *data is then aggregated* at a few nodes from where it may be sent to a center for further analysis. Surveillance applications require that data collection be done in a reliable and timely manner. One such application arises in networks used for rescue operations, where several mobile nodes collect data and aggregate them at one or more rescue centers. Text messages, low bit rate streaming data and voice would be the primary type of traffic in such scenarios. Due to the nature of the application, data aggregation over multiple hops becomes a necessity. Computationally non-intensive algorithms and protocols are preferred in such situations. The criticality of such applications and the constrained operational environment would further benefit if a minimal set of protocols that target the tasks in an efficient manner were used.

Some *desirable features* of such MANETs for surveillance can thus be listed as: 1) Multi-hop clustering for efficient data aggregation at few designated nodes; 2) robust connectivity and redundant paths to minimize data loss 3) a minimal protocol stack with low processing complexity to support timely data delivery. The *challenges* however to achieve the desirable are; 1) most clustering algorithms are single hop; multi-hop clustering require complex algorithms 2) proactive routing protocols, which result in reduced lead latency suitable for timely data delivery normally do not support redundant paths; 3) random access MAC protocols face high collisions and loss of data when nodes are mobile especially where packets have to be forwarded across multiple hops; 4) use of different algorithms for clustering and routing, and a MAC protocol that works independently results in protocol interaction issues and inefficiencies.

In this article, we introduce a novel MANET architecture that is highly suitable for critical surveillance applications that use mobile nodes. The architecture is built on the framework offered by an algorithm called the *Multi Meshed Tree* (MMT) algorithm [1]. This algorithm allows efficient coordination and operation of clustering, routing and MAC protocols in an integrated manner using a single address both at layers 2 and 3. A new protocol stack where clustering, MAC and routing operation are viewed as processes operating at a single layer is used. MMT algorithm allows creation of multiple multi-hop clusters, where in each cluster, the cluster head (CH) is the root of a meshed tree, and the cluster clients (CC) simultaneously reside on several tree branches originating from the root to create meshed trees. The random MAC protocol send bursts of data packets from a CC to the CH using sessions resulting in timely data aggregation.

We model the above using Opnet simulation tool and evaluate its performance in comparison with Optimized Link State Routing Protocol (OLSR) [8] and Ad hoc On demand Distance Vector Routing (AODV) protocol [4], operating over WLAN 802.11 at layer 2. The MMT routing protocols used in this work was the proactive version [1, 2]. Hence it was felt appropriate to compare with OLSR a standard proactive routing protocol. AODV is a reactive routing protocol and is supposed to have low overhead has been included in the studies to show the reduced control overhead with MMT routing protocol.

The rest of the article is organized as follows. In section II, we highlight related work in the different topics covered in the integrated solution. Section III presents the integration framework. Section IV briefly provides the simulation details and models based on Opnet simulation tool. Section V provides the graphs and performance analysis. Section VI concludes this paper by providing the relative performance improvements across the three schemes and their rationale highlighting the significance of the integrated approach.

II. RELATED WORK

To the best of our knowledge, there is no published work that integrates clustering, routing and MAC to operate based off a single algorithm and using a single address for surveillance MANETs. In this section, we hence present some related work conducted separately in the areas of random access based MAC protocols, routing protocols for

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

large MANETs and clustering techniques and conclude by highlighting the advantages of an integrated approach.

Clustering or zoning can be efficiently employed for the type of convergecast traffic encountered in surveillance networks, were the primary traffic flow is from CC to CH [9, 13]. In such cases proactive routing approaches are recommended as the routing is limited to the cluster or zone. However proactive routing algorithms require the dissemination of link state information to all routers in the zone, which can introduce latency in realizing or breaking a route, and high overhead. In the *Zone Routing Protocol* (ZRP) [10], each node pre-defines a zone centered at itself. and a framework is proposed, where any proactive routing protocol can be adopted within the zone. *Multi path distance vector zone* routing protocol [11] is an implementation of ZRP that uses multi path *Destination Sequence Distance Vector* for proactive routing. Another proactive approach, which works for groups of nodes, is *LANMAR* [12], which uses *Fisheye State Routing* [8].

Multi Hop clustering techniques such as the d-hop or k-hop clustering [13, 14] algorithms can offer flexibility in terms of controlling the cluster size and cluster diameter, but are often complex to implement.

Medium Access Control Protocols can be broadly categorized as scheduled and random access. Scheduled protocols require algorithms to schedule transmission turns for the nodes in the network, which could be achieved in a distributed or centralized manner. However in the case of surveillance networks with mobile nodes moving in a random manner, scheduling algorithms can be complex. Hence MAC protocols based on 802.11 are preferred [15].

Advantages of Single Algorithm and Interacting Modules: From the above discussions it would be clear that clustering and routing are normally treated separately and are based on different algorithms. Thus, to combine clustering and routing for an application it becomes essential to define an interworking mechanism that adds processing complexity and control overhead. When the MAC protocol operates independent of other protocols, efficient handling of time and loss sensitive packets diminishes. However, if all these operations can be based off a single algorithm, the complexity and overhead can be reduced considerably resulting in a protocol set that is highly efficient.

A similar approach was investigated for airborne surveillance networks of unmanned aerial vehicles travelling in circular trajectories [3]. The work was subsequently extended to an optimized MAC for ground surveillance networks where nodes are moving using random waypoint mobility models. Given the ground surveillance type of MANETs, the number of aggregation nodes is reduced by half in this article and the data traffic is streaming data instead of one MByte data files in [3].

III. THE INTEGRATION FRAMEWORK

A. The Multi Meshed Tree Algorithm

The MMT algorithm [1-2] to support the integrated approach will be briefly explained first. The formation of a single *meshed tree* based on the MMT algorithm is described with the aid of Fig. 1. The dotted lines connect nodes that are in communication range with one another at the physical layer. The node designated as CH is the root of the meshed tree. For ease in explanation, the meshed tree formation is kept simple and restricted to nodes that are connected to the CH by a maximum of 3 hops. At each node several values or IDs have been noted. These are the virtual IDs (VIDs) assigned to the node when it joins a tree branch in the meshed tree. Assume that the CH has a VID '1'. All nodes connected to this CH will have '1' as the first digit in their VIDs. Extending the above logic, a node gets a VID, which will inherit as its prefix the VID of the node upstream in the tree branch (the parent node), followed by a single (or multiple) digit(s) which indicates the child number under that parent. In Fig. 1, each arrow from CH is a tree branch that connects nodes to the root.

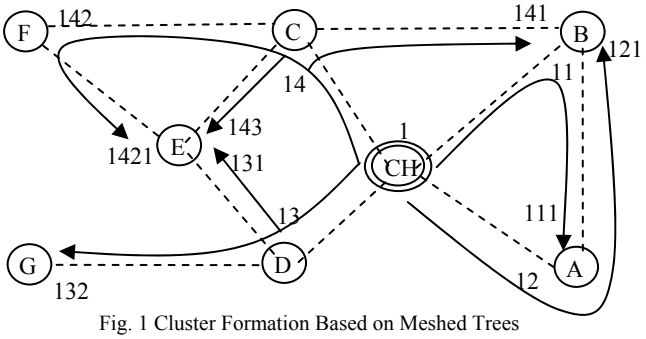


Fig. 1 Cluster Formation Based on Meshed Trees

Flexible Multi hop Cluster Formation: The size of the tree branch can be limited by limiting the length of the VID, which in turn allows control of cluster diameter. Each node that joins the cluster registers with the CH. This allows the CH to accept /reject a joining node to control the cluster size. The number of VIDs allowed for a node can control the amount of meshing in the tree branches of the cluster.

Multiple Dynamic Proactive Paths: The branches of the meshed tree provide the ***route*** to send and receive data and control packets between the CCs and the CH. The branch denoted by VIDs 14, 142 and 1421 connects nodes C (via VID 14), F (via VID 142) and E (via VID 1421), respectively, to the CH. Consider packet forwarding based on VIDs in which the CH has a packet to send to node E. If the CH decided to use E's VID 1421, it will include this as the destination address and broadcast the packet. Enroute nodes C and F will pick up the packet and forward to E. The VID of a node thus provides a virtual path vector from the CH to itself. Note that the CH could have also used VIDs 143 or 131 for node E, in which case the path taken by the packet would have been CH-C-E or CH-D-E respectively. Thus, between the CH and node E there are multiple routes as identified by the multiple VIDs. The support for multiple

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

proactive routes through the multiple VIDs allows for *dynamic route adaptability* to topology changes, as nodes request for new VIDs and joins different branches as their neighbors change.

Scalability: Lastly, a surveillance network can comprise of several tens of nodes; hence the solutions for surveillance networks have to be scalable [12]. We assume that several 'data aggregation nodes' are uniformly distributed among the non-data aggregation nodes during deployment of the surveillance network. Meshed tree clusters can be formed around each of the data aggregation nodes by assuming them to be roots of the meshed trees. Nodes bordering two or more clusters are allowed to join the different meshed trees and thus reside in branches originating from different CHs. Such border nodes will inform their CHs about their multiple VIDs under the different clusters. When a node moves away from one cluster, it can still be connected to other clusters, and thus the surveillance data collected by that node is not lost. By allowing nodes to belong to multiple clusters, the single meshed tree cluster can be extended to *multiple* overlapping *meshed tree* clusters that can collect data from several tens of nodes deployed over a wider area with very low probability of losing the captured data.

B. Burst Forwarding (BF) MAC

The VIDs acquired by a CC defines a path from the CC to the CH. Given that the paths in a MANET are transient and have short life times, the proposed MAC forwards several data packets (a burst) in a sequence (a session) over multiple hops. BF_MAC opens multi-hop data sessions using VIDs issued by the MMT algorithm. This allows for BF_MAC operation without additional 'address' overhead. The access is similar to the CSMA/CA protocol (WLAN 802.11), except that a data session between a CC and CH is started when a node succeeds in getting the channel. An exponential back off process is adopted. Contention window (CW) is handled as explained below.

EXP_ACK - The explicit acknowledge mode. A node that is either the final destination for a data session or is unable to forward a Request To Send (RTS) is in this mode.

CLEAR_TO_SEND: When a node senses an idle medium after its backoff count down, it goes into this mode. It continues in this mode till it senses a busy medium.

NOT_CLEAR_TO_SEND: A node that overhears transmissions in its neighborhood will go into this mode

DATA_SEND: when a node has data to send that originated from its application and receives an explicit or implicit CTS (explained below) is in this mode.

DATA_FORWARD: when a node that is not the originator for the data session receives a CTS for the RTS packet that it forwarded will enter this mode.

Session Establishment Procedure: When a node has data that arrives from its application layer it will first initiate the backoff process. On countdown to zero it will send an RTS packet, which contains the source VID, destination VID, and a hold time which is the total session time. This is calculated to include the time that the nodes in the path will be sending and forwarding packets as well as the time to establish the session.

If a node receives an RTS packet and determines that it is the next hop in the data session then it will forward the RTS onto the next hop node in the path specified by the VID. An RTS packet thus forwarded is overheard by the previous node on the session path and is treated as an Implicit CTS (IMP_CTS) packet. If a node receives an RTS packet and determines that it is the final destination for the session then it will send an explicit CTS (EXP_CTS). If a node receives an RTS and determines that it doesn't have the VID to involve in the session it will go into NOT_CLEAR_TO_SEND mode and remain silent for the hold time specified in the RTS packet.

Fig. 2 is used to explain the use of the modes of operation in BF_MAC. Node 'A' which has a VID 124 which defines its path to the CH (VID=0) has data to send to the CH. First node 'A' sends an RTS packet to node B (with VID 12). Node B receives the RTS as it is the next hop node in the session in the path (124) towards the CH. B forwards the RTS packet to node C. Node 'A' hears the forwarded RTS packet from 'B' and treats it as IMP_CTS. On receiving the IMP_ CTS, node A calculates the time it will take for the rest of the session to be established until an EXP_CTS is issued from CH and queues the first data packets to be sent. It then goes to DATA_SEND mode.

When node 'C' receives RTS from B it determines that it is on the path but is not the final destination so it forwards the RTS packet. When node 'B' receives the IMP_CTS it enters the DATA_FORWARD mode and is ready to receive and forward data packets. When the CH receives the RTS packet from node 'C' it sends an EXP_CTS, and the session is considered to be established.

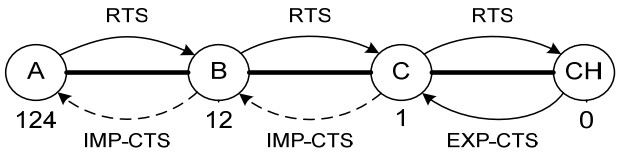


Fig. 2 BF-MAC Multi-hop Multi-packet Data Session

Non-Session and Non-Source Nodes: A node that overhears activity by it neighbors enters the NOT_CLEAR_TO_SEND mode. It also resets its contention window (CW), to the lowest value of 31. A node that is on a data session path but isn't the source node for the session will also set its CW to the minimum value at the end of the session for which it is currently forwarding the packets. This gives the non-session and non-source nodes a fair chance to get the media next time.

Data Sending and Forwarding: When the session's source node receives an IMP_CTS or EXP_CTS from the next node in the data session path it knows the data session is open to the next hop. If the packet was an IMP_CTS packet then the node calculates the time for session establishment and queues the first data packet for transmission at that point. If the packet was an EXP_CTS packet then the node will

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

begin sending the data packets because it knows the entire session path is just one hop.

The source node sends its first data packet and then waits for an EXP/IMP_ACK from the next node in the session path. When the next node in the data session receives the data packet it will modify the packet changing the sender's VID and the sender's UID to its VID and UID. It then checks it's mode, and if in DATA_FORWARD mode the node will forward the packet onto the next hop. When the previous node in the data session receives the forwarded data packet it interprets the packet as an IMP-ACK packet. A node in the session path that is in EXP_ACK mode will send an EXP_ACK packet back to the previous hop in the session path. If a node is the final destination then the BF_MAC will send the packet to the upper protocol layers. Else BF_MAC will queue the data packet in its own queue with a higher priority than it own data packets.

A source node will continue the above pattern of sending data packets until it has sent every packet for the session or has sent all of the remaining packets in the session. When the session ends at the source node it will double its CW. By doubling the contention window at the source node and resetting the contention window at all other nodes in range of the session, the BF_MAC effectively gives non source nodes a higher priority to begin their own data sessions.

Partially Established Sessions: Nodes in a session path will respond to RTS if they are in the CLEAR_TO_SEND mode only. Using the same example as above if node 'C' was in NOT_CLEAR_TO_SEND mode, perhaps from a session that was already established on the other side of the CH, node 'C' wouldn't respond to the RTS that node 'B' sent. Node 'B' would then timeout waiting for IMP_CTS and would hence enter EXP_ACK mode. The session would still be established from node A's perspective so it would send data packets to node 'B'. However, since 'B' is now in EXP_ACK it wouldn't forward the packets onto node 'C'. Instead it would modify the data packet's sender UID and VID fields just the same way it would if it were forwarding the packet, but it would put the packets on the top of its data queue and subsequently try to establish a session to the CH to forward the packets. Once node 'A' receives EXP_ACK for its data packets it would clear all related information.

Priority Queues: The BF_MAC maintains three queues and in each queue packets are inserted based on their type to provide a second level of priority within that queue.

- One queue stores the configuration or hello packet and has the highest priority. Only the latest packet is stored.
- *Route Break* packets, initiated by a parent node on discovering that one of its children is not connected are placed at the top of a high priority queue followed by disconnect packet, generated by a child node on detecting disconnection from its parent nodes. Next in the queue are data packets that are in transit and have to be forwarded for other nodes. Last in this queue are the Registration Reply used for MMT cluster formation operations. In surveillance application most of the traffic is travelling from a CC to a CH, while the Registration Reply packets go from a CH to a CC i.e., in the opposite direction. Hence these packets were included in the high priority queue.
- Data packets originating from a node are placed at the top of the normal priority queue, which is the third queue. This is followed by Registration Requests, followed by Registration Update packets sent by nodes to inform other CHs, when they join a new branch in a cluster, and finally Registration Acknowledgements by a newly joining node, confirming to the CH its success in joining a cluster. The MMT routing protocol will only pass a data packet down to the BF-MAC layer if a route exists. Hence data packets were given a higher priority over creating new routes to prevent the already established routes from expiring while new routes were being created.

SIFS Timer: BF_MAC uses short inter-frame space (SIFS), between the end of transmission of a packet and the beginning of the next packet to avoid collisions. In Fig. 3, when node 'B' sends an RTS packet and node 'A' receives the RTS, it will wait for SIFS time before forwarding the RTS onto the CH. CH waits SIFS time before sending EXP_CTS. This applies through the entire data session.

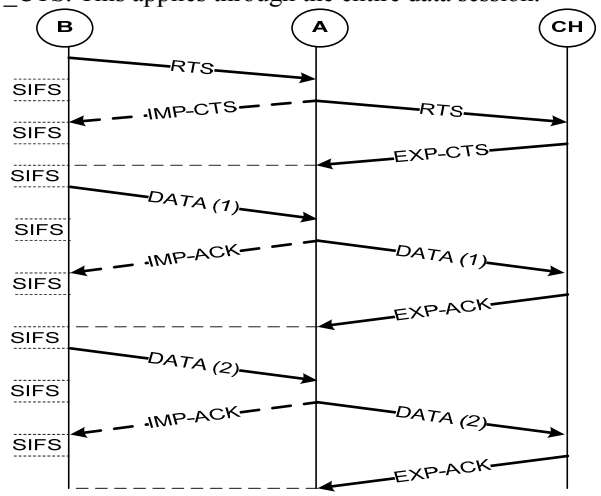


Fig. 3 Use of SIFS in BF_MAC

IV. SIMULATIONS

The proposed solution was modeled using the Opnet. Aggregation nodes were designated and allowed to move within a coverage area of surveillance network as shown in Fig. 4, to limit path distance between data collection and aggregation nodes. This would provide a best case scenario of data collection for all schemes.

Random walk mobility model was used for node movement as the study focused on ground surveillance. Node speeds were varied from 3 m/s, 5 m/s to 10 m/s. 'Hello' interval for all schemes was set to 2 seconds. Three different scenarios, one with 20 nodes and 2 aggregations nodes, second with 40 nodes and 4 aggregation nodes and lastly 80 nodes and 8 aggregation nodes.

The graphs presented capture the performance when all data collection nodes are each sending 10 Kbyte files in intervals of 1 second for the 20 node scenario and 2 seconds for the 40 and 80 nodes scenario. The 2 second interval was selected for the higher node scenario to reduce congestion.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

Due to the sequential file transfers, the traffic pattern resembles streaming data, with the exception that packet arrival was modeled as five packets (size 2 Kbytes) per one (or two) seconds. At the physical layer, packets with single *bit error rates* were dropped and *forward error correction* was not enabled. All other physical layer parameters were as provided in the standard Opnet 802.11 WLAN models. The data rates were maintained at 11 Mbps.

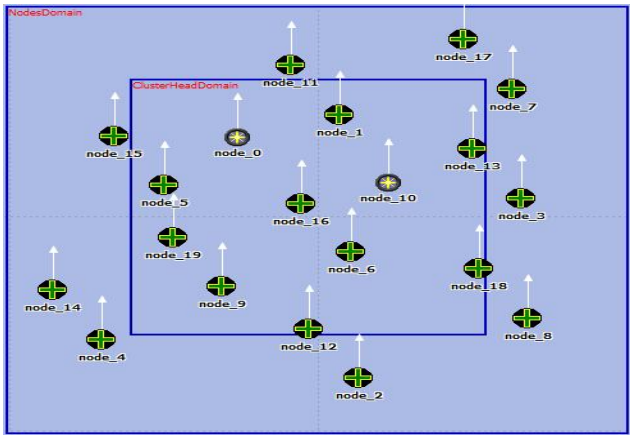


Fig. 4 CHs and CCs in a 20 node scenario

V. PERFORMANCE ANALYSIS

Success rate and latency in packet delivery and control overhead were recorded. Control overhead was calculated as percentage of control traffic to total traffic in bits.

Success Rate: Figures 5A, B and C respectively are plots of success rate achieved with OLSR, AODV and MMT when the mobile node speeds were varied from 3, 5 to 10 m/s. On the x axis is the number of nodes in the scenario. So the plot shows the performance variations as the number of nodes increase in the network and thus its scalability.

With node speeds of *3 m/s,* the success rate of MMT based solution drops from 98% to 96% as network size increases from 20 to 80 nodes. Success rate for OLSR drops from 94 to 91%, while AODV dropped from 93% to 85%.

As the scenario is one of data aggregation, and the data aggregation nodes are explicitly identified as the destination nodes and zone restricted OLSR scales better that AODV. Moreover reactive routing schemes do not perform well when the number of sending nodes is high – and the tests conducted in these cases had all data collection nodes sending files simultaneously to the data aggregation nodes.

When the speeds of the nodes were increased to *5 m/s,* the gap between MMT and OLSR based schemes shows an increase. MMT still maintains a success rate between 97% with 20 nodes to 93% with 80 nodes. While OLSR drops from 90 % scenario to 86% for the 80 node scenario, AODV success rate dropped down to 81% for the 80 node scenario.

OLSR performance degrades faster with increasing node speeds compared to MMT and ADOV, which is further noticed in the plot where node speeds were increased to *10 m/s* in Fig. 5C. The plot for OLSR gets closer to the plot for

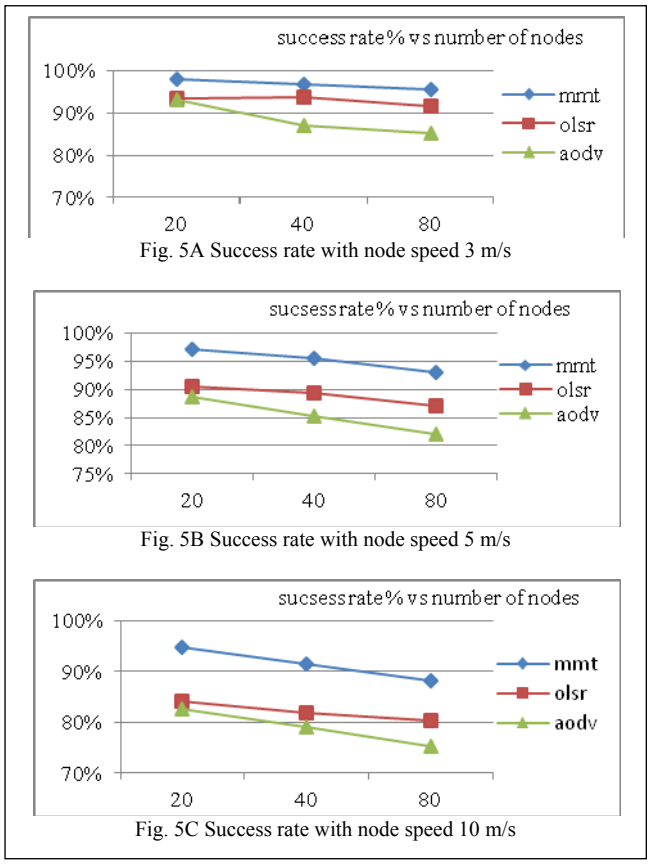


Fig. 5A Success rate with node speed 3 m/s

Fig. 5B Success rate with node speed 5 m/s

Fig. 5C Success rate with node speed 10 m/s

AODV, while the gap between MMT plot and OLSR plot increases. To summarize in Fig. 5C, MMT success rate is 8% higher than OLSR success rate for the 80 node scenario.

The inference from the three plots would be that with the same settings when node speeds (in this case all nodes) increase MMT performance deteriorates by 3 to 5% (20 nodes to 80 nodes), OLSR deteriorates by 9 to 11% (20 nodes to 80 nodes) while AODV deteriorates around 10%.

Overhead: Figures 6A, B and C are respectively the plots for control overhead expressed as a percentage to the total traffic as node speeds were varied from 3, 5 to 10 m/s. MMT based solutions show an overhead of 10% maximum with 80 nodes with node speeds of 3 m/s which goes to 19% when the node speeds were increased to 10 m/s. OLSR shows a consistent overhead of nearly 60% even in the 80 node scenario. This is because OLSR is not adaptive to dynamic topology changes as MMT i.e., OLSR sends hello packets and TC packets at a certain interval, which are independent of node movement or topology changes. AODV exhibits an overhead that varies from 60 to 68%, because it is also an adaptive protocol. However it is unable to cope in successful packet delivery as the node speeds and the network size increases, which was apparent in Fig. 5.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

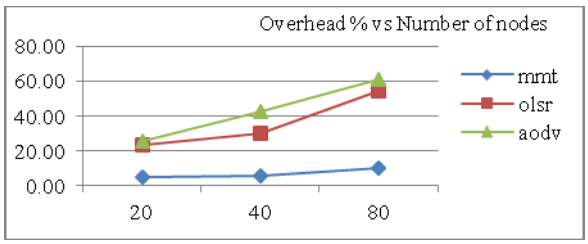


Fig. 6A Overhead % with node speed 3 m/s

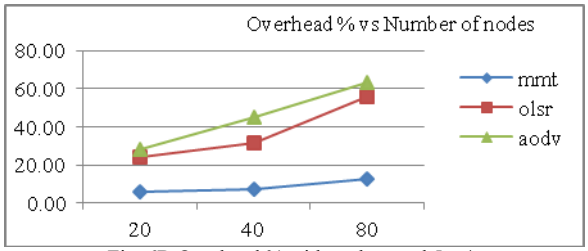


Fig. 6B Overhead % with node speed 5 m/s

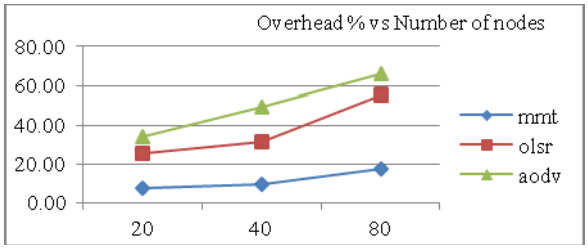


Fig. 6C Overhead % with node speed 10 m/s

Latency: Figures 7 A, B and C are the plots of average end to end latency incurred by the successfully delivered packets for varying node speeds.

MMT based solutions have an average end to end packet delivery latencies of 0.01 seconds when the node speeds were maintained at 3 m/s. OLSR exhibits slightly higher latencies. However ADOV latency varies from 0.04 seconds to 0.12 seconds when the network size increases from 20 nodes to 80 nodes. This can be explained when the average hops encountered in each case is considered next.

The average end to end packet delivery latency with OLSR is slightly better than that of MMT when the node speed is increased to 10 m/s. However t MMT successfully delivered 8 to 12 % more traffic. The increase in latency can also be accounted for when one looks at the average number of hops that the packets were delivered over in the case of MMT and OLSR. AODV on the other hand is significantly disadvantaged at higher node speeds and larger network sizes due to the fact that all non-aggregation nodes are sending data traffic. Fig. 7D is the plot for the maximum latency encountered when the node speed was maintained at 3 m/s. This graph has been provided just to show that while OLSR and MMT still maintain low maximum latencies AODV packets experience very high latencies – up to 30 seconds. However when the network size increases, this latency drops which can be attributed to the fact that the traffic delivered successfully has dropped considerably and in the case of a large network as the path length increases (discussed next) many packets are not delivered.

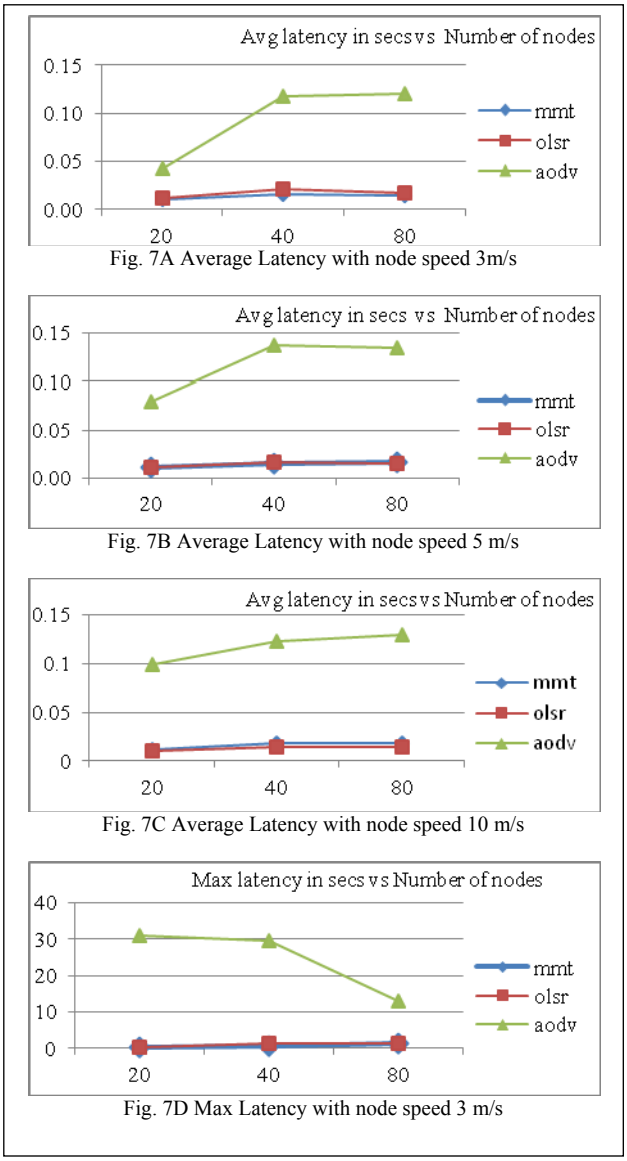


Fig. 7A Average Latency with node speed 3m/s

Fig. 7B Average Latency with node speed 5 m/s

Fig. 7C Average Latency with node speed 10 m/s

Fig. 7D Max Latency with node speed 3 m/s

Path Length in Hops: Figures 8A and B are the plots for path lengths encountered in the three schemes. These plots help understand the performance trends in previous graphs. We include only one set in this case when the node speed was maintained at 3 m/s and record the average and maximum hops encountered. Fig. 8A is a plot of maximum hops for the three schemes, when the node speed was maintained at 3 m/s. AODV records maximum hops of 12 with a network size of 80, which goes to show the lower success rate and high latencies encountered by AODV.

Fig. 8B on the other hand is the plot of the average hops. MMT recorded an average of 1.5 hops for the 80 node scenario, OLSR recorded average hops of 1.2, while AODV recorded 2.1 hops. It is worth noting that OLSR recorded and used the shortest paths among the three schemes. This is

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

because it collects the topology information and uses Dijkstra's algorithm to compute the shortest path. MMT on the other hand focuses on route robustness and quick dynamic adaption to topology changes with an intention to keep nodes connected. Hence while in OLSR the routing table was not updated due to the lack of timely collection of topology information, MMT continued to dynamically maintain multiple routes for every node, which were updated as nodes moved and the neighbors changed (with lower overhead). AODV recorded a high value in maximum path lengths but, the low average value indicates that such situations were rare.

VI. CONCLUSIONS

In this article, we introduced a new architecture for MANET use in critical surveillance applications. The goal is include the minimal set of functions across the protocol layers to facilitate timely and reliable delivery of data at a few aggregation nodes. For this purpose an integration framework was developed based on the MMT algorithm. The algorithm allowed integrated operation of clustering, proactive routing and MAC using a single address. A new technique of random access is introduced.

The proposed solution was evaluated and compared with standard OLSR and AODV routing protocols operating over WLAN 802.11 MAC. To the best of our knowledge this is the first article that compares such MANET performance under stressful operating conditions. Furthermore the article also brings to light the good performance achievable with OLSR for surveillance type MANETs, if the operating conditions are set accordingly. Reasons for AODV's performance deterioration under such scenarios are also explained. Lastly the proposed MMT based solution is shown to outperform both AODV and OLSR when node speeds and the network size increases, given that the operational parameters are maintained constant.

ACKNOWLEDGMENT

This work was supported by funding from Office off the Naval Research (ONR), USA.

REFERENCES

[1] Nirmala S., Pan Y., Narayan D., Ross D. and Lutzer C., "Route Robustness of a Multi-meshed Tree Routing Scheme for Internet MANETs", Proceeding of IEEE Globecom 2005. 28 Nov – 2nd Dec. 2005 St Louis, pp 3346-3351.

[2] Martin N., Al-Mousa Y. and Shenoy N, "An integrated routing and medium access control framework for surveillance networks with mobile nodes", ICDCN 2010, Bangalore, India. pp 315-323.

[3] Abolhasan M., Wysocki T. and Dutkiewicz E., "A review of routing protocols for mobile ad hoc networks", Journal of ad hoc networks, Elsevier publications, 2004, pp 1-22.

[4] Perkins C., E., Royer E. M., and Das S. R., "Ad Hoc On-Demand Distance Vector (AODV) Routing", IETF Mobile Ad Hoc Networks Working Group", IETF RFC 3561

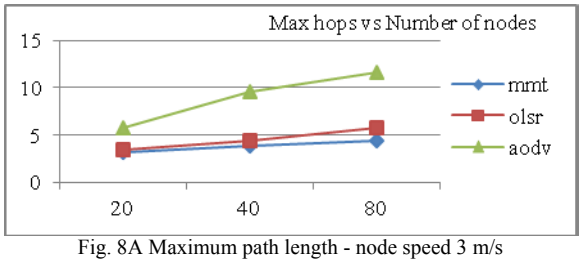


Fig. 8A Maximum path length - node speed 3 m/s

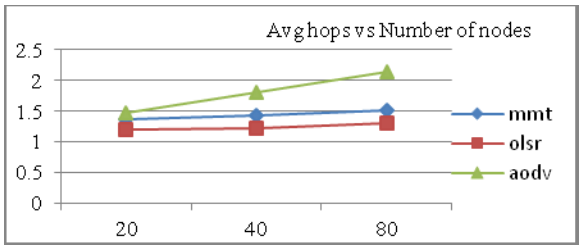


Fig. 8B Average path length – node speed 3 m/s

[5] Clausen T., Ed. and Jacquet P., Optimized Link State Routing Protocol (OLSR), Network Working Group, RFC: 3626

[6] Basagni S., Chlamtac I., and Farago A., "A generalized clustering algorithm for peer-to-peer networks", Workshop on Algorithmic Aspects of Communication, July 1997, pp 1-15.

[7] Hong X., Xu K. and Gerla M., "Scalable Routing Protocols for Mobile Ad Hoc Networks", IEEE Network Journal, July/Aug 2002, Vol 16, issue 4, pp 11-21.

[8] Pei G., Gerla M., and Chen T.-W., "Fisheye State Routing: A Routing Scheme for Ad Hoc Wireless Networks," IEEE International Conference on Communications, 2000, Vol 1, pp 70-74.

[9] Pei G., Gerla M., Hong X., and Chiang C. -C., "A Wireless Hierarchical Routing Protocol with Group Mobility," in Proceedings of IEEE WCNC'99, New Orleans, LA, Sept. 1999. pp 1583-42.

[10] Haas Z.J. and Pearlman M.R., "The Performance of Query Control Schemes for the Zone Routing Protocol," ACM/IEEE Transactions on Networking, vol. 9, no. 4, August 2001, pp. 427-438.

[11] Ibrahim, I. S., Etorban, A. and King, P.J.B., "Multipath Distance Vector Zone Routing Protocol for Mobile ad hoc networks (MDVZRP)", The 9th PG Net, Liverpool John Moores University, UK, pp. 171-176, 23-24 June 2008

[12] Pei G., Gerla M. and Hong X., "LANMAR: Landmark Routing for Large Scale Wireless Ad Hoc Networks with Group Mobility," in Proceedings of IEEE/ACM MobiHOC 2000, Boston, MA, Aug. 2000, pp. 11-18.

[13] Lin, C.R. and Gerla, M., "Adaptive clustering for mobile wireless networks," Selected Areas in Communications, IEEE Journal on , vol.15, no.7, pp.1265-1275, Sep 1997

[14] Basagni, S., "Distributed and mobility-adaptive clustering for multimedia support in multi-hop wireless networks," Vehicular Technology Conference, 1999. VTC 1999 - Fall. IEEE VTS 50th , vol.2, no., pp.889-893 vol.2, 1999

[15] Grönkvist J., A. Hansson A. and Nilsson J., "A comparison of access methods for multi-hop ad hoc radio networks", IEEE Vehicular Technology Conference, 2000, pp. 1435–1439.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

Resilience Issues for Application Workflows on Clouds

Toàn Nguyên
Project OPALE
INRIA Grenoble Rhône-Alpes
Grenoble, France
Toan.Nguyen@inria.fr

Jean-Antoine-Désidéri
Project OPALE
INRIA Sophia-Antipolis Méditerranée
Sophia-Antipolis, France
Jean-Antoine.Desideri@inria.fr

***Abstract*-Two areas are currently the focus of active research, namely cloud computing and high-performance computing. Their expected impact on business and scientific computing is such that most application areas are eagerly uptaking or waiting for the associated infrastructures. However, open issues still remain. Resilience and load-balancing are examples of such areas where innovative solutions are required to face new or increasing challenges, e.g., fault-tolerance. This paper presents existing concepts and open issues related to the design, implementation and deployment of a fault-tolerant application framework on cloud computing platforms. Experiments are sketched including the support for application resilience, i.e., fault-tolerance and exception-handling. They also support the transparent execution of distributed codes on remote high-performance clusters.**

Keywords-workflows-fault-tolerance; resilience; simulation; cloud computing; high-performance computing.

I. INTRODUCTION

The future of computing systems in the next decade is sometimes advertised as a combination of virtual labs running large-scale application workflows on clouds that operate exascale computers [40][41]. Although this vision is attractive, it currently carries some inherent weaknesses. Among them are the complexity of the applications, e.g., multi-scale and multi-disciplines, the technical layers barrier, e.g., the network infrastructures, the multicore HPC clusters and finally the overwhelming technicalities that rely on experts that are not the final users. The consequences are that important challenges still lay ahead of us, among which are error management, fault-tolerance and application resilience.

Error recovery has long been a difficult challenge for both the computer science engineers and the application users. Approaches dealing with errors, failures and faults have mostly been designed by system engineers [20]. The characterization of faults and failures is indeed made inside software systems [37]. The emergence and widespread use of high-performance multi-core systems is also increasing the concerns for error-prone infrastructures where the mean-time between failures is decreasing [44].

Operating and communication systems have long addressed the failure detection and recovery problems with sophisticated restart and fail-safe protocols, from both the theoretical and implementation perspectives [39][42].

However, the advent of high-performance computing systems and distributed computing environments provide opportunities for new challenging applications to be deployed and run in order to solve unprecedented complex problems, e.g., full 3D aircraft flight dynamics simulation [2]. This stimulates the design of large-scale and long-running multi-discipline and multi-scale applications. They are expected to be standard within the next decade.

This induces expectations from the designers and users of such applications, e.g., better application accuracy, better performance, high-level and user-friendly interfaces, and resilience capabilities.

Consequently, rising concerns appear questioning the characterization, tracing and recovery from errors in such complex applications [43].

Indeed, the number and variety of components invoked during the execution of these applications are increasing:

- Operating system components (system libraries)
- Network components (virtual nodes, servers, backbones, protocols, messaging, duplication, etc.)
- Middleware components (resource allocation, authentication, authorization, load-balancing, etc.)
- Application components (software libraries for synchronization, results storage and migration, computation, user interfaces, etc.).

This results in several layers of software where the early detection of errors and their effective recovery are crucial with respect to resource allocation, usage cost, performance, system survivability, application consistency and user satisfaction [6][8]. Therefore, the software stack includes several different logics that must be carefully taken into account, i.e., identified and coordinated, in case of errors [20][44].

This paper explores the design, implementation and use of cloud infrastructures from the application perspective. It proposes specific techniques to handle application errors and recovery. The cloud infrastructure includes heterogeneous hardware and software components. Further, the application codes must interact in

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

a timely, secure and effective manner. Additionally, because the coupling of remote hardware and software components is prone to run-time errors, sophisticated mechanisms are necessary to handle unexpected failures at the infrastructure, system and application levels [19][25]. Consequently, specific management software is required to handle unexpected application and system behaviors [9][11][12][15][45].

The paper is focused on reactive approaches to occurring errors. It does not address error prevention and proactive approaches, e.g., preventive data and code migration and duplication [44]. Neither does it address prevention issues based on statistical evaluation and prediction of error occurrences and log analysis.

Indeed, the paper follows the position mentioned in [44]: "*This limited comprehension of root causes makes fault effect avoidance (the capability to avoid the effects of faults) difficult. Without a good understanding of root causes, it seems illusory to design and validate fault prediction mechanisms. Without good fault prediction systems, research on proactive actions is almost useless. In addition, even if at some point, we are capable of predicting errors accurately, we still have to find: 1) acceptable solutions to handle false negatives, and 2) how to handle predicted software errors (process or virtual machine migration is not a response for software errors)*".

This paper focuses on application resilience, i.e., survivability mechanisms to ensure the consistent termination of the applications, in the case of unexpected faulty behavior. Section II is an overview of related work. Section III is a description of open issues and gives an overview of running testcases. Section IV is a conclusion.

II. RELATED WORK

A. Definitions

Application resilience uses several notions that need to be detailed:

- Errors
- Faults
- Failures
- Exceptions
- Recovery
- Fault-tolerance
- Robustness
- Resilience

The generic term *error* usually encompasses different types of abnormal situations and behaviors. These might originate in system, middleware and application unexpected discrepancies.

In systems such as Apache's ODE [37], system *failures* and application *faults* address different types of errors.

A *failure* to resolve a DNS address is different from a process fault, e.g., a bad expression. Indeed, a system failure does not impact the correct logics of the application process at work, and should not be handled by it, but by the system error-handling software instead: "*failures are non-terminal error conditions that do not affect the normal flow of the process*" [37].

However, an activity can be programmed to throw a *fault* following a system *failure*, and the user can choose in such a case to implement a specific application behavior, e.g., a number of activity retries or its termination.

Application and system software usually raise *exceptions* when faults and failures occur. The exception handling software then handles the faults and failures. This is the case for the YAWL workflow management system [46][47][48], where specific *exlets* can be defined by the users [4]. They are components dedicated to the management of abnormal application or system behavior (Figure 1). The extensive use of these exlets allows the users to modify the behavior of the applications in real-time, without stopping the running processes. Further, the new behavior is stored as a component workflow which incrementally modifies the application specifications. The latter can therefore be modified dynamically to handle changes in the user requirements.

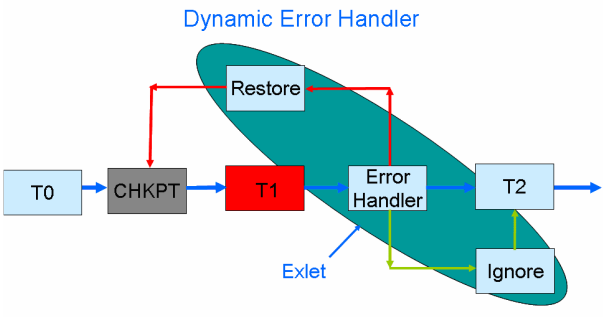


Figure 1. Error-handler.

Fault-tolerance is a generic term that has long been used to name the ability of systems and applications to handle errors. Transactional systems for example need to be fault-tolerant [38]. Critical business and scientific applications need to be fault-tolerant, i.e., to resume consistently in case of internal or external errors.

Therefore *checkpoints* need to be designed at specific intervals to backtrack the applications to consistent points in the application execution, and *restart* be enabled from there. They form the basis for *recovery* procedures.

Application *robustness* is the property of software that are able to survive consistently from data and code errors. This area is a major concern for complex numeric software that deal with data *uncertainties*. This is particularly the case for simulation applications [24].

This is also a primary concern for the applications faced to system and hardware errors. In the following, we include both (application external) fault-tolerance and (internal) robustness in the generic term *resilience* [1].

Therefore we do not follow here the definition given in [44]: "*By definition a failure is the impact of an error itself caused by a fault.*"

But, we fully adhere to the following observation: "*the response to a failure or an error depends on the context*

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

and the specific sensitivity to faults of the usage scenarios, applications and algorithms" [44].

B. *Overview*

Simulation is a prerequisite for product design and for scientific breakthrough in many application areas ranging from pharmacy, biology to climate modeling and aircraft design [25]. They all require extensive simulation and testing. This requires often large-scale multidiscipline experiments, including the management of petabytes volumes of data and large multi-core supercomputers [10].

In such application environments, various teams usually collaborate on several projects or part of projects. Computerized tools are often shared and tightly or loosely coupled [23]. Some codes may be remotely located and non-movable. This is supported by distributed code and data management facilities [29]. And unfortunately, this is prone to a large variety of unexpected errors and breakdowns [30].

Data replication and redundant computations have been proposed to prevent from random hardware and communication failures [31], as well as failure prediction [32], sometimes applied to deadline-dependent scheduling [12].

System level fault-tolerance in specific programming environments is also proposed, e.g., CIFTS [15]. Also, middleware usually support mechanisms to handle fault-tolerance in distributed job execution, usually calling upon data replication and redundant code execution [9][15][22][24].

Also, erratic application behavior needs to be supported [34]. This implies evolution of the application process in the event of such occurrences. Little has been done in this area [33][35]. The primary concerns of the application designers and users have so far focused on efficiency and performance [36]. Therefore, application unexpected behavior is usually handled by re-designing and re-programming pieces of code and adjusting parameter values and bounds. This usually requires the simulations to be stopped and restarted.

The concerns focus therefore on application resilience, although intra-node fault-tolerance is also a major concern [39].

Studies have focused on reducing checkpoint sizes and frequency, as well writing overheads [40]. Examples are diskless checkpointing [43], compressed checkpoints [44] and incremental checkpointing [41].

An extensible approach for petascale and future exascale systems is proposed in [45], based on a multi-level checkpointing scheme called *Scalable Checkpoint/Restart* (SCR) which proves to be effective. It provides an explicit checkpoint model to compute the optimal number of checkpoint levels and frequency of checkpoints at each level. The model and strategy are used to predict the checkpointing overhead and performance of the systems targeted. They are assessed by experiments on thousands of run hours on several production HPC clusters. This results in a thorough analysis of the impact of checkpoint intervals on overall system efficiency with respect to failure rate, compute intervals and file systems costs.

A dynamic approach is presented in the following sections. It support the evolution of the application behavior using the introduction of new exception handling rules at run-time by the users, based on occurring (and possibly unexpected) events and data values. The running workflows do not need to be suspended in this approach, as new rules can be added at run-time without stopping the executing workflows.

This allows on-the-fly management of unexpected events. This approach also allows a permanent evolution of the applications that supports their continuous adaptation to the occurrence of unforeseen situations [35]. As new situations arise and data values appear, new rules can be added to the workflows that will permanently take them into account in the future. These evolutions are dynamically hooked onto the workflows without the need to stop the running applications. The overall application logic remains therefore unchanged. This guarantees a constant adaptation to new situations without the need to redesign the existing workflows. Further, because exception-handling codes are themselves defined by dedicated component workflows, the user interface remains unchanged [14].

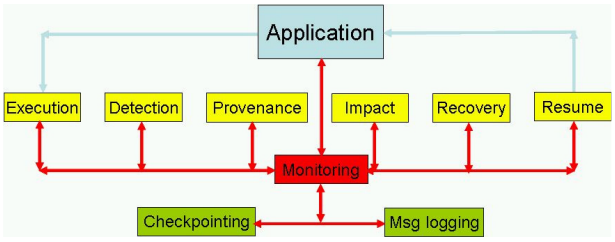


Figure 2. Architecture of a resilience sub-system.

III. OPEN ISSUES

A. *Error management*

Many open issues are still the subject of active research concerning application resilience. The paradigm ranges from code and data duplication and migration, to the monitoring of application behavior, and this includes also quick correctness checks on partial data values, the design of error-aware algorithms, as well as hybrid checkpointing-message logging features (Figure 2).

The baseline is:

- The early detection of errors,
- Root cause characterization,
- Characterization of transient vs. persistent errors,
- The tracing and provenance of faulty data,
- The identification of the impacted components and their associated corrupted results,

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

- The ranking of the errors (warnings, fatal, medium) and associated actions (ignore, restart, backtrack),
- The identification of pending components,
- The identification and purge of transient messages,
- The secured termination of non-faulty components,
- The secure storage of partial and consistent results,
- The quick recovery of faulty and impacted components,
- The re-synchronization of the components and their associated data,
- The properly sequenced restart of the components.

Each of these items needs appropriate implementation and algorithms in order to orchestrate the various actions required by the recovery of the faulty application components.

B. *Error detection*

The early characterization of errors is difficult because of the complex software stack involved in the execution of multi-discipline and multi-scale applications on clouds. The consequence is that errors might be detected long after the root cause that initiated them occurred. Also, the error observed might be a complex consequence of the root cause, possibly in a different software layer.

Similarly, the exact tracing and provenance data may be very hard to sort out, because the occurrence of the original fault may be hidden deep inside the software stack.

Without explicit data dependency information and real-time tracing of the components execution, the impacted components and associated results may be unknown. Hence, there is a need for explicit dependency information [38].

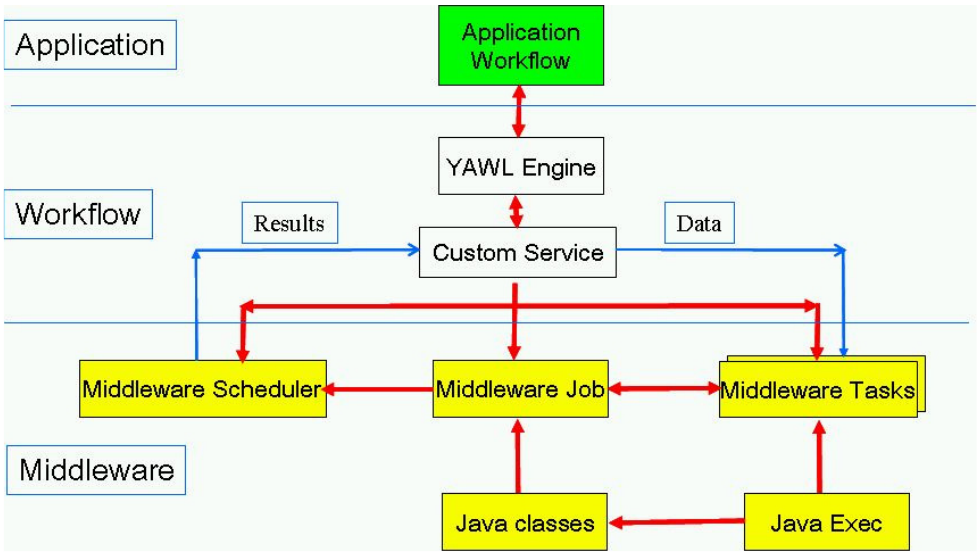


Figure 3. The YAWL workflow and middleware interface.

The ranking of errors is dependent on the application logic and semantics (e.g., default values usage). It is also dependent on the logics of each software layer composing the software stack. Some errors might be recoverable (Unresolved address, resource unavailable, etc.), some others not (Network partition, etc.). In each case, the actions to recover and resume differ: ignore, retry, reassign, suspend, abort.

In all cases, resilience requires the application to include four components:

- A monitoring component for (early) error detection,
- A (effective) decision system, for provenance and impact assessment,
- A (low overhead) checkpointing mechanism,
- An effective recovery mechanism.

Further, some errors might be undetected and transient. Without explicit data dependency information and real-time tracing of the components execution, the impacted components and associated results may be unknown. Hence there is a need for explicit dependency information between the component executing instances and between the corresponding result data [38].

A sub-system dedicated to application resilience includes therefore several components in charge of specific tasks contributing to the management of errors and consistent resuming of the applications (Figure 2). First, it includes an intelligence engine in charge of the application monitoring and of the orchestration of the resilience components. This engine runs as a background process in charge of event listening during the execution of the applications. It is also in charge of triggering the periodic checkpointing mechanism, depending on the policy defined for the applications being monitored. It is also in charge of triggering the message-logging component for safekeeping the messages exchanged

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

between tasks during their execution. This component is however optional, depending on the algorithms implemented, e.g., checkpointing only or hybrid checkpoint-message logging approaches. Both run as background processes and should execute without user intervention. Should an error occur, an error detection component that is constantly listening to the events published by the application tasks and the operating system raises the appropriate exceptions to the monitoring component. The following components are then triggered in such error cases: an optional provenance component which is in charge of root cause characterization, whenever possible. An impact assessment component is then triggered to evaluate the consequences of the error on the application tasks and data, that may be impacted by the error. Next, a recovery component is triggered in charge of restoring the impacted tasks and the associated data, in order to re-synchronize the tasks and data, and restore the application to a previous consistent state. A resuming component is finally triggered to deploy and rerun the appropriate tasks and data on the computing resources, in order to resume the application execution.

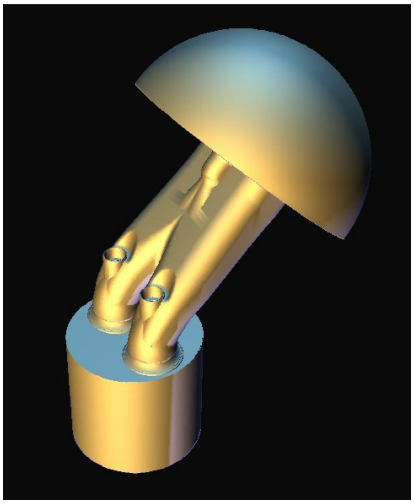


Figure 4. Cylinder optimized input pipes – courtesy Lab. Roberval, Université de Technologie de Compiègne (France).

In contrast with approaches designed for global fault-tolerance systems, e.g., CIFTS [15], this functional architecture describes a sub-system dedicated to application resilience. It can be immersed in, or contribute to, a more global fault-tolerance system that includes also the management of system and communication errors.

C. *Experiments*

A distributed platform featuring the resilience capabilities described above is developed [30], based on the YAWL workflow management system [46][48]. Experiments are connecting the platform to the FAMOSA optimization suite [24] developed at INRIA by project OPALE [29].

The experiments are deployed on the Grid5000 infrastructure [13]. This involves five different locations throughout France (Figure 5), including two locations near Paris for CAD data and mesh generation. In addition, another location near Nantes involves CFD calculations, and another one in Sophia-Antipolis near Nice is dedicated to optimization. The last location in Grenoble is for application deployment, monitoring and result visualization (Figure 5).

The first experiments simulated this deployment scenario by duplicating the application with two identical parallel sequences running on Lyon and Grenoble clusters respectively, then on Sophia-Antipolis and Grenoble respectively.

This allowed for performance assessment of the various clusters implied on Lyon, Sophia-Antipolis and Grenoble.

Because Grid5000 infrastructure does not currently serve Nantes, a further experiment will invoke the clusters in Rennes instead, Lille instead of Paris2, and Orsay instead of Paris1.

An extension will invoke one more cluster in Lyon instead of one of the Sophia-Antipolis instances. A total of six remote HPC clusters will therefore be invoked (Figure 6). The reason for this is that most application codes are proprietary and are located at the various partners offices.

Data transfers between clusters use a 10 Gbps IP network infrastructure dedicated to Grid5000 (Figure 10).

All the locations involve HPC clusters and are invoked from a remote workflow running on a Linux workstation in Grenoble.

The various errors that are taken into account by the resilience algorithm include run-time errors in the solvers, inconsistent CAD and mesh generation files, and execution time-outs [3].

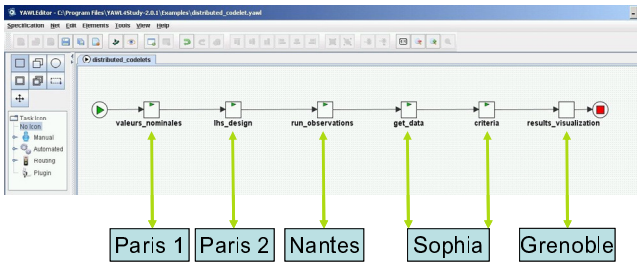


Figure 5. The workflow experiment schema.

FAMOSA is currently tested for car rear mirrors optimization (Figure 7) and by ONERA (the French National Aerospace Research Office) for aerodynamics optimization.

FAMOSA is an acronym for "*Fully Adaptive Multilevel Optimization Shape Algorithms*" and includes C++ components for:

- CAD generation,
- Mesh generation,
- Domain partitioning,
- Parallel CFD solvers using MPI, and
- Post-processors.

The input is a design vector and the output is a set of simulation results. The components also include other software for mesh generation, e.g., Gmsh [26], partitioning,

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

e.g., Metis [27] and solvers, e.g., Num3sis [28]. They are remotely invoked from the YAWL application workflow by shell scripts [30].

The FAMOSA components are triggered by remote shell scripts running for each one on the HPC cluster. The shell scripts are called by YAWL custom service invocations from the user workflow running on the workstation [30].

Other testcases implemented by academic and industry partners include the optimization of cylinder input pipes and valves for car engines (Figure 4 and 8) and vehicle aerodynamics (Figure 9).

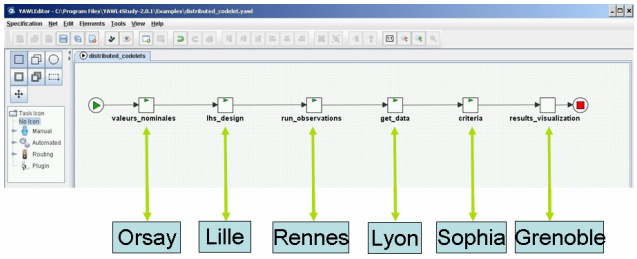


Figure 6. The final distributed workflow.

IV. CONCLUSION

The requirements for large-scale simulations make it necessary to deploy various software components on heterogeneous distributed computing infrastructures [10][33]. These environments are often remotely located among a number of project partners for administrative and collaborative purposes.

An overview of resilience is given with open issues. A workflow distributed platform and running testcases are briefly described. The underlying interface to the distributed components is a middleware providing resource allocation and job scheduling [13]. Besides fault-tolerance provided by the middleware, which handles communication and hardware failures, the users can define and handle application errors at the workflow level. Application errors may result from unforeseen situations, data values and boundary conditions. In such cases, user intervention is required in order to modify parameter values and application behavior. Complex error characterization is then invoked to assess the impact on the executing tasks and the data involved. The approach presented uses dynamic rules and constraint enforcement techniques, combined with asymmetric checkpoints. It is based on the YAWL workflow management system.

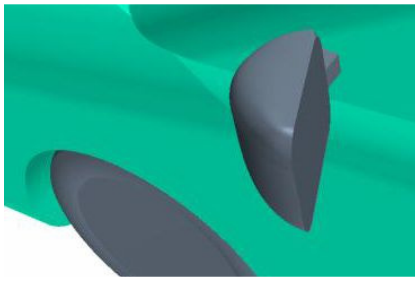


Figure 7. Optimized rear mirror (courtesy CD-adapco).

ACKNOWLEDGMENTS

This work is supported by the European Commission FP7 Cooperation Program "Transport (incl. aeronautics)", for the *GRAIN* Coordination and Support Action *("Greener Aeronautics International Networking"*), grant ACS0-GA-2010-266184. It is also supported by the French National Research Agency ANR (*Agence Nationale de la Recherche*) for the OMD2 project (*Optimisation Multi-Discipline Distribuée)*, grant ANR-08-COSI-007, program COSINUS (*Conception et Simulation*).

The authors wish to thank Laboratoire Roberval (Université de Technologie de Compiègne, France), and also CD-adapco, for the testcase implementations and the images.

REFERENCES

[1] T. Nguyên, L. Trifan and J-A Désidéri . "A Distributed Workflow Platform for Simulation". Proc. 4th Intl. Conf on Advanced Engineering Computing and Applications in Sciences (ADVCOMP2010). pp. 375-382. Florence (I). October 2010.

[2] A. Abbas, "High Computing Power: A radical Change in Aircraft Design Process", Proc. of the 2nd China-EU Workshop on Multi-Physics and RTD Collaboration in Aeronautics. Harbin (China) April 2009.

[3] T. Nguyên and J-A Désidéri, "Dynamic Resilient Workflows for Collaborative Design", Proc. of the 6th Intl. Conf. on Cooperative Design, Visualization and Engineering (CDVE2009). Luxemburg. September 2009. Springer-Verlag. LNCS 5738, pp. 341–350 (2009)

[4] W. Van der Aalst et al., Modern Business Process Automation: YAWL and its support environment, Springer (2010).

[5] E. Deelman et Y. Gil., "Managing Large-Scale Scientific Workflows in Distributed Environments: Experiences and Challenges", Proc. of the 2nd IEEE Intl. Conf. on e-Science and the Grid. pp. 165-172. Amsterdam (NL). December 2006.

[6] M. Ghanem, N. Azam, M. Boniface and J. Ferris. "Grid-enabled workflows for industrial product design", Proc. of the 2nd Intl. Conf. on e-Science and Grid Computing. pp. 285-294. Amsterdam (NL). December 2006.

[7] G. Kandaswamy, A. Mandal and D.A. Reed., "Fault-tolerant and recovery of scientific workflows on computational grids", Proc. of the 8th Intl. Symp. On Cluster Computing and the Grid. pp. 415-428. 2008.

[8] H. Simon. "Future directions in High-Performance Computing 2009-2018". Lecture given at the ParCFD 2009 Conference. Moffett Field (Ca). May 2009.

[9] J. Wang, I. Altintas, C. Berkley, L. Gilbert and M.B. Jones., "A high-level distributed execution framework for scientific workflows", Proc. of the 4th IEEE Intl. Conf. on eScience. pp. 147-156. Indianapolis (In). December 2008.

[10] D. Crawl and I. Altintas, "A Provenance-Based Fault Tolerance Mechanism for Scientific Workflows", Proc. of the 2nd Intl. Provenance and Annotation Workshop. IPAW 2008. Salt Lake City (UT). June 2008. Springer. LNCS 5272. pp 152-159.

[11] M. Adams and L. Aldred, "The worklet custom service for YAWL, Installation and User Manual, Beta-8 Release", Technical Report, Faculty of Information Technology, Queensland University of Technology, Brisbane (Aus.), October 2006.

[12] L. Ramakrishna, D. Nurmi et al.., "VGrADS: Enabling e-Science workflows on grids and clouds with fault tolerance", Proc. ACM SC'09 Conf. pp. 369-376. Portland (Or.), November 2009.

[13] Grid5000 project home. Last accessed: 23/11/2011. https://www.grid5000.fr/mediawiki/index.php/Grid5000:Home.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

[14] Dongarra, P. Beckman et al. "The International Exascale Software Roadmap". Volume 25, Number 1, 2011, International Journal of High Performance Computer Applications, pp. 77-83. Available at: http://www.exascale.org/ Last accessed: 03/31/2011.

[15] R. Gupta, P. Beckman et al. "CIFTS: a Coordinated Infrastructure for Fault-Tolerant Systems", Proc. 38th Intl. Conf. Parallel Processing Systems. pp. 145-156. Vienna (Au). September 2009.

[16] D. Abramson, B. Bethwaite et al. "Embedding Optimization in Computational Science Workflows", Journal of Computational Science 1 (2010). Pp 41-47. Elsevier.

[17] A.Bachmann, M. Kunde, D. Seider and A. Schreiber, "Advances in Generalization and Decoupling of Software Parts in a Scientific Simulation Workflow System", Proc. 4th Intl. Conf. Advanced Engineering Computing and Applications in Sciences (ADVCOMP2010). Pp 247-258. Florence (I). October 2010.

[18] R. Duan, R. Prodan and T. Fahringer. "DEE: a Distributed Fault Tolerant Workflow Enactment Engine for Grid Computing", Proc. 1st. Intl. Conf. on High-Performance Computing and Communications. pp. 255-267. Sorrento (I). LNCS 3726. September 2005.

[19] Sherp G., Hoing A., Gudenkauf S., Hasselbring W. and Kao O., "Using UNICORE and WS-BPEL for Scientific Workflow execution in Grid Environments", Proc. EuroPAR 2009. pp. 133-148. LNCS 6043. . Springer. 2010.

[20] B. Ludäscher, M. Weske, T. McPhillips and S. Bowers, "Scientific Workflows: Business as usual ?", Proc. BPM 2009. pp. 269-278. LNCS 5701. Springer. 2009.

[21] Montagnat J., Isnard B., Gatard T., Maheshwari K. and Fornarino M., "A Data-driven Workflow Language for Grids based on Array Programming Principles", Proc. SC 2009 4th Workshop on Workflows in Support of Large-Scale Science. pp. 23-35. WORKS 2009. Portland (Or). ACM 2009.

[22] Yildiz U., Guabtni A. and Ngu A.H., "Towards Scientific Workflow Patterns", Proc. SC 2009 4th Workshop on Workflows in Support of Large-Scale Science. pp. 135-145. WORKS 2009. Portland (Or). ACM 2009.

[23] Plankensteiner K., Prodan R. and Fahringer T., "Fault-tolerant Behavior in State-of-the-Art Grid Workflow Management Systems", CoreGRID Technical Report TR-0091. October 2007. http://www.coregrid.net Last accessed: 03/31/2011.

[24] Duvigneau R., Kloczko T., and Praveen C.., "A three-level parallelization strategy for robust design in aerodynamics", Proc. 20th Intl. Conf. on Parallel Computational Fluid Dynamics (ParCFD2008). pp. 241-252. May 2008. Lyon (F).

[25] E.C. Joseph, et al. "A Strategic Agenda for European Leadership in Supercomputing: HPC 2020", IDC Final Report of the HPC Study for the DG Information Society of the EC. July 2010. Available at: http://www.hpcuserforum.com/EU/ Last accessed: 03/31/2011.

[26] Gmsh. https://geuz.org/gmsh/ Last accessed: 03/31/2011.

[27] Metis. http://glaros.dtc.umn.edu/gkhome/metis/metis/overview Last accessed: 03/31/2011.

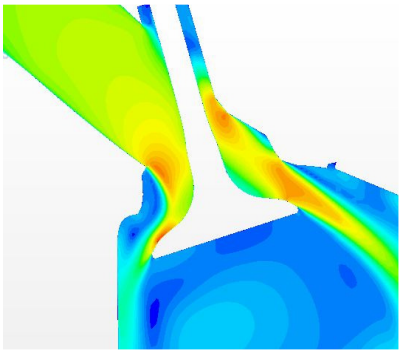


Figure 8. Cylinder flow simulation (courtesy CD-adapco).

[28] Num3sis. http://num3sis.inria.fr/blog/ Last accessed: 03/31/2011.

[29] OPALE project at INRIA. http://www-opale.inrialpes.fr and http://wiki.inria.fr/opale Last accessed: 05/03/2011.

[30] T. Nguyên, L. Trifan, J.A. Désidéri. "A Workflow Platform for Simulation on Grids", Proc. 7th Intl. Conf. on Networking and Services (ICNS2011). pp. 295-302. Venice (I). May 2011.

[31] Plankensteiner K., Prodan R. and Fahringer T., "A New Fault-Tolerant Heuristic for Scientific Workflows in Highly Distributed Environments based on Resubmission impact", Proc. 5th IEEE Intl. Conf. on e-Science. Oxford (UK). December 2009. pp 313-320.

[32] Z. Lan and Y. Li. "Adaptive Fault Management of Parallel Applications for High-Performance Computing", IEEE Trans. On Computers. pp. 45-56. Vol. 57, No. 12. December 2008.

[33] S. Ostermann, et al. "Extending Grids with Cloud Resource Management for Scientific Computing", Proc. 10th IEEE/ACM Intl. Conf. on Grid Computing. Pp. 266-278. 2009.

[34] E. Sindrilaru, A. Costan and V. Cristea. "Fault-Tolerance and Recovery in Grid Workflow Mangement Systems", Proc. 4th Intl. Conf. on Complex, Intelligent and Software Intensive Systems. pp. 162-173. Krakow (PL). February 2010.

[35] S. Hwang and C. Kesselman. "Grid Workflow: A Flexible Failure Handling Framework for the Grid", Proc. 12th IEEE Intl. Symp. on High Performance Distributed Computing. pp. 369-374. Seattle (USA). 2003.

[36] The Grid Workflow Forum. Last accessed: 06/21/2011. http://www.gridworkflow.org/snips/gridworkflow/space/start

[37] The Apache Foundation. http://ode.apache.org/bpel-extensions.html #BPELExtensions-ActivityFailureandRecovery Last accessed: 08/25 /2011.

[38] W. Zang, M. Yu, P. Liu. "A Distributed Algorithm for Workflow Recovery", Intl. Journal Intelligent Control and Systems. Vol. 12. No. 1. March 2007. pp 56-62.

[39] P. Beckman. "Facts and Speculations on Exascale: Revolution or Evolution?", Keynote Lecture. Proc. 17th European Conf. Parallel and Distributed Computing (Euro-Par 2011). pp. 135-142. Bordeaux (F). August 2011.

[40] P. Kovatch, M. Ezell, R. Braby. "The Malthusian Catastrophe is Upon Us! Are the Largest HPC Machines Ever Up?", Proc. Resilience Workshop at 17th European Conf. Parallel and Distributed Computing (Euro-Par 2011). pp. 255-262. Bordeaux (F). August 2011.

[41] R. Riesen, K. Ferreira, M. Ruiz Varela, M. Taufer, A. Rodrigues. "Simulating Application Resilience at Exascale", Proc. Resilience Workshop at 17th European Conf. Parallel and Distributed Computing (Euro-Par 2011). pp. 417-425. Bordeaux (F). August 2011.

[42] P. Bridges, et al. "Cooperative Application/OS DRAM Fault Recovery", Proc. Resilience Workshop at 17th European Conf. Parallel and Distributed Computing (Euro-Par 2011). pp. 213-222. Bordeaux (F). August 2011.

[43] Proc. 5th Workshop INRIA-Illinois Joint Laboratory on Petascale Computing. Grenoble (F). June 2011. http://jointlab.ncsa.illinois.edu/events/workshop5/ Last accessed 09/05/2011.

[44] F. Capello, et al. "Toward Exascale Resilience", Technical Report TR-JLPC-09-01. INRIA-Illinois Joint Laboratory on PetaScale Computing. Chicago (Il.). 2009. http://jointlab.ncsa.illinois.edu/

[45] Moody A., G.Bronevetsky, K. Mohror, B. de Supinski. Design, "Modeling and evaluation of a Scalable Multi-level checkpointing System", Proc. ACM/IEEE Intl. Conf. for High Performance Computing, Networking, Storage and Analysis (SC10). pp. 73-86. New Orleans (La.). Nov. 2010. http://library-ext.llnl.gov Also Tech. Report LLNL-TR-440491. July 2010. Last accessed: 09/12/2011.

[46] Adams M., ter Hofstede A., La Rosa M. "Open source software for workflow management: the case of YAWL", IEEE Software. 28(3): 16-19. pp. 211-219. May/June 2011.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

[47] Russell N., ter Hofstede A. “Surmounting BPM challenges: the YAWL story.”, Special Issue Paper on Research and Development on Flexible Process Aware Information Systems. Computer Science. 23(2): 67-79. pp. 123-132. March 2009. Springer 2009.

[48] Lachlan A., van der Aalst W., Dumas M., ter Hofstede A. “Dimensions of coupling in middleware”, Concurrency and Computation: Practice and Experience. 21(18):233-2269. pp. 75-82. J. Wiley & Sons 2009.

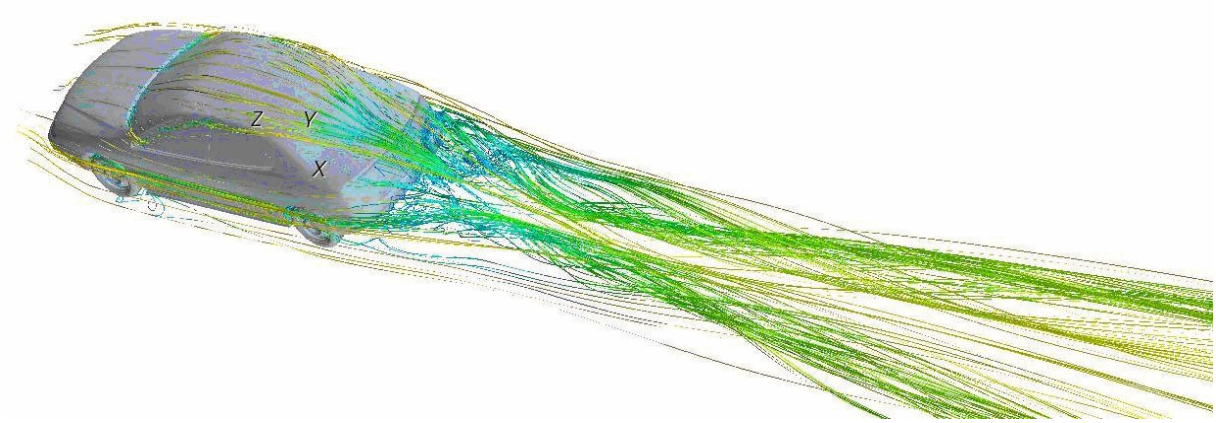


Figure 9. Vehicle aerodynamics simulation (courtesy CD-adapco).

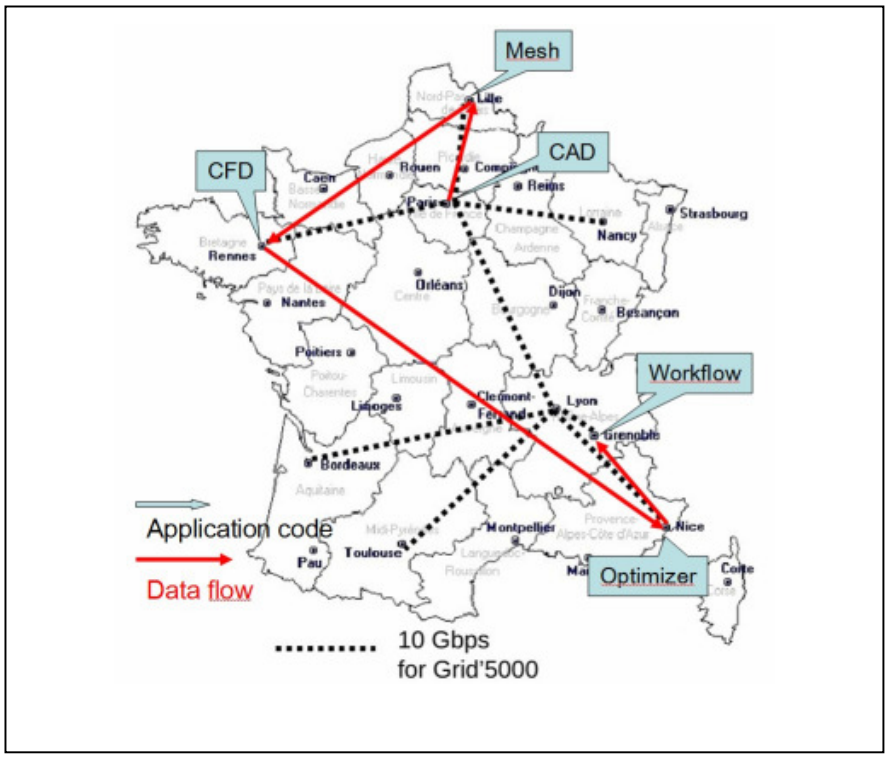


Figure 10. The testcase deployment on the Grid5000 infrastructure.

A Routing Method for Cooperative Forwarding in Multiple Wireless Sensor Networks

Junko Nagata, Kazuhiko Kinoshita, Koso Murakami

Department of Information Networking,
Graduate School of Information Science and Technology,
Osaka University, Japan
Email: {nagata.junko,kazuhiko,murakami}@ist.osaka-u.ac.jp

Yosuke Tanigawa, Hideki Tode

Department of Computer Science and Intelligent Systems,
Graduate School of Engineering,
Osaka Prefecture University, Japan
Email: {tanigawa,tode}@cs.osakafu-u.ac.jp

***Abstract*—In recent years, the number of applications of wireless sensor networks (WSNs) has been increasing. Multiple WSNs can be constructed within the same geographic area. We propose a routing method for cooperative forwarding in such multiple WSNs that will extend their lifetime. For multiple WSNs, each sink location will differ from the others, and some nodes around a sink in one WSN may be far from a sink in another WSN. We focus on this issue in the proposed method, with a node that is far from a sink in its own network and near to a sink in another network being able to forward packets from a node in another WSN to the corresponding sink.**

Keywords*-*multiple sensor network; cooperative forwarding; load balancing.

I. INTRODUCTION

Wireless sensor networks (WSNs) are composed of tiny battery-powered sensor nodes that have limited storage and radio capabilities [1], [2]. Therefore, for WSNs to remain operational for a long time, much attention has to be paid to energy consumption in the nodes.

In a typical WSN, sensor nodes acquire and send data to a processing center called the *sink*. Because all data are forwarded to a sink, nodes around the sink tend to transmit many more packets than the others [3]. In this case, the energy of such nodes will exhaust earlier than that of other nodes, causing an "energy hole" to appear around the sink. No more data can be delivered to the sink after the hole appears. Consequently, the energy remaining in the rest of the network is wasted, and the network lifetime is shorter than it could be [4].

In some applications, a WSN may comprise several thousand sensor nodes within an extended area (e.g., agriculture and environmental monitoring). In these cases, the diameter of the WSN may be some kilometers. To enable networks to be scalable, a WSN is typically subdivided into clusters and the data collected by cluster heads are sent to a sink. Clustering also supports data aggregation. This is a method by which data from multiple sensors are combined to eliminate redundant information and transmission, thereby reducing energy consumption [5].

From another point of view, WSNs can be classified into two types, namely homogeneous and heterogeneous sensor networks. In a homogeneous WSN, all nodes have the same capabilities. In recent years, however, heterogeneous WSNs have attracted much attention. These have a small number of "high-end" sensor nodes, with a wider range of radio communication capabilities and/or a larger battery compared with the "normal" nodes. A clustering method to achieve effective use of these high-end nodes has been proposed [6]. However, a clustering method alone is not sufficient to prolong the network lifetime for a heterogeneous WSN, and a clustering and multi-hop hybrid routing method has therefore been proposed [7].

In recent years, multiple WSNs have been constructed within the same geographic area [8], [9]. For such cases, researchers have been investigating cooperation between the WSNs. Some routing protocols for multiple WSNs have been proposed [10], [12], [11], [13].

In this paper, we propose a routing method for cooperative forwarding to prolong network lifetime by reducing the load on nodes around sinks in a multiple-WSN environment. By reducing the load around a sink, we aim to overcome the problem of some nodes becoming "bottlenecks". Our method decides how much other WSNs with different sink locations can help such "heavy-load" situations.

The remainder of the paper is organized as follows. We first introduce some related work in Section Ⅱ. In Section Ⅲ, we propose a routing method for cooperative forwarding to reduce the load around a sink. In Section Ⅳ, we evaluate the performance of the proposed method. Finally, Section Ⅴ concludes this paper and indicates directions for future work.

II. RELATED WORK

In recent years, the number of applications of WSNs has been increasing, with multiple WSNs attracting much attention. For such environments, some protocols have been proposed, as follows.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

The *Virtual Sensor Network (VSN)* is an emerging concept for supporting multipurpose, collaborative and resource-efficient WSNs by enabling, for example, dynamic variations to the subsets of sensors and users [10]. A VSN is formed as a logical network of cooperative nodes. For cases where applications overlap geographically, transmitting data for applications among a variety of devices enables the nodes to reduce redundant paths. Nodes are classified into an appropriate VSN based on the phenomena they are tracking (e.g., container tracking or corrosion-rate monitoring). It is expected that VSNs will provide protocols for the construction methods, maintenance and usage of subsets of sensors, providing a way to communicate efficiently between intermediate nodes or other VSNs.

Poorter et al. [11] propose to construct an overlay network for different WSNs. However, this protocol has the problem that differences in operating policies and radio communications are not considered.

Steffan et al. [12] focus on a general concept for the creation and maintenance of network-wide node subsets and describe a flexible and modular architecture that meets the requirements of multipurpose WSNs. However, the creation of these subsets is not very scalable. Most applications specify their own purpose and construct independent WSNs, but it is not necessary to offer acceptable cost performance and coverage.

Instead of many-to-one routing, many-to-many routing is proposed in [13]. This merges paths with simultaneous traffic, thereby minimizing the number of nodes involved in many-to-many routing and balancing their forwarding load. However, this proposal has the problem that the nodes along the merged paths have a heavier load and consume more energy.

Elhawary et al. [14] model a cooperative transmission link in wireless networks in terms of a transmitter cluster and a receiver cluster. They propose a cooperative communication protocol for the establishment of these clusters and for the cooperative transmission of data. The nodes calculate the link cost between these paths and select the most efficient path. This enables the nodes to reduce their energy consumption, but the nodes in a cluster need to be synchronized and each node has the increased overhead of forming many paths.

Taking another point of view, we propose to achieve load balancing by cooperative forwarding in multiple WSNs.

III. Proposed Method

A. Concept

It is assumed that there are n WSNs in the same area. These WSNs have different applications. In addition, their start and finish times may differ, depending on each network's requirements.

As shown in Figure 1, the locations of the sinks in multiple WSNs are separated. Some nodes around a sink in one WSN may therefore be far from a sink in another WSN. We focus on this fact in the proposed method, whereby a node that is far from a sink in its own network, but near a sink in another network, can forward a packet from a node in another WSN to the corresponding sink. In this paper, we call the former network the *home network* and the latter network the *visitor network*. The method achieves load balancing between a heavy-load node in a home network and a light-load node in a visitor network. As a result, the lifetime of both networks can be extended.

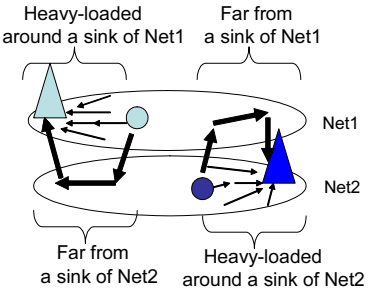


Figure 1. Heavy-load node around a sink

Specifically, each network constructs a path along which a node can't forward a packet from a node in another WSN in advance. It is based on the well-known Ad hoc On-Demand Distance Vector (AODV) protocol [15]), making it easy to implement. In addition, some nodes construct routes to the sinks of visitor networks.

B. Node Function

As described above, the proposed method enables a node that is far from a sink in its home network, but near a sink in a visitor network, and can forward a packet from a node in the visitor network.

Each node has a routing table that includes not only an entry for a sink in its home network but also an entry for a sink in the visitor network. When a node overhears a data packet from its visitor network, it decides whether to receive and forward it or to ignore it. This procedure is explained later in more detail.

C. Routing Table Creation

This subsection explains how to create the routing table.

Initially, each node sends an AODV-based route request packet to create an entry in its routing table for a sink in its home network. After this creation process, each node broadcasts an additional route request packet named *B-REQ* to the sinks of all visitor networks. (Nodes on the path from the node to the sink will create an entry in their routing table to the sink.)

In addition, as a metric to decide the next hop, *minEnergy* is also notified. This refers to the minimum residual energy of nodes along the path to the sink.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

D. Cooperative Routing Method

In the proposed method, when a node sends its sensing data, it attaches the value of its residual energy in a header field of the packet. When a node relays a packet, it compares the residual energy of the node itself and that recorded in the packet, and the recorded value is replaced by smaller one. As a result, the minimum energy along the path from its source node is recorded. According to this procedure, a node can record a value of minimum energy along the paths every networks and select the path for the maximum value of this energy.

Figure 2 demonstrates how the proposed method works. After node P has the created path in its home network (Sink1), it broadcasts a B-REQ to the visitor networks Net2 and Net3. When node Q and node R receive this B-REQ, they write their network ID in the header of the B-REQ and transmit it to their sink. After this procedure, the routes from node P to the sink via Net2 and Net3 are created, as shown in Tables I and II, respectively.

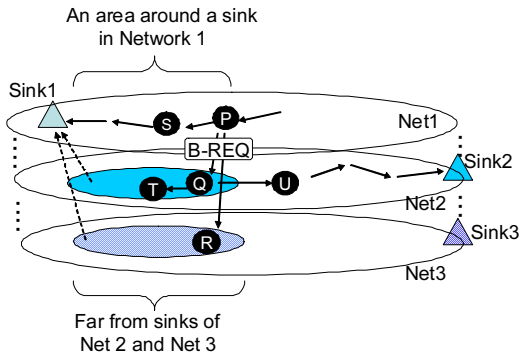


Figure 2. Multiple WSNs

Table I
NODE P'S ROUTING TABLE

Relay network	Destination	Next hop	minEnergy
1	Sink1	S	30
2	Sink1	Q	45
3	Sink1	R	65

Table II
NODE Q'S ROUTING TABLE

Relay network	Destination	Next hop	minEnergy
2	Sink2	U	55
2	Sink1	T	40

When node P receives a data packet for Sink1, it selects a suitable route from the entries in its routing table as shown in Table I. In this case, node P selects node R, which has the maximum value for minEnergy, as the next-hop node.

As we described below, the proposed method tries to extend the lifetime of each network by cooperative forwarding. However, it may result in a case where a network shortens its lifetime by the burden of forwarding for visitor networks. To avoid such a situation, in the proposed method, a node which has less residual energy does not relay packets from visitor networks.

Specifically, we define a value of *cooperation threshold* in each network as a metric to decide whether to forward packets from visitor networks or not. For this metric, each sink broadcasts the minimum value of residual energy among all the nodes in its home network to the nodes in its home network. When a node acquires the value, it compares with its own residual energy. If its own residual energy is smaller, it refuses to forward packets from visitor networks and applies itself to relay packets in its home network.

IV. PERFORMANCE EVALUATION

We evaluated the performance of the proposed method by simulation experiments using QualNet4.5.1 [16]. We simulated two WSNs (Network 1 and Network 2) as follows. Each WSN had 49 nodes based on a grid topology, as shown in Figure 3. The sensing field was a 60 m $\times$ 60 m square. The maximum range of radio transmission for each node was 15 m. One sink was located at the top left corner and another was at the bottom right corner. Each node sent data packets asynchronously, at intervals of 5 s.

For a comparison method, we simulated two WSNs coexisting in the same area, but with each WSN communicating independently (i.e., there was no cooperation and no sharing).

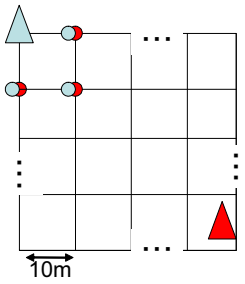


Figure 3. Simulation model

Figures 4 and 5 show the packet receiving rate of the sinks in Network 1 and Network 2, respectively. The receiving rate means the percentage of the packets which the sinks have successfully received for all sent packets by sensor nodes. It was observed every 400 seconds. The horizontal axis is elapsed time.

These figures show that the proposed method extends the lifetime of both Network 1 and Network 2. It indicates that the proposed method achieves good load balancing, with both networks benefitting.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

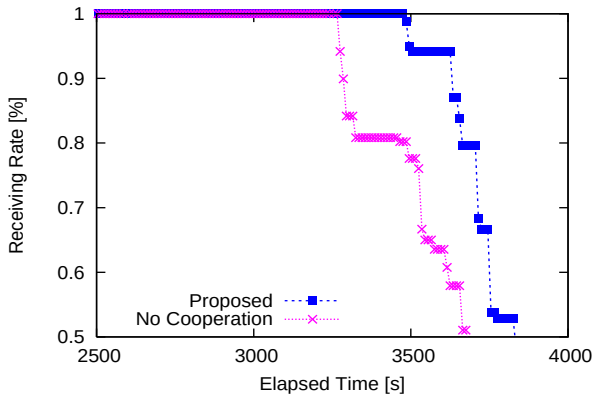


Figure 4. Receiving rate (Network 1)

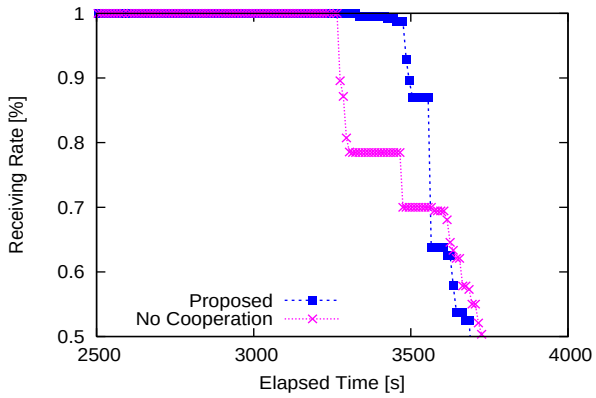


Figure 5. Receiving rate (Network 2)

V. Conclusion and Future Work

In this paper, we focused on the situation where multiple WSNs operate simultaneously in the same area, and aimed to extend the lifetime of all WSNs by introducing cooperation between them.

To achieve this, we proposed a cooperative forwarding method. This enables a node near a sink in a visitor network to forward a packet from a node in the visitor network to achieve load balancing.

Simulation results showed that the proposed method can extend the lifetime of all WSNs working in the same area.

This study still has work in progress, and we recognize the following problems.

The method for choosing the next hop is too simple. We should at least take the number of hops to the sink into account. The simulation model is elementary, and we continue to evaluate the proposed method in a wider variety of situations.

References

[1] I. Akyildiz, W. Su, Y. Sankarasubramaniam and E. Cayirci, “A survey on sensor networks,” IEEE Communications Magazine, Vol. 40, No. 8, pp. 102–114, Aug. 2002.

[2] J. Yick, B. Mukherjee and D. Ghosal, “Wireless sensor network survey,” Computer Networks 52, pp. 2292–2330, Apr. 2008.

[3] M. Perillo, Z. Cheng and W. Heinzelman, “On the problem of unbalanced load distribution in wireless sensor networks,” Proceedings of the IEEE GLOBECOM Workshops on Wireless Ad Hoc and Sensor Networks, pp. 74–79, Dec. 2004.

[4] X. Wu, G. Chen and S. Das, “Avoiding Energy Holes in Wireless Sensor Networks with Nonuniform Node Distribution,” IEEE Transactions on Parallel and Distributed Systems, Vol. 19, No. 5, pp. 710–720, May 2008.

[5] A. A. Abbasi and M. Younis, “A survey on clustering algorithms for wireless sensor networks,” Computer Communications, Vol. 30, No. 14–15, pp. 2826–2841, 2007.

[6] S. Priyankara, K. Kinoshita, H. Tode and K. Murakami, “A Clustering Method for Wireless Sensor Networks with Heterogeneous Node Types,” IEICE Transactions on Communications, Vol. E94-B, No. 8, pp. 2254–2264, Aug. 2011.

[7] S. Priyankara, K. Kinoshita, H. Tode and K. Murakami, “A Clustering/multi-hop Hybrid Routing Method for Wireless Sensor Networks with Heterogeneous Node Types,” Proceedings of the IEEE GLOBECOM Workshops on Heterogeneous Multi-hop Wireless and Mobile Networks, pp. 207–212, Dec. 2010.

[8] “e-Sense project”, http://www.ist-e-sense.org/ (Jan 2012, date last accessed).

[9] “U-City project”, http://ucta.or.kr/en/ucity/concept.php (Jan 2012, date last accessed).

[10] A. P. Jayasumana, Q. Han and T. H. Illangasekare, “Virtual sensor networks – a resource efficient approach for concurrent applications,” Proceedings of the International Conference on Information Technology (ITNG’07), pp. 111–115, 2007.

[11] E. Poorter, B. Latre, I. Moerman and P. Demeester, “Symbiotic Networks: Towards a New Level of Cooperation Between Wireless Networks,” International Journal of Wireless Personal Communications, pp. 479–495, Jun. 2008.

[12] J. Steffan, L. Fiege, M. Cilia and A. Buchmann, “Towards Multi-Purpose Wireless Sensor Networks,” Proceedings of the International Conference on Sensor Networks (IEEE SENET’05), 2005.

[13] L. Mottola and G. P. Picco, “MUSTER: Adaptive Energy-Aware Multi-Sink Routing in Wireless Sensor Networks,” IEEE Transactions on Mobile Computing, Vol. 10, pp. 1694–1709, 2010.

[14] M. Elhawary and Z. J. Haas, “Energy-Efficient Protocol for Cooperative Networks,” IEEE/ACM Transactions on Networking, Vol. 19, No. 2, Apr. 2011.

[15] C. E. Perkins, E. M. Royer and S. R. Das, “Ad Hoc On Demand Distance Vector (AODV) Routing,” http://www.ietf.org/internet-drafts/draft-ietf-manet-aodv-07.txt, Nov. 2000.

[16] QualNet Simulator version 4.5.1, Scalable Network Technologies, http://www.scalable-networks.com (Jan 2012, date last accessed).

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

Virtual Environment for Next Generation Sequencing Analysis

Olivier Terzo, Lorenzo Mossucca
Infrastructure and Systems for Advanced Computing (IS4AC)
Istituto Superiore Mario Boella (ISMB)
Torino, Italy
Email: (terzo,mossucca)@ismb.it

Andrea Acquaviva, Francesco Abate, Rosalba Provenzano
Department of Control and Computer Engineering
Politecnico di Torino
Torino, Italy
Email: (andrea.acquaviva,francesco.abate)@polito.it
rosalba.provenzano@studenti.polito.it

***Abstract*—Next Generation Sequencing technology, on the one hand, allows a more accurate analysis, and, on the other hand, increases the amount of data to process. A new protocol for sequencing the messenger RNA in a cell, known as RNA-Seq, generates millions of short sequence fragments in a single run. These fragments, or reads, can be used to measure levels of gene expression and to identify novel splice variants of genes. The proposed solution is a distributed architecture consisting of a Grid Environment and a Virtual Grid Environment, in order to reduce processing time by making the system scalable and flexible.**

***Keywords*-grid computing; cloud computing; virtual; next generation sequencing; hybrid architecture.**

I. INTRODUCTION

Next Generation Sequencing (NGS) technologies, also known as second generation, have revolutionized research in the field of biology and genomic with the ability to draw from a single experiment a larger amount of data sequence with the previous technology known as Sanger Sequencing [1] and [2]. The main novelty introduced by the NGS platform is to obtain from the molecules of DNA/RNA of smaller fragments, called read, which are sequenced in parallel thus reducing the processing time. Aberrant mutations in the RNA transcription, as chimeric transcripts, are on the base of various forms of disease and NGS proved to be extremely helpful in making the detection of these events more accurate and reliable. However, even if from the biological point of view NGS technology leads to new exciting perspectives spreading an incredible amount of data, on the other hand it raised new challenges in the development of tools and informative infrastructures. An NGS machine produces millions of reads in a single run that must be successively elaborated and analyzed. TopHat is a program that aligns RNA-Seq reads to a genome in order to identify exon-exon splice junctions. It is built on the ultrafast short read mapping program Bowtie [5]. TopHat finds splice junctions without a reference annotation. By first mapping RNA-Seq reads to the genome, TopHat identifies potential exons, since many RNA-Seq reads will contiguously align to the genome. Using this initial mapping, TopHat builds a database of possible splice junctions, and then maps the reads against this junction to confirm them. The goal is to offer to biologist private infrastructures to conduct their research and to respond to the ever evolving needs of NGS users. Cloud Computing is rapidly emerging as an alternative platform for the computational and data needs of our community. Biologists are already using the Amazon Elastic Cloud Computing (EC2) infrastructure for their research. In some situations, it is preferable to use a number of instances of a tailored Virtual Machine (VM) than submitting jobs to the own existing infrastructure. Grids appear mainly in high performance computing environments. In this context, several of off-the-shelf nodes can be linked together and work in parallel to solve problems, that, previously, could be addressed sequentially or by using supercomputers. Grid Computing is a technique developed to elaborate enormous amounts of data and enables large-scale resource sharing to solve problem by exploiting distributed scenarios. The main advantage of Grid is due to parallel computing, indeed if a problem can be split in smaller tasks, that can be executed independently, its solution calculation speed up considerably.

The paper is organized as follows: Section 2 the motivation is discussed. Section 3 explains the biological background and software used. Section 4 shows the architecture design: grid and virtual environment and schedulers functionalities. Section 5 is related to the test performances. The last section draws the conclusion and direction for future work.

II. MOTIVATION

The amount of data produced with NGS technology is a positive factor that on a hand contributes to make studies more accurate and reliable identification of mutations in aberrant splicing events, fused genes, on the other hand, open new challenges in the development tools and infrastructure that are able to do the post-processing of data produced in a powerful and timely fashion [3]. A NGS data sample consists of millions of reads, and in a classic situation, with only one workstation available, the time needed to obtain the output increases significantly. In such a context, this computing infrastructure allows to improve overall system performance optimizing the use of resources and increasing

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

the system scalability. Recall also that the alignment is a process in which each mapping reference is made to read independently from the other reads, and this means that you can perform a parallel analysis of the data. Even if the alignment is a very basic operation, due to the great number of data involved in the process, the computational effort in this phase is very high. This scenario recalls for the need of developing computing infrastructures presenting high performances CPU capability and memory availability.

III. Next Generations Sequencing

A. TopHat algorithm

TopHat [7] is a fast splice junction mapper for RNA-Seq reads. It aligns RNA-Seq reads to mammalian-sized genomes using the ultra high-throughput short read aligner Bowtie, and then analyzes the mapping results to identify splice junctions between exons. TopHat is a collaborative effort between the University of Maryland Center for Bioinformatics and Computational Biology and the University of California, Berkeley Departments of Mathematics and Molecular and Cell Biology. TopHat receives as input reads produced by the Illumina Genome Analyzer, although users have been successful in using TopHat with reads from other technologies. The input samples consist of two files of about 37 million of reads each. The two files are FASTA formatted paired-end reads. Dealing with paired-end reads means that the reads are sequenced by the sequencing machine only on the end of the same DNA/RNA molecule, thus the sequence in the middle part is unknown. Each sequenced end of the same read is also referred as mate. It results in two distinct files, the first one consists in the first mate of the same reads and the second one consists in the opposite mate. TopHat finds junctions by mapping reads to the reference in two phases. In the first phase, the pipeline maps all reads to the reference genome using Bowtie. All reads that do not map to the genome are set aside as initially unmapped reads. Bowtie reports, for each read, one or more alignment containing no more than a few mismatches in the 5'-most s bases of the read. The remaining portion of the read on the 3' end may have additional mismatches, provided that the Phred-quality-weighted Hamming distance is less than a specified threshold. TopHat allows Bowtie to report more than one alignment for a read, and suppresses all alignments for reads that have more than this number. This policy allows so called multireads from genes with multiple copies to be reported, but excludes alignments to low-complexity sequence, to which failed reads often align and then assembles the mapped reads. TopHat extracts the sequences for the resulting islands of contiguous sequence from the sparse consensus, inferring them to be putative exons. TopHat produces a compact consensus file containing called bases and the corresponding reference bases in order to generate the island sequences. TopHat uses the reference genome to call the base. Because most reads covering the ends of exons will also span splice junctions, the ends of exons in the pseudoconsensus will initially be covered by few reads, and as a result, an exons pseudoconsensus will likely be missing a small amount of sequence on each end. In order to capture this sequence along with donor and acceptor sites from flanking introns, TopHat includes a small amount of flanking sequence from the reference on both sides of each island. TopHat has a number of parameters and options, and their default values are tuned for processing mammalian RNA-Seq reads also it can be used for another class of organism.

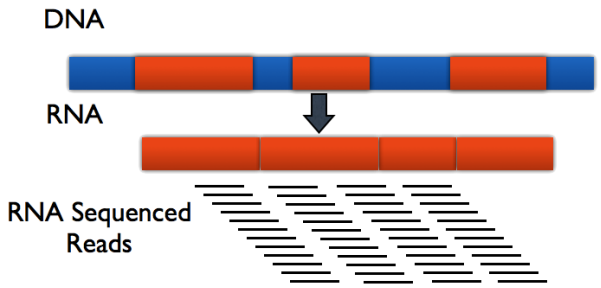


Figure 1. Alignment Phase.

B. Alignment Tools: Bowtie

The short reads alignment is surely the most common operation in RNA-Seq data analysis. The purpose of the alignment is to map each short read fragment onto a genome reference (see Figure 1). From the computational point of view, each short read consists in a sequence of four possible characters corresponding to the DNA bases and the sequence length depends on the sequencing machine adopted for the biological experiment [6]. The main novelty introduced by NGS technology is the capability of sequencing small DNA/RNA fragments in parallel, increasing the throughput and producing very short reads as output. However, this feature make the computational problem more challenging because of the higher amount of read produced and the accuracy in the mapping (the shorter the sequence length, the higher the probability of having multiple matches). For this reason many alignment tools specifically focussed on the alignment of short reads have been recently developed. In the present work, we are interested in characterizing the performances of alignment tools on real NGS data. On the wave of this remark, Bowtie has been chosen, a wide diffused alignment program particularly aimed at align short reads. In order to detect the actual limitation of the alignment phase, we considered real NGS data coming from the analysis of Chronic Myeloid Leukemia. In our analysis flow, the HG19 assembly produced in the 2007 is considered as reference genome the last human genome assembly produced to now. In order to increase the computational performances during

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

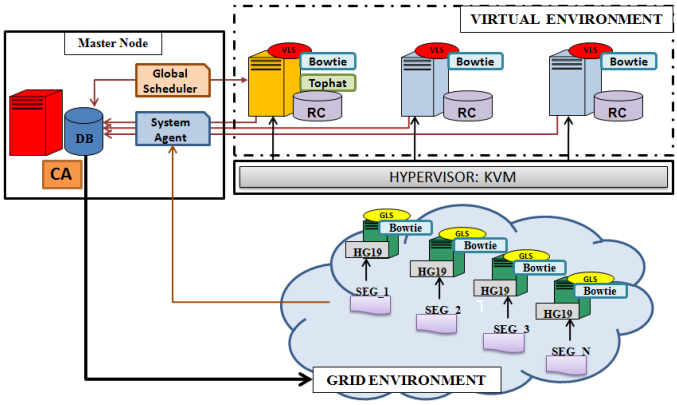


Figure 2. VirtualBio Architecture.

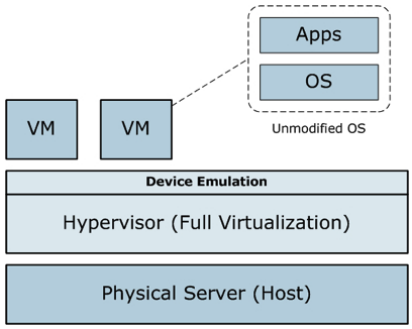


Figure 3. Full Virtualizzation Approach.

the read mapping, Bowtie program creates an index of the provided human genome reference. This operation is particularly straightforward from the computational point of view, but it must be performed only one time for the human genome reference and it is independent on the mapping samples. The alignment phase itself is particularly suitable to be parallelized. In fact, each mapping operation is applied to each read independently on the other read mapping.

C. Distributed Approach

In a preliminary phase of reverse engineering, studying TopHat, blocks of transactions have been highlighted that were executed sequentially. We identified 3 main blocks, that can be executed independently: (a) left and right mate mapped with HG19, (b) segments mapped with HG19, (c) segments mapped with segment juncs. A feature of these 3 blocks is that they are performed by a external software, called Bowtie, as explained before. In steps (a) and (c), since the files involved in the development are significant, we created a common repository that contains the temporary folder used by TopHat. Although the use of a common repository is slightly increased processing time, due to the SSH protocol connection, this time is less than the time of transfer of the entire set of files. Instead the step (b) uses small files these can be performed on a grid, both physical and virtual, because the transfer times are lower. Only difference that the input files are transferred to worker node through Globus Toolkit. These worker nodes when the process is terminated, re-send the output file to the node that requested execution.

IV. VirtualBio Infrastructure

The proposed architecture allows to manage RNA data, prepared by the version of TopHat in Grid, but not only, it could also handle other processing flows that use software and other tools e.g., original version of TopHat or only Bowtie (see Figure 2). The architecture, called VirtualBio, is composed of three main components: a Master Node (MN), a part consists of the Physical Worker Nodes (PWN) that set the grid environment while a part consists of Virtual Worker Nodes (VWN) that set the virtualized environment [9]. The MN is a physical machine with quite good hardware characteristics, is responsible for CA, contains the database, where all information about the nodes belonging to the infrastructure have stored, the node status, the flow of the various biological analysis that can be made in the system and system monitoring. Moreover, on it has been configured the common repository, using the Network File System (NFS). NFS is a protocol developed by Sun Microsystems in 1984 to allow computers to share files and folders over a network [13]. NFS is an open standard, defined in RFCs, and allows any to implement the protocol. Both environments are configured with the middleware Globus Toolkit [11], since it allows obtaining a reliable information technology infrastructure that enables the integrated, collaborative use of computers, networks and databases. The Globus Toolkit is a collection of software components designed to support the development of applications for high performance distributed computing environments, or computational grids.

A. Grid Environment

The Grid environment consists of machines with high computing power, this allows to use own machines and also machines belonging to different virtual organization. The only requirement is to have the necessary software installed for the processing (Bowtie, TopHat and Globus Toolkit). On each worker node of the grid environment is installed the Grid Local Scheduler, an essential component for performing biological tests.

1) Grid Local Scheduler: The Local Grid Scheduler (GLS) is a scheduler active physical machines, has been developed for the design phase (b), it aligns the segment with respect to the human genome (HG19) through Bowtie. Since the transfer of the input file is not influential, the worker nodes do not need to be in the same subnet as the Master Node, but may also belong to different virtual organization, so system can have greater scalability and can use machines powerful performance [10].

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

Reads	1 CPU	2 CPU	3 CPU	4 CPU	5 CPU	6 CPU	7 CPU	8 CPU
100	31	24	24	24	24	25	32	26
1000	26	38	24	24	24	25	13	20
10000	24	24	22	23	22	23	22	22
100000	28	25	25	24	24	24	24	24
1 E06	62	49	50	42	41	41	38	38
10 E06	424	258	230	185	176	165	158	154
85 E06	3425	2044	1791	1432	1342	1247	1180	1048

Table I
BOWTIE EXECUTION TIME (SECONDS)

B. Virtual Environment

Virtualized environment also helps to improve infrastructure management, allowing the use of virtual node template to create virtual nodes in a short time, speeding up the integration of new nodes on the grid and, therefore, improving the reactivity and the scalability of the infrastructure. The open source KVM has been used as hypervisor. It allows to create Fully Virtualized machines. The kernel component of KVM is included in mainline Linux [12]. KVM allows a Full Virtualization solution for Linux on x86 hardware containing virtualization extensions (Intel VT or AMD-V). KVM is implemented as a module within the Linux kernel. A hypervisor hosts the virtual machine images as regular Linux processes, so that each virtual machine image can use all of the features of the Linux kernel, including hardware, security, storage and applications. Full Virtualization provides emulation of the underlying platform on which a guest operating system and application set run without modifications and unaware that the platform is virtualized (see Figure 3). It implies that every platform device is emulated with enough details to permit the guest OS to manipulate them at their native level. Moreover, it allows administrators to create guests that use different operating systems. These guests have no knowledge about the host OS since they are not aware that the hardware they see is not real but emulated. The guests, however, require real computing resources from the host, so they use a hypervisor to coordinate instructions to the CPU. The main advantage of this paradigm concerns the ability to run virtual machines on all popular operating systems without requiring them to be modified since the emulated hardware is completely transparent. The virtualized environment has pre-installed images, which contain all software and libraries needed for running Bowtie and TopHat. The pre-configured images allow an ease instantiation of the machines when needed, and can be easily shutdown after the use.

1) Virtual Local Scheduler: The Virtual Local Scheduler (VLS) is a scheduler active virtual machines. Its purpose is to draw up the steps (a) and (c) of TopHat. As the GLS, the VLS performs the mapping files for input received through Bowtie. The step (a) allows the alignment with respect to the human genome (HG19) and step (c) allows the alignment with respect to the segment juncs previously constituted by TopHat. Since the considerable size of the files involved in these two steps, the VLS works directly on the temporary folder that is located in the common repository, allowing to avoid wasting time due to the transfer of data. Even in this case the interaction with the database is essential and very frequent, network problems may affect the entire biological analysis.

V. PERFORMANCE CONSIDERATIONS

In a preliminary work, we have introduced two case studies, based on Bowtie execution, from two different points of view. The first is the fragment size, while the other one is the CPUs number on the worker node. In Table I a summary of the calculations obtained changing CPUs number are presented. We can notice for reads between 100 and 1000000 no gain of time has occurred, so for our studies only reads from 1000000 to 85000000 are considered. This is because the files have limited data, thus the processing times are already reduced at this stage and then having multiple processors is irrelevant. As we explained before, during an analysis phase of the algorithm, 3 main blocks have been identified, (a) left and right mate aligned with HG19, (b) segments aligned with HG19, (c) segments aligned with segment juncs. The processing time of each segment depends on parameter pthread that is specified in command of Bowtie and refers to the number of parallel processes that can run. In Figure 4, processing time of a single segment of the variation of the parameter pthread is depicted. The test was run on a machine with the following hardware characteristics: Intel Xeon CPU X5660 @ 2.80 GHz, 12 CPUs and 20 GB of RAM, it is worth noting that in order to gain the maximum time the number of pthread must be equal to the number of CPUs. Once past this threshold, the trend is no longer regular, this is due to the scheduling allocation of the CPU operating system. This test allowed to have a vision on the processing time will have access to machines with different power and CPUs number, opening to a more accurate scheduling policy adapted to the needs of time of the biologist. The Table II instead, depicts the processing time of entire flow of TopHat, comparing the original algorithm (sequential version) with the modified algorithm (parallel version). We obtain a considerable gain

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

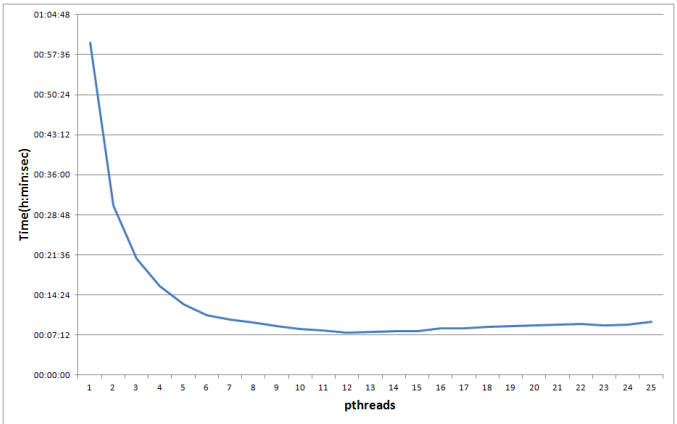


Figure 4. Bowtie Execution Time.

of time that varies depending on the power of machines available. In our tests in order to make homogenous system, we have used machines with the same hardware (12 CPUs).

TopHat	Time [h:min:s]
Sequential Version	03:47:00
Parallel Version	02:10:00

Table II
TOPHAT EXECUTION TIME.

VI. CONCLUSIONS AND FUTURE WORKS

VirtualBio is a tool for NGS analysis, in particular for the alignment through TopHat and Bowtie. The tool wants to offer to biologist private infrastructure to conduct their studies. The novelty of this solution covered both the field of infrastructure and the optimization algorithm of TopHat. Infrastructure is based of two environments: grid and virtual, using a common repository and a set of job schedulers. The TopHat algorithm has been optimized making parallel some sections that were sequential. The architecture allows to reduces the elaboration time by at least 40%. Future work includes the optimization of scheduling policies that are also open to a scenario multisample and implementation of the architecture in Cloud environment, thus increasing the system scalability.

REFERENCES

[1] De Magalhes J.P., Finch C.E. and Janssens G., *Next-generation sequencing in aging research: Emerging applications, problems, pitfalls and possible solutions.*, Ageing Research Reviews, 2010 Jul; Vol. 9(3), pp. 315-323

[2] Sanger F., Nicklen S. and Coulson A.R., *DNA sequencing with chain-terminating inhibitors*, Proc. Natl. Acad. Sci. USA 74, 1977), pp. 5463-5467

[3] Kircher M. and Kelso J., *High-throughput DNA sequencing concepts and limitations.*, Bioessays. 2010 Jun, Vol. 32(6), pp. 524-356

[4] Maher C.A., Palanisamy N., Brenner J.C., Cao X., Kalyana-Sundaram S., Luo S, Khrebtukova I., Barrette T.R., Grasso C., Yu J., Lonigro R.J., Schroth G., Kumar-Sinha C., Chinnaiyan Y., *Chimeric transcript discovery by paired-end transcriptome sequencing*, AM. Proc Natl Acad Sci USA, 2009 July, Vol. 28

[5] Pop M. and S.L. Salzberg, *Bioinformatics challenges of new sequencing technology*, Trends Genet. Vol. 24, 2008

[6] Langmead B., Trapnell C., Pop M. and Salzberg Steven L. *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome*. Genome Biology 10:R25

[7] Trapnell C., Pachter L. and Salzberg Steven L. *TopHat: discovering splice junctions with RNA-Seq*, Bioinformatics (2009), Vol. 25, doi:10.1093/bioinformatics/btp120, pp. 1105-1111

[8] Langmead B., Hansen K. and Leek J. *Cloud-scale RNA-sequencing differential expression analysis with Myrna* Genome Biology (2010) 11:R83

[9] Berman F., Fox G. and Hey A.J.G. *Grid Computing Making the Global Infrastructure a Reality*, Wiley, 2005

[10] Kurowski K., Nabrzyski J., Oleksiak A. and Weglarz J. *Scheduling jobs on the Grid-Multicriteria approach*, Computational Methods in Science and Technology, 12(2), pp. 123-138, 2006

[11] Globus Toolkit, http://www.globus.org/toolkit/, January, 2012

[12] Kernel-based Virtual Machine, http://www.linux-kvm.org/, December, 2011

[13] Network File System, http://wiki.ubuntu-it.org/Server/Nfs, December, 2011

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

A Meshed Tree Algorithm for Loop Avoidance in Switched Networks

Nirmala Shenoy
Networking Security and Systems Administration Department
Rochester Institute of Technology,
Rochester, NY, USA
nxsvks@rit.edu

***Abstract*—Loop free forwarding and routing is a continuing challenge in networks that have link and path redundancy. Solutions to overcome looping in bridged or switched networks are addressed by special protocols at layer 2, which block ports in the bridges to build a logical forwarding spanning tree. In this paper a meshed tree algorithm that aids in building and maintaining multiple overlapped tree branches from a single root node without blocking any ports is presented. Its potential use in bridged networks for loop avoidance is discussed. Some of its salient features are compared with spanning tree-based protocols and TRILL (Transparent Interconnection of Lots of Links) on RBridges (router bridges), another solution proposed for resolving loops in bridged networks.**

Keywords- Loop Avoidance; Switched Networks; Meshed Trees.

I. INTRODUCTION

Link redundancy is introduced in bridged (switched) networks to provide backup paths in the event of failure of an active link. This results in a physical network topology that has loops. The physical loops in turn cause broadcast storms when forwarding broadcast packets. Implementing a loop free logical topology over the physical topology is one way to avoid broadcast storms. The first of such logical loop free forwarding solutions called the Spanning Tree Algorithm (STA) was proposed by Radia Perlman [1]. Spanning tree in bridged networks was constructed by blocking some of the bridge ports. Based on STA the Spanning Tree Protocol (STP) was developed, and is an the specification for this protocol are available IEEE standard [1]. Rapid Spanning Tree protocol (RSTP) was then developed to overcome the high convergence times during topology changes in the basic STP. TRILL (transparent interconnection of lots of links) on Rbridges (router bridges) was subsequently proposed by the same researcher to overcome the disadvantages of STA based loop avoidance at the cost of some overhead and implementation complexity by adopting the IS-IS (intermediate system to intermediate system) routing protocol, where IS-IS related messages are encapsulated in special frames by the Rbridges. This is currently an *ietf* (Internet Engineering task Force) draft [4].

The premise of the above solutions is that a single logical tree from a root node that operationally eliminates physical loops is necessary to resolve the conflicting requirements of physical link redundancy and loop free forwarding. In the event of a link failure the tree has to be recomputed. While spanning tree is a single tree constructed from a single elected root node, with Dijkstra algorithm a tree is constructed at every node, assuming itself to be a root node, thus every node has a tree that it can use to forward. Dijkstra algorithm requires the connectivity information about all segments in the networks to compute the tree, while in the case of spanning tree, nodes join the tree branches based on the information they receive from their neighbors.

In this paper, we introduce a *meshed tree* (MT) algorithm that allows creation and maintenance of *multiple* overlapping tree branches from *one* root node. The multiple branches mesh at nodes, and in the event of failure of a link (or branch) the node can immediately fall back on another branch without the necessity for renewed tree resolution. This eliminates intermittent inconsistent topologies, which ensue during tree reconstruction.

Fig. 1 is provided to illustrate the difference between normal and meshed trees constructed over a given physical network topology. The circles are the nodes and the dotted lines are the physical links. Picture (a) shows a normal logical tree (thick line), which can be created either using the spanning tree or the Dijkstra algorithm. Picture (b) shows three tree branches (two originating from the root and the third from another node) that mesh at the nodes. The meshed tree branches thus formed have a single root node

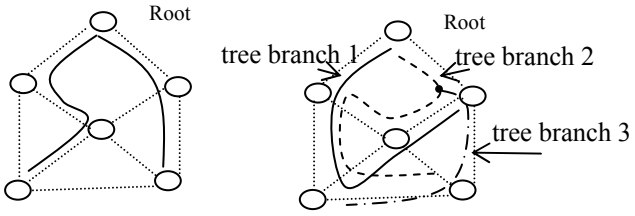


Figure 1 Single vs meshed trees

which is the principle of meshed trees. As each node in the network is on multiple tree branches, packets to the nodes can be forwarded using any branch.

MT Rationale: It is necessary to have a logical tree that spans all the nodes for forwarding broadcast packets. Let us call this the primary tree. But, that should not preclude the construction of multiple tree branches simultaneously or the overlapping of the tree branches, if achieved without loop formations. Tree branches other than those belonging to the primary tree will thus take over packet forwarding seamlessly in the event of link failure in the primary tree.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

Meshed trees are implemented through a simple numbering scheme that will be used to assign virtual IDs (VID) to a node (in this case bridge) in the network. A VID in a compact manner defines a tree branch and hence a logical packet forwarding path from the root node to the node with the VID. A node can acquire several VIDs as it is allowed to join multiple tree branches. Meshed trees thus leverage the redundancy in meshed topologies to set up several loop free logical forwarding paths without blocking bridge ports.

Meshed tree creation and maintenance requires simple processing while enabling an easy transition from STA based protocols as described in this article. The goal in this paper is to provide the operational details of the MT algorithm and describe its features in comparison with existing loop resolution algorithms used in bridged networks. One optimization that is possible because of the unblocked bridge ports is also presented.

The rest of the paper is organized as follows. Section II discusses related work. In Section III, we present the operational details of the MT algorithm (MTA), as applicable to bridged networks. Section IV provides details on one optimization possible with MTA for forwarding unicast packets. Section V compares the performance and features of the MTA based solution against both STA based and the TRILL protocol. Section VI provides conclusions.

II. RELATED WORK

In this section, we focus on the two primary techniques adopted for loop resolution in bridged networks. The first of these is based on ST algorithm and the second is the TRILL on RBridges. The presentation in this section focuses on some distinct features of these techniques, without describing operational details as they are publicly available.

A. Protocols on Spanning Tree Algorithm

The STP is based on the STA. To avoid loops in the network while maintaining access to all the network segments, the bridges collectively elect a root bridge and then compute a spanning tree from the root bridge. In STP, each bridge first assumes that it is the root and announces its bridgeID. The information is used by the bridges to elect the root bridge. The bridgeID besides carrying the uniqueID of the bridge which is its MAC (medium access control) address also has a priority field to override the lowest MAC address bridge from being elected as the root bridge. Once a root bridge is elected, the other bridges then resolve their connection to the root bridge, by listening to messages from their neighbors to form a spanning tree.

STP has high convergence times subsequent to topology changes. To reduce the convergence times the *Rapid Spanning Tree* protocol (RSTP) was proposed [2]. RSTP is a refinement of the STP and therefore shares most of its basic operation characteristics, with some notable differences. The differences are; the detection of root bridge failure is done in 1 'hello' time; response to Bridge Protocol Data Units (BPDUs) sent from the direction of the root bridge; allowing RSTP bridges to 'propose' their spanning tree information on their designated ports; allowing the receiving RSTP bridge to determine if the root information is superior, and set all other ports to 'discarding' and send an 'agreement' to the first bridge; whereupon the first bridge, can rapidly transition that port to forwarding state bypassing the traditional listening/learning states, and thus allowing faster convergence; maintain backup details regarding the discarding status of ports to avoid timeouts if the current forwarding ports were to fail.

Advantages: STA based implementation is simple as the spanning tree is executed with the exchange of BPDUs at layer 2, where a BPDU carries the 'tree formation' information in multicast Ethernet frames.

Disadvantages: Several disadvantages of STA based protocols are noted in [2]. Traffic is concentrated on the spanning tree path, and all traffic follows that path even when other more direct paths are available, resulting in inefficient use of the link topology and reduction in aggregate bandwidth and causing traffic to take circuitous paths. Spanning tree is dependent on the way a set of bridges is interconnected. Small changes in this topology can cause large changes in the spanning tree. Changes in the spanning tree take time to propagate and converge especially for non-RSTP protocols. Though 802.1Q support for multiple spanning trees helped, it also required additional configuration. The number of trees is limited, and the defects apply within each tree regardless [3].

B. TRILL Protocolon RBridges

TRILL on Rbridges overcomes the shortcomings of the STP as it combines the functionality of layer 3 by using the IS-IS routing protocol [4] at layer 2 to compute pair-wise optimal paths between two Rbridges based on a link state algorithm. The computed pair-wise optimal paths will be used for forwarding the frames at layer 2. The solution is transparent to layer 3 protocols. IS-IS allows for the inclusion of information such as layer 2 addresses of reachable end nodes. Inconsistencies and loop formations during topology change are overcome by the 'hop count' used in TRILL frame headers for inter-bridge forwarding.

C. Operation of TRILL Protocolon

- *Election* of a Designated Rbridge (DR), which is the only bridge allowed to learn the membership of end nodes on that link, and to forward traffic destined to that link.
- The egress Rbridge from a link, usually the DR, *encapsulates* the frame with an additional header that contains, at the minimum, a hop count, and preferably also a destination Rbridge identifier. Frames in transit are distinguished from originating frames, since they contain the encapsulation header.
- Rbridges additionally calculate a spanning tree-based on the link state database used by IS-IS for purposes of delivering layer 2 multicast frames, and frames to unknown

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

destinations. Frames to be handled by the spanning tree use an encapsulation header with a destination 'Rbridge ID=0'.

- Use of End station address distribution (ESADI) protocol by the RBridge to distribute addresses of end nodes on its link to enable all Rbridges to know which Rbridge is the appropriate destination Rbridge for an end node.

Advantages over 802-style bridging [4]: Frames travel via an optimal path. As transit frames are routed, with a header that contains a hop count, temporary loops will not result in frame proliferation, and will quickly be discarded on the hop count reaching 0. Routing changes can be made instantaneously and safely based on local information

Loop Avoidance: An appointed forwarder for a link is responsible for loop avoidance [4]. It inhibits itself for a configurable time from 30 to zero seconds, which defaults to 30 second, after it sees a root bridge change on the link. An inhibited appointed forwarder for a port drops any native frames it receives and does not transmit any native frames in the LAN for which it is appointed. The forwarder will inhibit itself, as described above, if, within the past five Hello times, it has received a Hello in which the sender asserts that it is appointed as the forwarder. Optionally, they may not de-encapsulate a frame from ingress RBridge say 'RBm' unless it has RBm's Link State PDUs and the root bridge on the link it is about to forward onto is not listed in RBm's list of root bridges for that LAN. This is known as "de-encapsulation check" or "root bridge collision check".

III. THE MESHED TREE ALGORITHM

The 'meshed tree' algorithm allows construction of logically 'meshed trees' from a single root node in distributed fashion and with local information [12-14]. In the discussion presented in this article the election of a root bridge is not included as the focus is on the loop resolution / avoidance capability of MT algorithms. However a process similar to that adopted by STA can be used to elect a root bridge or a bridge can be designated to be a root bridge.

Bridge ID: For the operation of the MT algorithm bridgeIDs are necessary. These have to be unique only within the bridged network. Hence, a simple MAC address derivative can be used. This would be useful to keep the tree VIDs

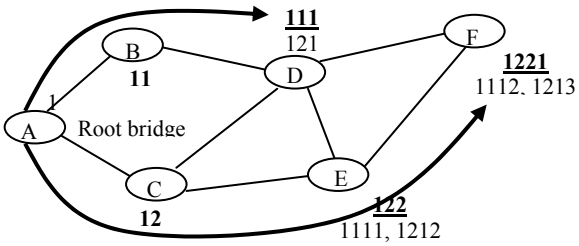


Figure 2: Meshed Tree Creation

simple, as the first value in the tree VID is the root bridgeID. One approach is to allocate a single digit ID to the root bridge once it is elected or designated. This however calls for a process of resolution if the root bridge fails which would be out of scope for this article.

Operation: The creation of a single meshed tree using VIDs is described first. In Fig. 2 a meshed tree originating at bridge 'A' is shown. Let bridge A have VID = 1. After being elected the root bridge, at regular intervals, bridge 'A' will announce its VID in a BPDU packet. Bridges B and C listen to the advertised VID and request to join as branches of bridge A. Bridge A allocates B a VID=11and C a VID=12 (by appending single digit value to its ID - the rationale for using a single digit is provided at the multiple digits can be used) and thus bridges B and C have now joined in tree originating from bridge A. Bridges B and C now advertise their VIDs.

Multiple VIDs from different parents: D hears advertisement from B and C and decides to join the branches from both B and C, while E hears only from C and decides to join the branch from C. D gets assigned a VID of 111 from B and 121 from C, while E gets assigned 122 by C.

Multiple VIDs under same parent: When E hears D announcing its VIDs, it can request a VID under each of D's

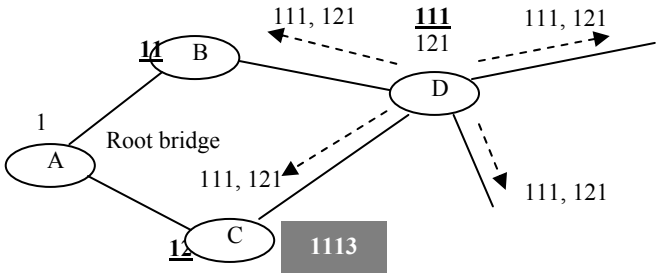


Figure 3 Loop Avoidance

VIDs and thus acquires VIDs 1111 and 1212.

To complete the explanation, F acquires VIDs 1221 from E and 1112 and 1213 from D. (*Note that D could also have acquired VIDs under E's other VIDs.)* Though the tree branches can mesh to the maximum limited only by the actual physical connections, they can however be controlled by limiting the number of VIDs that each bridge can acquire. In Fig. 2, only partial meshing is shown for clarity purposes. The 'meshed tree' thus formed by the VIDs provides a simple yet robust scheme to set up several redundant logical paths for packet forwarding without blocking ports [5-7].

Loop Avoidance in the algorithm is explained with Figure 3, which captures a partial topology from Figure 2. Assume C hears the VID 121 and 111 advertised by bridge D. It will not request to join the tree branch under the VID 121, as 'C' sees its VID sequence '12', in the advertised VID 121 and thus avoid loop formation.

Primary VID Tree: This is the tree that will be used for forwarding broadcast packets. Except for the root bridge, each of the other bridges will maintain one VID as the primary VID under the meshed trees and other VIDs as backup to be used in the failure of the primary VID. In

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

Figure 2, the VIDs that are underlined and in bold are the primary VIDs. The criteria to determine a primary VID may be predefined i.e. it could be based on link costs or on hops, as shown in the examples. The thick arrows identify the primary VID tree originating from bridge A.

Broadcast Packets: For forwarding broadcast packets or packets to unknown destinations the bridges should associate the VIDs to the ports through which they were acquired, so when using a VID, they are aware of the port on which the packet should be forwarded. This information is omitted in the figure for picture clarity. For simplicity and without loss of generality port 1 has been assumed to be the port from which the primary VID was heard by the bridges.

The rule for forwarding broadcasts packets by non-root bridges is: if received from the port of primary VID, then send out on all ports that have a VID derived from the

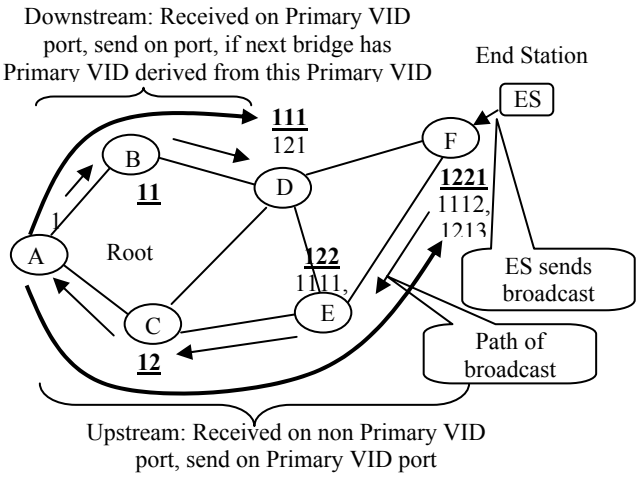


Figure 4: Broadcast Packet Forwarding

primary VID. However, if the broadcast packet is received from any other port send out on port with the primary VID.

We will illustrate this with an example using Fig. 4. When bridge F receives a broadcast packet on one of its other ports let us say the one in which end station ES is connected it will send the packet out of port 1, its primary VID port. E receives the packet and will forward to its port 1, C will receive the packet and forward on its port 1 to A. This is the process that will be adopted when forwarding a broadcast packet upstream along the Primary VID tree.

The root bridge receives the packet on one of its ports, and will send it out on the other ports – in this case there is only one other port. The broadcast packet is then picked up by bridge B on its primary VID port. The broadcast packet now has to be forwarded downstream on the primary VID tree. B will send the packet out on all ports, if there are bridges on those ports that have a VID derived from its primary VID. To forward packets destined to end stations an optimized approach is discussed in Section IV. A general case is also explained in this section.

Link Failures: Let us assume that link CE failed as shown in Fig. 5. Bridge E will detect this and invalidate its VID 122, and fallback on VID 1111 as the primary VID. E may announce to bridges that have VIDs derived from 122 about the failure of VID 122 based on which bridge F invalidates VID 1221 to fallback on VID 1112 as the primary VID. The new primary VID tree is shown by the thick arrows. Frame forwarding will continue as usual except that E will use the path via D to forward to other bridges as none of the ports are blocked. On the revival of link CE the VID 122 may be restored.

The time for a bridge to identify a link failure is limited by the time it takes the protocol to recognize that the link is down. The bridge that recognizes this will immediately fallback to its backup VID and propagate the information only to those bridges that have a VID derived from the failed VID. The tree will thus get pruned if necessary but new tree reconstruction will not be necessary.

Other Features: Each single digit following the root bridge VID indicates a hop from the root bridge. End nodes connected to the bridge ports are not included in this count. Based on the root bridge VID other bridges will acquire

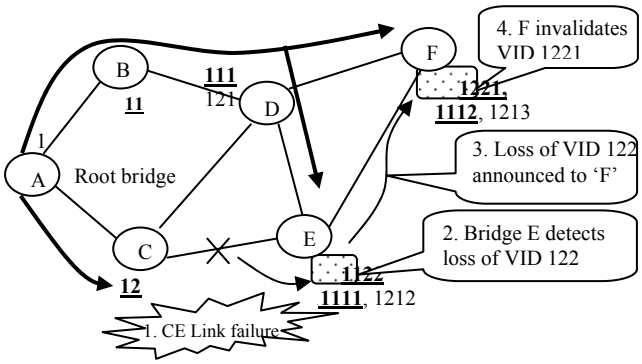


Figure 5: Primary VID Tree after Failure of Link CE

their VIDs, which they then use for packet forwarding, thus the first digit in a VID carries the 'root' bridge VID.

Construction of meshed trees requires bridges to advertise their VIDs at regular intervals. As the VID carries the branch information, bridges that hear the advertised VIDs can decide to join any or all of branches advertised. The joining decision can be based on criteria like shortest hops, best cost, and diversity in branch among others. Thus there is a local tree joining process for nodes, which makes the meshed trees to be constructed with local information.

IV. OPTIMIZED FORWARDING

Bridges can learn of hosts connected to them from source address in the frames they forward. Optimized forwarding of unicast packet with the MT algorithm is described in this section. In the MT approach as ports are not blocked, each bridge can advertise it's SAT (Source Address Table) of hosts directly connected, to neighboring bridges. Each receiving bridge can then populate their SAT with this information. This will require a bridge to use its

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

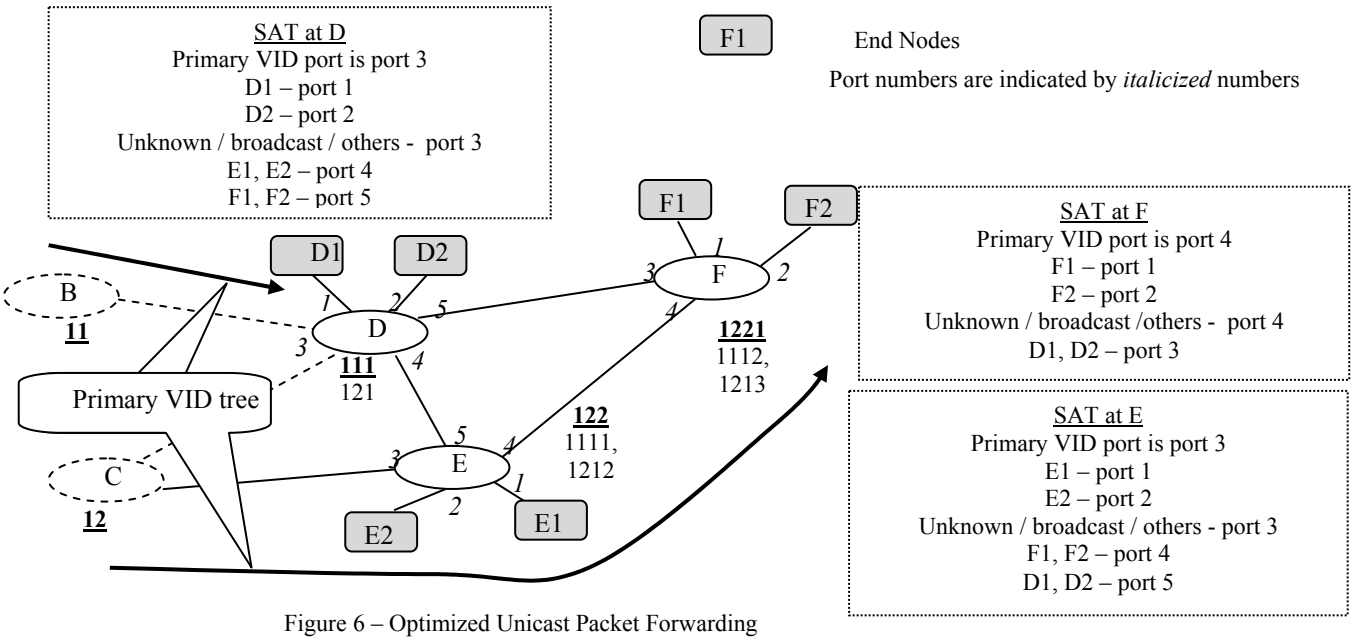


Figure 6 – Optimized Unicast Packet Forwarding

primary VID tree for broadcast frames or frames to unknown destinations. Frames destined to known MAC addresses can be forwarded on the most suitable port

In Fig. 6, we illustrate this with only a partial topology of the bridged network discussed thus far. Bridge F learns of end nodes F1 and F2 connected to it. Similarly bridge E learns of E1 and E2 and bridge D learns of D1 and D2. The local SAT information is advertised by the bridges on all their ports to neighboring bridges as these ports are not blocked. Bridge F has learned of hosts D1 and D2 which can be accessed on port 3. If it receives frames addressed to these hosts it will forward to port 3. However if it receives frames to hosts connected to bridges B and C the packets would be forwarded along the primary VID tree towards the root. Along the path if a bridge knows about the end nodes through its SAT, the packets would be directly forwarded to that port without going through the root bridge.

V. CONCLUSIONS

Loop free forwarding in networks with redundant paths have been addressed on the premise that a logical single tree topology originating from a root is essential. This resulted in the spanning tree algorithm, which faced high convergence delays later resolved by RSTP. Several disadvantages outlined in this article persisted, resulting in introduction of a more complex solution using the IS-IS on a protocol above layer 2. This article describes a simple solution that can replace STA algorithm at layer 2, without its disadvantages, but at the same time avoid the complex implementation requirements of TRILL on Rbridges. While the documents on STP and TRILL on Rbridges provide detailed specifications, it was possible only to highlight certain feature of meshed trees, to emphasize its loop free forwarding capabilities, without the complexity of the solutions being investigated to replace STP.

The current state of the work is as described above, where different implementation and optimization approaches for the meshed tree algorithm have been investigated. It is planned to model these details and compare for performance with Spanning tree and Rapid Spanning tree implementation in switched networks of varying topologies using Opnet simulation tool [7].

REFERENCES

[1] LAN/MAN Standards Committee of the IEEE Computer Society, ed. (1998). *ANSI/IEEE Std 802.1D, 1998 Edition, Part 3: Media Access Control (MAC) Bridges*. IEEE

[2] Wodjek W., "Rapid Spanning Tree Protocol: A new solution from old technology", http://www.compactpci-systems.com/articles, March 2003.

[3] Perlman R., Eastlake D., Dutt G. D. and Gai, A. G., "Rbridges: Base Protocol Specification", http://www.ietf.org/internet-drafts/draft-ietf-trill-rbridge-protocol-16.txt, March 3, 2010

[4] Touch J., Perlman R., "Transparent Interconnection of Lots of Links (TRILL): Problem and Applicability Statement", RFC 5556.

[5] Shenoy N., Pan Y., Narayan D., Ross D. and Lutzer C. (2005), "Route robustness of a multi-meshed tree routing scheme for internet MANETs", Proceeding of IEEE Globecom 2005. 28 Nov – 2nd Dec. 2005 St Louis, pp. 3346-3351.

[6] Book chapter on "Multi Hop routing and load balancing in Mobile Ad hoc networks", Encyclopedia on Ad Hoc and Ubiquitous Computing, Chapter editor Nirmala Shenoy and Sumita Mishra, Published by World Scientific Book Company, 2008.

[7] www.opnet.com

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

Node repositioning method based on topology information in IEEE 802.16j relay networks

Takafumi Shigefuji*, Go Hasegawa†, Yoshiaki Taniguchi† and Hirotaka Nakano†
*Graduate School of Information Science and Technology, Osaka University
1–5, Yamadaoka, Suita, Osaka, Japan
Email: t-sigefj@ist.osaka-u.ac.jp
†Cybermedia Center, Osaka University
1–32, Machikaneyama-cho, Toyonaka, Osaka, Japan
Email: {hasegawa,y-tanigu,nakano}@cmc.osaka-u.ac.jp

***Abstract*—IEEE 802.16j relay networks can provide wide-area wireless broadband service. In these networks, the locations of relay nodes and the topology are important factors in achieving high performance. In particular, radio wave interference must be considered when establishing node locations. However, nodes are generally deployed in limited areas and cannot be moved to arbitrary positions. In this paper, we propose a node repositioning method that works within these constraints to improve network performance and that utilizes only topology information. Since the computational cost is high for assessing all possible node positions, the proposed method limits the number of candidate nodes to be repositioned on the basis of topology information. We examine the effectiveness of the proposed method through simulation experiments and show that a performance improvement of up to 28% is realized and nearly optimal results are obtained, but at much lower computational cost than an exhaustive search.**

Keywords- IEEE 802.16j; relay network; node repositioning; network performance improvement.

I. INTRODUCTION

Wireless relay networks based on IEEE 802.16j [1] are now attracting considerable attention since they can provide wide-area network service at a low cost to metropolitan areas or areas in which the construction of wired networks is difficult [2]. An IEEE 802.16j network consists of two types of nodes: gateway nodes and relay nodes [3]. A gateway node is wired to an external network. On the other hand, a relay node is not directly wired to the external network and instead connects via wireless links that form a multi-hop relay network whose root is the gateway node (Figure 1). One advantage of this type of network is that the service area can be extended easily and the network capacity can be increased by adding relay nodes [4].

In wireless networks, a general problem is that communication links that interfere with each other cannot communicate at the same time [5], [6]. IEEE 802.16j networks solve the radio wave interference problem by adopting the Orthogonal Frequency Division Multiple Access (OFDMA) scheme [7], which is a scheduling mechanism based on the Time Division Multiple Access (TDMA) and FDMA schemes [8]. In this scheduling method, time slots are assigned to each communication link as transmission opportunities according to its traffic demands. The links that interfere with each other use different time slots and the links that do not interfere with each other use the same time slots to increase spatial reuse of radio resources [9], [10]. To decrease the number of assigned time slots is important for achieving high performance in these networks and radio wave interference must be taken into account when positioning the nodes to reduce the number. In general, however, nodes are deployed in limited areas and cannot be moved to arbitrary positions. An example is shown in Figure 1, where nodes are set up on the roofs of buildings in an urban area and a node can be moved within only the roof area.

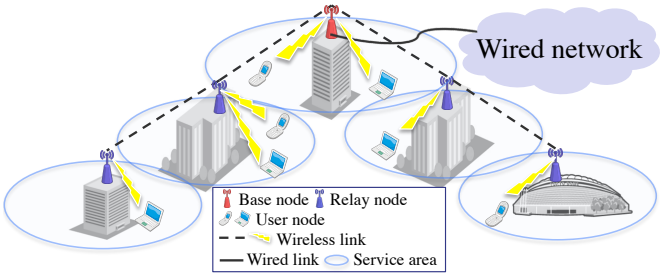


Figure 1. IEEE 802.16j relay network

In this paper, we propose a node repositioning method that works within such constraints to improve the performance of IEEE 802.16j relay networks. Our proposed method utilizes only topology information to determine which nodes are repositioned. Furthermore, we limit the number of nodes to be repositioned since the computational cost is high for assessing all possible node positions.

We evaluate the effectiveness of the proposed method through simulation experiments where the number of time slots assigned to all links in the network is taken as the assessment criterion. We also show the results of evaluating

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

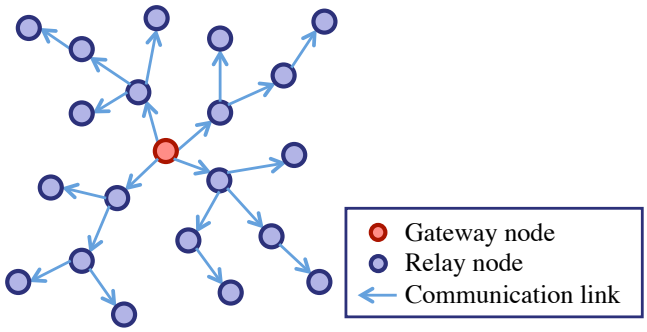


Figure 2. Network model

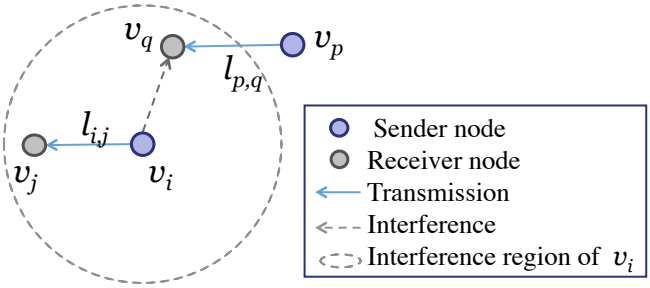


Figure 3. Radio wave interference model

computational cost.

The rest of this paper is organized as follows. In Section II, we explain the network model and the time slot assignment algorithm. Then, we propose the node repositioning method in Section III. In Section IV, we evaluate the effectiveness of the proposed method through simulation experiments. Finally, we present the conclusions of this paper and areas of future research.

II. System model

In this section, we describe the IEEE 802.16j relay network model, radio wave interference model, and time slot assignment method which we employ in this study.

A. IEEE 802.16j relay network model

The network model consists of one gateway node and multiple relay nodes. These nodes form a multi-hop relay network whose root is the gateway node (Figure 2). The network topology is constructed such that all relay nodes reach the gateway node via the minimum number of hops. After determining the topology, each node sets its own transmission range to the minimum value for maintaining its link. This adjustment of transmission range helps in reducing power consumption and radio wave interference. In this paper, we consider only downstream transmission from the gateway node to relay nodes.

B. Radio wave interference model

Here, we explain the model of radio wave interference between links. Figure 3 shows the situation where link $l_{i,j}$ between nodes v_i and v_j interferes with link $l_{p,q}$ between nodes v_p and v_q. Here, v_i and v_p are sender nodes, and v_j and v_q are receiver nodes. The transmission range of v_i is t_i. We define the interference ratio as γ, and thus the interference range r_i of v_i is represented as $\gamma \cdot t_i$.

In Figure 3, v_q is located within the interference range r_i of v_i, which is expressed with parameters as $||v_i - v_q|| < r_i$ ($||v_i - v_j||$ means the physical distance between v_i and v_j). When $l_{i,j}$ and $l_{p,q}$ transmit data at the same time, v_q cannot correctly receive the signal from v_p since v_q receives radio waves from both v_i and v_p simultaneously. In this case, $l_{i,j}$ interferes with $l_{p,q}$. Additionally, we define the interference relationship between $l_{i,j}$ and $l_{p,q}$ as the situation where either $l_{i,j}$ interferes with $l_{p,q}$ or $l_{p,q}$ interferes with $l_{i,j}$. These two conditions are represented as $||v_i - v_q|| < r_i$ and $||v_p - v_p|| < r_p$. In other words, when $||v_i - v_q|| < r_i$ or $||v_p - v_p|| < r_p$ is satisfied, the radio wave interference occurs between $l_{i,j}$ and $l_{p,q}$.

C. Time slot assignment algorithm

IEEE 802.16j relay networks resolve the radio interference problem by employing the TDMA scheme. In such networks, time slots are assigned to links as transmission opportunities. For high network performance, it is important to assign different time slots to links that interfere with each other, and to assign the same time slots to links that do not interfere with each other. In this paper, we utilize the scheduling algorithm proposed in [10]. The algorithm assigns time slots to links in the network in accordance with their traffic demands by treating the time slot assignment problem as a vertex coloring problem [11].

III. Node repositioning method

In this section, we explain the method to reposition relay nodes under movement range constraints. We assume that the initial positions of relay nodes are determined beforehand, and that we can move certain relay nodes to improve network performance. We take the movement of a node to be constrained within a certain range centered on the node's initial position. In what follows, we first explain the algorithm to determine the repositioning of one node. We then describe how to reposition multiple nodes to further improve network performance.

A. One-node repositioning

In repositioning one node, the selection of the node to be repositioned is important. The target node is determined as follows. We first select a candidate node to be repositioned and find its position that gives the best network performance from all the possible positions. We repeat this process for all

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

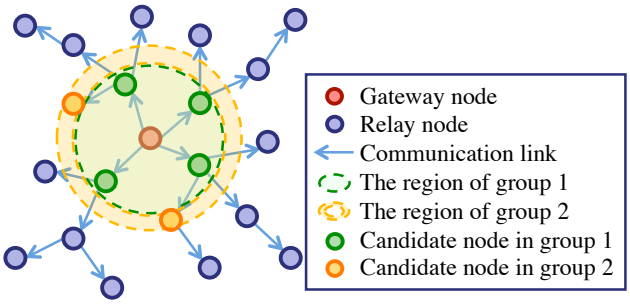


Figure 4. Groups of node repositioning

candidate nodes and find the node and its movement which lead to the best network performance.

However, the computational cost is too large to check all nodes as candidate nodes. Here, we propose limiting the number of candidate nodes according to their distance from the gateway node, as shown in Figure 4. We categorize the candidate nodes as follows. First, nodes that are directly connected to the gateway node are categorized as group 1. Then, the nodes that are not directly connected to the gateway node, but that can be connected to the gateway node by repositioning, are categorized as group 2.

By setting this constraint on the nodes to be checked, we can substantially decrease the computational cost, especially when the number of relay nodes in the network is large.

B. Repositioning of multiple nodes

When we reposition multiple nodes to further improve the network performance, we can consider two approaches: parallel repositioning and serial repositioning.

In parallel repositioning, we consider all the possible positions of multiple nodes. Doing so gives the optimal solution for the repositioning of multiple nodes but at a high computational cost that increases greatly with the number of repositioned nodes.

On the other hand, in serial repositioning, we sequentially apply the one-node repositioning method described in Section III-A. This strategy can be regarded as a simple hill-climbing heuristic[12]. Therefore, the computational cost is lower in comparison with parallel repositioning, but the global optimal solution might not be found.

In Section IV, we compare the serial and parallel methods and show that serial repositioning is effective in terms of both computational cost and network performance.

IV. Performance Evaluation

In this section, we evaluate the performance of the node repositioning method proposed in Section III through simulation experiments.

A. Simulation settings and performance metrics

We perform simulation experiments under the following conditions. A gateway node is deployed at the center of a 1×1 field, and 9, 29 or 49 relay nodes are randomly deployed. We refer to these three setups as the 10-node network, 30-node network, and 50-node network, respectively. The maximum transmission range of the gateway node and the relay nodes are set to 0.56 for the 10-node network, 0.32 for the 30-node network, and 0.25 for the 50-node network, so that no relay node is disconnected from the network in the initial setup. The radio interference ratio γ is set to 2.

The movement range is set to 1/10 of the maximum transmission range. When moving one node, we consider three methods for establishing the candidate positions as shown in Figure 5. These three methods differ in the resolution of the positions. We can expect that a finer resolution will give better network performance but at a higher computational cost.

The traffic demand from each node depends on the size of the Voronoi diagram [13] of the node. The number of time slots necessary for each link is determined according to the traffic demands. Time slots are assigned to the links by the algorithm proposed in [10]. The number of time slots assigned to all links in the network is called the frame length. The frame length is an important metric of network performance, and a smaller value corresponds to higher network performance.

The change in network performance from repositioning is evaluated in terms of frame length ratio, which is the ratio of the frame length of the repositioned network divided by that of the original network. We conducted 100 simulation experiments for each parameter setting and evaluate the distribution of the frame length ratio. We also measured the time required to perform the calculation in order to evaluate the computational cost.

B. Effect of movement resolution

We first evaluate the effect of movement resolution (Figure 5) on the performance of one-node repositioning. Here, we move only one node in a 10-node network.

Figure 6 shows the distribution of frame length ratio for 100 simulation experiments. Nearly the same result was obtained for all movement resolutions. This means that the coarsest resolution is adequate when calculating one-node movement. Therefore, we use the coarsest resolution in the following experiments.

C. Performance of one-node repositioning

We evaluated the performance of one-node repositioning in more detail. Specifically, we investigated the effect of limiting the number of candidate nodes, as shown in Figure 4. The frame length ratios obtained using the proposed method with only group 1, only group 2, and both groups 1 and 2 are compared with the ratio obtained using the method where

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

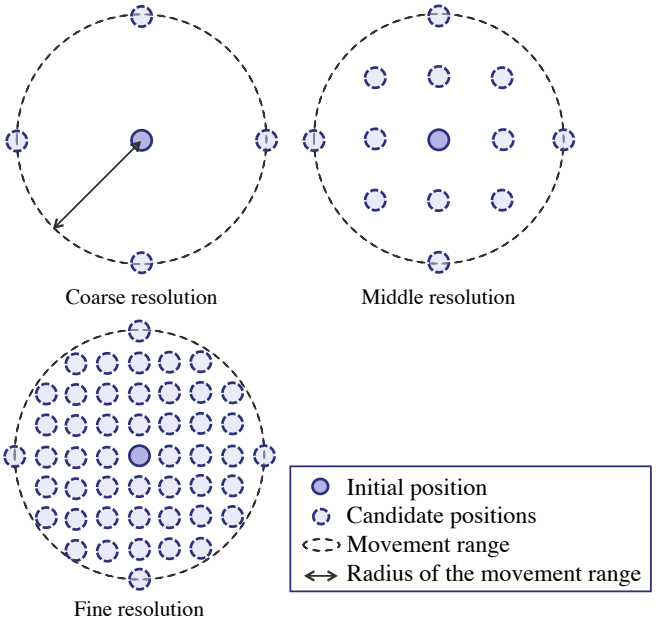


Figure 5. Resolution of movement positions

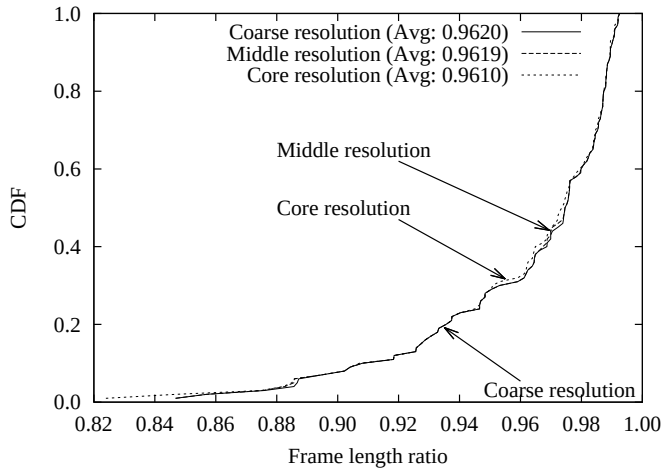


Figure 6. Effect of movement resolution

every node is checked. Note that checking all nodes gives the global optimal solution for one-node repositioning.

Figure 7 shows the distributions of frame length ratio from 100 simulation experiments for the 10-node network, 30-node network, and 50-node network. The result of one-node repositioning using only group 1 or 2 is far from optimal, but repositioning with both groups 1 and 2 is nearly optimal. Furthermore, the result remains unchanged, 5–7% average improvement in the frame length ratio, regardless of the number of relay nodes in the network.

These results indicate that compared with the exhaustive search method, the proposed one-node repositioning method can decrease the frame length substantially at a lower computational cost.

D. Multiple node repositioning

We next evaluate the performance of multiple node repositioning. In detail, we compare the performance of parallel repositioning with all nodes being candidate nodes (giving optimal results), and the serial repositioning with candidate nodes limited to groups 1 and 2. By using the 30-node network, we move 2–3 nodes in the parallel method and 2–9 nodes in the serial method. Note that we cannot execute the simulation experiment of the parallel method using more than 4 repositioned nodes since the calculation time is too large.

Figure 8 shows the distributions of frame length ratio from 100 simulation experiments with 2–3 nodes being repositioned. We can observe from this figure that as the number of repositioned nodes increased, the results from the serial repositioning method with both groups 1 and 2 became slightly worse than the optimal results. However, considerably better results were obtained from serially positioning using both groups 1 and 2 than from serial repositioning with group 1 or group 2 alone. These results show that limiting the candidate nodes to both groups 1 and 2 is still effective, even in multiple node repositioning.

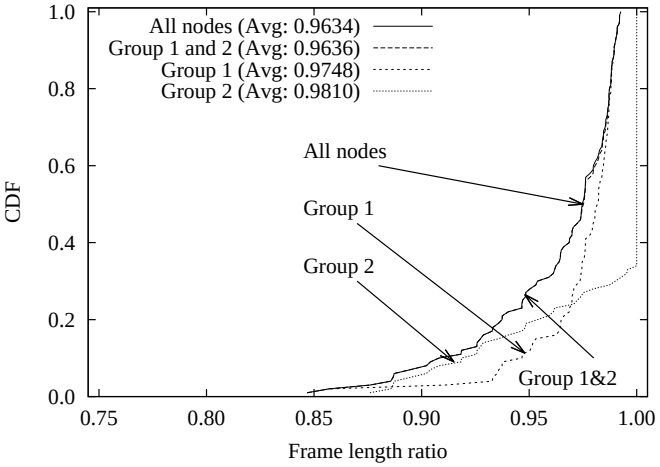
Furthermore, serial repositioning has a large advantage in terms of computational cost. In Figure 9, we plot the calculation times of the parallel and serial repositioning methods, which were executed on a PC with four 2.93 GHz quad-core CPUs and 64 GB of memory. This figure clearly shows the small computational cost of the serial repositioning method.

Figure 10 shows the distribution of frame length ratio from 100 simulation experiments using the serial repositioning method with 1–9 nodes being repositioned in the 30-node network. The results show little change for more than 5 nodes being repositioned. In addition, by using the serial repositioning method with 9 nodes, the frame length ratio was improved by up to 28% and 10% on average, which is almost the same as the result of the parallel repositioning method with 3 nodes (Figure 11), while the calculation time is quite small as shown in Figure 9. We can conclude from these results that, by the proposed method, serial repositioning with a portion of the nodes is sufficient for realizing a notable performance improvement.

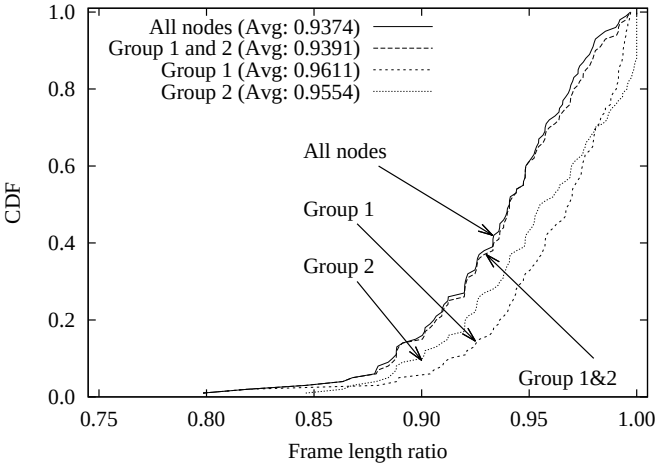
V. Conclusions and Future Work

In this paper, we proposed a node repositioning method to improve the performance of IEEE 802.16j relay networks. A low computational cost in multiple node repositioning was realized by limiting the number of candidate nodes and by repositioning multiple nodes in series. The proposed method achieves sufficient effectiveness by repositioning a small number of relay nodes and brings about a nearly optimal

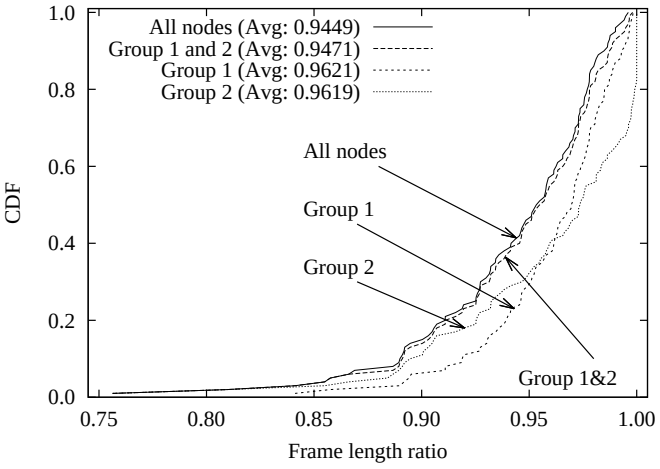
Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1



(a) 10-node network

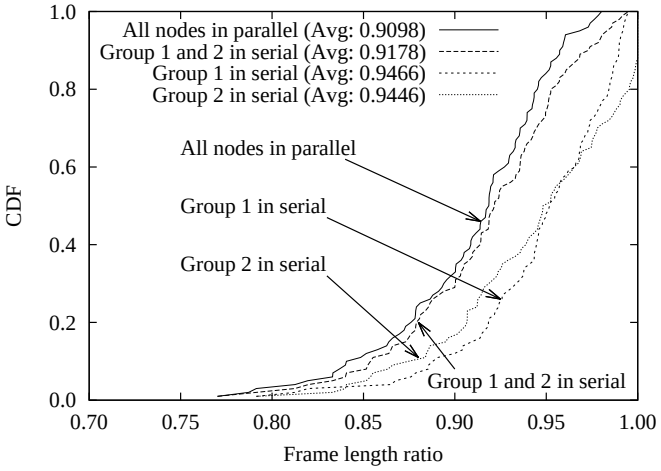


(b) 30-node network

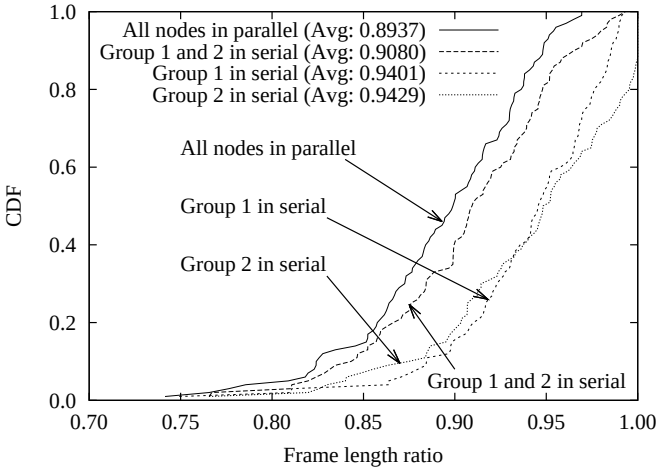


(c) 50-node network

Figure 7. Performance of one-node repositioning



(a) 2 nodes repositioning



(b) 3 nodes repositioning

Figure 8. Frame length ratio from multiple node repositioning

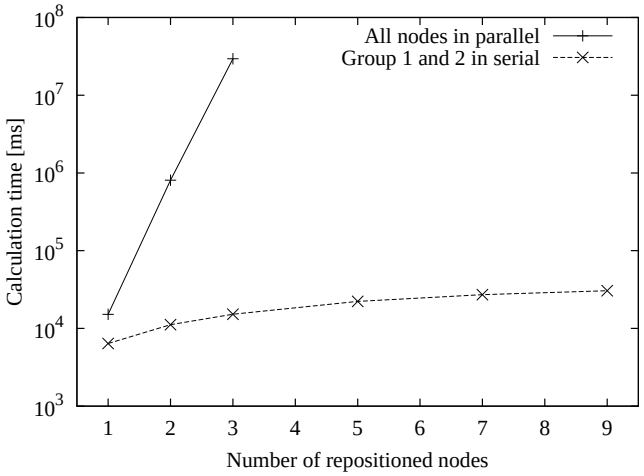


Figure 9. Calculation time for multiple node repositioning

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

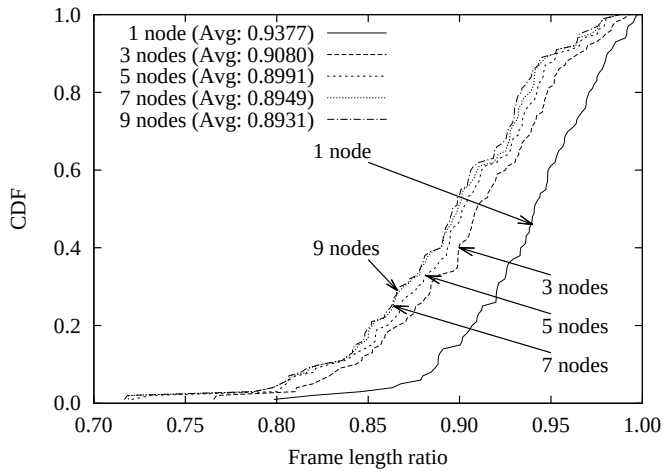


Figure 10. Frame length ratio from multiple node repositioning using both groups 1 and 2

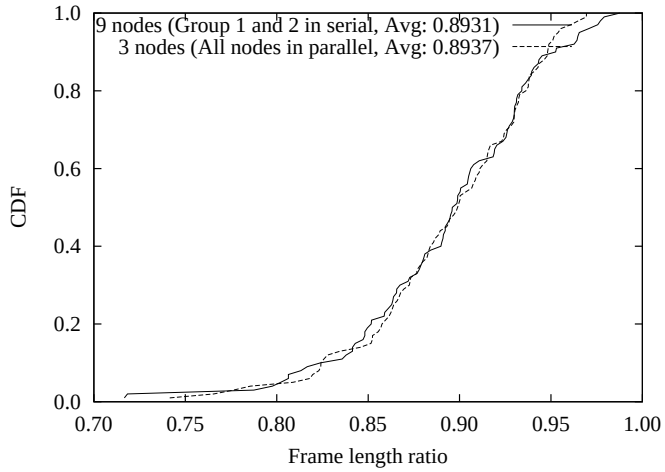


Figure 11. Comparison of multiple node repositioning

performance improvement, but at much lower computational cost than an exhaustive search.

In future work, we will consider ways to further improve the serial repositioning method. We will also evaluate the effects on upstream link and the effectiveness of the proposed method with more accurate radio interference models such as the signal-to-interference noise ratio model [14].

REFERENCES

[1] IEEE LAN/MAN Standards Committee, "IEEE standard for local and metropolitan area networks part 16: Air interface for broadband wireless access systems," *IEEE Std 802.16j-2009*, June 2009.

[2] M. Kuran and T. Tugcu, "A survey on emerging broadband wireless access technologies," *Computer Networks*, vol. 51, pp. 3013–3046, Aug. 2007.

[3] N. Athanasopoulos, P. Tsiakas, K. Voudouris, D. Manor, A. Mor, and G. Agapiou, "An IEEE 802.16j prototype relay station architecture," in *Proceedings of IEEE MELECON 2010*, pp. 1247–1252, IEEE, June 2010.

[4] V. Genc, S. Murphy, Y. Yu, and J. Murphy, "IEEE 802.16j relay-based wireless access networks: An overview," *IEEE Wireless Communications*, vol. 15, pp. 56–63, Oct. 2008.

[5] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, pp. 388–404, Mar. 2000.

[6] H. Zhu and K. Lu, "On the interference modeling issues for coordinated distributed scheduling in IEEE 802.16 mesh networks," in *Proceedings of IEEE BROADNETS 2006*, pp. 1–10, Oct. 2006.

[7] V. Roman and P. Consulting, "Broadband wireless access solutions based on OFDM access in IEEE 802.16," *IEEE Communications Magazine*, vol. 40, pp. 96–103, Apr. 2002.

[8] Z. Tao, A. Li, K. Teo, and J. Zhang, "Frame structure design for IEEE 802.16j mobile multihop relay (MMR) networks," in *Proceedings of IEEE GLOBECOM 2007*, pp. 4301–4306, Nov. 2007.

[9] W. Wang, Y. Wang, X. Li, W. Song, and O. Frieder, "Efficient interference-aware TDMA link scheduling for static wireless networks," in *Proceedings of MobiCom 2006*, pp. 262–273, Sept. 2006.

[10] R. Ishii, G. Hasegawa, Y. Taniguchi, and H. Nakano, "Time slot assignment algorithms in IEEE 802.16 multi-hop relay networks," in *Proceedings of ICNS 2010*, pp. 265–270, Mar. 2010.

[11] M. V. Marathe, H. Breu, H. B. Hunt, S. Ravi, and D. Rosenkrantz, "Simple heuristics for unit disk graphs," *Networks*, vol. 25, pp. 59–68, Oct. 1995.

[12] B. Bonet and H. Geffner, "Planning as heuristic search," *Artificial Intelligence*, vol. 129, pp. 5–33, June 2001.

[13] F. Aurenhammer, "Voronoi diagrams - a survey of a fundamental geometric data structure," *ACM Computing Surveys*, vol. 23, pp. 345–405, Sept. 1991.

[14] D. Son, B. Krishnamachari, and J. Heidemann, "Experimental study of concurrent transmission in wireless sensor networks," in *Proceedings of ACM SenSys 2006*, pp. 237–250, ACM, Nov. 2006.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

Evaluation of IEEE 802.16j Relay Network Performance Considering Obstruction of Radio Wave Propagation by Obstacles

Yuuki Ise*, Go Hasegawa†, Yoshiaki Taniguchi† and Hirotaka Nakano†
*Graduate School of Information Science and Technology, Osaka University,
1–5 Yamadaoka, Suita, Osaka, Japan
Email: y-ise@ist.osaka-u.ac.jp
†Cybermedia Center, Osaka University,
1–32 Machikaneyama, Toyonaka, Osaka, Japan
Email: {hasegawa, y-tanigu, nakano}@cmc.osaka-u.ac.jp

***Abstract*—In IEEE 802.16j networks using a time division multiple access protocol, radio wave interference between links must be taken into account when time slots are assigned to the links. The protocol model, which defines the transmission and interference ranges as circles, has often been utilized in related studies for determining the interference between links. However, previous studies have not considered the presence of obstacles, which can affect radio wave propagation between wireless links and consequently network performance. In this paper, we investigate the performance of IEEE 802.16j networks considering the effects of obstacles. First we define an obstacle model where radio wave propagation is obstructed by the obstacles, after which we evaluate network performance by simulation experiments based on this model. The detailed simulation results reveal that the deployment of additional nodes improves the service ratio more effectively than an increase of the transmission range of nodes. Additionally, we present a method for estimating the performance of IEEE 802.16j networks on basis of regression analysis. By evaluating the accuracy of the performed analysis, we find that the relative error in the service ratio for 95% of the results is within 0.1 and the relative error in the power-to-throughput ratio for 90% of the results is within 0.2.**

Keywords-IEEE 802.16j; WiMAX; relay network; obstacle; radio wave obstruction.

I. INTRODUCTION

The IEEE 802.16j protocol [1] utilizes multi-hop wireless networks for extending the network service area [2]. Generally, IEEE 802.16j wireless multi-hop relay networks (hereinafter, relay networks) consist of two types of nodes: gateway and relay nodes. As shown in Fig. 1, there is a wired connection between the gateway node and an external network, while relay nodes communicate with the gateway node or with neighboring relay nodes through wireless links. These nodes construct a tree topology where the root is the gateway node and there is a wireless multi-hop transmission path from any relay node to the gateway node. A client terminal can access the external network by connecting to one of the relay nodes whose service area covers the client terminal.

Relay networks use a time division multiple access (TDMA) protocol [3], which gives transmission opportunities for wireless links between relay nodes. Therefore, radio wave interference between links should be taken into account when time slots are assigned to the links. Previous studies on relay networks have focused on preventing radio wave interference by introducing concepts such as link scheduling and

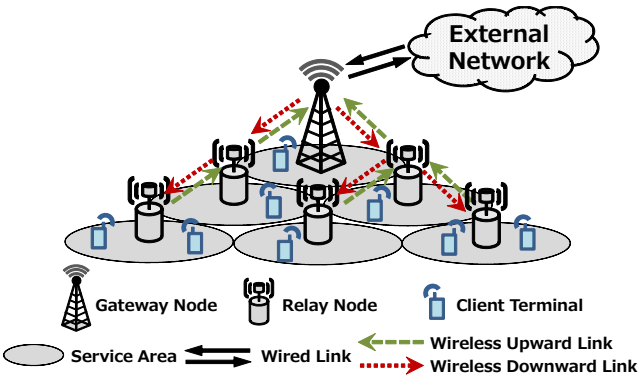


Figure 1. IEEE 802.16j wireless multi-hop relay networks

time slot assignment [4], [5]. In addition, the protocol model [6], which defines the transmission and interference ranges as circles, has often been utilized in previous works [7]–[10]. In the protocol model, whether a transmission succeeds or encounters interference depends on only the distance between the sender and receiver nodes, which are obtained through comparison with the transmission and interference ranges of other nodes. Therefore, the connectivity and the occurrence of radio wave interference in the network can be easily determined.

However, so far, the presence of obstacles in the field has not been considered. Here, obstacles refer to physical objects, such as buildings, that obstruct radio wave propagation and might substantially alter the connectivity and radio wave interference in a relay network. For example, even if a transmission between two nodes in a network is possible in the absence of obstacles, the addition of obstacles can prevent the nodes from communicating with each other. Furthermore, obstacles can obstruct radio wave propagation, thus reducing the size of the service area. On the other hand, in terms of radio wave interference, the presence of obstacles can increase network performance by reducing the occurrence of radio wave interference. Therefore, it is important to consider the presence of obstacles and their influence on radio wave propagation.

In this paper, we investigate the performance of relay networks by considering the presence of obstacles. First, we define an obstacle model based on the protocol model. In the obstacle model, rectangular obstacles are deployed in the field in various patterns. Although typically obstacles

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

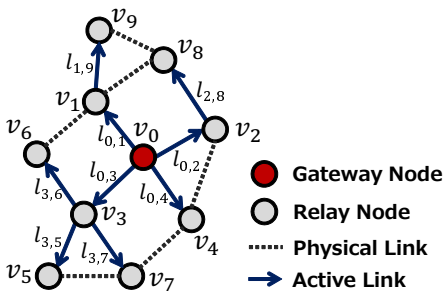


Figure 2. Network topology

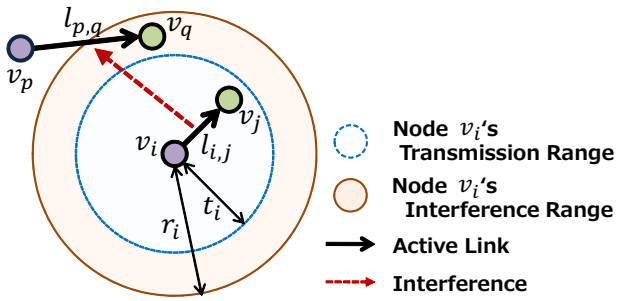


Figure 3. Radio interference based on the protocol model

have several effects on radio waves, such as obstruction, reflection, and diffraction, the obstacle model in this work considers only obstruction since its effects on relay network performance is the most pronounced. Using the obstacle model, we evaluate the performance of relay networks through simulation experiments. We also conduct simulation experiments using actual obstacles on the Osaka University campus.

Additionally, we use multiple regression analysis based on the simulation results and construct a regression equation for estimating the performance of relay networks.

The rest of this paper is organized as follows. The relay network model is introduced in Section II, the obstacle model is described in Section III, and the simulation results are presented in Section IV. Furthermore, a method for estimating the performance of relay networks is presented in Section V. Finally, our conclusions are presented in Section VI together with a description of future research directions.

II. Network Model

In this section, we briefly describe the network model and the protocol model utilized for the performance evaluation. We also explain time slot assignment mechanisms based on TDMA.

A. Network topology

The network is assumed to consist of N nodes, where v_i $(0 \leq i \leq (N-1))$ denotes the i-th node. One node in the network, denoted as v_0, serves as the gateway node, and the remaining nodes function as relay nodes, forming a network topology that describes the communication between all nodes in the form of a directed graph. In relay networks, the gateway node is connected to an external network, and the relay nodes communicate with the gateway node either directly or via other relay nodes along the path between the relay node and the gateway node. The path is determined by a routing algorithm, and the directed graph is constructed as a tree structure whose root is the gateway node v_0. When two nodes can communicate directly with each other, the link between them is referred to as a physical link, and a link used that forms the network topology is referred to as an active link. In this paper, an active link from node v_i to node v_j is denoted as $l_{i,j}$.

Figure 2 shows an example of a network topology where a gateway node and nine relay nodes are deployed. In the figure, the red circle indicates the gateway node, gray circles indicate relay nodes, dashed lines indicate physical links, and solid arrow lines indicate active links.

B. Protocol model

In this paper, the propagation and interference of radio waves are modelled by the protocol model [6], which defines the transmission and interference ranges of a node as circles centered at the node, as shown in Fig. 3. Here, t_i and r_i represent the transmission range and the radio interference range of node v_i, respectively. In general, $r_i > t_i$, and the ratio of the interference range to the transmission range for node v_i is set to be between 2 and 4 depending on the environment [11].

In the protocol model, transmission and radio wave interference depend on the distance between nodes. Node v_j can receive a transmission from node v_i when $||v_i - v_j|| \leq t_i$ is satisfied, where $||v_i - v_j||$ is the distance between nodes v_i and v_j. Figure 3 also shows an example of radio interference between nodes in the protocol model. In this figure, there are four nodes maintaining two active links ($l_{i,j}$ and $l_{p,q}$). The protocol model defines the interference between $l_{i,j}$ and $l_{p,q}$ based on the distances between the four vertices v_i, v_j, v_p, and v_q. When $||v_i - v_q|| \leq r_i$ is satisfied, link $l_{i,j}$ interferes with link $l_{p,q}$.

C. Time slot assignment

The IEEE 802.16j protocol uses the TDMA mechanism to control the ability of nodes to transmit by assigning time slots for transmission. In the TDMA mechanism, different time slots are assigned to wireless links that interfere with each other in order to prevent radio wave interference. In other words, multiple links can communicate simultaneously within the same time slot as long as the time slot is assigned to multiple links that do not interfere with each other. This mechanism is known as spatial reuse [12]. The throughput of relay networks can be substantially improved by spatial reuse with concurrent transmissions since such an approach reduces the total number of time slots assigned to all active links in the network, which is referred to as *frame length* in this paper. The time slot assignment problem with consideration of spatial reuse is regarded as a vertex coloring problem [13] of the conflict graph [14]. In the conflict graph, a vertex represents a link in the network and an edge between two vertices is constructed when the corresponding links interfere with each other, and time slots can be assigned to links in the network by allocating different colors to adjacent vertices in the conflict graph. However, since the vertex

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

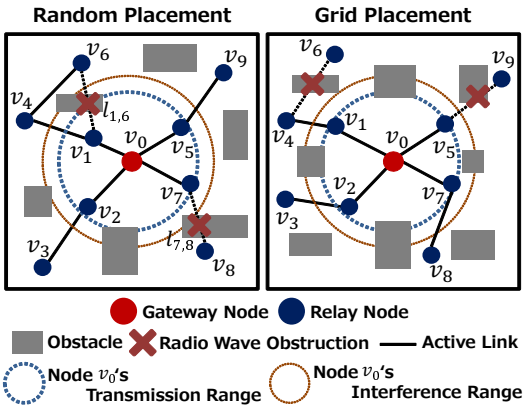


Figure 4. Obstacle model

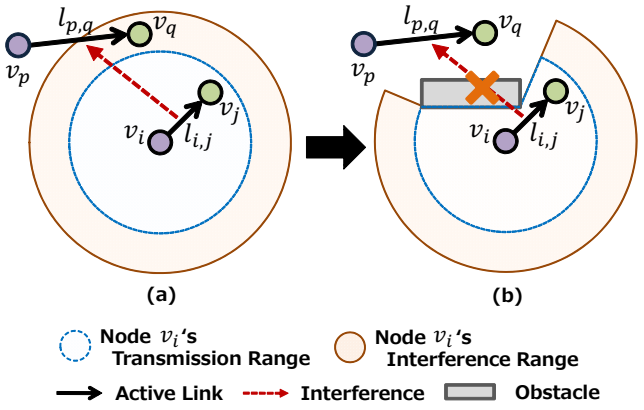


Figure 5. The influence of obstacles on radio wave interference

coloring problem is known to be NP-hard [15], heuristic algorithms have been proposed for solving the problem [11], [13], [16]. In this paper, we use the method proposed in [11] to assign time slots to links for performance evaluation.

III. Obstacle Model

In this section, we introduce an obstacle model in which obstacles are deployed in the network described in Section II. We also examine the influence of obstacles on network performance.

A. Obstacle model characteristics

Figure 4 shows the obstacle model, where rectangular obstacles are deployed in the field following two placement patterns. In random placement, the obstacles are deployed at random in the field. In grid placement, on the other hand, the obstacles are deployed in a grid pattern, with the exception of the center of the field, where the gateway node is located. The side lengths of the obstacles are chosen at random from within a certain range of values, and the obstacles are placed parallel to the field. The gateway node and the relay nodes cannot be deployed at locations occupied by obstacles. Although the influence of obstacles on radio waves includes obstruction, reflection, and diffraction, in this paper we consider only the obstruction of radio waves since its effects on the performance of relay networks is the most pronounced. We also ignore the height of the obstacles since the network model is constructed in a plane.

B. Effects of obstacles in relay networks

The obstacle model in this work considers only the effects of radio wave obstruction, which are described in this subsection. When the propagation of radio waves is obstructed by obstacles, the connectivity and radio interference in the network can change, either improving or degrading the performance of the relay network. For example, in the left panel of Fig. 4, $l_{1,6}$ and $l_{7,8}$ become disconnected due to radio wave obstruction by an obstacle. In this case, v_6 can connect to the network via v_4. On the other hand, v_8 is completely disconnected from the network by another obstacle. This is a negative aspect of obstacles in relay networks, owing to the increased number of isolated nodes and the higher average hop count between relay nodes and the gateway node.

On the other hand, Fig. 5 shows an example of a beneficial effect of radio wave obstruction. Although in Fig. 5(a) $l_{i,j}$ and $l_{p,q}$ interfere with each other, by adding obstacles as shown in Fig. 5(b), the interference range of v_i is limited and v_q is not affected by interference from v_i. As a result, the addition of an obstacle allows these two links to transmit simultaneously, which is a positive aspect of obstacles.

As described above, deploying obstacles entails both advantages and disadvantages in terms of network performance. Therefore, it is important to evaluate the performance of relay networks in the presence of obstacles.

IV. Performance Evaluation

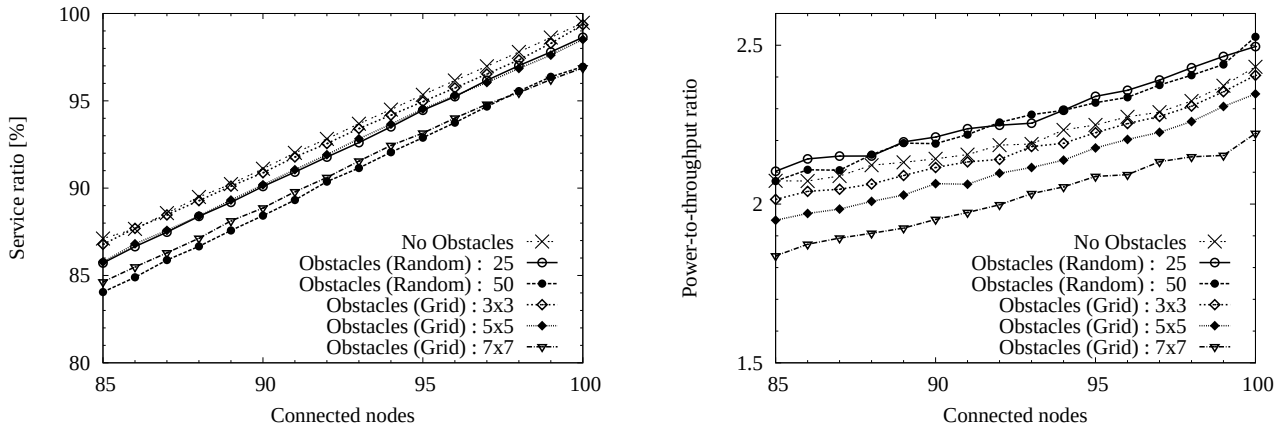
In this section, we evaluate the influence of obstacles on relay network performance by using the obstacle model described in Section III.

A. Evaluation settings

In the simulation experiments, one gateway node was placed at the center of a 1×1 square area, and 99 relay nodes were distributed at random. Based on the protocol model, the transmission range for all nodes was set to 0.15, 0.20, 0.25, 0.30, or 0.35 in consecutive experiments. The ratio of the interference range to the transmission range was set to 2 for all nodes. A directed transmission graph (Fig. 2) was constructed with a tree topology rooted at the gateway node such that the hop count between the gateway node and each relay node was minimized. The number of obstacles was set to 0, 25, or 50 in the case of random placement. For the grid placement, we chose one the following placement patterns: 3×3 (8 obstacles), 5×5 (24 obstacles), and 7×7 (48 obstacles). The length of the sides of each obstacle was set to a random value between 0.01 and 0.1. The transmission range of a node and the length of the sides of each obstacle are relative value based on the size of simulation area. In the simulation, the detail of distribution function is not taken into account.

We monitored the *service ratio* and the *power-to-throughput ratio* as measures of network performance. The service ratio is the ratio of the area where the relay network can provide service to the overall field area, excepting the

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1



(a) Service ratio (b) Power-to-throughput ratio

Figure 6. Effect of connected nodes on network performance

area of the obstacles, and the power-to-throughput ratio is the value of the total power consumption of the nodes divided by the gateway throughput. Here, the total power consumption is defined as the sum of the squares of the transmission ranges of all connected nodes. In addition, the gateway throughput is defined as the ratio between the number of time slots assigned to all links to the gateway node and the frame length. We conducted 100,000 iterations of the simulation experiments for each set of parameter settings, where all results were divided according to the number of connected nodes, and the average values were used for performance evaluation.

B. Impact of obstacle placement pattern

Figure 6 shows the service ratio and the power-to-throughput ratio as a function of the number of connected nodes when the transmission range of the nodes is set to 0.15. The x-axis of both graphs means the number of connected nodes. Not that the value of less than 100 means that there are some nodes disconnected from the network due to radio wave obstruction by obstacles. The figure shows that both measures increase as the number of connected nodes increase, but there are differences between the two plots. The service ratio in Fig. 6(a) decreases as the number of obstacles increases, regardless of the placement pattern, since radio wave propagation is obstructed by obstacles. On the other hand, the power-to-throughput ratio in Fig. 6(b) does not display such a simple trend. In the case of grid placement, the power-to-throughput ratio decreases as the number of obstacles increases, whereas in the case of random placement, the ratio increases together with the number of obstacles. The reason for this is as follows.

In random placement, the obstacles are deployed at random in the field, and therefore a link to the gateway node is likely to become disconnected due to the radio wave obstruction by obstacles. Therefore, the number of nodes connected to the gateway node decreases, which results in a decreased gateway throughput. On the other hand, in grid placement, since the obstacles are regularly spaced, the number of nodes connected to the gateway node is barely affected. As a result, the gateway throughput is improved

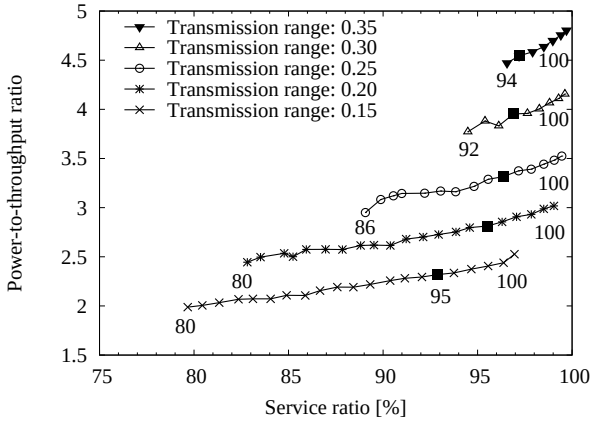


Figure 7. Effect of transmission range on network performance

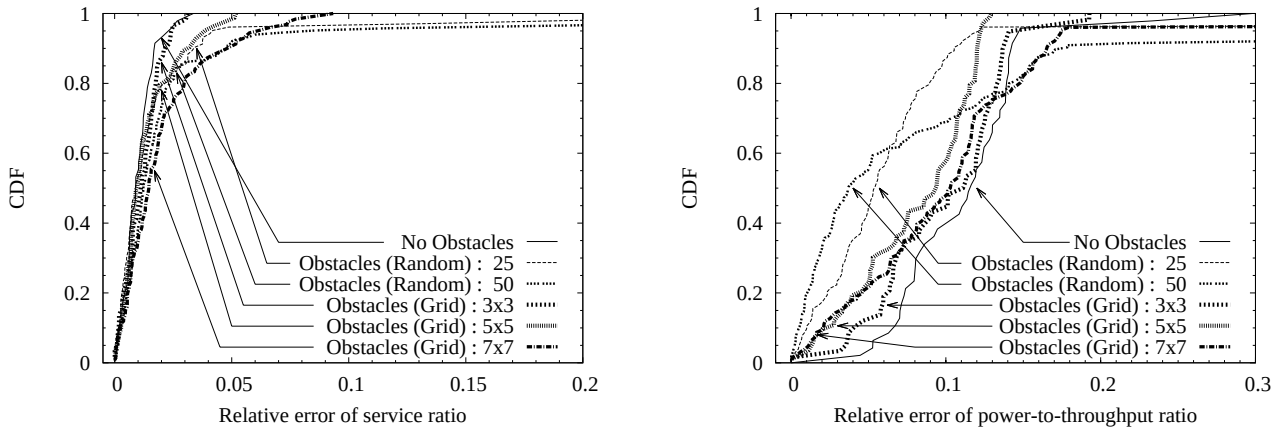
since the interference is reduced effectively due to the presence of obstacles. Therefore, we find that the service ratio is unaffected by the placement pattern of obstacles, but the power-to-throughput ratio is sensitive to the obstacle placement pattern.

C. Effect of the transmission range

Figure 7 shows the relation between the service ratio and the power-to-throughput ratio when the transmission range of the nodes is set to 0.15, 0.20, 0.25, 0.30, and 0.35. In this case, 50 obstacles are placed at random. In the graph, the number of connected nodes is indicated for each plot, where the number of connected nodes increases from left to right and the rightmost point indicates the average value when the number of connected nodes is 100.

We focus on the plots denoted with squares in the figure when the number of connected nodes is 95. Here, we consider two methods for improving the service ratio. One involves increasing the transmission range of the nodes, and the other involves deploying additional nodes in the network. As shown in Fig. 7, when the transmission range of the nodes is increased, the power-to-throughput ratio increases rapidly as the service ratio increases. On the other hand, by deploying additional nodes in the field, the service ratio

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1



(a) Service ratio (b) Power-to-throughput ratio

Figure 8. Relative error between regression equation results and simulation results

can be enhanced with a small increase of the power-to-throughput ratio. Therefore, the deployment of additional nodes improves the service ratio more effectively than an increase of the transmission range of the nodes.

V. METHOD FOR ESTIMATING NETWORK PERFORMANCE

In this section, we present a method for estimating the performance of relay networks on the basis of simulation results and regression analysis.

A. Regression equations

In the process of constructing a network, the ability to estimate network performance before the network is actually constructed is highly valuable. For example, such performance estimation enables the number of nodes deployed in the network or the transmission range of the nodes to be adjusted in order to achieve pre-determined performance goals such as service ratio, network throughput, and power consumption. Therefore, as a method for estimating network performance, we derived regression equations based on the simulation results presented in the previous section. Specifically, the equations for the service ratio and the power-to-throughput ratio are denoted as $S(n,t,d)$ and $P(n,t,d)$, where n, t, and d represent the number of connected nodes in the network, the transmission range of the nodes, and the distribution density of obstacles, respectively. The distribution density of obstacles is defined as the ratio of the area of obstacles to the overall area. All parameters are normalized to fall within the range between 0 and 1. The following regression equations can be derived from the simulation results. The details of the regression analysis are omitted due to space limitations.

$$S(n,t,d) = 92.8n + 22.8t^2 - 14.3d + 7.87 \quad (1)$$

$$P(n,t,d) = 2.34n + 11.3t - 2.28d - 1.30 \quad (2)$$

The performance of relay networks can be estimated by using the above equations. In order to examine the accuracy of the equations, the estimation values from the equations are compared with the simulation results in the following subsections, where the relative error is used as a measure of accuracy.

B. Evaluation of the regression equation accuracy

Figure 8 shows the respective distributions of the relative error for the service ratio and the power-to-throughput ratio with several values for the number of obstacles in the case of both placement patterns. The accuracy of the equations is clearly high regardless of the obstacle placement pattern and the number of obstacles in the network. In particular, for the service ratio (Fig. 8(a)), the relative error for 95% of the results is within 0.1. For the power-to-throughput ratio in Fig. 8(b), the relative error for about 90% of results is within 0.2. These results show that Eqs. (1) and (2) can provide accurate estimates of network performance without simulation experiments.

C. Campus model

Finally, to evaluate the accuracy of the equations in a more realistic situation, simulation experiments and performance estimation with Eqs. (1) and (2) were conducted with respect to actual obstacles on the Osaka University campus, as shown in Fig. 9.

Figure 10 shows a comparison between the estimation and simulation results for different transmission ranges. In the graph, lines without plotted points represent the estimation results, and ones with plotted points represent the simulation results. When the number of connected nodes is between 75 and 85, the results of estimation and simulation are close; however, as the number of connected nodes increases further, the error between the two results becomes large, because in the campus model, radio wave propagation is obstructed more frequently as a result of the large number of obstacles and the difference in shape and size of each obstacle in the field as compared with those in the obstacle model with random and grid placement. Therefore, although the service ratio is difficult to improve, the number of connected nodes increases. The another reason is the heterogeneous distribution of obstacles in the field as shown in Fig. 9, while both obstacle placement patterns in this paper assume the homogeneous distribution.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

Figure 9. Campus model

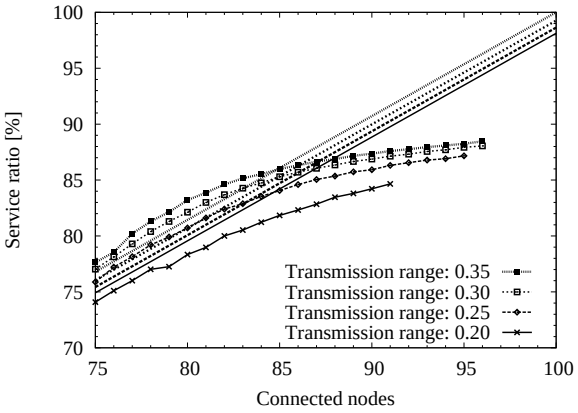


Figure 10. Comparison between estimation and simulation results

VI. Conclusions and Future Work

In this paper, the performance of IEEE 802.16j multi-hop networks was investigated by considering the presence of obstacles, where the obstacle model was defined as an extension of the protocol model. Simulation experiments based on using the obstacle model revealed that the obstacle placement pattern does not affect the service ratio, but does greatly affect the power-to-throughput ratio. A method for estimating the performance of relay networks on the basis of regression analysis was also provided in this study, and a comparison between the simulation results and the estimation results derived from the regression equations confirmed that the equations can yield an accurate estimation of network performance. However, in testing the model with respect to a real-world environment (a university campus), the accuracy of the equations was found to be lower than in the case of random and grid placement due to differences in the characteristics of the obstacles, such as shape and size, and the distribution characteristics of obstacles.

Future work will be directed toward applying other radio interference models using signal-to-interference-plus-noise ratio in order to consider other effects of obstacles, such as reflection and diffraction of radio waves. Moreover, additional research will be conducted with the aim to increase the accuracy of the regression equations, especially with respect to real-world networks.

Acknowledgment

This work was partly supported by KAKENHI 23700079.

References

[1] IEEE Std 802.16j, *IEEE standard for local and metropolitan area networks, Part 16: Air interface for fixed broadband wireless access systems*, Jun. 2009.

[2] S. W. Peters and R. W. H. Jr, "The future of WiMAX: Multihop relaying with IEEE 802.16j," *IEEE Communications Magazine*, vol. 1, pp. 104–111, Jan. 2009.

[3] D. Gohsh, A. Gupta, and P. Mohapatra, "Scheduling in multihop WiMAX networks," *ACM SIGMOBILE Mobile Computing and Communication Review*, vol. 12, pp. 1–11, Apr. 2008.

[4] C. Y. Hong and A. C. Pang, "3-approximation algorithm for joint routing and link scheduling in wireless relay networks," *IEEE Transactions on Wireless Communications*, vol. 8, no. 2, pp. 856–861, Feb. 2009.

[5] H. Y. Wei, S. Ganguly, R. Izmailov, and Z. J. Haas, "Interference-aware IEEE 802.16 WiMAX mesh networks," in *Proc. IEEE VTC 2005*, pp. 3102–3106, May 2005.

[6] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, pp. 388–404, Mar. 2000.

[7] H. Venkataraman, A. K. P. Kalyampudi, J. McManis, and G. M. Muntean, "Clustered architecture for adaptive multimedia streaming in WiMAX-based cellular networks," in *Proc. WCECS 2009*, pp. 753–758, Oct. 2009.

[8] P. Thulasiraman and X. S. Shen, "Interference aware subcarrier assignment for throughput maximization in OFDMA wireless relay mesh networks," in *Proc. ICC 2009*, pp. 14–18, Jun. 2009.

[9] C. Cicconetti, I. F. Akyildiz, and L. Lenzini, "Bandwidth balancing in multi-channel IEEE 802.16 wireless mesh networks," in *Proc. INFOCOM 2007*, pp. 6–12, May 2007.

[10] Y. Lu and G. Zhang, "Maintaining routing tree in IEEE 802.16 centralized scheduling mesh networks," in *Proc. ICCCN 2007*, pp. 240–245, Aug. 2007.

[11] W. Wang, Y. Wang, X. Y. Li, W. Z. Song, and O. Frieder, "Efficient interference-aware TDMA link scheduling for static wireless networks," in *Proc. MobiCom 2006*, pp. 262–273, Sep. 2006.

[12] L. Kleinrock and J. Silvester, "Spatial reuse in multihop packet radio networks," in *Proc. the IEEE*, vol. 75, pp. 156–167, Jan. 1987.

[13] M. V. Marathe, H. Breu, H. B. Hunt, S. S. Ravi, and D. J. Rosenkrantz, "Simple heuristics for unit disk graph," *Networks*, vol. 25, pp. 59–68, Sep. 1995.

[14] K. Jain, J. Padhye, V. N. Padmanabhan, and L. Qiu, "Impact of interference on multi-hop wireless network performance," *Wireless Networks*, vol. 11, pp. 471–487, Jul. 2005.

[15] S. Khanna, N. Linial, and S. Safra, "On the hardness of approximating the chromatic number," *Combinatorica*, vol. 20, no. 3, pp. 393–415, Mar. 2000.

[16] W. Klotz, "Graph coloring algorithms," *Mathematical report TU-Clausthal*, vol. 5, pp. 1–9, May 2002.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

Distribution of the Frequency of Connections in Academic WLAN Networks

Enrica Zola
Department of Telematics
Universitat Politècnica de Catalunya (UPC)
Barcelona, Spain
enrica@entel.upc.edu

Francisco Barcelo-Arroyo
Department of Telematics
Universitat Politècnica de Catalunya (UPC)
Barcelona, Spain
barcelo@entel.upc.edu

***Abstract*—Understanding the trends in the use of wireless local area networks (WLANs) (i.e., how much, when and where traffic is present) is an important issue for modeling the network and for optimizing the allocated resources. Interesting results can be extracted by analyzing traces from real scenarios. In previous works, the authors studied three buildings belonging to two campuses in Barcelona (Spain) and its surroundings. Similar common trends were observed in the three buildings, despite the different amount of users, purpose of the building, geographical location and size of the campus. In this work, the fidelity of the users in accessing the WLAN in different days is analyzed in depth. The population accessing the networks is mostly composed of infrequent users: less than half of the devices access the WLAN more than four days during the studied period. Special insight is given to the underlying distribution. It is shown that, in contrast to previous studies in the same environment, the distribution of the frequency of reconnections to the WLAN is not uniform. The main difference among different buildings is the fidelity of users: users on a small campus are more likely to reappear on different days than on a large campus, where the population is more heterogeneous. The results of this analysis provide general tools for characterizing campus-wide WLAN and a better understanding of usage and performance issues in a mature wireless network in Europe.**

Keywords-WLAN traces; syslog; user behavior; frequency of connections.

I. INTRODUCTION

Among other advantages, wireless communications provide flexibility in the deployment of the network and allow users to move around while connected. Users are increasingly interested in taking advantage of the flexibility of wireless technology, and a boom in its implementation to local area networks has been seen. There are many applications of wireless local area networks (WLAN) to universities, from the classical communications (including email and web browsing) to special e-learning applications. Accordingly, universities have pioneered the development of infrastructures to provide connectivity. An example is the Wireless Andrew at the Carnegie Mellon University campus [1], an enterprise-wide broadband wireless network developed in 1993.

In the last decade, research has been carried out to understand the use of WLAN (typically using standard IEEE 802.11) networks in different scenarios, like campus-wide universities [2]-[9], corporate networks [10], and metropolitan area networks [11]. All these works differ in the way in which data are collected; *tcpdump*, the Simple Network Management Protocol (SNMP) and *syslog* are the most common tools. *Tcpdump* is normally used to sniff the traffic and analyze the applications run by users and the amount of data managed in the network [2], [4], [5], while SNMP is applied to periodically poll the access points (APs) of the network [2], [10] and to obtain information regarding the authenticated/associated users and their approximate location (i.e., the coverage range of the AP to which a user is associated). Information on the authenticated/associated devices at each AP can also be obtained through *syslog* [4], [7], a standard for forwarding log messages in an IP (Internet Protocol) network. As found in [12], although polling-based trace collection is suitable for usage statistics, it is not very suitable for deriving the association patterns of users because they tend to overlook details of association changes due to the polling interval. Differences may also be found in the duration of the period analyzed; some span a period of a week [3], in which case weekly cycles cannot be observed, others one month [10], and others three to five months [2], [4], [5], [7], in which case monthly patterns can also be observed. Each work can also involve different numbers of users. In previous studies, number of users ranged from 74 users in the earliest study [2] to 8 thousand in wider environments [4], [5], [7], [10].

In this work, we process data recently collected at the Technical University of Catalonia (UPC), Spain, from March to May 2009. UPC is composed of eight campuses in Barcelona and its surroundings. Wireless local area network traffic is analyzed at each building separately to deal with homogeneous data and not to mix behaviors from populations with different features. Two libraries were selected. The first one, BRGF, is located on a large campus in Barcelona, while the second one, EDSE, is on a campus in Castelldefels. Results from another building (EETAC) in Castelldefels, a faculty building with classrooms and studying rooms, are also provided. To complete the work presented in [7], this paper analyzes the frequency in the connections to the WLAN. A fitting distribution is found which can be used to describe how often the same user will appear in the network during a given period of time. Differences are found with respect to previous works in a wide university [4], while the pattern is more similar to that found in a corporate network [10].

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

The paper is organized as follows. Section 2 provides a summary of the related literature. Section 3 provides details of the methodology followed to collect the data from the WLAN and of the environment where the analysis was conducted. Main findings on users' activity are given in Section 4. The frequency of user's connections is examined in Section 5. Section 6 concludes the paper.

II. RELATED WORK

Several studies have been performed to analyze network traffic and user behavior in different WLAN environments. The first study was performed at the Computer Science Department of Stanford University in 1999 [2]. In this work, the authors prepared a testbed with 74 laptops. Users moved freely inside a building and the authors analyzed their movements and traffic during three months. Data were collected using three different techniques (i.e., *tcpdump* traces, SNMP polling and authentication logs). During the day, they observed higher activity in the afternoon, while during the week they found lower activity in the weekends. They found out that most users did not move much within the building; however, a few users were highly mobile. Regarding the total number of days that users are active (present in the network) during the traced period, they found that while some users rarely connect their laptops to the network (17 users do so on 5 days or less), others connect their laptops frequently (14 users are active at least 37 days during the traced period).

The use of the WLAN at the Saskatchewan University Campus was presented in [3]. The campus is composed of 40 buildings covering public spaces (i.e., lounges, libraries, coffee shops, etc.), classrooms, laboratories and offices. A traffic trace was collected during one week in January 2003 using EtherPeek, a software package that allows researchers to record MAC (i.e., Medium Access Control) addresses and traffic load information. MAC addresses were matched with the authentication logs obtained from each of the 18 APs of the campus. In total, 134 unique users connected to the network during the week under study. Individual users visited at most 8 different APs. Regarding the number of authentications, they observed that more than half of the total number of users authenticated more than 50 times during the week. According to the method used for collecting traces, they stated that authentications cannot be literally interpreted as distinct sessions of network usage.

Generalization of the results presented in these works is difficult due to the low number of users observed (74, 134 and 195, respectively). Moreover, the fact that users knew about the tracing study may have perturbed the user behavior in [2], while in [3] authors are aware that no effort was made to ensure that the week they analyzed was representative of overall usage patterns.

Kotz and Essien analyzed the wireless network at Dartmouth College during three months [4] and extracted information about users' mobility, card and building activity, traffic load and protocols. The main results are that network activity shows clear patterns: almost half of the users were active on a typical day, and about one third of those were mobile. The weekly pattern showed a typical student's pattern of activity, with lower activity on Fridays and Saturdays, and a pick up on Sundays. Users varied in the number of days that they used their cards, from only once to every day in the 77-day trace (many users lived on campus and could be always on-line). The distribution of the frequency in the user activity is roughly uniform between one and 77 days, with a median of 28 days. In 2004, they revisited the WLAN [5] and found that, despite a drastic increase in traffic, users were mainly non-mobile.

Similar user patterns were found in a corporate network from July 20 to August 17, 2002 [10]. Despite mobility results report higher mobility than on university campuses, users still spend a large fraction of time at only one location. The results regarding the daily and weekly patterns are similar to those observed on university campuses. The number of days that users connected to the WLAN varies greatly: only 12% to 25% of users are present more than 18 out of the 20 work days, whereas 22% to 38% of users appear only during one or two days (i.e., outside visitors mostly from other sites that the company has in the same metropolitan area). The presence of visitors and the absence of employees were uniformly distributed. In terms of the fraction of days that users access the network, the distribution is similar to a single building on a university campus [2]. Compared with a whole campus [4], more users appeared only one or two days (visitors) and fewer users appeared more than 2/3 of the days. The authors observed that the higher uniformity of a campus wide distribution might be related to the fact that the study tracked many users for prolonged periods of time (i.e., students living on campus) and not only when they went to work in specific buildings.

Mc Nett and Voelker analyzed the mobility patterns of users of wireless handheld PDAs in a campus wireless network using trace belonging to 11 weeks of wireless network activity. They also observed the frequency of connections in their traces: 50% of the users initiated more than 77 sessions over the trace period. This means that the median user initiated an average of one session per day over the trace period. Still, 20% of the users initiate roughly three sessions per day, and 10% initiated roughly four connections per day. To understand user activity on a day to day basis, they analyzed the number of days the users actually turned on their PDAs. Half the users turned on their devices less than 21 days during the trace. This is lower than the median number of days from [4]. Moreover, the distribution is not uniform as observed in [4]; 8% of the users only used their PDAs one day during their trace period. The number of users for each number of active days tends to drop from there. 20% of the users used their PDAs 60% of the total days; 10% used them more than 75% of the total days; and there were a few users who used their PDAs nearly every day.

The trends regarding the frequency of connections observed at the three buildings of the UPC are presented in the following sections. Not only the users' behavior will be related to the environment and compared with previous similar works. A deeper insight will be given to the underlying distribution, in order to model the fidelity of the user in the WLAN network.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

III. TRACE COLLECTION AND SCENARIO

This paper analyzes data collected during 3 months from March 2009. Users were not informed that the study was performed. The only sensitive information gathered were the MAC and IP addresses of network cards connected to the network and the name assigned to each AP; these data are not shown.

The data were collected with syslog [14], which is a standard client/server protocol for forwarding log messages in an IP network: the client sends a small textual message to the syslog server through User Datagram Protocol and/or Transport Control Protocol connections. The access points were configured to send syslog messages to a central server whenever they received authentication, deauthentication, association or disassociation IEEE 802.11 messages. Each message contains the AP name, the MAC address of the card, the time stamp at which the AP received the message to 1-second precision, and the type of message. From the MAC address, it is possible to relate a given address to a given device; however, the same user may have multiple cards, or the same device can be used by different persons. In the rest of the paper, we will use the term "device" to refer to a given MAC address.

According to the IEEE 802.11 standard, after authentication, a device chooses the best AP among a list of nearby APs and associates with it. When a device no longer needs to use the network, it disassociates with its current AP. Disassociation can be due to the device moving into another cell (i.e., handover), to authentication problems, or to the device leaving the network. The aim of this study is to analyze users' frequency of connections to the WLAN and extract the underlying distribution. For this purpose, only IEEE 802.11 association frames are considered. A device can only associate to an AP to which it has authenticated before, so authentication messages can be ignored. In order to catch users' frequency, the number of days an association response frame is received by a given device is calculated. The same for each device observed in the trace (see last row in Table I)

A. Library in Barcelona (BRGF)

The campus in Barcelona houses the faculties of Information and Communication Technologies and of Civil Engineering, with a total of 4,339 students. On this campus, one can also find research centers, the library, student associations, and the UPC Foundation, which is dedicated to lifelong learning and professional retraining. The central library (BRGF) is housed in a five-story building of 6,300 m^2, four above-ground floors and one basement. People access the library from the ground floor, where the loan service is housed. The first and second floors house the library collection, which is divided according to the specific subjects taught on the campus; students spend their time there to work on their subjects and to study. The third floor provides specific documentation for PhD students and researchers. In the basement, there are two studying rooms and two rooms for foreign languages teaching. BRGF provides room for about 700 users and provides 63 desktop PCs and 9 laptops for library users. BRGF is open from 9 am to 9 pm from Monday to Friday, and during the exam period, the opening hours are extended until 2:30 am and during weekends.

The WLAN infrastructure at BRGF is composed of eight APs on the first four floors and four APs in the basement, which provide good coverage all over the building. A total of 5,917 devices associated to the wireless network during the whole trace period. Table I summarizes the main parameters of the three buildings.

B. Library in Castelldefels (EDSE)

The Castelldefels campus houses the Castelldefels School of Technology (EETAC), the School of Agricultural Engineering of Barcelona and a library; it is located 30 km from Barcelona. The library is housed in a three-story building. It is open from 8 am to 9 pm from Monday to Friday. Three APs are located on the ground floor (edse002 to edse004) where the loan service is located, four on the first floor (edse101 to edse104) and two on the second floor (edse201 and edse202), where the library collection and the studying rooms are located.

The infrastructure provides good coverage inside the building. In this analysis, the APs in the basement are also considered (edses01 to edses04) because students go to the bar and connect to the WLAN from there as well. A total of 1,419 devices associated to the wireless network during the whole trace period.

C. Classrooms (EETAC)

EETAC is a Higher Education School specializing in technical and scientific courses in Aeronautics and Telecommunications. It is housed in the Castelldefels campus. The total number of students attending classes at EETAC is about 1,500 persons. The building houses 20 classrooms, 25 laboratories, professors' offices and 2 studying rooms, all distributed on 3 floors. It is open from 8 am to 9 pm from Monday to Friday.

The WLAN infrastructure at EETAC is composed of twelve access points, which provide good coverage throughout the building. A total of 1,417 devices associated to the wireless network during the whole trace period.

IV. COMPARISON OF THE ACTIVITY

In our previous study [7], the association pattern was observed during the three-month period. Table I reports the main parameters of the three scenarios. The number of APs providing the WLAN infrastructure is the same in the three buildings, despite the EETAC building is more than three time bigger than EDSE building. The number of students in the Barcelona campus is 3 times that in the Castelldefels campus, while the number of detected devices at BRGF is higher than the number of students due to the proximity to other university (e.g., students from outside the UPC are using the WLAN). Since the Castelldefels campus is far away from other universities, the number of detected devices nearly equals the number of students.

Details on the user behavior at each building are given in [7]. A decrease in the usage of the WLAN was observed at each building during Easter holidays (one week in April),

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

during weekends at the two buildings in Castelldefels, during one week in April at EETAC building due to exam period. In order to deal with homogeneous data, only the working days in March and May are taken into account in the rest of the paper (no holidays, nor weekends). April is skipped due to inconstant activity.

V. FREQUENCY OF CONNECTIONS

The empirical cumulative distribution (CDF) of devices according to the number of days per period that they appear as active is shown in Figure 1. At BRGF, more than 50% of the devices connect just once or twice in two months. This proportion of infrequent users decreases for EDSE (35%) and EETAC (24%). This is explained by the proximity between the library at BRGF and other campuses that cause a higher proportion of occasional visits; in a smaller campus, on the other hand, users tend to re-appear more often. 10% of the users at each building appear: more than 12 days at BRGF; more than 21 days at EDSE; and more than 24 days at EETAC. Again, more similar trends are observed in the Castelldefels campus. The number of very frequent users (i.e., those who appear more than 30 days over the whole period) is 1% at BRGF and 4% at both buildings in Castelldefels. The building housing classrooms (EETAC) displays the highest fidelity among the three buildings (lower number of infrequent users, and higher number of very frequent users).

TABLE I. PARAMETERS OF THE THREE SCENARIOS.

	BRGF	EDSE	EETAC
Nº of APs	12	13	12
Sq. meters	6,300	3,000	10,000
Students in the campus	4,439	1,524	1,524 (same as EDSE)
Detected devices (whole period)	5917	1417	1419

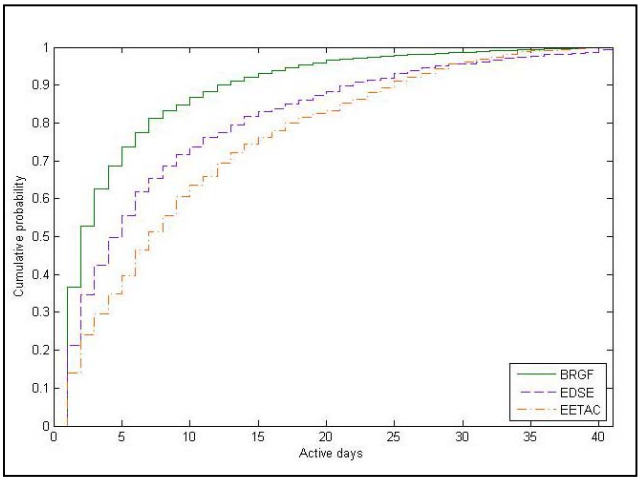


Figure 1. Fraction of active devices per period.

Table II resumes the trends observed in previous similar works. Results in [4] and [6] show very different behaviors if compared with our trace, despite all refer to traces taken from university campuses. In these other works, the percentage of users who connect once or twice during the whole period is lower than in our traces; half of the users connect more often; 10% of the users connect more than 91% of the days in [4] and more than 75% in [6], which are higher than in our trace. The distribution in [4] is roughly uniform between one and 77 days. On the other hand, the figure observed in a corporate network [10] is more similar to the one observed in our trace. EDSE can be compared to the small building (SBldg) in [10], while EETAC to the large building (LBldg). This similarity with a corporate building is due to the fact that, at UPC, students do not spend the night inside the campus (which is instead typical in the campuses in the US) but go home at night and return the next day they have classes; this behavior is similar to that of a worker. BRGF shows quite different trends, with the highest percentage of infrequent users (connected up to 2 days) and the lowest percentage of constant users.

The probability mass function of devices that are active a given number of days is shown in black in Figures 2 to 4 for BRGF, EDSE and EETAC building, respectively. Three theoretical distributions have been proposed and tested through the Chi-Square goodness of fit (gof) test: Zipf, geometric and negative binomial. The latter is not represented in the figures since it gives the worst matching. Moments' matching has been applied to estimate the parameters for each distribution, as shown in Table III. Again the lowest mean belongs to the BRGF due to its proximity to other similar campuses.

TABLE II. MOMENT ESTIMATORS FOR EACH CANDIDATE DISTRIBUTION.

Ref.	Environment	Buildings	% of users up to 2 days:	50% of users connect up to [% of total days]:	10% of the users connect more than [% of total days]:
This work	Campus	BRGF EDSE EETAC	53 35 24	5 10 15	31 51 59
[4]	Campus		10	36	91
[6]	Campus		11	25	75
[10]	Corporate network	Large Medium Small	24 22 38	12 27 30	67 63 60

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

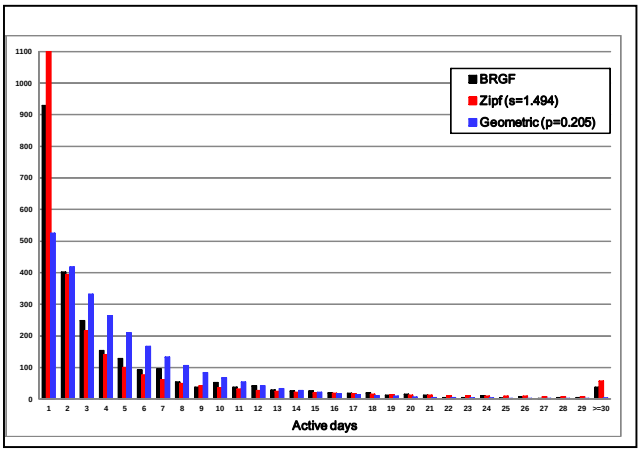


Figure 2. Probability mass function of the active devices for a given number of days (BRGF).

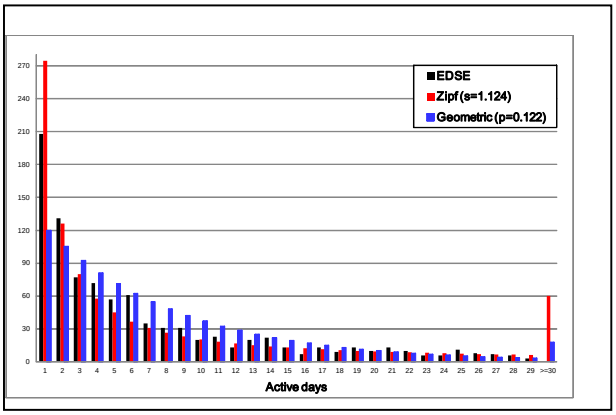


Figure 3. Probability mass function of the active devices for a given number of days (EDSE).

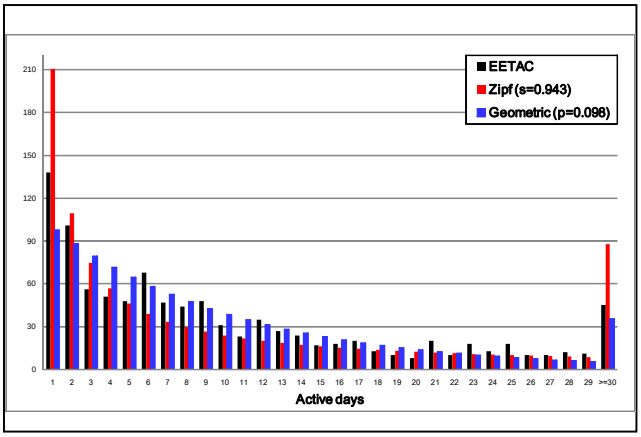


Figure 4. Probability mass function of the active devices for a given number of days (EETAC).

The Chi-Square gof test with 5% significance was performed for each building and distribution. Given the null hypothesis that the fitting distribution is F, this test compares the number O_i of observed elements of the empirical distribution in category C_i (for i = 1,2,…,k), with the expected number E_i of elements of F in category C_i. As a measure of comparison the test uses [15]:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}. \quad (1)$$

χ^2 is known as the Chi-Square test statistic: if it is greater or equal to the critical value χ^2_α, then the null hypothesis that the fitting distribution is F must be rejected at level α. The critical value depends on the level of significance α and on the sample size n (see Table III), and can be computed as:

$$\chi^2_\alpha = n\left\{1 - \frac{2}{9n} + z_\alpha \sqrt{\frac{2}{9n}}\right\}^3, \quad (1)$$

where z_α can be found in [15]. In this paper, we use $z_{0.05}$ = 1.6449. The test statistic χ^2 and the critical value χ^2_α are displayed in Table IV.

For all the buildings, the Zipf distribution provides a good fit. According to the Zipf law's, the frequency of active days is inversely proportional to its rank in the frequency table. Thus the most frequent number of active days will occur approximately twice as often as the second most frequent number of active days, three times as often as the third most frequent number of active days, etc. For the EETAC building, the best fit is provided by the geometric distribution. Despite the similarities shown in Table II, the distribution for the two buildings at the Castelldefels campus is not the same. Instead, the same distribution characterizes the frequency in the connection to the WLAN in the libraries (BRGF and EDSE), while a different behavior is shown in a building housing classes.

TABLE III. MOMENT ESTIMATORS FOR EACH CANDIDATE DISTRIBUTION.

Distribution	BRGF	EDSE	EETAC
Sample size	2524	980	994
Mean	4.88	8.18	10.18
std	6.27	9.10	9.12
Zipf	1.494	1.124	0.943
Geometric	0.205	0.122	0.098
Neg. binomial	p=0.876 r=0.692	p=0.901 r=0.897	p=0.878 r=1.420

TABLE IV. CHI-SQUARE TEST STATISTIC (X^2) FOR EACH THEORETICAL DISTRIBUTION.

Distribution	BRGF	EDSE	EETAC
Zipf	**119.29**	**65.64**	146.00
Geometric	980.19	160.02	**91.39**
Neg. binomial	3092.56	4102.36	2035.40
Critical value (χ^2_α)	2641.99	1053.94	1068.46

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

VI. CONCLUSIONS

The frequency of the connections at three different buildings in two different campuses of the UPC in Barcelona (Spain) has been investigated. It has been shown that the trends are more similar to those observed in a corporate network [10] than to those reported from other universities in the USA [4][6]. This is mainly due to the fact that the students in Barcelona do not live inside the campus, so their pattern is more similar to that of workers. The distribution of the active days has been analyzed through the Chi-Square goodness of fit test: the same distribution (Zipf) can characterize the behavior at the libraries of both campuses, while the geometric distribution better fit the behavior in the building with classrooms. The results presented in this paper provide general tools for characterizing campus-wide WLAN and a better understanding of usage and performance issues in a mature wireless network in Europe. These findings may be useful both for those researchers interested in simulations under realistic scenarios and for optimal planning of a WLAN infrastructure in similar environments.

ACKNOWLEDGMENT

The authors would like to thank UPCnet for providing WLAN traces. This work was supported by the Spanish Government and ERDF through CICYT projects TEC2009-08198.

REFERENCES

[1] A. Hills, "Wireless Andrew [mobile computing for university campus]," IEEE Spectrum, vol. 36, no. 6, June 1999, pp. 49-53, doi: http://dx.doi.org/10.1109/6.769269.

[2] D. Tang, and M. Baker, "Analysis of a Local-Area Wireless Network," Proc. of the 6th Annual International Conference on Mobile Computing and Networking (MobiCom'00), ACM, New York, August 6-11, 2000, pp. 1-10, doi: http://dx.doi.org/10.1145/345910.345912.

[3] D. Schwab, and R. Bunt, "Characterising the Use of a Campus Wireless Network," Proc. of the 23rd Annual Joint Conference of the IEEE Computer and Communications Societes (Infocom'04), March 7-11, 2004, pp. 862-870, vol. 2, doi: http://dx.doi.org/10.1109/INFCOM.2004.1356974.

[4] D. Kotz, and K. Essien, "Analysis of a Campus-Wide Wireless Network," Wireless Networks, vol. 11, no. 1-2, January 2005, pp. 115-133, doi: http://dx.doi.org/10.1007/s11276-004-4750-0.

[5] T. Henderson, D. Kotz, and I. Abyzov, "The Changing Usage of a Mature Campus-wide Wireless Network. Computer Networks," The International Journal of Computer and Telecommunications Networking, vol. 52, no. 14, October 2008, pp. 2690-2712, doi: http://dx.doi.org/10.1016/j.comnet.2008.05.003.

[6] M. McNett, and G.M. Voelker, "Access and Mobility of Wireless PDA Users," ACM Sigmobile Mobile Computing and Communications Review, vol. 9, no. 2, April 2004, pp. 40-55, doi: http://dx.doi.org/10.1145/1072989.1072995.

[7] E. Zola, and F. Barcelo-Arroyo, "A Comparative Analysis of the User Behavior in Academic WiFi Networks," Proc. of the 6th Performance Monitoring, Measurement and Evaluation of Heterogeneous Wireless and Wired Networks Workshop (ACM PM2HW2N 2011), October 31-November 4, 2011.

[8] E. Zola, F. Barcelo-Arroyo, and M. Lopez-Ramirez, "User Behaviour in a WLAN Campus: a Real Case Study," Proc. of the third Ercim Workshop on eMobility, May 27-28, 2009, pp. 67-77.

[9] R. Hutchins, and E.W. Zegura, "Measurements of a Campus Wireless Network," Proc. of the IEEE International Conference on Communications (ICC'02), April 28 – May 2, 2002, pp. 3161-3167, vol. 5, doi: http://dx.doi.org/10.1109/ICC.2002.997419.

[10] M. Balazinska, and P. Castro, "Characterizing Mobility and Network Usage in a Corporate Wireless Local-Area Network," Proc. of the 1st International Conference on Mobile Systems. Applications and Services (MobiSys'03), ACM, New York, May 5-8, 2003, pp. 303-316, doi: http://dx.doi.org/10.1145/1066116.1066127.

[11] D. Tang, and M. Baker, "Analysis of a Metropolitan-Area Wireless Network," Wireless Networks, vol. 8, no. 2-3, March-May 2002, pp. 107-120, doi: http://dx.doi.org/10.1023/A:1013739407600.

[12] W.J. Hsu, and A. Helmy, "On Modeling User Associations in Wireless LAN Traces on University Campuses," Proc. of the 4th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks, April 3-6, 2006, pp. 1-9, doi: http://dx.doi.org/10.1109/WIOPT.2006.1666494.

[13] A. Balachandran, G.M. Voelker, P. Bahl, and P.V. Rangan, "Characterizing User Behavior and Network Performance in a Public Wireless LAN," Proc. of the International Conference on Measurements and Modeling of Computer Systems (ACM SIGMETRICS'02), June 15-19, 2002, pp. 195-205. doi: http://dx.doi.org/10.1145/511334.511359.

[14] C. Lonvick, "The BSD Syslog Protocol," RFC 3164 (rfc3164). IETF. 2001.

[15] A.O. Allen, "Probability, Statistics, and Queueing Theory with Computer Science Applications," Academic Press Professional, Inc., San Diego, CA, USA, 1990, isbn: 0-12-051051-0.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

Power Consumption Analysis of Data Transmission in IEEE 802.11 Multi-hop Networks

Wataru Toorisaka†, Go Hasegawa‡, and Masayuki Murata†
† Graduate School of Information Science and Technology, Osaka University
1-5 Yamadaoka, Suita-shi, Osaka, 565-0871 Japan
Email: {t-wataru,murata}@ist.osaka-u.ac.jp
‡ Cybermedia Center, Osaka University
1-32 Machikaneyama-cho, Toyonaka-shi, Osaka, 560-0043 Japan
Email: hasegawa@cmc.osaka-u.ac.jp

***Abstract*—The issue of power consumption in wireless networks is becoming increasingly important due to the rapid development of various wireless devices such as sensors, smartphones, and tablet PCs. The IEEE 802.11 wireless LAN standard defines multiple data transmission protocols, each of which has various characteristics such as power consumption, data rate, modulation method, and transmission distance. Therefore, it is important to choose the optimal data rate in terms of power consumption as well as throughput, especially when considering data transmission over multi-hop networks. In this paper, we present a mathematical analysis of power consumption in data transmission over IEEE 802.11-based wireless multi-hop networks to investigate the effect of data rate selection on power consumption. The analysis results show that there are some situations where a low data rate should intentionally be selected in order to minimize power consumption. Our analysis indicates that power consumption can be decreased by up to 13% when the symbol error rate is comparatively small.**

Index Terms*—IEEE 802.11; wireless multi-hop networks; power consumption; modulation method.

I. INTRODUCTION

Internet access via wireless networks has become very popular due to the rapid development of the wireless devices. These devices are mostly battery-driven, and wireless communication accounts for around 10% to 50% of their total power consumption [1-3]. Therefore, decreasing the power consumption in wireless communication is an important issue, especially when considering wireless multi-hop networks such as sensor networks and wireless mesh networks in which energy efficiency is essential. In this paper, we focus on the power consumption in wireless multi-hop networks based on IEEE 802.11 wireless LAN (WLAN), which is the most popular for implementing wireless multi-hop networks.

The IEEE 802.11 WLAN standard has multiple data rates that can be used, each of which has various characteristics such as modulation method, maximum transmission distance, and power consumption. Many rate adaptation algorithms have been proposed in the literature, such as automatic rate fallback (ARF) [4], receiver-based auto rate (RBAR) [5], and adaptive ARF (AARF) [6]. In ARF and AARF, each sender attempts to use a higher transmission rate after a fixed number of successful transmissions at a given rate and switches back to a lower rate after some consecutive failures. RBAR requires to change some MAC control frames and include a new header field. However, these algorithms are designed for maximizing the throughput of applications and they do not focus on energy efficiency. In addition, these existing algorithms do not consider multi-hop networks. On the other hand, the authors in [7-10] present the mathematical analysis on the power consumption in data transmission over WLAN. However, those analyses do not take multi-hop networks into account.

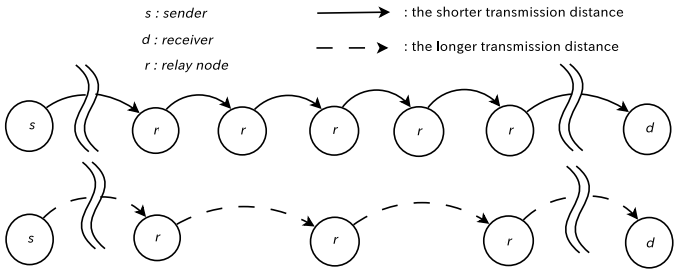


Fig. 1. Effect of transmission distance.

In wireless communication, in general, when a node lowers its transmission power, the transmission distance becomes shorter, resulting in a reduction in power consumption. However, when considering wireless multi-hop network, a shorter transmission distance may increase the total power consumption, since the shorter transmission distance requires greater node density and increases the hop count between a sender and a receiver, as shown in Figure 1. Using a higher data rate can decrease the air time of a packet, which may in turn decrease power consumption. However, a higher data rate generally has a shorter maximum transmission distance, and thus may increase the hop count for data transmission. Note that such an increase in hop count would lower energy efficiency, since the number of packet transmissions would rise.

In addition, some data rates in IEEE 802.11 WLAN employ different modulation methods, which may affect energy efficiency. In general, a modulation method used at a higher data rate can transmit more bits per transmitted symbol, but may result in a higher symbol error rate in a poor wireless

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

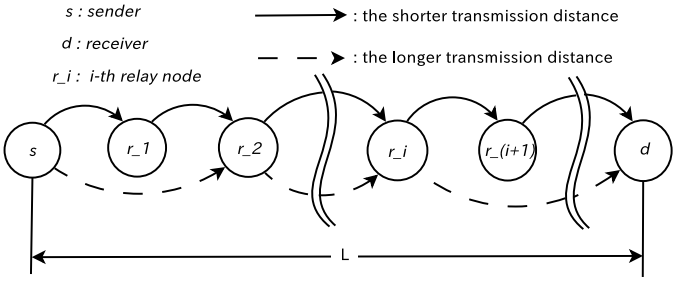


Fig. 2. Network model

environment due to noise and interference. A higher symbol error rate would then increase frame losses and retransmissions, lowering energy efficiency.

The complicated situations described above warrant special attention when considering the energy efficiency of wireless multi-hop networks. Therefore, in this paper, we present a mathematical analysis of power consumption in data transmission over IEEE 802.11-based wireless multi-hop networks. In particular, we consider the detailed behavior of the carrier sense multiple access with collision avoidance (CSMA/CA) method, and the complicated trade-off relationships described above. We show numerical examples of the analysis based on the specifications of an existing WLAN interface device and clarify the effect of data rate on energy efficiency in wireless multi-hop networks. In particular, we show that there are some situations where a low data rate should be intentionally selected in order to minimize power consumption.

The rest of this paper is organized as follows. In Section II, we describe our mathematical analysis of power consumption in data transmission over IEEE 802.11 wireless multi-hop networks. In Section III, we show numerical examples of the analysis and discuss the effect of data rate selection on energy efficiency. Finally, in Section IV, we give our conclusions and discuss directions of future research.

II. Mathematical Analysis

In the analysis, we assume a CSMA/CA MAC with RTS/CTS. There are multiple data rates that can be used in IEEE 802.11 WLAN standard, each of which is different in terms of transmission power, transmission distance, and modulation method. Therefore, the distance of one hop has a large impact on energy efficiency when data are transmitted over a multi-hop network.

Figure 2 shows the network model for the analysis. The network has a linear topology, where data are transmitted from a sender (node s) to a receiver (node d), which are separated by distance L. r_i $(i = 1, 2, ...)$ is a relay node located between the sender and receiver. For simplicity, we do not consider the effects of radio wave interference and overhearing on power consumption. The number of hops between the sender and receiver is determined when we choose the distance for one-hop transmission. In other words, when the one-hop transmission distance is D, the number of hops between the sender and receiver becomes $\lceil L/D \rceil$. This corresponds to the situation where we have an infinite number of relay nodes between the sender and receiver and we can select some of them to be used according to the transmission distance, as shown in Figure 2. Under this assumption we can explicitly evaluate the effect of data rate and its characteristics on the energy efficiency of multi-hop networks.

In Section II-A, we explain the detailed behavior of CSMA/CA in IEEE 802.11 WLAN. In Section II-B, we describe the power consumption of one-hop transmission of a single data frame between two relay nodes, and in Section II-C, we analyze multi-hop transmission of a total data whose size is S_{DATA} in Section II-C.

A. Frame exchange with CSMA/CA with RTS/CTS

Let us first look at one-hop data transmission. Figure 3 illustrates the frame exchange in the data transmission from r_i to r_{i+1} by CSMA/CA with request to send/clear to send (RTS/CTS). Figure 3(a) shows the case where no frame loss occurs, and Figure 3(b) shows the case where successive frame losses occur.

As shown in Figure 3(a), when a transmission demand occurs at r_i, it transmits an RTS command frame to r_{i+1} after a distributed coordination function (DCF) interframe space (DIFS) and a random backoff (BO_1). Then, r_{i+1} waits for a short interframe space (SIFS) and transmits a CTS command frame to r_i. When r_i receives the CTS command frame, it begins to transmit a data frame after an SIFS. After r_{i+1} finishes receiving the data frame, it transmits an acknowledgment (ACK) frame to r_i after an SIFS. When r_i receives the ACK frame from r_{i+1}, the transmission of one data frame is completed.

In Figure 3(b), when r_{i+1} fails to receive a data frame from r_i, it does not transmit an ACK frame to r_i. In this case, r_i waits a retransmission time out (RTO) and retransmits the data frame after a DIFS and a random backoff (BO_2). This cycle continues until the data frame successfully reaches r_{i+1} . Note that BO_j in Figure 3(b) is the random backoff for the $(j-1)$ th retransmission of the data frame. In the analysis, we assume that RTS, CTS, and ACK frames are never lost in the entire data transmission process.

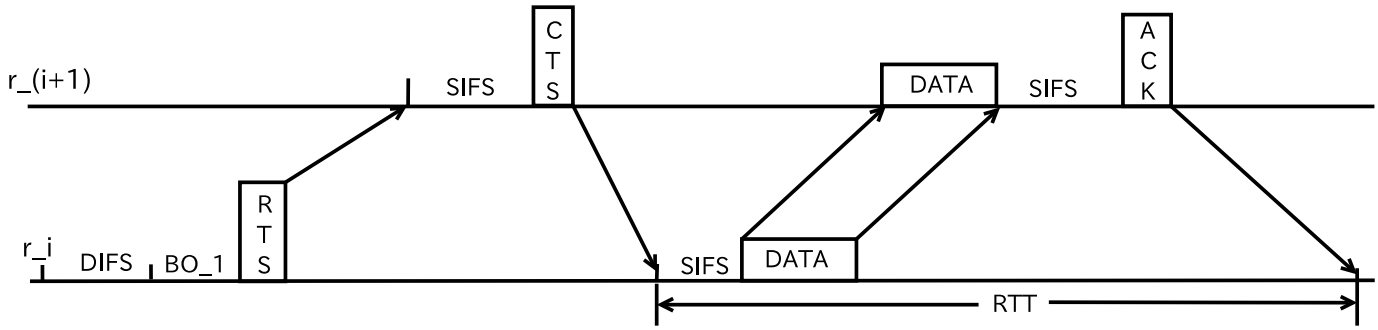
B. Power consumption in one-hop data transmission

Based on the behavior shown in Section II-A, we calculate the power consumption in one-hop transmission of a single data frame. We denote the total size of the data to be transmitted as S_{DATA} and the size of one data frame as S_f. Then, the number of data frames to be transmitted, n_f, is calculated as

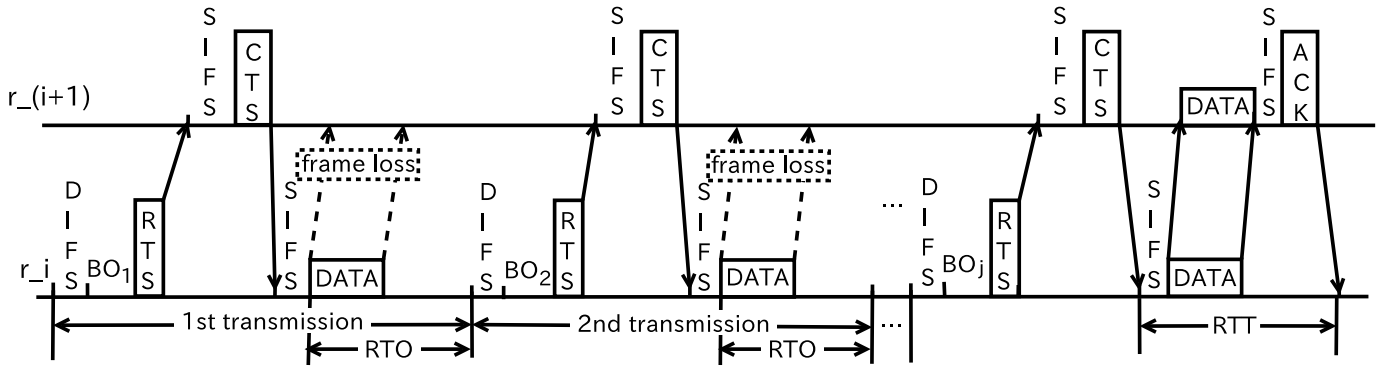
$$n_f = \left\lceil \frac{S_{DATA}}{S_f} \right\rceil . \quad (1)$$

The modulation method used for a data rate is defined to modulate l bit(s) per transmitted symbol. Then, the number

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1



(a) Case of no frame loss



(b) Case of successive flame losses

Fig. 3. Frame exchange based on IEEE 802.11 with RTS/CTS

of symbols in a data frame is

$$n_s = \left\lceil \frac{S_f}{l} \right\rceil. \quad (2)$$

The value of l is different in each modulation method. For example, $l = 1$ in binary phase shift keying (BPSK) and $l = 2$ in quadrature phase shift keying (QPSK). Note that a single symbol error corresponds to multiple bit errors when $l > 1$. In the numerical evaluations in Section III, we treat the symbol error rate as the probability that a transmitted symbol is not received successfully. We then assume that a loss of a data frame occurs due to one or more bit errors when the frame is transmitted from r_i to r_{i+1}. By denoting the symbol loss rate as p_s, the probability with which a data frame fails to be transmitted successfully, is

$$p_f = 1 - (1 - p_s)^{n_s}. \quad (3)$$

The average number of transmissions until r_{i+1} successfully receives the data frame, denoted by e, is given by

$$e = \lim_{x \to \infty} \sum_{i=1}^{x} i p_f^{i-1}(1 - p_f). \quad (4)$$

This Equation (4) is solved as follows:

$$\begin{aligned}
(1-p_f)e &= \lim_{x\to\infty}\left(\sum_{i=1}^{x} i p_f^{i-1}(1-p_f) - \sum_{i=1}^{x} i p_f^{i}(1-p_f)\right), \\
e &= \lim_{x\to\infty}\left(\sum_{i=1}^{x} i p_f^{i-1} - \sum_{i=2}^{x+1}(i-1)p_f^{i-1}\right), \\
&= \lim_{x\to\infty}\left\{\left(1 + \sum_{i=2}^{x} i p_f^{i-1}\right)\right. \\
&\quad \left. - \left(\sum_{i=2}^{x} i p_f^{i} + (x+1)p_f^{x} - \sum_{i=2}^{x+1} p_f^{i-1}\right)\right\}, \\
&= \lim_{x\to\infty}\left(1 + \sum_{i=2}^{x} p_f^{i-1} + p_f - (x+1)p_f^{x}\right), \\
&= \lim_{x\to\infty}\left(\sum_{i=1}^{x} p_f^{i-1} - x p_f^{x}\right), \\
&= \frac{1}{1-p_f}. \quad (5)
\end{aligned}$$

Next, we examine the backoff time (BO_j in Figure 3(b)). The backoff time is a waiting period before a data frame is transmitted to prevent collisions between multiple transmitting nodes. The length of the backoff time is determined at random

within the range $[0, CW]$ multiplied by slot time, denoted by T_{slot}. The value of CW varies according to the number of successive retransmissions. The value of CW for the backoff time at the j th retransmission, CW_j, is calculated as

$$CW_j = \min\left(2^{j-1}CW_{min}, CW_{max}\right), \quad (1 \leq j). \tag{6}$$

Then the length of backoff time on j th retransmission is obtained as $CW_j \cdot T_{slot}$. For the sake of simplicity, we assume that CW_{max} is given as follows:

$$CW_{max} = 2^m CW_{min}. \tag{7}$$

where m in an integer value. Then, we can compute the average value of the sum of the backoff times for the transmission of one data frame, T_{BO}, from Equations (5) and (7):

$$\begin{aligned} T_{BO} &= \lim_{x\to\infty}\left(\sum_{j=1}^{m}\left\{\frac{2^{j-1}CW_{min}T_{slot}}{2}\cdot p_f^{j-1}(1-p_f)\right\}\right. \\ &\quad \left. + \sum_{j=m+1}^{x}\left\{\frac{CW_{max}T_{slot}}{2}\cdot p_f^{j-1}(1-p_f)\right\}\right). \end{aligned} \tag{8}$$

In what follows, the first term in Equation (8) is denoted as Q_1 and the second term is denoted as Q_2. These terms are calculated as

$$\begin{aligned} Q_1 &= \lim_{x\to\infty}\frac{1}{2}\sum_{j=1}^{m}\left\{(2p_f)^{j-1}CW_{min}T_{slot}(1-p_f)\right\}, \\ &= \frac{1}{2}\sum_{j=1}^{m}\left\{(2p_f)^{j-1}CW_{min}T_{slot}(1-p_f)\right\}, \\ &= \frac{CW_{min}T_{slot}(1-p_f)}{2}\sum_{j=1}^{m}(2p_f)^{j-1}, \\ &= \begin{cases} \frac{CW_{min}T_{slot}(1-p_f)}{2}\cdot\frac{(2p_f)^m-1}{2p_f-1} & (\frac{1}{2} < p_f < 1) \\ \frac{mCW_{min}T_{slot}(1-p_f)}{4} & (p_f = \frac{1}{2}) \\ \frac{CW_{min}T_{slot}(1-p_f)}{2}\cdot\frac{1-(2p_f)^m}{1-2p_f} & (0 < p_f < \frac{1}{2}), \end{cases} \end{aligned} \tag{9}$$

$$\begin{aligned} Q_2 &= 2^{m-1}CW_{min}T_{slot}(1-p_f)\sum_{j=m+1}^{x}p_f^{j-1}, \\ &= 2^{m-1}CW_{min}T_{slot}(1-p_f)\left(\sum_{j=1}^{x}p_f^{j-1} - \sum_{j=1}^{m}p_f^{i-1}\right), \\ &= 2^{m-1}CW_{min}T_{slot}(1-p_f)\left(\frac{1-p_f^x}{1-p_f} - \frac{1-p_f^m}{1-p_f}\right), \\ &= 2^{m-1}CW_{min}T_{slot}(p_f^m - p_f^x), \\ &\to 2^{m-1}p_f^m CW_{min}T_{slot} \quad (x\to\infty). \end{aligned} \tag{10}$$

Consequently, from Equations (8)- (10), T_{BO} is given by

$$T_{BO} = Q_1 + 2^{m-1}p_f^m CW_{min}T_{slot}. \tag{11}$$

We now calculate the power consumed for the transmission of a data frame. We denote the period for bit transmission, bit reception, and the idle time of the sender (node r_i in Figure 3) as T^s_{send}, T^s_{recv}, and T^s_{idle}, respectively. Similarly, for the receiver (node r_{i+1} in Figure 3), we use the variables T^r_{send}, T^r_{recv}, and T^r_{idle}. In reference to Figure 3, these variables can be used to formulate the following equations:

$$\begin{aligned} T^s_{send} &= T^r_{recv}, \\ &= \frac{1}{d^{(k)}(1-p_f)}(S_{RTS}+S_{DATA}+S_{head}), \end{aligned} \tag{12}$$

$$\begin{aligned} T^s_{recv} &= T^r_{send}, \\ &= \frac{1}{d^{(k)}}\left(\frac{S_{CTS}}{1-p_f}+S_{ACK}\right), \end{aligned} \tag{13}$$

$$\begin{aligned} T^s_{idle} &= T^r_{idle}, \\ &= \frac{1}{1-p_f}\{T_{DIFS}+(1-p_f)T_{BO} \\ &\quad +(3-2p_f)T_{SIFS} \\ &\quad +p_f\left(T_{RTO}-\frac{S_{DATA}+S_{head}}{d^{(k)}}\right)\}. \end{aligned} \tag{14}$$

T_{SIFS} and T_{DIFS} are respectively an SIFS and a DIFS. T_{RTO} is the time of RTO. S_{RTS}, S_{CTS}, and S_{ACK} are respectively the size of an RTS frame, a CTS frame, and an ACK frame. S_{head} is the sum of the physical layer convergence protocol (PLCP) preamble and the PLCP header added at the physical layer. $d^{(k)}$ is the data rate to be used. We assume that the number of available data rates in IEEE 802.11 WLAN is K. The power consumption in one-hop transmission with the data rate of $d^{(k)}$ is

$$\begin{aligned} E_1^{(k)} &= P_i\times(T^s_{idle}+T^r_{idle}) + P_t\times(T^s_{send}+T^r_{send}) \\ &\quad +P_r\times(T^s_{recv}+T^r_{recv}), \\ &= 2P_s\times T^s_{idle} + (P_t+P_r)(T^s_{send}+T^s_{recv}). \end{aligned} \tag{15}$$

P_t and P_r are the power needed in bit transmission and reception per unit time, respectively. P_i is the power consumed in the idle period.

C. Power consumption in multi-hop data transmission

We now calculate the power consumed in the entire data transmission process over the multi-hop network depicted in Figure 2. The transmission power and transmission distance of the k th data rate are denoted as $P_t^{(k)}$ and $r^{(k)}$, respectively. We also introduce the maximum transmission power and the maximum transmission distance at the k th data rate, denoted by $\hat{P}_t^{(k)}$ and $\hat{r}^{(k)}$, respectively. We assume that when a data frame is transmitted at less than the maximum power, the relation between the transmission power and transmission distance is expressed as

$$P_t^{(k)} = \hat{P}_t^{(k)}\cdot\left(\frac{r^{(k)}}{\hat{r}^{(k)}}\right)^{\alpha}, \tag{16}$$

where α is the parameter that describes the attenuation [11, 12]. The above equation can be transformed for $r^{(k)}$ as follows.

$$r^{(k)} = \hat{r}^{(k)}\cdot\left(\frac{P_t^{(k)}}{\hat{P}_t^{(k)}}\right)^{\frac{1}{\alpha}}. \tag{17}$$

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

TABLE I
PARAMETER SETTINGS

item	size
S_{ACK}	40 [bytes]
S_{RTS}	40 [bytes]
S_{CTS}	40 [bytes]
S_f	1000 [bytes]
S_{header}	24 [bytes]

(a) Frame size and physical-layer overhead

item	length
T_{DIFS}	34 [μs]
T_{SIFS}	16 [μs]
T_{slot}	9 [μs]
T_{RTO}	5 RTT

(b) IEEE 802.11 parameters

TABLE II
TRANSMISSION DISTANCE AND POWER OF CISCO AIRONET IEEE 802.11/A/B/G WIRELESS CARDBUS ADAPTER

data rate [Mbps]	1	6	11	18	54
maximum transmission distance [m]	610	396	304	183	76
maximum transmission power [mW]	100	100	100	50	20

Since the distance between the sender and receiver is L, the hop count is given by

$$h^{(k)} = \left\lceil \frac{L}{r^{(k)}} \right\rceil. \quad (18)$$

Finally, the power consumption for the transmission of S_{DATA} data over the multi-hop network $E_M^{(k)}$ is given as follows:

$$E_M^{(k)} = n_f \cdot E_1^{(k)} \cdot h^{(k)}. \quad (19)$$

III. NUMERICAL EVALUATION

A. Parameter settings

We set the distance L between the sender and receiver to 1000 [m] and the total data size S_{DATA} is set to 100 [Kbytes]. As the parameters for determining the backoff time in Equation (11), we set $m = 10$ and $CW_{min} = 15$. The parameter α in Equation (17) is set to 2. Frame sizes and physical-layer overhead are listed in Table II(a), and IEEE 802.11 parameters are shown in Table II(b). T_{RTO} is set at five times the round trip time (RTT), according to the implementation of FreeBSD [13]. RTT, which is shown in Figure 3, is calculated based on frame size and data rate by ignoring the propagation delay between relay nodes. We utilize the specifications shown in Table II for a Cisco Aironet IEEE 802.11a/b/g Wireless CardBus adapter [14] for the maximum transmission distance and corresponding power of each data rate.

B. Numerical results and discussions

Figure 4 shows the power consumption for various data rates as a function of symbol error rate when we set the transmission power to 20 [mW]. Here, we assume that the symbol error rate remains unchanged when we change the data rate. We can see from this figure that power consumption can be decreased simply by using a higher data rate. This is because the main contribution to reducing power consumption is from the decreased air time of a packet.

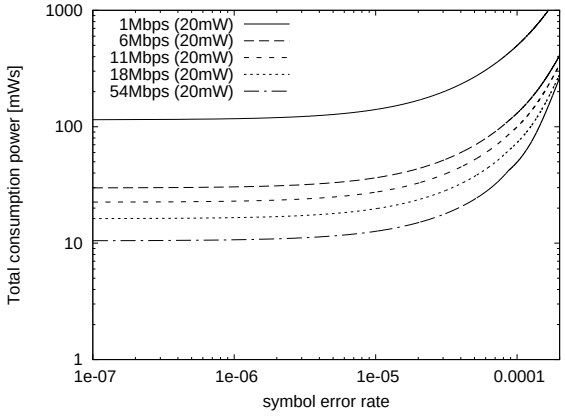


Fig. 4. Power consumption at various data rates

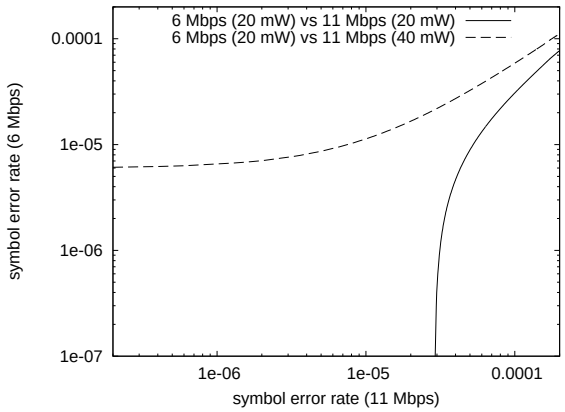


Fig. 5. Symbol error rates at two data rates that give equal power consumption

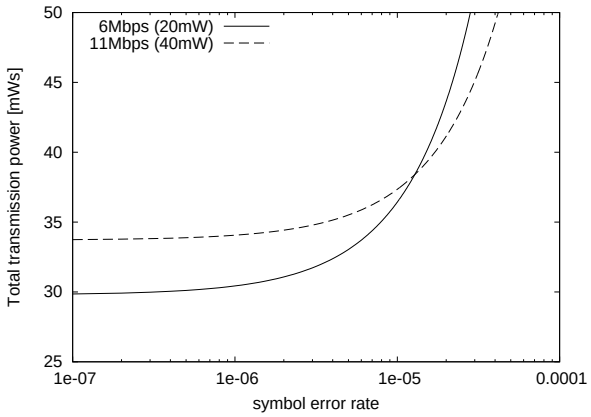


Fig. 6. Power consumption comparison between 6 [Mbps] (BPSK) and 11 [Mbps] (QPSK)

When we consider the transmission power configuration, the situation changes notably. In Figure 5, we plot the relationships between symbol error rates when we use a 6 [Mbps] data rate and when we use a 11 [Mbps] data rate, for certain transmission powers where the power consumption at the two data rates become equal. For example, observing the plot for

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

6 [Mbps] (20 [mW]) versus 11 [Mbps] (40 [mW]), using 11 [Mbps] has smaller power consumption in the upper-left region of the plot and 6 [Mbps] has an advantage in the lower-right region of the plot.

One possible way to decrease the symbol error rate at a higher data rate is to increase the transmission power. By comparing the two curves in the graph, we can observe that when the transmission power is increased at the 11 [Mbps] data rate from 20 [mW] to 40 [mW], the region where the 11 [Mbps] data rate has smaller power consumption decreases considerably. This means that we should carefully choose the data rate and transmission power according to the symbol error rate which is observed during the data transmission.

Finally, we show another example where we utilize a static relation between two modulation methods. Here we assume that QPSK consumes twice as much power as BPSK in order to obtain a given symbol error rate [15]. In other words, when we decrease the symbol error rate at higher data rates we should significantly increase the transmission power. With consideration these characteristics, BPSK at 6 [Mbps] and 20 [mW] transmission power is compared with QPSK at 11 [Mbps] and 40 [mW] transmission power in Figure 6. Here, when the symbol error rate is approximately 10^{-6} or lower, the lower data rate gives smaller power consumption, which is an opposite result to that in Figure 4. The reduction in power consumption reaches about 13% when the 6 [Mbps] data rate is chosen. This means that increasing transmission power to decrease the symbol error rate results in increased power consumption for total data transmission, although we can expect a longer transmission distance and smaller hop count with larger transmission power.

From the above results, we conclude that we should consider various factors that affect the power consumption of data transmission over wireless multi-hop networks.

IV. Conclusion

In this paper, we presented a mathematical analysis of power consumption in data transmission over IEEE 802.11-based wireless multi-hop networks to investigate the effects of data rate selection on energy efficiency. The analysis revealed that power consumption can be decreased by up to 13% when the symbol error rate is comparatively small.

For future work, we plan to consider other modulation methods such as quadrature amplitude modulation to provide further insight into energy efficiency at high data rates. Another plan for future research is to enhance the accuracy of the analysis, by including the effects of losses of ACK, RTS, and CTS frames, data frame collision, interference, and overhearing.

Acknowledgment

This work was supported in part by the Ministry of Internal Affairs and Communications (MIC), Japan, under the Promotion program for Reducing global Environmental loaD through ICT innovation (PREDICT).

References

[1] A. Communications, "Power Consumption and Energy Efficiency Comparisons of WLAN Products," tech. rep., Atheros, May 2003.

[2] V. Raghunathan, T. Pering, R. Want, A. Nguyen, and P. Jensen, "Experience with a low power wireless mobile computing platform," in *Proceedings of ISLPED 2004*, pp. 363–368, Aug 2004.

[3] Y. Agarwal, "Dynamic power management using on demand paging for networked embedded systems," in *Proceedings of the ASP-DAC 2005*, pp. 755–759, Jul 2005.

[4] A. Kamerman and L. Monteban, "WaveLAN-II: a high-performance wireless LAN for the unlicensed band," *Bell Labs Technical Journal*, vol. 2, pp. 118–133, summer 1997.

[5] G. Holland, N. Vaidya, and P. Bahl, "A rate-adaptive MAC protocol for multi-hop wireless networks," in *Proceedings of the 7th annual international conference on Mobile computing and networking*, pp. 236–250, Jul 2001.

[6] M. Lacage, M. H. Manshaei, and T. Turletti, "IEEE 802.11 rate adaptation: A practical approach," in *Proceedings of MSWiM 2004*, pp. 126–134, Oct 2004.

[7] G. Kuriakose, S. Harsha, A. Kumar, and V. Sharma, "Analytical models for capacity estimation of IEEE 802.11 WLANs using DCF for internet applications," *Wireless Networks*, vol. 15, pp. 259–277, Feb 2009.

[8] V. Baiamonte and C.-F. Chiasserini, "Saving energy during channel contention in 802.11 WLANs," *Mobile Networks*, vol. 11, pp. 287–296, Apr 2006.

[9] M. Ergen and P. Varaiya, "Decomposition of energy consumption in IEEE 802.11," in *Proceedings of ICC 2007*, pp. 403–408, Jun 2007.

[10] M. Hashimoto, G. Hasegawa, and M. Murata, "Modeling and analysis of power consumption in TCP data transmission over a wireless LAN environment," *the IEICE techreport*, vol. 110, pp. 1–6, Dec 2010.

[11] W. R. Heizelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proceedings of HICSS 2000*, pp. 3005–3014, Jan 2000.

[12] R. Bhatia and M. Kodialam, "On power efficient communication over multi-hop wireless networks: joint routing, scheduling and power control," in *Proceedings of INFOCOM 2004*, pp. 1457–1466, Mar 2004.

[13] Gary R. Wright and W. Richard Stevens, *TCP/IP Illustrated, Vol.2:The Implementation*. Addison-Wesley Professional, 1998.

[14] Cisco Aironet 802.11a/b/g wireless CardBus adapter. available at http://www.cisco.com/web/JP/product/hs/wireless/adapter/prodlit/cecbs_ds.html.

[15] Ishii Satoshi, *The wireless communication and the digital modulation and demodulation technique*. CQ publisher, 2005.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

Improving Quality of Service in Wireless Multimedia Communication with Smooth TCP

Michael Bauer
Department of Computer Science
The University of Western Ontario
London, Ontario, Canada, N6A5B7
Email: bauer@csd.uwo.ca

Md. Ashfaqul Islam
Department of Computer Science
The University of Western Ontario
London, Ontario, Canada, N6A5B7
Email: mislam59@uwo.ca

***Abstract*—Wireless multimedia communications are becoming mainstays of many applications and these applications are likely to continue to grow and become more demanding. TCP, though not designed for this kind of communication, is still commonly used. Smooth TCP (STCP) has been introduced in previous research as a variant of TCP with properties that can enable it to work better in environments in which TCP was not originally designed, such as wireless multimedia communications. While STCP has been compared to TCP is some situations, it has not been compared to TCP in wireless multimedia environments. In this paper, we briefly describe STCP and report on initial experiments that compare TCP and STCP through simulation. The the results suggest that STCP can provide better support than TCP for wireless multimedia communications.**

Keywords-wireless communication, multimedia, TCP, performance.

I. Introduction

The growth over the past decade in the computational power and wireless capabilities of consumer handheld devices has ushered in an era of new and exciting mobile applications and services. Many of these applications and services have created a demand for multi-modal, dynamic data. This data is delivered using TCP or UDP over the Internet, neither of which was designed with multi-modal, time sensitive data in mind. In particular, TCP was not designed to operate with time-sensitive data nor in wireless environments, though it remains a core protocol in use for wireless data traffic. A substantial challenge for TCP is when multi-modal traffic must be delivered to an application over a wireless connection.

Unlike general data communication, multimedia communication and time sensitive streaming applications rely on timely data delivery and are somewhat less concerned about the guaranteed end-to-end data delivery. In this regard, TCP is not suitable as because it chooses reliability over timeliness. A number of researchers have looked at ways to enhance TCP's timeliness properties. For example, some researchers [1] have proposed new protocols which ensures a minimum rate for multimedia traffic; this is, however, not feasible for wireless environments where bandwidth availability is a great concern.

In this paper, we examine the performance of Smooth TCP (STCP). STCP, introduced and studied in previous research [2] [3] [4]; is a variation of TCP which behaves much like TCP in general data transfer. STCP differs, however, in that it is first of all parameterized and with a suitable set of parameters can be set to have properties more suitable for communications in environments for which TCP was not originally designed. The performance of STCP has been compared to TCP in several scenarios, but there has been no comparison to TCP in wireless, multimedia communication. In this paper, we describe a parameterized version of STCP for wireless multimedia environments and compare it to TCP.

The paper is organized as follows. The following section reviews some previous work on addressing modifications to TCP to handle multi-modal data and to operate in wireless environments. In Section III, we briefly introduce STCP, describe its properties and introduce a parameterized version for wireless multimedia environments. Section IV outlines the simulation approach and tools used. Section V presents experiments and results. Finally, we draw conclusions on the potential use of STCP and outline some future work.

II. Related Work

Researchers exploring variations in TCP have typically looked to address either the use of TCP in wireless environments or its use with multi-modal data, but not both. We briefly look at some of the previous approaches in addressing these challenges.

Some of the challenges with TCP in wireless multimedia communication environments have been identified by previous research [5]. The wireless medium is very susceptible to path loss or link failure, is based on shared bandwidth, requires hand-off of mobile devices between access points, etc. TCP was developed focusing on wired networks and was not designed to support Quality of Service in an unreliable wireless medium. Previous research has identified that unmodified standard TCP performs poorly in a wireless environment, as TCP can not distinguish packet losses caused by network congestion from those attributed to transmission errors. TCP misinterprets this loss as congestion and invokes congestion control. This leads to unnecessary retransmissions and loss of throughput [6]. The congestion control mechanism of TCP reacts adversely to packet losses due to temporarily broken routes in wireless networks [7], [8].

TCP has been designed to ensure that data arrives at its destination regardless of timing dependencies and this can

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

be challenging for TCP for the transmission of time-sensitive data [9]. To address the transfer of multimedia, a number of alternative strategies have proposed changes to TCP. SCTP (Stream Control Transmission Protocol) is a relatively new transport layer protocol which aims to transport telecommunication signaling messages over an IP based network [10]. SCTP can be said to have a blend of features of TCP and UDP, for example, it inherits TCP's congestion control scheme and connection oriented communication [11]. SCTP also offers two other distinct features: multi-homing and multi-streaming [2] [10] [11]. A simulation study of the performance [12] showed TCP outperformed SCTP in some cases because of extra overhead present in SCTP.

Other approaches have looked to provide protocols at the application layer in order to compensate for TCP's deficiencies as well as to provide other time-based services. The Real-time Transport Protocol (RTP) is used extensively in communication and entertainment systems that involve streaming media such as audio and/or video tele-conference, Internet telephony, Internet TV etc. [13]. It provides identification and sequential orderings of data bits. It can also monitor the delivery of multimedia content. RTP does not address resource reservation and does not guarantee quality-of-service for real-time services. RTP provides data delivery monitoring services and stream control with the help of another supporting protocol - the Real-time Transport Control Protocol (RTCP) [13]. At the level of functionality, RTP and RTCP use two consecutive ports and carry data and control information side by side. This facilitates the option of using pause and play in audio /video streams.

RTSP (Real TIme Streaming Protocol) [14] functions similarly to HTTP but differs from it in that it requires a permanent connection. It uses a message identifier to monitor each data connection. The protocol is used to establish and control media sessions between end points. RTSP works with sessions rather than connections with the server. There is no notion of an RTSP connection; instead, a server maintains a session labeled by an identifier. While using RTSP as an application level protocol, the client can open and close several connections to the server and can use RTSP requests. RTSP also uses the underlying (transport) layer.

Alternatives to TCP have focused on the application layer or involve substantive changes to TCP. Moreover, most of the research efforts have focused on how to deal with multimedia applications over TCP and do address the challenges that TCP has in wireless environments. In practice, both need to be addressed.

III. SMOOTHTCP (STCP)

SmoothTCP (STCP) [2], [3], [4] is a very recent addition to the collection of TCP variations. Vieira [3] introduced STCP and compared it to standard TCP through various experiments. The results demonstrated that STCP can outperform TCP in some circumstances and is as good otherwise. One of the main advantages os STCP is that it works very naturally in a "normal" TCP environment.

STCP mainly differs from standard TCP in the way it works with the congestion control mechanism. STCP introduces the notion of *smoothness* within congestion control [4]. In order to create a smooth congestion control mechanism, additional parameters are used in conjunction with the basic parameters of TCP in order to smooth out the way congestion is managed.

Four algorithms [15] define TCP's congestion control scheme: 1) the *slow start* phase ensures a moderate increase in the size of sending window (named **cwnd**); 2) the *congestion avoidance* phase starts when **cwnd** exceeds the threshold and continues unless a congestion is detected; 3) the *fast retransmit* phase; and the *fast recovery* **cwnd** and the threshold are both halved. STCP differs from TCP in its approach to congestion control in that it introduces the notion of *smoothness* within congestion control. Though it uses the same slow start algorithm, it manages the size of **cwnd** differently. The change in the window size is done more *smoothly* and results in fewer frequent radical changes in the value of **cwnd**. It introduces smoothness to the congestion control of TCP. There are three key properties required for *smoothness*.

1. Smooth Curve Property: TCP's congestion avoidance scheme uses the Additive Increase Multiplicative Decrease (AIMD) algorithm [15]. The AIMD algorithm is generalized by the family of Binomial Congestion Control (BCC) Algorithms [16]. As per the definition of BCC, the functions belonging to this family introduce discontinuities in the graph of the congestion window which can affect the performance of TCP in certain circumstances. To achieve smoothness, the graph needs to be smoothed [3].

2. Vertical Smoothness Property: TCP uses retransmission timeouts (RTOs) and duplicate acknowledgements (dupACKs) to detect packet drops or congestion. However, there might be other reasons for a temporary timeout, e.g., a packet may have traveled on a redundant path or there might be a sudden increase in link latency in wireless network. Both could cause a timeout which could make TCP believe that there is congestion. These events do not always mean congestion or packet drop, yet TCP might take action to control the congestion, which would be unnecessary. These types of spurious events are called *Vertical Bursts* [3] and cause spikes in the average time for acknowledgements. To avoid TCP reacting too quickly, the occurrence of the "vertical bursts" needs to be smoothed.

3. Proactive Control Property: We know that TCP detects a congestion or a packet drop, but it takes almost 1 RTT plus 3 dupACK arrival times to inform the sender that a congestion has occurred. This is the time that TCP waits before taking any action against congestion. This relatively long time is known as a *temporal gap* and within this gap a sender may send other packets. The proactive control property of STCP reduces the temporal gap [3]. Three different methods were considered for reducing the temporal gap using different metrics: Variation in Round Trip Time (RTTVar), Number of Retransmission Time Outs (RTOs), Number of Retransmitted Packets. Our work here focuses on retransmission time outs.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

A. The STCP Function

SmoothTCP is defined as a subset of functions of the Smooth Congestion Control Algorithms [3]. The general form of the STCP function is defined as follows:

$$D = \frac{\delta w}{\delta p_1} dp_1 + ... + \frac{\delta w}{\delta p_i} dp_i + ... + \frac{\delta w}{\delta p_n} dp_n \quad (1)$$

Here, p_1, $p_2 ... p_n$ are the set of *state variables* which describe the network's status. For example, the variation in Round Trip Time (RTTVar), the number of Retransmission Time Outs (RTO), or the number of retransmitted packets are some of the variables that could be used in equation 1. D is the change in the congestion window **cwnd**. A positive value indicates an increment of D to the current **cwnd**, while a negative value indicates a decrease. The basic idea is that the change in the congestion window, namely D, is determined by changes in the state variables, hence the partial derivatives. In turn, the change in each parameter is modeled using a sigmoid function $f(u)$ to accommodate bursts, spikes and to adjust the window size to adapt the change; the proposed function is: $f(u) = A + C * tanh(B(u + M))$ [3]. Here, u is one of the state variables and A, B, C and M are coefficients providing weighting of each such state variable. Thus, each one of the partial derivatives are of the form:

$$f(u) = \frac{\delta w}{\delta p_i} = A_{p_i} + C_{p_i} tanh(B_{p_i}(p_i + M_{p_i})) \quad (2)$$

A is the lower asymptote and it determines the smallest value of $f(u)$. B controls the growth rate of $f(u)$, C is the width of the range set of $f(u)$ which determines the maximum amount of variation in the window size, and M can be used for controlling the position of the curve on the u-axis.

The coefficients A, B, C and M can be determined using environmental conditions and could be set as constants, set by an application (e.g. by a multimedia server prior to sending content) or even set dynamically based on operating conditions. For this study, we have chosen to use three *state variables*: Round Trip Time (*rtt*), the number of Retransmission Time-Outs (*rto*) and the number of retransmitted packets (*resend_pack* or *rsnd*) as the variables (p_i) of equation 1. These variables were initially selected since they are readily available within most implementations of TCP; a brief explanation of these variables and the coefficient values selected follows.

rtt: *rtt* refers to the time difference between the time a packet is sent from the sender and the time sender receives the associated acknowledgement of that packet. TCP keeps track of round trip time and uses it to detect any possible congestion. In a steady network, the round trip time remains steady in terms of variation. The larger the difference between a measured round trip time and the average as measured by TCP, the higher the possibilities of congestion in the network. Round trip time estimation was proposed by Van Jacobson in [17] and is used in our simulation. The form of the equation 2 for *rtt* is defined as:

$$\frac{\delta w}{\delta rtt} = A_{rtt} + C_{rtt} tanh(B_{rtt}(rtt + M_{rtt})) \quad (3)$$

$$\frac{\delta w}{\delta rtt} = 1452 \times tanh(-(rtt - 1)) \quad (4)$$

Thus, the values for these coefficient are:

$A_{rtt} = 0$
$C_{rtt} = maximumsegmentsize(mss) = 1452$
$B_{rtt} = -1$
$M_{rtt} = -1$

A_{rtt} is set to 0 to get a smallest value from overall calculation. C_{rtt}, being the width of the range set of the function, is set to the Maximum Segment Size (mss). It is logical to have a full mss as the multiplicative factor to the $tanh$ function so that in the best case it will achieve a complete mss increase. In other cases, mss will contribute to the outcome of $tanh$ function. Having B_{rtt} and M_{rtt} equal to -1 ensures that the *rtt* value will contribute towards window size increment while it has a value less than 1. When it reaches at 1, it will contribute 0 and afterward, it will contribute negatively. Previous research done to identify variability in TCP's round trip time [4] has found out that 90% of all the round trip time samples lie between $0.1s$ to $1s$. This is the reason we choose M_{rtt} equal to -1. B_{rtt} is used to make the result of $(rtt + M_{rtt})$ a negative value when rtt exceeds the value of 1 (second).

rto: TCP maintains a timer to trigger any retransmission of packets. Whenever the timer expires, TCP retransmits the packet from the top of retransmission queue. The *rto* value depends on *rtt*. Initially the timer is set to a low value which is closer to the average round trip time. Setting the time to a very low value would cause unnecessary retransmission. Whenever there is a retransmission time out, TCP resends the packet from the queue and at the same time the value of *rto* is increased. Karn's [18] algorithm suggests a doubling of RTO each time the retransmission timer expires. With a higher value of *rto*, it is understandable that the TCP sending rate should be decreased as keeping the same sending window size which would create further congestion. The coefficients and the form of equation 2 used with *rto* value is the same as the equation for *rtt*, namely:

$$\frac{\delta w}{\delta rto} = A_{rto} + C_{rto} tanh(B_{rto}(rto + M_{rto})) \quad (5)$$

$$\frac{\delta w}{\delta rto} = 1452 \times tanh(-(rto - 1)) \quad (6)$$

$$(7)$$

A_{rto} is set to 0 following the same reason as A_{rtt} and the same explanation goes for setting C_{rto} equal to snd_mss. M_{rto} is set so that it can contribute positively whenever the *rto* value is less than 1 and otherwise, whenever the *rto* value exceeds 1, $tanh$ will return a negative value which in turn will

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

decrease the sending window size by contributing a negative value.

rsnd: The number of retransmitted packets plays an important role in TCP's congestion control mechanism. There is a maximum value set for this parameter in each TCP connection, where exceeding that value will cause the connection to terminate; we use the default value (12). Whenever there is a retransmission, the connection needs to be slowed to avoid further retransmission. The equation form for this parameter is defined as:

$$\frac{\delta w}{\delta rsnd} = A_{rsnd} + C_{rsnd} tanh(B_{rsnd}(rsnd + M_{rsnd})) \quad (8)$$

$$\frac{\delta w}{\delta rto} = 1452 \times tanh(-(rsnd - 0.5)) \quad (9)$$

The coefficients of this function are set similarly to the previous variables. $A_{rsnd} and C_{rsnd}$ are set to 0 and mss as before. Here, M_{rsnd} is set to -0.5 to have a minimum effect when there is no retransmission. But whenever there is a retransmission, it will start decreasing the window size immediately. The higher the number of retransmissions the greater the decrement will be. This ensures a lower sending rate to try to bring the connection into a steady mode.

Thus, the new value for *cwnd* is:

$$cwnd + \frac{\delta w}{\delta rtt} + \frac{\delta w}{\delta rto} + \frac{\delta w}{\delta rsnd} \quad (10)$$

There is a scope for fine tuning the coefficients and selecting appropriate parameters to get the maximum efficiency from this algorithm, but we do not explore that in this paper. The parameterized characteristics of STCP makes it more flexible and potentially more efficient in handling different types of scenarios, such as scenarios involving multimedia contents. The nature of the traffic and the condition of the network can be used to tune the coefficients. Events related to wireless environment congestion can be managed by tuning parameters which are solely related to wireless environment. This aspect of STCP means that it has the potential of improving the overall performance of TCP by adapting to network types and traffic contents. In particular, we are interested in understanding how the parameters can be selected and what are the best ways to choose the coefficients' initial values and how to tune them further.

IV. Simulation Approach

Our specific interest is in understanding STCP in wireless networks and, in particular, traffic involving multimedia contents. As this is a comparative study of two different protocols (standard TCP and SmoothTCP), the research requires a platform where protocols can be implemented and observed in different scenarios. We use an available network simulator, OMNeT++ [19], [20]. OMNeT++ has been used for a variety of network research [21], [22], [23]. It also has available a well developed wireless library, INET [19], [20]. For our simulation, we needed to add several components and elements to OMNet++.

Multimedia Content: "Multimedia contents" will mean a bit stream and entail a continuous data transmission from server to client in response to a request from the client. Our "media files" were of specific sizes for the experiments and we only make use of dummy packets, not actual multimedia data, since we are only interested in the traffic delivery under differing network conditions. We do not need the client to process the data, but only to acknowledge the arrival, so it is enough to work with dummy data packets.

Multimedia Client Application: We use a client application which requests multimedia data from the media server (discussed below). It will only acknowledge the receipt of data and will not process it. From a technical point of view, a TCP client of the OMNeT++ simulator always issues a CLOSE command after receiving the response it requested. The client application in our case was modified so that it does not issue a CLOSE command unless the whole stream has been transmitted from server to client. After receiving the notification from server side about the end of stream, client application sends CLOSE command to close the connection.

Multimedia Server Application: An application on the server which receives requests and transmit the data. It stops only when it receives a CLOSE connection request from the client. Our server will use TCP and STCP, respectively.

The Media Server: A model of a server connected to our network that runs the Multimedia Server Application.

Wireless environment: A wireless environment will have wireless nodes and access points; it also has the unique features of a wireless network, including shared bandwidth, use of radio channels, variable transmitter power, etc. Successive experiments with different set of parameters and prior research [24] helped us determine the more significant parameters to study: the variables in the wireless environment we adjusted were the radio transmitter power and bit rate (described below).

More details on the simulation environment, the components modeled and their details, can be found in [25].

V. Experiments and Results

Our simulation includes both wireless and wired network components. Our topology consisted of a single access point with one Ethernet interface and wireless network interface (802.11). The access point was connected to a router via a 10Mb/s wired connection. A multimedia server was connected to the router via a 100Mb/s wired connection. Two identical laptops with the client application moved within this environment. The simulation environment of OMNeT++ provides a rectangular "playground" which contains the devices and in which the laptops "move". Our laptops start their movement from opposite edges of the playground and travel straight towards the opposite side, then return to their original point and repeat this through the duration of the simulation. They move in a straight line with a speed of 10 meters/second.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

The access point is stationary. Three factors are varied for the simulation:

- "Transmitter power": set to either 350mW or 450mW;
- "Bit rate": set to either 24Mbps and 54Mbps;
- "Data size": size of our multimedia file to be transferred: 30Mbytes and 60Mbytes.

Our basis of comparison is the time required to transmit all the data packets from server to the laptops and the number of retransmission time outs that occurs during the transmission. The transmission time results of the experiments are presented in Tables I and III and the results on the number of RTOs are presented in Tables II and IV.

Bitrate	Power (mW)	Protocol	Avg. Time (sec)
24Mbps	350mW	TCP	129.28
24Mbps	350mW	STCP	126.22
24Mbps	450mW	TCP	125.31
24Mbps	450mW	STCP	120.08
54Mbps	350mW	TCP	122.03
54Mbps	350mW	STCP	117.58
54Mbps	450mW	TCP	117.76
54Mbps	450mW	STCP	114.54

TABLE I
TIMING RESULTS FOR 30 MBYTES OF DATA.

Bitrate	Power (mW)	Protocol	Avg. RTOs
24Mbps	350mW	TCP	2344.5
24Mbps	350mW	STCP	2143.5
24Mbps	450mW	TCP	2348.0
24Mbps	450mW	STCP	2166.0
54Mbps	350mW	TCP	2346.0
54Mbps	350mW	STCP	2143.0
54Mbps	450mW	TCP	2350.0
54Mbps	450mW	STCP	2162.0

TABLE II
TIME OUTS FOR 30 MBYTES OF DATA.

Bitrate	Power (mW)	Protocol	Avg. Time (sec)
24Mbps	350mW	TCP	261.19
24Mbps	350mW	STCP	256.07
24Mbps	450mW	TCP	248.71
24Mbps	450mW	STCP	242.35
54Mbps	350mW	TCP	247.47
54Mbps	350mW	STCP	240.80
54Mbps	450mW	TCP	235.59
54Mbps	450mW	STCP	228.28

TABLE III
TIMING RESULTS FOR 60 MBYTES OF DATA.

For data of 30Mbytes and 60Mbytes, STCP consistently outperforms TCP, though by only a little in some cases. STCP also clearly outperforms TCP in the number of RTOs, substantially reducing the number of RTOs.

While looking at the results suggests that the differences in transfer time and in terms of RTOs might be attributable to STCP, it is certainly not clear. To understand the factors

Bitrate	Power (mW)	Protocol	Avg. RTOs
24Mbps	350mW	TCP	4683.5
24Mbps	350mW	STCP	4324.0
24Mbps	450mW	TCP	4690.5
24Mbps	450mW	STCP	4333.5
54Mbps	350mW	TCP	4694.5
54Mbps	350mW	STCP	4321.5
54Mbps	450mW	TCP	4693.5
54Mbps	450mW	STCP	4324.0

TABLE IV
TIME OUTS FOR 60 MBYTES OF DATA.

impacting the measured results, we performed an analysis of variance on the data. We considered the bitrate, power, protocol (TCP, STCP) and data size as factors and considered transmission time and number of timeouts as dependent variables. Thus, it was a four factor, two level analysis. For both the transmission time and the retransmission timeouts, the data size was the primary factor, explaining 98% of the variability. In looking at the raw results, this is clear - the data size dictates the time and number of timeouts.

We then considered the results separately for the different sizes of data, i.e., did separate analyses for the 30MB and 60MB experiments. Thus, each of these was a three factor, two level analysis. For the 30MB data size, the bit rate had the greatest impact on the variability, explaining 59% in the variability of transmission times. The difference in protocols explained about 18% of the variability and the transmission power explained about 22%. For the 60MB data size, the results were similar, with the bit rate explaining 49%, the transmission power 40% and the protocol used explaining 10%. In terms of transmission time, the impact of the protocol used was not the dominant factor. This is not surprising, since a higher bit rate would have a major impact on the transmission times. Thus, these experiments do not show much difference in the total times; larger files for longer durations may have to be examined.

However, when one considers the number of RTOs, the results are different. The analysis of variance for both the 30MB and 60MB data size files shows that the protocol accounts for almost all the variability (99%) in both sets of results. The other factors have no little impact (less than .5%) on the variability of the number of timeouts. There is a clear advantage for STCP over TCP in reducing the number of retransmission timeouts.

VI. CONCLUSION

While the results presented in this paper are still early, they do show that STCP has some potential use in wireless environments. Even though other aspects of STCP have been studied elsewhere [25], there is still quite a bit to understand about its behavior. One of the advantages of STCP is that, as illustrated in the simulation, it works with TCP. In our simulation, only the multimedia server used STCP, other components, such as the laptops, used the standard TCP. Since

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

STCP adjusts the congestion window on the sender's side, it does not need to exist everywhere. This is a definite advantage.

The parameterized smoothing functions can also be used in other novel ways. As mentioned, parameters could be set by applications or by a server depending on context, such as settings for a video on demand server. These can then be changed as the environment changes, e.g. more users, etc. How to do this is unexplored. An alternative is to look at a more "dynamic" version of STCP, where the coefficients are dynamically changed depending on the network environment. We are currently exploring how this might be done. There is also a need to compare STCP to other TCP variants, such as RTP and UDP. This is something we are also exploring.

REFERENCES

[1] I. Kim, Y. Kim, M. Kang, J. Mo, and D. Kwak, "TCP-MR: Achieving end-to-end rate guarantee for real-time multimedia," in *Proceedings of the 2nd International Conference on Communications and Electronics (ICCE)*, 2008, pp. 80–85.

[2] E. Vieira and M. Bauer, "Proactively controlling round-trip time variation and packet drops using smoothTCP-q," in *QShine'06: Proceeding of the 3rd International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks*, 2006, pp. 39–44.

[3] E. Vieira, "Smooth congestion control algorithms," PhD Thesis, The University of Western Ontario, 2006.

[4] E. Vieira and M. M. Bauer, "The variation in RTT of smooth TCP," in *Proceedings of the 3rd Consumer Communications and Networking Conference*, 2006, pp. 361 – 365.

[5] G. Holland and N. Vaidya, "Analysis of TCP performance over mobile ad hoc networks," in *MOBICOM*, 1999, pp. 219–230.

[6] T. Dyer, D. Thomas, R. Boppana, and V. Rajendra, "A comparison of TCP performance over three routing protocols for mobile ad hoc networks," in *MobiHoc '01: Proceedings of the 2nd ACM international symposium on Mobile ad hoc networking & computing*. New York, NY, USA: ACM, 2001, pp. 56–66.

[7] K. Chandran, S. Ragbunathan, S. Venkatesan, and R. Prakash, "A feedback based scheme for improving TCP performance in ad-hoc wireless networks," in *Proceedings of the 18th International Conference on Distributed Computing Systems*, 1998, pp. 472–479.

[8] W.-T. Chen and J.-S. Lee, "Some mechanisms to improve TCP/IP performance over wireless and mobile computing environment," in *Proceedings of the 7th International Conference on Parallel and Distributed Systems*, 2000, pp. 437–444.

[9] A. Chodorek and R. Chodorek, *Applicability of TCP-friendly protocols for real-time multimedia transmission*, Faculty of Electronics and Telecommunicaitons, Poznan University of Technology, 2007.

[10] J. Shi, Y. Jin, H. Huang, and D. Zhang, "Experimental performance studies of SCTP in wireless access networks," in *Proceedings of the International Conference on Communication Technology*, 2003, pp. 392 – 395.

[11] R. Fracchia, C. Casetti, C.-F. Chiasserini, and M. Meo, "A wise extension of SCTP for wireless networks," in *IEEE International Conference on Communications*, 2005, pp. 1448 – 1453.

[12] A. Kumar, L. Jacob, and A. L. Ananda, "SCTP vs TCP: Performance comparison in MANETs," in *Proceedings of the 29th IEEE International Conference on Local Computer Networks*, 2004, pp. 431 – 432.

[13] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, *RFC 3550: RTP: A Transport Protocol for Real-Time Applications*, IETF Network Working Group, July 2003.

[14] H. Schulzrinne, A. Rao, and R. Lanphier, "Rfc 2326: Real time streaming protocol (rtsp)," IETF Network Working Group, United States, 1998.

[15] M. Allman, V. Paxson, and W. Stevens, *RFC 2581: TCP Congestion Control*, IETF Network Working Group, April 1999.

[16] D. Bansal and H. Balakrishnam, *TCP-Friendly Congestion Control for Real-time Streaming Applications*, May 2000.

[17] V. Jacobson, "Congestion avoidance and control," in *SIGCOMM '88: Symposium Proceedings on Communications Architectures and Protocols*. New York, NY, USA: ACM, 1988, pp. 314–329.

[18] P. Karn and C. Partridge, "Improving round-trip time estimates in reliable transport protocols," in *SIGCOMM '87: Proceedings of the ACM Workshop on Frontiers in Computer Communications Technology*. New York, NY, USA: ACM, 1987, pp. 2–7.

[19] OMNeT++, "http://www.omnetpp.org/index.php."

[20] A. Varga and R. Hornig, "An overview of the OMNeT++ simulation environment," in *Proceedings of the 1st International Conference on Simulation Tools and Techniques for Communications, Networks and Systems*. ICST, Brussels, Belgium, Belgium: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2008, pp. 1–10.

[21] A. Varga, "Using the OMNeT++ discrete event simulation system in education," *IEEE Transactions on Education*, vol. 42, no. 4, pp. 11–11, November 1999.

[22] J.-C. Maureira, O. Dalle, and D. Dujovne, "Generation of realistic 802.11 interferences in the OMNET++ INET framework based on real traffic measurements," in *Proceedings of the 2nd International Conference on Simulation Tools and Techniques*. ICST, Brussels, Belgium, Belgium: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2009, pp. 1–8.

[23] M. Bredel and M. Bergner, "On the accuracy of IEEE 802.11g wireless LAN simulations using OMNeT++," in *Proceedings of the 2nd International Conference on Simulation Tools and Techniques*. ICST, Brussels, Belgium: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2009, pp. 1–5.

[24] K. Rmachandran, R. Kokku, H. Zhang, and M. Gruteser, "Symphony: Synchronous two-phase rate and power control in 802.11 WLANs," in *Proceedings of the 6th International Conference on Mobile Systems, Applications and Services (MobiSys'08)*. New York, NY, USA: ACM, 2008, pp. 132–145.

[25] M. Islam, "Smooth TCP: A solution for wireless multimedia communication," MSc Thesis, The University of Western Ontario, 2010.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

A Novel Probabilistic Deadline Scheduling Mechanism for DCCP

Daniel Wilson
School of Information Technology
Murdoch University
Perth, W.A., Australia, 6150
Email: Daniel.Wilson@murdoch.edu.au

Mike Dixon
School of Information Technology
Murdoch University
Perth, W.A., Australia, 6150
Email: m.dixon@murdoch.edu.au

Terry Koziniec
School of Information Technology
Murdoch University
Perth, W.A., Australia, 6150
Email: t.koziniec@murdoch.edu.au

***Abstract*—This paper introduces a novel cross layer probability based deadline scheduling mechanism designed specifically for real time Data Congestion Control Protocol (DCCP) flows. Scheduling in this mechanism is determined based on the probability of a data packet being received within its useful lifespan. In order to predict this probability, DCCP is modified to access routing table information used by CISCO Systems Inc.'s Enhanced Interior Gateway Routing Protocol (EIGRP), to estimate the approximate forward path delay period. Once the packet's probability of arriving within its useful lifetime is determined, the scheduling algorithm then places the packet into one of three predefined queues to ensure all packets that are received are given the highest chance of being delivered within their useful lifespan period. In addition to describing the design of the mechanism, this paper will also present proof of concept modelling carried out to quantify the effectiveness of the mechanism. The results presented in this paper show the mechanism described is able to predict the time a packet will need to traverse the network using EIGRP's metrics with greater than 90 percent accuracy (on average) in the tested topologies. The results will also show the mechanism is stable and able to operate in medium sized networks with marginal overhead.**

***Index Terms*—congestion control; real time; stale packets; queuing; scheduling;**

I. Introduction

This research highlights how real-time data possesses a finite lifespan and describes a mechanism that allows this lifespan property to be incorporated into DCCP [1] to allow more efficient transportation of DCCP-Data packets carrying real-time data. To achieve this, this paper will present a mechanism that can be employed on intermediate Bandwidth Optimization Devices (BODs), that will calculate a packets probability of arriving within its useful lifespan to determine if the packet should be sent or not, and the priority needed to send it. In order to do this, information relating to the network located between the device where the forwarding decision is being made and the end point, is derived from EIGRP [2]. This paper will begin with a high level overview of the mechanism. Next, specific details describing the mechanism will be presented to show how EIGRP metrics are used to calculate a packet's probability of arriving within its usefil lifespan. This is followed by description of a scheduling scheme that prioritizes packets based on their probability of arriving in time. Following this, proof of concept experimentation that was used to quantify the effectiveness of the scheme is presented. The paper concludes with discussion of the experiments, their associated limitations and finally a conclusion of the paper and its findings.

II. Related Research

Gurtov and Ludwig [3] are credited as being the first to introduce packet lifetime discard mechanisms to the research community. In their model, each packet has a packet life variable embedded into a custom built IP header option field. Based on this information, packets that have existed beyond their useful time are purged on intermediate routers placed between the sender and the receiver. They also propose that packets not likely to be delivered within the valid time should also be purged, further increasing efficiency of bandwidth utilization.

Gurtov and Ludwigs' research showed that purging stale packets on the network was able to improve overall performance for all applications utilizing the network. Specifically, the reduction in the overall delay in delivery times experienced by fresh packets, when stale packets were purged, was significant. Performance of the application for which packets were purged appeared to be unaffected by the purging process, as the packets would likely have been dropped by the application once received if the purging action had not taken place.

While Gurtov and Ludwigs' study was, and remains, the most significant work in the area, there have been a number other researchers who have carried out similar research. Yuen and Yue in [4], present a scheme of purging stale packets and then subsequently prioritizing real time packets in a data queue. Their research was of a mathematical nature and no mention was made regarding the practical implementation of the scheme. Their findings showed that by purging stale packets from queues, fresh packets in those same queues would experience shorter delays and less contention for available bandwidth.

Gurtov and Ludwigs' foundation idea was also used by Chebrolu's and Rao in [5]. In this work, they examine bandwidth optimization specifically for MPEG video where bandwidth was severely restricted. Using a combination of the Gurtov and Ludwigs packet discard mechanism and knowledge of how MPEG video standards operate, they created Minimal Cost Drop (MC Drop) [5]. MC Drop is used firstly to determine

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

which packets already are, or are likely to become stale, due to the high levels of congestion on a network link. Packets that are already stale are dropped immediately. Packets that are likely to be become stale are applied to a policy to determine which of these packets should be dropped first to ensure the best video quality at the receiver side. The results obtained from experiments showed that MC Drop allowed for higher overall video quality in instances where there was a severely restricted bandwidth link, when compared to conventional non discard techniques.

TCP-RTM [6] was designed by Liang and Cheritan in 2002, and proposes a number of extensions that make TCP more suitable to real time data. One of these extensions involves marking stale packets to speed up the delivery of fresh packets. Radovanovic et al., in [7], took Liang and Cheritan's work one step further towards Gurtov and Ludwigs' ideology by using the SNOOP protocol to purge packets identified by the TCP-RTM extension as being stale. In addition, they implemented a mechanism for suppressing packet retransmission and congestion events resulting from the purging process. Results gathered from a NS2 simulation found that the packet discarding scheme improved overall network goodput and reduced the overall delay experienced by packets when compared to conventional non discard techniques. This technique offered improved performance for real time applications using the TCP protocol and has much promise.

All the above findings show unequivocally the potential improvement possible by implementing a packet discarding scheme for real time traffic. There does not however appear to be a packet life discard scheme proposed or implemented specifically for the DCCP protocol, a protocol which appears so well suited to such a scheme. Given that the DCCP protocol has been designed for the application of real time data, adding such a scheme could potentially be of major benefit to the protocol. Additionally, a number of problems which have hindered the wide spread adoption of packet discarding schemes, such as the loss of congestion control integrity and having to suppress retransmissions, are removed in DCCP as it is an unreliable protocol. Using Gurtov and Ludwigs' packet discard principles, this paper presents an efficient and effective packet discard policy designed specifically for the nuances of the DCCP protocol.

III. High level overview

In this section a high level overview of how the probability based scheduling (PBS) mechanism operates is presented. In the PBS mechanism, the application layer on the sending device is configured to specify the packet's maximum lifespan value concurrently with the passing down of the payload data. This value represents the maximum time the data is of use to the receiving application. Upon receiving data from the application layer, DCCP on the sending device timestamps the packet with a birthtime variable. The birthtime variable is extracted from a NTP synchronized system clock and specifies the exact time when the packet was created. The payload data will then be encapsulated into a generic DCCP-Data packet with the two above-mentioned variables appended as additional lifespan option fields. The packet will then be passed to the relevant network layer protocol for transportation across the network.

Located within the network, between the sender and the receiver, will be a number of intermediate BODs. These devices are modified/enhanced layer 3 network devices, such as routers and layer 3 switches (switched networks). In order to convert these devices into BODs, the PBS mechanism described in this paper, or other similar mechanisms, are added to the devices to allow them to utilize the lifespan variables in the DCCP data packets. In this scheme, when the BOD receives a packet, it also queries/extracts information from the routing protocol in order to determine the approximate time it will take for the packet to reach its intended destination. Once this time is known, the PBS mechanism then determines if the packet has sufficient lifetime remaining to reach the destination. Packets are then be assigned a probability of arriving on time, and categorized into four distinct categories. This allows advanced scheduling decisions to be made in order to offer the greatest good to all packets attempting to traverse the network. In addition, packets that have no possibility of arriving in time are also removed from the network making additional bandwidth available to other flows which would be impacted by the transmission of these stale packets.

IV. Definition of terms

In the following sections a number of acronyms will be used in order to keep the specification concise. In this section these acronyms will be defined with an accompanying description of where they are derived from. *BT* will be used in place of birth time. This variable is created when the sender receives data from the application layer and encapsulates the payload into a DCCP-Data packet. This variable is obtained from the system clock and is the exact time that the DCCP packet comes into existence. When the application layer passes the information down to the DCCP layer, it also passes down a *Maximum Time-to-Live (MTTL)* variable. This variable represents, in milliseconds, the amount of time the real-time data is valid for. In order to determine if a packet has expired or become stale, the MTTL value is added to the BT value to get the *Explicit NTP Expiry Time (ENET)*. At any point in the transmission, an intermediate device can query the NTP clock on the system for the *Current NTP Time (CNT)*. If the $ENET > CNT$ then the packet's contents are stale. In order to find out how much life time a packet has remaining the *Time-left-to-live (TLTL)* variable is used. To calculate the TLTL the ENET is subtracted from the CNT i.e. $(TLTL = ENET - CNT)$.

The next acronym that will be used is the *Time-Left-In-Queue (TLIQ)*. This variable represents the minimum amount of time it will take for a packet to be transmitted in the current queue if no change is made to the position of any packets in that queue. The final acronym used is the *Time-Needed-To-Cross-Network (TNTCN) variable*. This value is derived from the routing protocol and represents the estimated time that will be needed for a given packet to be delivered from

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

the current device to the destination network based on current network conditions. How the TLIQ and TNTCN variables are calculated will now be presented in greater detail.

A. Time left in Queue (TLIQ)

From the moment the PBS mechanism is initiated, it is vital that queue depth information is monitored and updated constantly for each of the queue categories. In addition to this, the current expected queue delay for newly arriving packets must also be calculated continuously. To achieve this, the *Time-Left-In-Queue (TLIQ)* variable is calculated every time a packet is received by the BOD device. The TLIQ value represents the time a packet will remain on the device for while in the outgoing interface queue and is critical in determining whether a packet should be placed in the critical queue or the normal queue. The formula for calculating the TLIQ value is shown in Equation 1.

$$TLIQ = \frac{TSQ}{IS * BA} \quad (1)$$

In order to calculate the TLIQ value for a newly arriving packet, the following steps take place. First, the interface service rate is calculated by taking the interface speed (IS) and multiplying it by the bandwidth percentage available to that queue (BA). Next, the queue depth information is accessed to determine the cumulative size of all existing packets in the queue in kilobits (TSQ). From these two variables, the time that is required for the queue to be serviced is deduced. When a new packet arrives, the TLIQ value must be readily available to the packet categorization mechanism so it can determine whether the packet belongs in the critical queue or the normal queue.

B. Time Needed to Cross Network (TNTCN)

The novel element of this scheme is provided through the addition of the *Time-Needed-to-Cross-Network (TNTCN)* variable. This variable is derived using information obtained from a routing protocol and is used to provide an estimate of the time that will be required for the packet to reach the destination network. From this formula, the probability of the packet arriving at the intended destination within its useful lifespan is then calculated. EIGRP is used in the PBS mechanism to calculate the TNTCN variable. There are a number of reasons why EIGRP was selected to provide this information. Firstly, EIGRP uses a number of variables to calculate its routing and topology tables. These variables include Delay, Reliability, Load, Bandwidth, Hop count and MTU. Each of these variables represent unique conditions that exist in the network between the queue and the destination network.

To reduce the computational overheads of the PBS mechanism, these EIGRP variables are utilized for the purposes of determining the probability of a packet arriving within its lifespan. The ethos behind this research is to promote cross layer cohesion and reduce computational overheads by alleviating task repetition wherever possible. To configure the PBS mechanism to calculate the variables EIGRP already provides was deemed unnecessary when these variables were already readily available. To add to this, the algorithms used to derive these variables in EIGRP are well matured and have been proven to work efficiently.

Another reason why the EIGRP routing protocol was selected was because it offers fast convergence times through its Diffusing Update Algorithm (DUAL)[8]. Fast convergence times equate to higher levels of accuracy in relation to actual network conditions, which is a prerequisite for the PBS mechanism. If changes to the network occur, the routing protocol must be able to detect these quickly in order to provide accurate information to this mechanism. Failure to do so in a timely manner will lead to incorrect scheduling and prioritization of packets. The drawbacks to using the EIGRP protocol are that the choice to do so limits the PBS mechanism purely to topologies that are configured with CISCO Systems Inc.'s devices. In addition, EIGRP is an interior routing protocol and therefore, this limits the scale of the PBS mechanism to single autonomous systems and not the broader Internet in its current form. Finally, it is also important to mention that EIGRP does utilize bandwidth in order to communicate routing updates to other routing devices on the network. This paper will assume that the EIGRP protocol has been installed on the network device for core routing functionality and not for the express purposes of this mechanism.

V. Mechanism Details

There are two main components that make up the PBS mechanism. Firstly, there is the categorization component that is used to determine which category and subsequently which queue is appropriate for an incoming packet. The second component is a packet schedule, which services the respective queues in such a way as to ensure strict adherence to the queue's allocated bandwidth percentage. These components will now be discussed in greater detail.

A. Categorization of Packets

1) Queue Structure: The first phase of this mechanism involves the categorization of incoming packets into one of three unique queues based on information gathered from the BT and MTTL variable in the DCCP header option fields. This section will commence by describing the function/role of each of the three queue classes.

Discard queue

The first packet queue class is for packets that have expired, or for packets that will expire before they reach their destination. This queue, known as the discard queue, will employ techniques that remove stale packets from queues. In order to qualify for this queue the following two criteria are checked. Firstly, if the packet arrives stale $(ENET < CNT)$, then the packet is placed in the discard queue. Secondly, if the packet will not reach its destination before becoming stale $(TLTL < TNTCN)$, then the packet will also be placed in

the discard queue.

Critical Queue

The second queue class, named the critical queue, is for packets that will expire unless a prioritization action takes place to prevent them from doing so. Specifically, if the process of passing through the queue on the BOD will directly lead to the packet becoming stale, then the packet is placed into this queue. The critical queue is allocated a predetermined amount of bandwidth to service packets at a prioritized rate. To qualify for this queue a packet must meet the following criteria. The packet must have sufficient lifespan remaining to allow it to travel across the network to its intended destination without becoming stale, i.e. $(TLTL > TNTCN)$. In addition, to qualify for this queue the packet must be in a position where the cumulated time needed for queuing, (TLIQ) and the time needed for the packet to cross the network, is greater than the time the packet has remaining before becoming stale (TLTL). The notion behind this queue is that packets that can reach the destination if they are not delayed excessively by the queuing process on the BOD device, are given the best chance to do so by being placed in a smaller more rapidly serviced queue. If this action does not occur, the packets become stale before reaching their destination due to the queue delay on the BOD.

Normal Queue

The final queue class is for packets that will likely reach the destination network within their respective lifespan provided there are no abnormal network fluctuations or extreme changes in network conditions. To qualify for this queue, the packet must possess a lifespan (TLTL) greater than the time needed for queuing and the time needed for the packet to traverse the network to the destination device. I.e. $TLTL > (TLIQ + TNTCN)$. Packets in this queue will be offered a guaranteed position and guaranteed bandwidth allocation ensuring the queue time is deterministic. Packets that are placed into this queue are placed there on a first-in-first-out (FIFO) basis and no packet is ever to be placed in front of a pre-existing packet. As this queue is given the majority share of available bandwidth and offers guaranteed service rates, it creates incentive for the application selecting the TTL values to actively aim to deliver packets that are placed into this queue as the likelihood of packets in this queue being delivered on time is probabilistically higher than the other two queues.

In the Figure 1, the categorization scheme described above is demonstrated graphically.

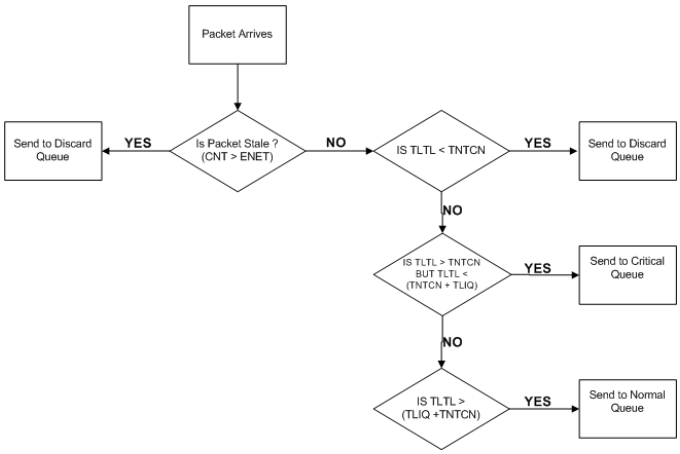


Fig. 1. Flow Diagram Representing Queue Categorization Scheme

B. Probability Calculation

When a packet is received for transmission on an interface, the PBS mechanism parses the packet's DCCP and IP header to obtain three variables. The first two variables that are extracted are derived from the additional lifetime fields, namely the BT and the MTTL. In addition to this, the IP header is also accessed to obtain the packet's intended destination network. Having obtained these three variables, the probability of the packet arriving in time is calculated by performing a TNTCN lookup task. This lookup task will provide a value, in milliseconds, of the estimated time that it will take for the packet to traverse the network and reach its destination network. In the next section, the way EIGRP metrics are used to calculate the TNTCN value is presented.

1) Using EIGRP to calculate TNTCN: In order to determine the delay between the BOD and the destination network there are three elements that make up the TNTCN latency value. The first of these elements is commonly referred to as *wire line latency*. This value is the time it takes for the data signal to travel across wired or wireless medium between intermediate and end point devices including the serialization of packets into the necessary electrical signal format. This latency is governed by principals of physics and is almost always deterministic. Typically, compared to the queuing delay latency that will be discussed later, this value is negligible.

The second element that contributes to the TNTCN value will be referred herein to as *switching latency*. This value represents the minimum latency that is added by the intermediate device if no queuing takes place. This latency is created by the internal activities that occur on intermediate devices such as switches and routers with the exception of any queuing related activities. Examples of these activities include transferring packets between interfaces, packet encapsulation, MAC table and routing table lookup activities to name but a few. These values are also all typically deterministic and will normally remain constant throughout a flow. This value is the minimum latency that the intermediate device will add to the TNTCN.

The third component that contributes to the TNTCN value is *queuing latency*. Although queuing takes place on intermediate devices, this value will be treated separately from the switching latency variable described above. The reason for this is that incorporation of queues into the network path introduces a non-deterministic element into the delay calculation. Typically, queue sizes will fluctuate and cause variance in queuing latency values.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

In order to provide an accurate TNTCN value to the PBS mechanism, all three of these values must be taken into consideration. The TNTCN value is therefore calculated using the following formula: $TNTCN = SwitchingLatency + WireLatency + QueuingLatency$

Having defined the three components that make up the TNTCN variable, the way in which the EIGRP metrics are used to calculate these elements will now be described. To estimate the time that will be required for the packet to travel across the remaining portion of the network, four of EIGRPs metrics are used by the PBS mechanism. These four metrics are EIGRP Delay, Load, Bandwidth and Hop count.

2) Use of EIGRP delay metrics: The EIGRP delay is the total delay that exists between the router and the final destination network. To calculate this value, each EIGRP device along the path assigns a predefined delay metric to all its interfaces based on the interfaces speed. Each of these calculated values along the network path is then combined in order to determine the EIGRP delay metric for the destination network. EIGRP assigns the delay value to the interface using a predefined table of values. For example, EIGRP will assign a 100Mbps Ethernet link a delay value of 100microseconds. A 1000Mbps link will be assigned a 10microsecond value and so on. The values used equate roughly to the propagation time that would be required to transmit a 1250byte packet across a network.

The issue with using CISCO Systems Inc.'s predefined delay values is that when smaller packets are transmitted across the network, the predefined delay values used by EIGRP are not indicative of the actual latency smaller packets would experience. As DCCP will be used for real-time data, which will typically use smaller more frequent data packets, the standard EIGRP delay values are not by default suitable.

In order to make the reported EIGRP Delay value more suitable to smaller packets, the delay value will be adjusted in this mechanism according to packets packet size, in order to provide more accurate and representative delay values. To do this the following formula is used.

$$Delay = \frac{PS}{1250} * RD \tag{2}$$

By factoring the reported EIGRP delay value (RD) by the actual packet size (PS), as shown in the formula above, realistic wire delay latency values are obtained.

3) Use of EIGRP Load Metric: Using the delay metric alone constitutes only the wire latency portion of the total expected latency that will occur between the BOD and the destination network. As mentioned earlier, the incorporation of queues in the network path introduces an indeterminable amount of latency due to queue size fluctuations. The inability to predict how long packets will be in queues for during transfer across the network makes the task of predicting their probability of arriving within their useful lifespan problematic. One approach that could be utilized to make queuing latency deterministic is to adopt a worst case scenario view and assume queues are always full and apply the maximum latency the queue can produce into the probability equation. While this could work, this approach was not deemed feasible because it would require that all devices have knowledge of upstream device queue sizes. If this approach was taken, it would be more beneficial to simply create a separate communication channel between devices to exchange queue depth information.

In order to overcome this non deterministic element in a more eloquent manner than described above, this mechanism makes use of the EIGRP load metric. The EIGRP load metric is calculated dynamically by EIGRP and represents the level of saturation that exists on links along the network path to the destination network. The load value is reported as a value between 1 and 255 with 1/255 representing a completely unsaturated network and 255/255 signifying a completely saturated network. In order to calculate the load value, EIGRP uses a moving average of saturation levels over a 5 minute period with sampling occurring every 5 seconds. This large sampling period prevents sudden saturation increases in network load from causing instability in this mechanism. While the frequent 5 second sampling rate ensures that there is accurate reflection of network conditions in the load value.

EIGRPs load metric is ultimately indicative of the level of congestion occurring in the network. Wherever congestion occurs, the subsequent result is virtually always an increase in queuing at various points along the network path. The larger the queue sizes become along the network path, the longer packets take to reach their intended destination. As queue sizes and link saturation values begin to increase, so too does the EIGRP load value. The PBS mechanism uses a combination of the EIGRP bandwidth metric, and the EIGRP load metric to estimate the queue time that a packet can expect to take in order to pass through the queues on the forward path to the destination network.

To calculate the queuing latency (QL), the following formula is used.

$$QL = \frac{LD}{255} * \frac{\frac{DQS*PKS}{EB}}{1} \tag{3}$$

From Formula 3 above, it can be seen that the first step taken is to obtain the default queue size of the slowest upstream device (DQS). On CISCO Systems Inc.'s devices, the default outgoing queue size is set at 40 packets. The next step taken is to calculate the size in bits that could reside in the DQS queue. This is problematic as the packets sizes can vary and therefore the exact value is not known. Assuming all the packets are of the MTU size would increase the queuing delay value inaccurately. For this reason, a bias that will result in smaller than potentially possible TNTCN values is implemented. The size of the current packet (PKS), in bits, is used in place of the MTU. If the packet is 180 bytes, the PBS mechanism will assume all the packets at the bottleneck are 180 bytes in size. In addition to this, the queue size on the local BOD is used instead of what the actual queue size may be on the device with the lowest bandwidth (slowest upstream device). There is strong support showing it would be beneficial for EIGRP to be modified to transmit queue size information of the devices in the topology.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

Once the queue size in bits is determined, the PBS mechanism then takes the slowest link in the network path, obtained from EIGRP bandwidth metric (EB), and calculates the maximum time it will take a queue to be emptied based on the interface speed. Finally, this time is weighted against the reported EIGRP load value (LD) to provide a realistic queue delay value. In the experiment section of this paper, the accuracy of the above formula will be examined, and it will be shown that this formula and the weighting used produces a sufficiently accurate estimate of the actual delay that occurs as a result of queuing.

4) Use of EIGRP Hop Count Metric: The final element of latency that must be taken into consideration is switching latency. This latency is added through the processes that take place on intermediate devices along the network path such as routing and re-encapsulation. To calculate the switching latency value, the EIGRP hop count metric is used by the PBS mechanism. The value produced by this metric represents the number of layer 3 routing devices that exist between the BOD and the destination network. The difficulty here is determining what latency each hop or layer 3 device will add to the total latency as there can be substantial differences in layer 3 router performance. This value can unfortunately never be exactly deduced due to variations in router hardware and configurations.

In order to provide a future proof mechanism that allows the switching time to be accurately calculated when faster switching technologies eventuate, it is recommended that each BOD should perform a calculation of its own switching time during initialization. Once this is done, this value is multiplied by the EIGRP hop count value to determine the total switching latency. In practice, this value is minuscule compared to queuing latency and therefore completely accurate determination of this value is deemed unnecessary.

5) Final TNTCN Formula: Having described where the various elements that make up the total latency value are sourced from using the EIGRP metrics, the final equation used to calculate the TNTCN variable is shown in two variations below:

$$TNTCN = (WireLatency + QueuingLatency + SwitchingLatency)$$

OR

$$TNTCN = \frac{PS}{1250} * RD + \frac{LD}{255} * \frac{\frac{DQS*PKS}{EB}}{1} + HC * 8.2M \quad (4)$$

It is important to remember that these metrics are used to merely estimate the time a packet will need to reach its destination network. As network conditions are constantly changing, there is no way this scheme will provide perfect results. This scheme takes a moderate to best case scenario approach in determining the TNTCN. This means the TNTCN value is slightly ambitious because doing this gives the packets a higher probability of arriving in their specified lifespans. Adopting a worst case scenario approach, which provides higher estimated times, would mean that packets would have lower probabilities of being deemed to arrive in time and as a result a higher number of packets would be placed in the incorrect queue if network condition were not exactly as the calculation predicted. Being too lenient and providing too much leeway in the calculation, would cause reduce the PBS mechanism's effectiveness. The experimentation below will show that the weighting and selection of the various metrics described above provide a TNTCN value that is sufficiently accurate for use in the PBS mechanism.

6) Packets destined for unknown networks: If a packet is received and there is no information relating to the packet's intended destination in the EIGRP routing table, then the packet is placed into the normal queue and no priority is given to it. This ensures that all packets are serviced, even when they are destined for unknown networks. In addition, applications can append a 99999 millisecond lifespan to packets they do not wish to have categorized. Upon receiving such a packet, the BOD will automatically place the packet into the normal queue.

C. Scheduling implementation of BOD

For the purposes of implementing the mechanism, statistical time division multiplexing was selected as the method for scheduling packet delivery on outgoing interfaces. With statistical multiplexing, each of the queues essentially became a channel in the scheduling algorithm. Each channel was then assigned the predetermined amount of bandwidth and serviced at a fixed rate accordingly. The advantage of statistical multiplexing over traditional time division multiplexing is that where a particular channel does not need to transmit during its pre-allocated slot, this slot can then be used by the next channel waiting to transmit. In traditional time division multiplexing this does not occur and the slot remains unused rather than being allocated to the next channel. For the purposes of the experimentation, the Opnet Modeler simulation toolkit provided statistical time division multiplexing modules as well as the queuing modules that were added and configured to service the queues on outgoing interfaces on the BOD device. For the purposes of this scheme, each of the three queues were configured as an individual channel and serviced based on a strict bandwidth allocation percentage.

VI. Experimentation

In this section, two experiments that were conducted to examine and test the scheme described above are presented. The primary purpose of these experiments is to test the accuracy of the TNTCN formula and ensure the action of placing packets in the critical queue is of benefit.

A. Experiment 1: Accuracy of TNTCN Formula

In this experiment, the accuracy of the formula used to derive the TNTCN value was explored. This experimentation compares the computed TNTCN value to the actual time it took for the packets to reach their destination networks. This comparison determines if the weighting used in the formula is

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

suitable for the purposes of the PBS mechanism by ensuring the TNTCN and actual values are within acceptable range. In addition to this, where there was variance in the times, the TNTCN formula should provide a bias towards lower TNTCN values than the actual time it took for packets to cross the network. If the TNTCN formula produces excessively large TNTCN values then packets could inadvertently be placed into the critical queue, or even worse be deemed unable to reach the destination within the remaining lifespan and placed in the discard queue. By maintaining a bias towards smaller TNTCN values, the PBS mechanism will produce fewer incorrectly categorized packets.

1) Hypotheses: It was hypothesized that the TNTCN and actual times needed to cross the network would typically be within a 10% threshold of difference during normal network operation periods. In addition to this, it was also hypothesized that the TNTCN formula would provide a bias towards generating smaller TNTCN values than the actual time that is needed by packets to cross the network.

2) Experiment Methodology: In order to carry out this experimentation 20 unique simulated topologies were created. These topologies were modelled based on a medium sized wide area network (WAN). In each of these topologies between 10 and 30 CISCO Systems Inc. 2600 series routers were configured. A variety of LAN and WAN links were used to connect the devices with varying speeds. These speeds ranged from 128kbps to 44736kbps in order to add complexity to the topology. In addition to this, each of the experimental topologies were configured to initiate multiple simultaneous flows across the network throughout the simulated period. As a result, at any given time 10, 20, 50 or 100 simultaneous flows existed on the network. This complexity was necessary due to the inclusion of EIGRP into the mechanism and the need to have a network complex enough to provide large routing tables and traffic loads that were capable of creating congestion events. In order to gather the data needed for this analysis, five randomly selected flows were chosen from each simulation (where more than five flows existed). The experiments were run over a simulated period of five minutes. The statistical results data from the experiment was only collected during the final 30 seconds of the five minute simulated period. This was done to allow sufficient time for the routing protocol to converge and network load or congestion levels to reach stabilized levels. Using global variables in the Opnet Modeler simulator, each time a packet left the BOD, a timestamp entry was created in a predefined global variable array. In addition to this, the TNTCN value calculated by this scheme was also stored in the same global variable array. Once the packet arrived at the destination network, its sequence number was used to access the corresponding global variable and the current NTP time was entered into the variable store.

The last routing devices along the network path were configured to add the simulation_time_pkt_arrive value to the global variable array when the packet arrived at the incoming interface on the router. The reason for configuring the last router to perform this task and not the destination workstation, was because the EIGRP metric is based on the premise that it calculates the time for the packet to reach the destination network only and not the actual endpoint device. Although in practice the difference is likely to be negligible, in the interests of rigour, it must be noted the actual_time variable was recorded when the packet arrived at the destination network and not the final node. At the end of the simulation period the group of all global variable arrays were collected. From these variables the actual time the packets took to traverse the network to the final destination network was calculated using: $simulator\_pkt\_arrive simulator\_pkt\_depart$. The actual time that was taken and the calculated time needed (TNTCN) were then compared.

In a few instances some of the packets were discarded by an upstream BOD device where the packet was deemed unable to traverse the network within time or where the packet had become stale. In order to accommodate this phenomenon in the results presented below, wherever the Simulator_time_pkt_arrive value was not recorded, that entire array relating to that particular packet was discarded.

3) Results: In the experiment 20 simulated topologies were tested and 5 flows from each of these topologies were monitored specifically. This led to a collection of 100 separate result sets. In the interests of keeping this paper concise not all 100 flows will be discussed in great detail. Instead a summarized version of the results will be presented.

In Table 1, the percentage of average variation found between the TNTCN calculation and the actual time it took packets to reach the destination network is shown. To calculate the values shown in Table 1, the percentage of difference was first calculated from each packet flow statistic array ($TNTCActualtime * (100/1)$). This value represents the percentage difference in times between that occurred between the calculated TNTCN value and the actual time that was measured. Once this had been done for all the flows, the average of these values was then calculated. To do this all the percentage of differences values calculated in the previous step were summed and then averaged ($Sum(percentageofdifference)/numberofcalculations$). This value represents the average difference in the calculated TNTCN and actual times that occurred for the flow. Note, a + symbol indicates that the average TNTCN value was larger than the actual time taken value and a symbol indicates the TNTCN value was found to be smaller on average than the actual time taken.

In topologies 14 and 17, a T3 serial link was placed into the network topology which resulted in a major issue with the PBS mechanism being discovered. The delay value automatically assigned to serial links in EIGRP is 20000 microseconds by default, irrespective of the speed of the serial link. When this value was used in the TNTCN formula, there was a large discrepancy between the calculated TNTCN value and the actual time it took for the packet to reach the destination network. Note, for the remainder of this narrative results from topologies 14 and 17 will be excluded because of this anomaly. As can be seen from the results in Table 1, in the

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

Top No.	Flow 1	Flow 2	Flow 3	Flow 4	Flow 5
1	-3.210%	-2,922%	-3.749%	-4.542%	-2.8924%
2	-3.433%	-3.879%	-3.773%	-4.185%	-3.860%
3	-4.371%	-3.823%	-4.087%	-4.592%	-3.376%
4	-3.483%	-3.653%	-3.833%	-4.042%	-3.325%
5	-6.084%	-6.339%	-6.545%	-6.423%	-6.284%
6	-1.218%	-0.924%	-1.188%	-1.255%	-0.827%
7	-0.313%	-0.267%	-0.231%	+0,109%	-0.119%
8	-4.520%	-5.643%	-5.012%	-5.221%	-4.845%
9	-6.767%	-6.289%	-6.133%	-6.934%	-6.459%
10	-6.001%	-5.847%	-6.210%	-6.387%	-6.113%
11	-2.329%	-2.367%	-2.753%	-2.958%	-2.164%
12	-3.562%	-3.324%	-3.491%	-3.762%	-3.445%
13	-3.011%	-2.872%	-2.992%	-3.019%	-2.691%
14	+17.601%	+17.938%	+17.614%	+18.478%	+17.737%
15	-7.968%	-7.758%	-7387%	-8.544%	-7.891%
16	-8.021%	-8.762%	-8.539%	-8.893%	-8.142%
17	+12.116%	+12.529%	+12.418%	+12.933%	+12.674%
18	-2.985%	-2.754%	-2.531%	-2.997%	-2.655%
19	-3.749%	-3.564%	-3.285%	-3.857%	-3.651%
20	-0.871%	-0.654%	-0.211%	+0.097%	-0.054%

TABLE I
TNTCN ACCURACY - VARIATION OF TNTCN AND ACTUAL TIMES RECORDED (CORRECT TO 3 DECIMAL PLACES)

vast majority of flows the calculated TNTCN produced, on average, a TNTCN value that was within the 10% threshold that was hypothesized. In addition, in all flows the calculated TNTCN value had a distinct bias towards generating lower TNTCN values than the actual transmit times that occurred. Delving deeper into the results, there was clear evidence that in more complex topologies such as in topologies 7 and 20, the mechanism did at times begin to lose its bias towards generating smaller TNTCN values. Exploring the results from these topologies, it became apparent that where EIGRP load metric value increased above 200/255, the PBS mechanism began generating larger volumes of TNTCN values that were higher than the actual time it took for packets to cross the network. Where the EIGRP load value was below 200/255, the TNTCN formula had a clear bias towards generating smaller TNTCN values than the time packets took to cross the network, and only very rarely generated higher values. However, once the EIGRP load value exceeded 200 the TNTCN values showed no clear bias towards being lower than the actual time it was taking for packets to arrive at the destination network.

The next grouping of statistics took the results one step further to determine the percentage of times the packets calculated TNTC values were higher than the actual time taken for the packet to arrive at the destination network. This value shows the percentage of packets in the particular flow that are given higher than necessary TNTCN values. The danger with having a higher TNTCN value is that packets may be categorized into the critical or discard queue incorrectly because of this higher value. The reality for these packets is that they have arrived in a faster time than the TNTCN value had predicted. Using the same statistics collated in the first part of the experiment, all instances where the TNTCN value was larger than the actual time that was taken for the packet to reach the destination network were identified. The number of instances where this was found to be the case was then represented as a percentage of the total number of packets that were send during the collection period (no. of young packets / total number of packets).These results are shown in the Table 2.

Top No.	Flow 1	Flow 2	Flow 3	Flow 4	Flow 5
1	<0.001%	<0.001%	<0.001%	<0.001%	<0.001%
2	<0.001%	<0.001%	<0.001%	<0.001%	<0.001%
3	<0.001%	<0.001%	<0.001%	<0.001%	<0.001%
4	<0.001%	<0.001%	<0.001%	<0.001%	<0.001%
5	<0.001%	<0.001%	<0.001%	<0.001%	<0.001%
6	0.001013%	0.002112%	0.00129%	0.00135%	0.00285%
7	0.005282%	0.006001%	0.00609%	0.00724%	0.00585%
8	<0.001%	<0.001%	<0.001%	<0.001%	<0.001%
9	<0.001%	<0.001%	<0.001%	<0.001%	<0.001%
10	<0.001%	<0.001%	<0.001%	<0.001%	<0.001%
11	<0.001%	<0.001%	<0.001%	<0.001%	<0.001%
12	<0.001%	<0.001%	<0.001%	<0.001%	<0.001%
13	<0.001%	<0.001%	<0.001%	<0.001%	<0.001%
15	<0.001%	<0.001%	<0.001%	<0.001%	<0.001%
16	<0.001%	<0.001%	<0.001%	<0.001%	<0.001%
18	<0.001%	<0.001%	<0.001%	<0.001%	<0.001%
19	<0.001%	<0.001%	<0.001%	<0.001%	<0.001%
20	0.00341%	0.00244%	0.00407%	0.00553%	0.00285%

TABLE II
TNTCN ACCURACY - PERCENTAGE OF PACKETS WITH EXCESSIVELY HIGH TNTCN VALUES(CORRECT TO 3 DECIMAL PLACES)

In the all but three topologies, namely topology 6, 7 and 20, the percentage of packets that were given a higher TNTCN value than the actual time taken were below 0.001% of the total number of packets sent. In topology 6, the TNTCN formula calculated a higher TNTCN value than that was actually experienced for approximately 0.002% of the packets. In topology 7 and 20 this value was approximately 0.006% and 0.004% respectively. As can be seen in Table 2, the number of times the TNTCN formula generated higher values than those that were recorded in the simulation were extremely low.

4) Analysis and discussion or results: The results show that the TNTCN formula used in the experimentation, in the majority of cases ($> 90\%$), was on average, able to provide values that fell within the specified threshold sought by the PBS mechanism. In addition to this, the PBS mechanism also provided a bias towards generating lower TNTCN values than were actually experienced in all but 8 of the 100 streams tested. This suggests there is strong evidence to support the hypothesis, and that the given formula provides suitable values needed in the PBS mechanism. The formula used to calculate the TNTCN value could potentially be improved in order to provide more accurate TNTCN values. One way this could be achieved would be to incorporate an additional 6th metric into EIGRP, whereby queue depths are advertised and exchanged among EIGRP devices. By doing this, the need to try to estimate queuing times would be eliminated from the formula

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

calculation which would make this mechanism more accurate. While the TNTCN value could potentially be improved, the results categorically show that for the purposes of the PBS mechanism and this proof of concept experimentation, the current formula is adequate. The final issue discovered through this experimentation is that EIGRP assigns a delay value of 20000 microseconds to all serial links in the topology when calculating the delay metric. Using this value creates a very high TNTCN value and would lead to a large number of incorrectly categorized packets. While this is problematic, the reality is such that if high speed serial links are employed on networks, typically the delay value in EIGRP is configured manually in order to ensure EIGRP is given an accurate representation of actual network conditions. In such a case this PBS mechanism would simply make use of the manually configured delay value and this issue would be resolved.

B. Experiment 2 Effectiveness of the critical queue

The second experiment in this paper tested the effectiveness of the critical queue by exploring how many of the packets sent via the critical queue were received within their useful lifespan. The reasons for doing so are to ensure the function of prioritizing these packets performs its intended duty of ensuring packets reach their destination networks within the necessary time frame. If this is not achieved and packets that are placed into the critical queue arrive at the destination stale on a consistent basis, it will bring into question the effectiveness and benefits of the scheme.

1) Hypotheses: It is hypothesized that the majority of packets placed into the critical queue will arrive at their destination within their valid lifespan window.

2) Methodology: To carry out this analysis, the three topologies where the highest number of packets that were placed into the critical queue out of the 20 topologies used in the previous experiment were selected. The BOD was then modified in these three topologies to mark all of the packets that were placed into the critical queue using the COS field in the IP header for that packet. All packets which were placed into the critical queue were given a COS value of 8. The workstations in the topology were then configured to detect the COS marking when incoming packets were received. When a packet was received with a COS value of 8 in the IP header, the workstation then calculated if the packet was fresh or stale. This was done by checking the current NTP time with the lifetime information in the DCCP header for that packet. If the packet was found to be fresh a $global_fresh_pkt_rcvd_counter$ variable was incremented by one. Alternatively, if the packet was found to be stale a global $stale_pkt_rcvd_counter$ variable was incremented by one. In addition to these two variables, a third variable was also used which recorded the total number of packets placed into the critical queues on the BOD. This variable was called $critical_pkt_sent_counter$. This variable was used to ensure all packets placed in the critical queue were received in either a stale state or a fresh state. Each time a packet was placed into the critical queue, the $critical_pkt_sent_counter$ value was incremented by one. To simplify the analysis of results, only one BOD was used in the topologies in which this experimentation was performed.

Topology	Total Pkts Received Fresh	Total Pkts Received Stale	Total Pkts Reported by BOD	Percentage of Pkts Arrived fresh
14	6080	830	6910	87.988%
15	9970	1180	11150	89.417%
16	15330	1850	17180	89.232%

TABLE III
PERCENTAGE OF PACKETS FROM CRITICAL QUEUE RECEIVED ON TIME

3) Results: In Table 3, it can be seen that in all three topologies the number of critical packets that were received fresh was significantly larger than the number of critical packets that were received stale. Specifically, between approximately 87 and 89 percent of the packets transmitted from the critical queue were able to reach the final endpoint within their useful lifespan.

4) Analysis and discussion of results: While the results shown here indicate the PBS mechanism allows the majority of packets placed in the critical queue to be received within their useful lifespans, it must be noted that the network conditions are conducive to this occurring. It is simply not possible to compare the results that occur when the PBS mechanism is enabled to results taken from a standard DCCP topology where categorization does not take place. This is because the act of prioritizing packets has a compounding effect on all other subsequent packets in that flow. Direct comparison between a standard topology and a topology where the PBS mechanism is enabled would not be accurate because of the infinite variation that would occur, particularly due to changes that occur in congestion window rate values.

C. Limitations of the experimentation

One of the limitations of this experimentation is that the performance gains suggested in the results section of the experimentation are very much based on topology selected. During the simulation analysis phase, the results showed categorically that the level of improvement, as well any adverse effects this scheme caused, became almost entirely dependent on the network topology. While all efforts were made to generalize the result findings to make them applicable to a large a range of topologies, the results still should be treated as proof of concept rather than exactly what could be expected in any given topology.

An additional limitation with the experiments is that they do not take into consideration the impact EIGRP protocol has on the overall performance of the various streams. While EIGRP is designed to be as minimalistic as possible, there are overheads associated with its function that should be considered. This impact is not measured in any of the results above as the complexity of the multiple streams and contention between the streams was not explored on a per stream basis. While there would be some impact caused through the addition of EIGRP needed by the scheme, this impact is likely to be extremely minimal. The final limitation that will be discussed

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

in regards to the experiments was that the workstations were designed in such a way that they generated packets as fast as the DCCP protocol would allow them to transmit. This is not what would be expected in normal real world network operations as the application layer would likely limit the rate of transfer, especially in real-time applications. In theory, the volume of traffic generated by the 5 to 20 workstations would be representative of a much larger number of real world flows than depicted in the experimental topology.

D. Experimentation conclusion

The experiments above serve two main purposes. The first purpose of the various experiments was to validate that the PBS mechanism worked effectively, reliability and in a stable manner in a number of different topologies. The second main purpose was to identify scenarios and network conditions where the PBS mechanism would be most beneficial in improving DCCP performance. The results from the various experiments showed that both of these occurred and there is potential benefit that can be obtained through the implementation of the mechanism. In terms of reliability, the networks tested remained stable and throughout the simulations all links and nodes remained up. In these instances, the BOD devices remained stable and were able to service up to 20 simultaneous packet streams transferring data at rates governed only by the CCID3 protocol. Having stated these findings, it is important to note that this research did not explore topologies where changes to the topology structure occurred due to device or link failure during the statistical collection. Finally these experiments show that the selection of the probability based TNTCN value falls within the acceptable threshold and produced improvement to overall network performance. Now that the reliability and benefits of the mechanism have been shown, it is hoped a more advanced and accurate formula can be developed in future research. This would allow the 10% acceptable threshold to be greatly reduced. This will also mean greater accuracy can occur in the categorization process of packets and thus improve the scheme.

VII. Conclusion

This paper has introduced a novel scheduling mechanism for the DCCP protocol. Specifically, the research has introduced a PBS mechanism that utilizes an array of routing protocol information to predict the time that it is likely to take for packets to reach their destination networks and pro-actively sorts and prioritizes packets based on this prediction. By placing packets into one of three queues, packets that are likely to become stale or that have already become so, are pruned or given a lower priority to ensure they do not have an adverse effect on fresh packets utilizing the same contended resources. Packets that need to be prioritized in order to avoid becoming stale are given the best chance of being delivered on time. Finally, the mechanism provides a deterministic queue that ensures the majority of normal packets remain largely unaffected by the actions occurring in the other two aforementioned queues. Detailed discussion as to how the mechanism should be implemented have been presented which was followed by two experiments that were carried out to examine the accuracy of the PBS mechanism. These experiments showed that the proposed mechanism is stable, reliable and capable of offering benefit to CCID3 controlled DCCP. This research concludes that through the results obtained in the proof of concept implementation of the mechanism, not only is the mechanism workable, it is also provides highly accurate TNTCN predictions. In addition, the results also show that the mechanism is able to ensure delivery of packets placed into the critical queue which would likely have become stale had the intervention of the mechanism not taken place. While there is still some work required to optimize the mechanisms efficiency, the objectives of this paper to showcase this novel probability based packet optimization mechanism through proof of concept implementation and modelling have been achieved.

References

[1] E. Kohler, M. Handley, S. Floyd, and J. Padhye. "RFC 4340: Datagram congestion control protocol (DCCP)." Technical report, IETF, Request For Comments, 2006.

[2] D. Farinachi, Introduction to enhanced IGRP (EIGRP). Cisco Systems Inc Press, 1993.

[3] A. Gurtov and R. Ludwig, "Lifetime packet discard for efficient real-time transport over cellular links." *ACM SIGMOBILE Mobile Computing and Communications Review*, 7(4) pp. 32-45. IEEE, 2003.

[4] C.W. Yuen and O.C. Yue, "Channel state dependent packet discard policy for 3g networks." *In Vehicular Technology Conference, 2006. VTC 2006-Spring.* IEEE 63rd, volume 1, pp. 405-409, 2006.

[5] K. Chebrolu and R.R. Rao, "Selective frame discard for interactive video." *In Communications, 2004 IEEE International Conference on*, volume 7, pp. 4097-4102. IEEE, 2004.

[6] S. Liang and D. Cheriton, "Tcp-rtm: Using tcp for real time multimedia applications" *In International Conference on Network Protocols*, 2002

[7] I. Radovanovic, R. Verhoeven, and J. Lukkien, "Improving tcp/ip performance over last-hop wireless networks for streaming video delivery." In *Consumer Electronics, 2007. ICCE 2007. Digest of Technical Papers*. International Conference on, pp. 1-2. IEEE, 2007.

[8] R. Albrightson, JJ Garcia-Luna-Aceves, and J. Boyle, "Eigrp-a fast routing protocol based on distance vectors." In *Proc. Networld/Interop*, volume 94, pp. 136-147, 1994.

[9] Low Latency Queueing. 2001 [cited 2010 05/07/2011]; *Available from*: $http : //www.cisco.com/en/US/docs/ios/12_0s/feature/guide/fsllq26.html$.

[10] S. Floyd, E. Kohler, and J. Padhye, "RFC 4342: Profile for datagram congestion control protocol (DCCP) congestion control id 3: Tcp-friendly rate control (TFRC)". Technical report, IETF, Request For Comments, 2006.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

IQMESH, Technology for Wireless Mesh Networks: Implementation Case Studies

Vladimir Sulc
MICRORISC s.r.o.
Jicin, Czech Republic
sulc@microrisc.com

Radek Kuchta | Radimir Vrba
Faculty of Electrical Engineering and Communication
Brno University of Technology
Brno, Czech Republic
kuchtar | vrbar @feec.vutbr.cz

***Abstract* – in this paper, IQMESH, a new networking technology for wireless mesh networks, its basic principles and related routing algorithms are presented. The presented technology was developed especially for applications in the field of buildings automation and telemetry. However, other applications such as smart grids or street lighting can also benefit from straightforward implementations and low system resources requirements. An implementation for IQRF communication platform, described at the end, shows actual system resources requirements in a specific scenario limited by 240 hops and 65 thousand devices.**

Keywords-wireless; mesh; networking; routing; algorithm; IQRF; IQMESH;

I. INTRODUCTION

IQMESH is a networking technology developed for Wireless Mesh Networks (WMN) utilizing packet transmissions. In such networks, messages are sent in smaller parts called packets. A packet has information about a recipient and, in general, the mesh network. This is transmitted from a sender to its recipient through nodes connected in the mesh network. A strategy for sending packets from one node to another is commonly known as routing, and its goal is to deliver packets efficiently and reliably.

A mesh network in which every node has a direct link to another node is a fully connected mesh network. In real WMN only partially connected mesh networks are used, which means that there is no universal direct link between devices. As an illustrative example of such mesh network topology and related routing is to compare it to vehicles traveling between cities. The whole path from the origin to the destination consists of numerous individual roads connecting cities. Searching for the best connection between two cities is similar to mesh networks finding the shortest and most efficient path, from the origin to the destination, between two selected nodes.

A link can be established between any two nodes in a mesh network. In a network consisting of n nodes and one coordinator, the number of possible links will always be lower or equal to N_{MAX} calculated as (1).

$$N_{MAX} = \frac{n\,(n-1)}{2} \quad (1)$$

Devices in WMN communicate with each other wirelessly, so communication between two devices is usually limited by the communication distance, or so called range limitation of these two devices. Positions of devices in general WMN are not known and the range limitation can depend on many conditions, therefore the routing is usually a great algorithmic challenge to find the best path the packet should travel along. More nodes result in more possible links and consequently to more combinations of links.

Many different routing algorithms are used practically. A flooding, routing based on tables and random routing are just a few basic examples of such algorithms. Unfortunately, due to the many specifics of WMN and limitations of the target, application is not possible to easily utilize such algorithms.

Flooding in a general mesh network is based on propagation of the packet over the whole network and is to be considered as the most reliable for WMN. Real implementations of WMN in industrial, scientific and medical ISM radio bands are limited physically by the connection speed, the so called bit rate, resulting in big delays and low network responsiveness.

Routing algorithms based on the sharing and distribution of routing tables or vectors are usually considered to be the most efficient. High memory demands and big overload in the case of distribution routing tables usually limit usage of such algorithms for larger WMNs where resources of nodes controllers are not limited by the economy of the project.

Possible packet collisions in connection with lower bit rates limits real implementations of random routing in WMN and practically disqualifies telemetry and control applications where the highest reliability should be achieved.

WMNs are nowadays considered and already used as a communication platform for many different applications in the field of telemetry and automation. Automatic meter reading AMR, street lighting control or smart grids are just few examples of such applications utilizing networks with hundreds or thousands of devices. Therefore both the cost of communication devices and high reliability of the routing should be priorities. Technology described in this paper provides both reliable and effective packet delivery solutions with minimal demands on system resources, it is an extension of the paper discovering technology at ICN2012.

II. RELATED WORK

Efficient and reliable packet delivery in large wireless networks consisting of hundreds or even thousands of devices and supporting up to several hundred hops is a big algorithmic challenge. Considering actual speed, output power and spectrum limitations, as well as economic factors, a flooding mechanism seems to be the most viable for most target applications.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

Therefore, flooding is commonly used in wireless ad-hoc networks. There are many techniques of flooding differing in control algorithms, efficiency, reliability and overhead.

The simplest flooding technique is based on re-transmitting only new, not yet registered packets. In this scheme every packet should be identified and is re-transmitted only once. This mechanism guaranties that a packet is delivered to all nodes at minimal costs. In the real environment of a WMN, collisions would affect functionality and result in high traffic [1]. Reducing flooding traffic is the goal of many approaches to make the flooding mechanism more reliable and efficient [1]. Proposing a probable flooding scheme, e.g., distance-based, location-based or cluster-based flooding.

Schemes in category 1-hop neighborhood are based on knowledge of the closest neighbors reachable directly in one hop. Different approaches [2-5] are based on 1-hop neighborhood knowledge. Cai et al. [2] propose adding the list of its 1-hop neighborhood to the packet, and recommends to the receiver not to forward the message if its complete 1-hop neighborhood is already included in the received list.

Schemes in the category 2+ hops neighborhood are based on storage of the neighborhood, which is limited by the number of hops from each node. In this category every node knows the network topology up to n-hops. In [6] Qayyum et al. propose a heuristic algorithm to compute multiple relays. Ko et al. in [7] propose improving broadcast operations for ad-hoc networks using 2-hop connected dominating sets.

Spohn et al. in [8] argue by simulations that protocols focused on making an optimal broadcast tree with the implicit assumption that all nodes should be reachable from the source may no longer be true because flooding protocols in wireless sensor networks are used to deliver data packets towards a single, or only subset, of destination nodes and proposes a new flooding protocol for utilizing directional information to achieve efficiency in data delivery.

Time Synchronized Mesh Protocol, described in [9] is based on TDMA and requires sharing time information and precise time synchronization of all nodes. Overall power consumption would benefit from the time synchronization, but on the other hand interference, collisions and environment influences would impact delivery reliability.

III. IQMESH

IQMESH is the networking technology developed for WMN with a coordinator and utilization of packet transmissions. Reliability is achieved through a flooding mechanism, collisions are avoided by TDMA and its routing efficiency is based on the virtual routing structure VRS, created by the coordinator during discovery. The following paragraphs provide a step by step analysis of particular parts of the IQMESH technology.

A. Basic principles

WMN is a general network of devices connected wirelessly. Every device in the network has some unique information (address) enabling addressing inside of the network - MAC, node ID, index, address, etc. Packets sent in such network include the address of the recipient. The principle of IQMESH technology is to extend this addressing space and define a new virtual routing structure in the network. A coordinator will dedicate to every device, found during discovery process, a unique Virtual Routing Number VRN, which will be used for future routing. Figure 1a shows an illustrative example of a standard network, where its nodes can be addressed by their address N1 – N5, after discovery, additional routing information is added as shown in figure 1b. Only VRN are used for the routing, while devices' addresses are used solely for addressing. Flooding and other routing algorithms can benefit from systematic indexing of nodes by VRN, e.g., if the VRN reflects distance by the number of routing hops from coordinator to the node.

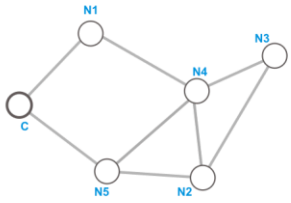


Fig. 1a Network example

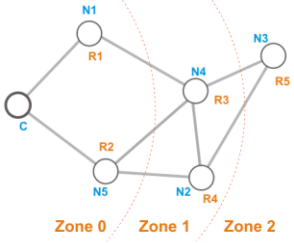


Fig. 1b Network after discovery

B. Discovery

During discovery, the coordinator seeks out nodes connected to the network and dedicates them to a unique VRN, reflecting their distance from the coordinator. For example, an incremental indexing can be used. The coordinator starts to search its neighbors. All devices responding to its "Answer Me" type message will receive their VRN. Based on received response it is assumed that the link between coordinator and responding nodes should be symmetrical, enabling future routing both to and from the coordinator. All nodes directly responding to the coordinator have a direct link to the coordinator, so they should belong to Zone 0, being directly accessible without routing. Then, the coordinator incrementally asks all nodes from Zone 0 to discover their 1-hop neighborhood and then dedicates a VRN to all newly found nodes which have not been found in the previous step, and thus not belonging to the Zone 0. Each node can also store some additional information in this step, e.g., respective zone number, parent's VRN, parent's network address or VRN of the first node in respective zone. This information would later be used for routing optimization. Processing all answers from nodes belonging to Zone 0, the coordinator will know all nodes belonging to Zone 1, which are nodes accessible to the coordinator by one routing hop. The same procedure will be then invoked recursively for all nodes belonging to zones Zone 1 and higher until all nodes are found or until there are some further zones available. At the end of discovery every found, and thus discovered, node has a unique VRN reflecting its distance from the coordinator. In typical applications such as smart buildings, telemetry systems and street lighting the discovery is made just once during the installation phase.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

C. Routing

The goal of packet routing in target applications is to reliably and efficiently deliver data over the network. In IQMESH based networks, the flooding mechanism is primarily used. VRS created during discovery process is directionally flooded. The network would be flooded from the coordinator to the node for all control purposes or from the node to the coordinator for data collection. A special order of VRS together with TDMA enables a directional, efficient and collision free flooding mechanism based on VRN. Every node routing packet in its dedicated time slot will also add to the packet its own VRN_X enabling other nodes to know and consequently synchronize to their respective time slot. The coordinator uses VRNC equal to 0. The network routing mechanism is illustrated in Fig. 2.

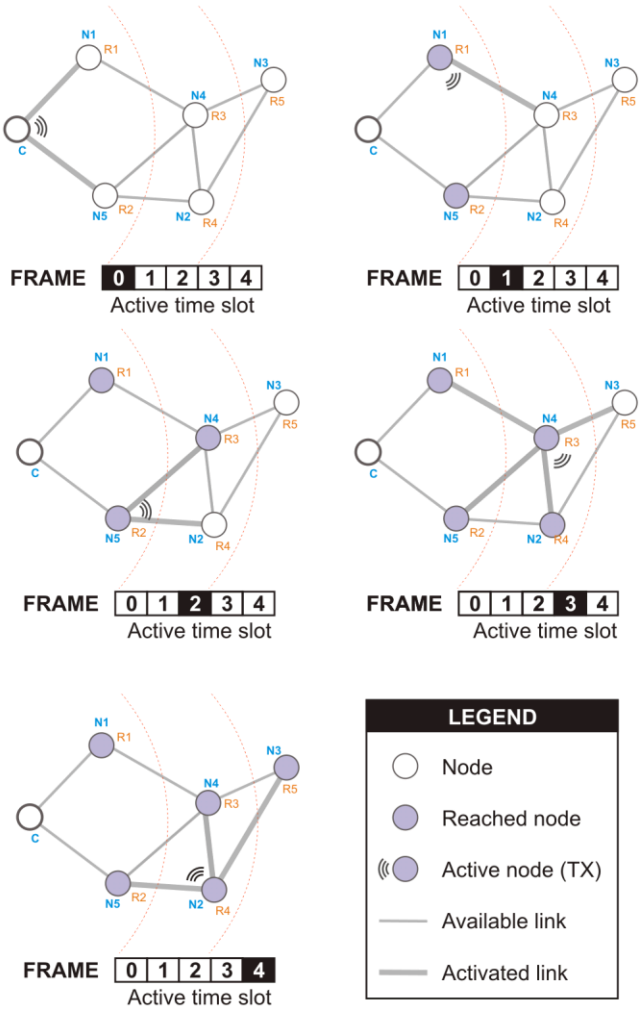


Fig. 2 Routing mechanism

D. Time synchronization

In contrast to many other techniques, e.g., [9], there is no need to share precise time information over the network. Routed packets keep track of number of hops made. This number corresponds with the respective time slot based on VRN for every device. This mechanism, together with adding information of length of time slot, enables efficient re-synchronization of all devices based on packet reception. In addition, dynamic timing based on packet length can be supported.

E. Resources

Generally, the coordinator needs to hold the information of the discovered nodes, whereas each node should keep VRN information only. This means no special HW resources are required to implement IQMESH technology. Storing some additional information recognized during discovery, as described in future paragraphs, can dramatically increase the efficiency of the routing. Specific IQRF implementation showing real system requirements will be mentioned further on.

F. Optimizations

Additional system resources dedicated to the coordinator or to each node can increase future routing efficiency. Storing the address or VRN of a parent node by each node could be mentioned as a good example. In such case, every node can reach the coordinator quite efficiently by tree topology via parent nodes. Also, the storing of minimal VRN in the respective zone, or equivalent saving of the maximal VRN of the previous zone by each routed node can be used to increase routing efficiency. Software techniques used during discovery or following discovery process can also increase the final efficiency of the routing, e.g., the coordinator can exclude all leaf nodes without child links from VRS during discovery.

G. Reliability

The described IQMESH technology ensures reliable and efficient directional flooding. Due to the expected redundant links in many real WMNs, it dramatically increases reliability. A temporarily lost link will obviously not cause failure of packet delivery. Routing mechanisms making use of underlying TDMA avoid conflicts as every routing node has just one dedicated time slot corresponding to its VRN. Many tests with environment noise simulations have confirmed reliability increase. Usually only noise generated during time slots dedicated for devices without redundant links can cause failure in packet delivery. Noise generated during the first time slot usually affects delivery, as no redundant links have yet to be created, which is a first time slot issue FTSI. Overall performance in standard office building environments was measured and a fail rate based on several weeks of experimentation resulted in 1 not delivered packet from 17 250 transmissions. Two additional slots for the coordinator were used to fix the FTSI.

H. Efficiency - routing from coordinator

As redundant paths resulting from the principle of VRS flooding are expected and packet collisions are eliminated by TDMA, it might not be obvious to use reception acknowledgment during flooding. However, this assures fair routing efficiency without any impact to reliability. Based on

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

the TDMA, flooding routing realized via VRS for nodes with $VRN_X < VRN_A$, where VRN_A is VRN of addressed node, and assumption that every node is addressed in the same frequency, the average frame will consist of time slots, where n is the number of nodes in a WMN:

$$T_{AVG} = \frac{\sum_{k=1}^{n} VRN_k}{n} = \frac{n}{2} \quad (2)$$

Generally, blind flooding efficiency in similar cases would be calculated as a number of links to ensure 100% reliability of packet delivery. Comparing (1) and (2), we can see dramatic efficiency increase.

For any addressed node within the zone Z_x, only nodes belonging to previous zone Z_{x-1} should make the routing without any strong impact to the reliability. The resulting efficiency based on this presumption will be higher, but always dependent on the topology of the specific WMN. The following formula reflects expected system efficiency for such a scenario:

$$T_{AVG} \leq \frac{n}{2} \quad (3)$$

Efficiency of this routing scenario, skipping redundant routes by nodes in the same zone Z_x for specific node X, can be expected as (4), where VRN_{Zx-1} is VRN of the last node related to the zone Z_{x-1}. Based on this principle, all nodes related to the zone Z_0 can be addressed directly without routing.

$$T_{AVG} = VRN_{Z_{x-1}} \quad (4)$$

I. *Efficiency - routing to coordinator*

As the matter of fact, there is information about parent nodes recognized and stored during discovery of every node. This information means that there is a tree topology available for routing packets from nodes to the network coordinator. In such a scenario, every routing node in its time slot sends a packet exclusively to its parent. Therefore in using TDMA, the number of time slots is equal to the number of hops and, for each frame, corresponds to the zone number Z_X for the node originating communication. Assuming Z_{MAX} as a maximum zone number in the network while indexed from 0, it would be generally proven:

$$T_{AVG} \leq Z_{MAX} \quad (5)$$

Reliability increase is mostly preferred in typical WMN applications. Oriented flooding with redundant backup paths can be used for such applications. In this case, each node originating communication to the coordinator will use its own VRN_X number as a limit of hops. For such routing, similar efficiency like (2) is expected.

Avoiding redundancy of routing by nodes from the same zone, routing efficiency for a specific node X can be expressed similarly like in formula (4), where is a VRN of the last node related to the zone Z_{X-1}.

IV. IQMESH IMPLEMENTATION IN IQRF PLATFORM

IQRF is the communication platform and related technologies [10]. The name IQRF stands for an Intelligent Radio Frequency. Basic specifications and early designs were made in 2004, when, in Malaga, Spain, the first integrated modules were introduced. IQRF is the platform integrating a variety of components for building LR-WPAN in an easy way, simplifying and shortening the design phase of a wireless communication system. Specific implementation of IQMESH routing technology will be described in following paragraphs.

A. *Network abstraction and limitations*

IQRF platform addresses mainly LR-WPAN applications, such as building automation systems and telemetry systems. In such applications many devices can be connected in a fixed infrastructure, usually created during the installation process, and provides permanent links to other devices in this infrastructure. There are commonly other devices connected in the network without permanent links to the other devices in the network, e.g., because of the mobility of such devices or because of power supply limitations. Based on these criteria, IQRF abstracted network topology is described in Figure 3.

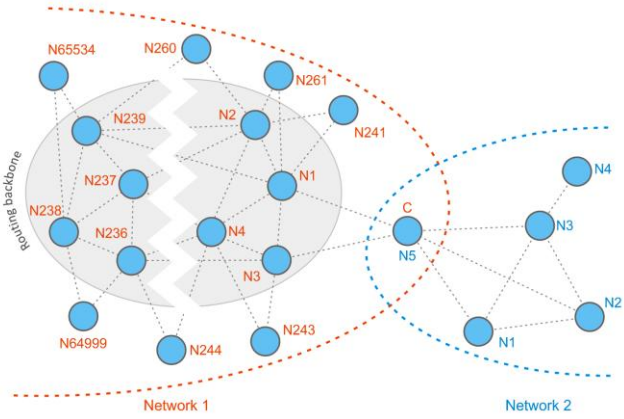


Fig. 3 IQRF abstracted network topology

The routing backbone usually consists of devices providing permanent links and bonded to the network during the installation process. Typical examples for such devices would be equipment like gateways, actuators or sensors mounted in specific locations. Such devices can be used as a routing backbone. IQMESH defines Virtual Routing Structure (VRS) to be used for directionally oriented flooding. VRS would be created from such infrastructure during the installation process (or later on). IQRF OS available on every transceiver module provides functionality for automation of such processes through function *discovery(x)* instigated by the coordinator and via servicing of system packets by node devices.

As visible from Figure 3 and description available in [10], several limitations were applied to increase efficiency of the network. Routing backbone consists of up to 239

devices, allowing very efficient one byte addressing, broadcast and also group messages. VRS is created from routing backbone devices very simply by calling *discovery(x)* function, where parameter *x* limits the number of zones, which is related to a depth of the network. Other devices will receive logical network address 0xFE from the coordinator during bonding. To extend the addressing scheme for addressing 65k devices, additional two byte user addresses are used. Every IQRF device can work simultaneously in two wireless networks, further extending the network or chaining of several networks allowing the possibility to build up larger systems.

IQRF specific implementation of IQRF OS 3.0x on TR-5xx transceiver modules supports several non-routing and routing schemes. A specific routing scheme is chosen in application based on requested efficiency and purpose by setting system variable *RTDEF*. Routing based on network logical address, tree routing to the coordinator and routing based on VRS are three basic routing schemes supported by current implementation of IQRF OS. Addressing in the network is realized via the logical network address obtained during bonding which is one byte long or via a two byte user address dedicated by the user by calling function *setUserAddress(x)*.

B. Dynamic time slots

To avoid conflicts within the network during routing, TDMA is used. One frame can include up to 240 time slots, allowing for up to 240 hops in the network. In IQRF, time slots are measured and set up in ticks, every tick is 10 ms long. As data load in the packet can consist from 1 byte up to 128 bytes, and 19.2 kb/s is the typical bit rate used for transmissions, 1.2 kb/s up to 115 kb/s are supported, the length of the frame would be too long if a fixed time slot is used.

Support of dynamic time slots based on the data load in the packet and requested purpose dramatically increased routing efficiency. Time slot is defined by setting the variable *RTSLOT* to the number of ticks convenient for a specific purpose. Polling request of the coordinator, e.g., can include just a specification of one or more nodes which should send data to the coordinator. In such a case, a minimum time slot 1 tick long can be used to propagate this request over the network, assuring delivery in 2.4 s in the worst case to any device. Time slot 5 ticks long together with simultaneous choosing of tree routing schemes can be used for a 128 byte long answer to assure maximum time efficiency.

Routing description, such as examples demonstrating routing and right parameters setup for specific purposes are available at User's and Reference guides, download-able from [10].

C. System resources

IQRF OS, including complete support both for the coordinator and nodes, is ported to TR-52Bx modules based on a PIC16F886 microcontroller. System resources used for routing and related services:

Program memory:	< 2k instructions
Data memory:	< 40 bytes (node mode)
	< 300 bytes (during discovery)
EEPROM:	< 40 bytes (node mode)
	< 2k bytes
Packet overload:	+ 6 bytes

D. Development, testing, results

The standard testing environment during development was based on a set of 200 node devices, each device including transceiver module TR-52BA inserted into a DK-EVAL-03 evaluation board and a GW-USB-04 In Circuit Wireless Programmer enabling bulk programming of all devices by one click and from one coordinator device consisting of transceiver module TR-52BA and CK-USB-02 universal programmer / debugger.

In-building applications mainly for lighting and dimming control realized in networks consisting of hundreds and thousands of devices confirmed reliability and the ability to work in real time. On the other hand, due to such a local environment, just a few hops were needed to cover the whole building.

The real challenge was street lighting control, covering large parts of towns, with networks composed of up to 200 devices with different networks using different channels to avoid spectrum concurrency. Fig. 4 shows one of such implementations realized in the suburb of Nitra, Slovakia, EU. Several kilometers range were covered by devices based on transceivers supporting only 3.2 mW of output power with small PCB antenna.

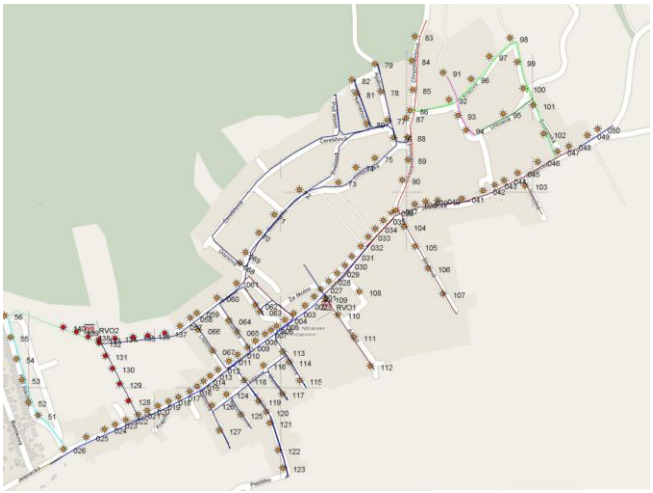


Fig. 4 Real implementation of street lighting application

V. IQMESH IMPLEMENTATION

Protocol implementation on specific hardware of TR-53BA transceiver modules will be briefly described and examples of the code will be presented to demonstrate easiness of use.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

A. TR-53BA module with built-in operating system

Transceiver module TR-53BA is a tiny electronic board with complete circuitry needed for realization of wireless RF connectivity. It is a basic communication component of the IQRF platform, used also in all IQRF gateways, routers and devices. Transceiver modules operate in the 868 MHz and 916 MHz license free ISM (Industry, Scientific and Medical) frequency bands. Modules can be used as a communication peripheral for any electronic device, or, due to the high integration, also as a controlling board for stand-alone applications.

High integration and functionality implemented in operating system dramatically reduce the time of application development, while ultra low power consumption predetermines modules for use in a battery powered applications. Mechanical concept of the module allows optional montage to any board equipped by inexpensive SIM card connector and also soldering to the mother boards when high mechanical stability is requested.

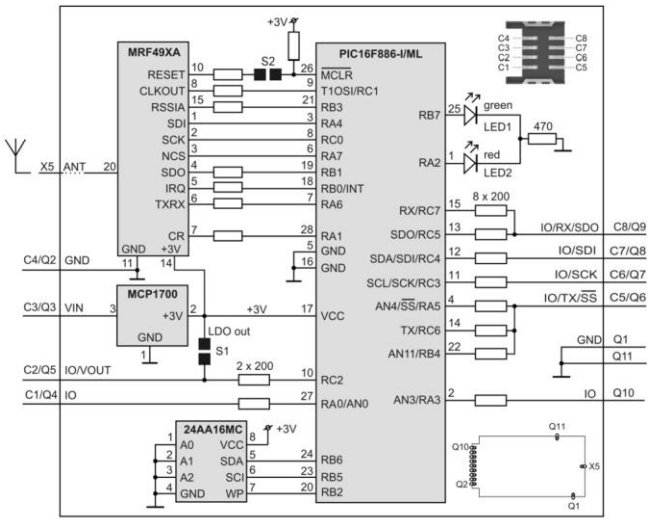


Fig. 5a TR-53BA simplified schematics

Fig. 5 shows simplified schematics and Fig 6 physical layout and main building blocks of the module. Microcontroller provides both input/output pins and basic interfaces for serial connection. Both RF and networking functionality is implemented in built-in operating system, so programmer can focus on his application only.

Detailed description of IQRF OS functionality, system architecture and both RF and networking related functionality are described in User's Guide [11] discovering packets description, bonding process, addressing techniques, description of implemented routing algorithms and examples. Reference Guide [12] provides detailed description of all available functions and shows their usage using basic examples.

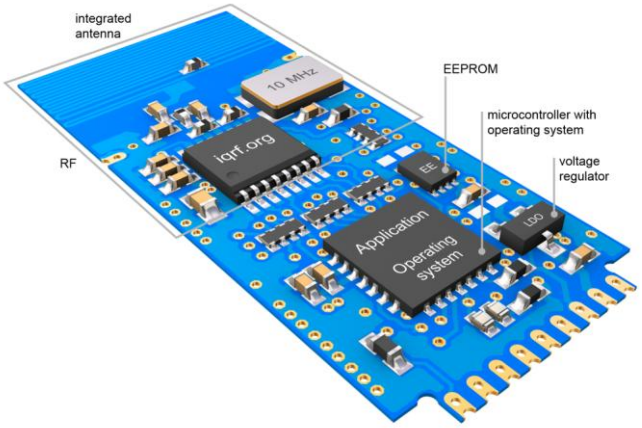


Fig. 6 TR-53BA physical layout

Operating system functionality related to wireless communication and networking will be described in future paragraphs and examples of the code will be provided for easier understanding.

B. Communication - packet transmission

As described in [11], IQRF OS has buffer oriented architecture. Sending a packet is realized by call of function *RFTXpacket()*, which takes *DLEN* bytes from *bufferRF* dedicated to RF communication and executes packet transmission. In the same way, packets receiving is realized via calling function *RFRXpacket()*. If some packet comes, data will be available in *bufferRF* and variable *DLEN* will hold the value indicating data length in the received packet. Code sequence Code 1 shows basic construction of packet transmission, code sequence Code 2 shows realization of simple receiver.

```
// Code 1 - packet transmission

PIN = 0;                // Peer-to-peer mode
DLEN = 10;              // 10 bytes TX definition
RFTXpacket();           // Transmits 10 bytes


// Code 2 - simple receiver

...

while (1)               // main loop
{
  if (RFRXpacket())
  {
     // ... code executed on packet reception
  }
  else
  {
     // ... code executed if timeout occurred
  }
```

C. Networking modes - Coordinator / Node

Besides peer-to-peer mode, networking mode is supported. Every device can work simultaneously in two different networks, in one network works like a Node, while in second network take roles of network Coordinator. This

feature can be helpful for building wider networks by chaining networks or their fragments together. Switching between networks is done automatically upon successful packet reception. In this case, the following transmission will be implicitly addressed to the network from which packet came. Code sequence Code 3 describes functionality related to the switching between networks and packets relaying.

```
// Code 3 - networking modes


if (RFRXpacket())
{
  if (!_NTWF)          // ignore non-networking
    continue;          //    packets

  x = getNetworkParams();

  if (x.7)             // Coordinator is polling data
  {
    bufferRF[0] = MYSTATUS;
    DLEN = 1;
    RX = 0;            // RX defines addressee (C now)
    RFTXpacket();      // data sent back
  }
  else                 // Nodes in Net1 sending data
  {                    // data relayed to Coordinator
    setNodeMode();
    RX = 0;
    RFTXpacket();
  }
}
```

D. Bonding to other networks

In *Coordinator mode*, transceiver is managing its network of bonded Nodes, while in *Node mode* it is controlled by different Coordinator. Nodes should be bonded to the network before being able to send packets addressed to this network. Bonding is based on Node's request confirmed by the Coordinator via exchanging RF system packets.

An example of bonding initiated by Node_01 is shown in Code 4a and 4b. In this example, Node with address 1, e.g. a specialized device as a remote controller, requests the Coordinator to bond a new device to the network. Bonding on Node's side is initiated by pressing a button in this example.

```
// Code 4a - Bonding on Coordinator's side

if (RFRXpacket())
{
  if (!_NTWF)          // ignore non-networking
    continue;          //    packets

  if (TX == 1)         // packet sent from Node_01
  {
    if (bondNewNode())
    {
      pulseLEDG();
     // ... code called if new bond created
    }
    else
    {
     pulseLEDR();
     // ... code called when bonding failed

    }
  }
}
```

```
// Code 4b - Node bonds to the network

if (buttonPressed)
{
  if (bondRequest())
  {
    pulseLEDG();
    // ... code called if successfully bonded
  }
  else
  {
    pulseLEDR();
    // ... code called when bonding failed
  }
}
```

E. Routing

IQMESH implementation in IQRF OS 3.00 supports up to 239 routing devices for a packet. Several routing algorithms are supported, allowing programmers to change them based on specific topology and needs, e.g. reliability, speed or response time. OS ensures that the packet is ignored by all devices except of the addressee and routing devices.

Routing algorithm should be specified with respect to reliability and speed requirements. Routing algorithm is specified by setting up variable *RTDEF*, defining requested routing algorithm.

Implicitly, every Node device can route packets. As it is not convenient at every case, e.g. devices with only one connection (leafs) or moving devices, routing can be selectively disabled or re-enabled for individual nodes. This functionality is realized via functions *setRoutingOn()* and *setRoutingOff()*.

Discovery is the process invoked by network Coordinator to search all devices belonging to its network and automatically dedicate Virtual Routing Numbers to Node devices. It is executed by calling function *discovery(x)*, where parameter *x* specifies maximum zone number of the network to be discovered. Discovery function is usually called during the installation process or for healing purposes, in the case of major topology changes. Standard mesh routing mechanisms used for routing uses redundant routes, so minor changes usually do not requests explicit healing process or rediscovery.

Code sequence shown in Code 5a demonstrates polling data from specific Node, while sequences in Code 5b and Code 5c show returning data from the Node back to the Coordinator. This example also shows loop mechanism used for addressed Node synchronization to avoid transmission conflicts.

```
// Code 5a - Coordinator polls data from the Node

setCoordinatorMode();
RX = 1;                // Node to be addressed
DLEN = 1;              // command sent to the Node
PIN = 0;
_ROUTEF = 1;           // routing requested
RTDEF = 2;             // DFM algorithm chosen
RTMAX = _HOPS;         // Number of hops specified
RTSLOT = 1;            // ticks dedicated for time slot
RFTXpacket();          // packet sent
```

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

```
// Code 5b - Node response by original routing scheme

if (RFRXpacket())
{
  if (!_NTWF)          // ignore non-networking
    continue;          //    packets

  if (TX != 0)         // allow only packets sent from
    continue;          //    Coordinator (address 0x00)

  if (bufferRF[0] != 1)// allow only polling command 1
    continue;          //    specified in the packet

  while (RTSLOT)       // synchronization to avoid
  {                    //    network conflicts
    waitDelay(RTDT1);
    RTSLOT--;
  }

  RX = 0;              // return back to Coordinator
  copyBufferCOM2RF();  // prepare data to the bufferRF
  DLEN = 64;           // 64 byte packet definition
  RTMAX = 0xFF;        // 0xFF translated to VRN
  RTSLOT = 5;          // each timeslot 5 ticks long
  RFTXpacket();        // sends packet
}


// Code 5c - Node response by TREE routing scheme

if (RFRXpacket())
{
  if (!_NTWF)          // ignore non-networking
    continue;          //    packets

  if (TX != 0)         // allow only packets sent from
    continue;          //    Coordinator (address 0x00)

  if (bufferRF[0] != 1)// allow only polling command 1
    continue;          //    specified in the packet

  while (RTSLOT)       // synchronization to avoid
  {                    //    network conflicts
    waitDelay(RTDT1);
    RTSLOT--;
  }

  RX = 0;              // return back to Coordinator
  RTDEF = _TREE;       // tree routing scheme
  RFTXpacket();        // sends packet
}
```

VI. CONCLUSION

IQMESH networking technology for wireless mesh networks, its basic principles and related routing algorithms were presented. Easiness of use and code efficiency was demonstrated on several examples. Specific implementation in IQRF communication platform was described. Like with any other technology or algorithm, IQMESH applications can benefit from technological advantages and would be affected by its limitations. The flooding scheme would be an excellent option for telemetry systems, e.g., AMR applications for water meters providing data just a few times a day or for street lighting applications. On the other hand, many redundant links and consequent time delays can generate difficulties in real time applications and missing support for node to node communication would create more programming work on the application layer.

VII. FUTURE WORK

Spreading over frequency spectrum instead of TDMA, a combination of both methods, achieving higher efficiency of the routing, increasing reliability in noisy environments and the usage of described technology are just few topics for future research. Future advanced data aggregation algorithms can benefit from VRS and routing schemes.

VIII. ACKNOWLEDGEMENT

This research has been supported and co-financed by the Ministry of Industry and Trade of the Czech Republic under contracts FR-TI1/058 "Project Smart House - Open Platform" and FR-TI3/27, project "Open Platform for Modern Cities".

[1] S. Y. Ni, Yu-Chee Tseng, Yuh-Shyan Chen, and Jang-Ping Sheu. The Broadcast Storm Problem in a Mobile Ad Hoc Network. ACM MOBICOM, pp. 51-162, Aug' 1999.

[2] Ying Cai, Kien A. Hua, and Aaron Phillips, “Leveraging 1-hop Neighborhood Knowledge for Efficient Flooding in Wireless Ad Hoc Networks”, 24th IEEE International Performance Computing and Communications Conference (IPCCC), April 7-9, 2005, Phoenix, Arizona.

[3] Xinxin Liu, Xiaohua Jia, Hai Liu, and Li Feng, “A Location Aided Flooding Protocol for Wireless Ad Hoc Networks”, MSN 2007, LNCS 4864, pp.302-313, 2007.

[4] H. Lim and C. Kim, “Multicast Tree Construction and Flooding in Wireless Ad Hoc Networks,” In Proc. of the ACM Int’l Workshop on Modeling, Analysis and Simulation of Wireless and Mobile System (MSWIM), pp 61-68, Aug. 2000.

[5] Hai Liu, Pengjun Wan, Xiaohua Jia, Xinxin Liu and Frances Yao, “Efficient Flooding Scheme Based on 1-hop Information in Ad Hoc Networks”, in proceedings IEEE infocom, Communications Society subject, 2006.

[6] A. Qayyum, L. Viennot, and A. Laouiti, “Multipoint Relaying for Flooding Broadcast Messages in Mobile Wireless Networks,” In Proceeding of the 35th Hawaii International Conference on System Sciences, 2002.

[7] M. A. Spohn, J.J. Garcia-Luna-Aceves, “Improving Broadcast Operations in Ad Hoc Networks Using Two-Hop Connected Dominating Sets,” In Proceedings of IEEE Global Telecommunications Conference Workshops, 2004.

[8] A New Directional Flooding Protocol for Wireless Sensor Networks Young-Bae Ko, Jong-Mu Choi and Jai-Hoon Kim, Information Networking. Networking Technologies for Broadband and Mobile Networks Lecture Notes in Computer Science, 2004, Volume 3090/2004, 93-102, DOI: 10.1007/978-3-540-25978-7_10

[9] Technical Overview of TSMP

[10] Datasheets and examples from http://www.iqrf.org [October 15, 2011]

[11] IQRF OS v3.00 User's guide at http://www.iqrf.org/ug [January 15, 2011]

[12] IQRF OS v3.00 Reference guide at http://www.iqrf.org/rg [January 15, 2011]

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

Automated Audio-visual Dialogs over Internet to Assist Dependant People

Thierry Simonnet
R&D department
ESIEE-Paris
Noisy le Grand, France
t.simonnet@esiee.fr

Gérard Chollet, Daniel Caon
CNRS-LTCI
TELECOM-ParisTech
Paris, France
{*chollet, caon*}*@telecom-paristech.fr*

Jérôme Boudy
TELECOM-SudParis
Evry, France
jerome.boudy@it-sudparis.eu

***Abstract*—An increasing number of people are in need of help at home (elderly, isolated and/or disabled persons and people with mild cognitive impairment). Several solutions can be considered to maintain social links while providing tele-care to these people. Many proposals suggest the use of automatic speech recognition (ASR) to control equipments and to maintain social link. In this paper, we will look at an environment constrained solution, its drawbacks (such as latency) and its advantages (e.g.flexibility, integration). A key design choice is to control equipments using a Voice over Internet Protocol (VoIP) solution with an ASR system, while addressing bandwidth limitations, providing good communication quality to obtain the best results for speech recognition. The resulting platform offers a powerful framework for setting up a virtual butler but also different services as voice controlled equipments and text transcriptions of medical talk.**

***Keywords*-ASR; Voice over IP (VoIP);**

I. INTRODUCTION

As part of ongoing projects, research has been conducted towards the implementation of efficient solutions for audio/video communications between people and system control, through an unified channel.

In Europe, there is an increasing demand [5], [9], [20], for maintaining dependent people at home, to reduce hospitals load, improve their quality of life and strengthen their social links. To this extent, a need for suitable communication systems and telecare technologies has arisen. Maintaining such people at home often requires medical assistance, excellent and reliable communication tools, handled by their relatives and the caregivers [17].

Nonetheless, several constraints have to be taken into account: These systems make use of an internet connection and as such they rely on bandwidth availability. Most European personal internet accesses use Asymmetric Digital Subscriber Line (ADSL) technology, offering a limited upload bandwidth. Still, in order to offer accurate and exploitable communication, image and sound quality must be maintained. Video quality is usually highly dependent on available bandwidth, and how different compression algorithms perform with limited bandwidth.

This article is organized as follows. We present environment, platform overview and explanation of our choices in Section 2, technicals descriptions in Section 3. A detail of technical integration is given in Section 4. Results are shown in section 5. Then the identified remaining issues and perspectives are discussed before concluding in Section 6.

II. REMOTE AUTOMATIC SPEECH RECOGNITION INTERFACE

A. Speech recognition

Speech is probably the most natural way that human beings employ to communicate between themselves, also being one of the most impressive system interfaces for human-computer interaction. The use of Automatic Speech Recognition (ASR) technologies becomes even more interesting when applied to the case of users who are not familiar with (or are physically/mentally unable to manage) the traditional computer interfaces. The potential of ASR Research includes domains from vocal commands to complex dialog systems, capable to identify one's state of mind and to detect distress situations.

B. Usage scenarios

Suppose that one always has access to a personal and collective memory-prosthesis, relayed by an audio-visual Butler.

There are different way to interact with the Butler. Here are samples of actions :

- find his way as the Butler embodied by a Smartphone, connected to GPS, knows our location and can guide us,
- record short video or photos to an album of his recent past,
- remember the name and other information about a person that we met, the Butler equipped with a camera takes a picture and finds this information,
- shopping, shopping list, prices, ...
- provide recipes, remember which menu were prepared for his friends, family, ...
- view and update their diary, appointments, bills to pay, planned parties,
- answer the phone, messaging, ...
- find informations on the web

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

- detect situations of distress, abnormal behavior through a wearable vital/actimetric sensors-based device [2], ...

Some of these features are already available on smartphones, others are being developed such as the Microsoft MyLifeBits project [8].

III. VoIP ARCHITECTURE AND SERVICES

A. Existing Platform

The current platform is composed of three parts: a master server, a smart home and a remote client. A master server, handling:

- Asterisk Internet Protocol Private Branch eX change (IP-PBX, or IPBX) for voice/video communications routing; [7]
- Julius ASR server (can be hosted by another server) [14].

Home, equipped with:

- a platform featuring a camera, a display and a VoIP client;
- various sensors for person monitoring;
- internet gateway (local IPBX).

A remote client system, basically a Personal Computer with a VoIP Client or a smartphone.

B. A Unified and standardized communication solution

Different kinds of media for different equipments need to be addressed. A VoIP solution offers a complete and unified communication infrastructure and then various services can be developed with it. This infrastructure can be used for a closed group of users but also be plugged on public VoIP network. This solution meets all criteria for medical/paramedical usage :

- supports various Internet infrastructures (e.g., public IP, private IP, ADSL box);
- interoperability with public and private (e.g., ekiga.net, google talk, skype) telecommunication networks;
- low latency (less than 100ms with H263 video) mandatory for remote control (robot, home automation);
- automatic internet bandwidth adjustment;
- single solution for videoconferencing, robot relay and the Smart Home control;
- support for various clients (e.g., softphones, IP phones, mobile phones, specialized softphones for remote control);
- choice of audio and video codecs;
- communication robustness;
- compatibility with IPBX call centers;
- ability to set up centralized services (low cost of deployment) as IVR (Interactive Voice Responce), ASR, multi-conferencing, voice and video messaging;
- unique identifier (phone number);
- centralization of data (voice, video);
- internationalization with customization of user language.

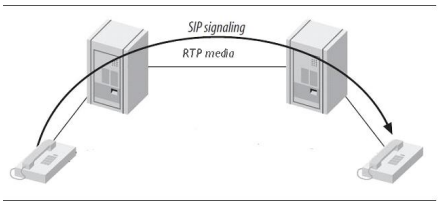


Figure 1. SIP trunking architecture

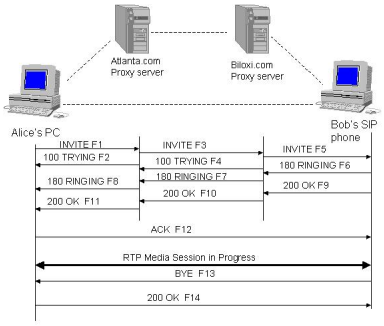


Figure 2. Call dialog

C. Communication infrastructure

Internet is the main communication media for the project. VoIP solutions imply the use of a PBX. The Asterisk PBX from DIGIUM Company will be used for the first version of the infrastructure. Asterisk PBX has standard configuration for classical communications but needs new and modified communication module for our purposes, repectively for voice and video transmissions. Patient network will use private IP addresses, then it will be necessary to have a local PBX to manage local communications and to act as a gateway to make or receive a call from public or private domains.

When a call is started, a SIP [23] request is sent to the PBX, which transmit it to the other client. When this signalling communication is done a direct one is established using RTP (Real Time protocol) [24]; (See Fig 2). This protocol, over UDP [25], keeps the packet order and drops old ones. Fig 3 shows how different components are set on ISO layers. To establish a communication between 2 private networks, it is necessary to use trunking services to allow all communications through PBXs (see Fig 1).

1) Codecs: A codec (Code-DECode) is a module that can Code and DECode an analog or a digital signal. For VoIP codec is used for norm but also for the module itself. X264 codec code and decode streams that use MPEG-4 AVC/H264 norm. PBXs are not designed for stream translation. A direct RTP communication is set between 2 clients and thus clients must have compatible codecs that respect norms.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

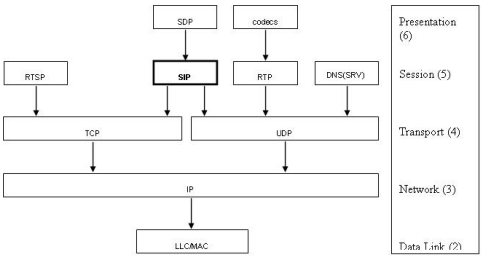


Figure 3. SIP and OSI

SIP Method	Description	RFC
ACK	Acknowledge final response to Invite	3261
BYE	Terminate a session	3261
CANCEL	Cancel a previous	3261
INFO	Mid-session signaling	2976
INVITE	Initiate a session	3261
MESSAGE	Allows the transfer of IMs	3428
NOTIFY	Event notification	3265
OPTIONS	Query to find the capabilities	3261
PRACK	Acknowledgement for Provisional responses	3262
PUBLISH	Publish event state	3903
REFER	Transfer user to a 3rd party	3515
REGISTER	Register with a SIP network	3261
SUBSCRIBE	Request asynchronous event notification	3265
UPDATE	Update parameters of a session	3311

Table I
SIP METHODS

Asterisk can handle :

- voice : ulaw, alaw, gsm, ilbc, speex [10], [22], g726, adpcm, lpc10, g729, g723;
- video : h261 [11], h263 [12], h263+, h264 [13], [19].

For our vAssist system, we can use: law, alaw, speex for voice and H261, H263 and H264 for video. The key point is using a well balanced setup between "compression", "delay" and "video quality". Increasing compression rate indeed increases the delay due to buffer use and a higher processing load per time unit.

2) Alarms: It is also possible to transmit alarms using SIP MESSAGE method. (see Table 1 for SIP Methods). Asterisk doesn't handle this method and it is necessary to implement RFC 3428 [26] for SIP channel. It is also possible to use T.140 (RFC 4103 [27]) method for Instant Messaging/Alarms communications. Both solutions could be implemented.

D. Services

1) Central PBX server: It is necessary to interconnect the central PBX to a call center with economical and security parameters. The following features have to be implemented:

- direct connection with an auto-connected client and using specific dial number;
- security in the possibility of encryption (e.g., VPN or stream encryption), OSP, closed group of subscribers for confidentiality. Management by phone number and not by identity;
- dialplan will transfer any specific local calls to the right call centre ;
- depending on partners' needs, we will develop dedicated modules for PBX (e.g., MP4, Interactive voice/video respond solution, RTSP (Real-Time Stream Protocol), Speech to text - ASR);
- no data duplication;
- ability to centralize all the data for exploitation or study purposes;
- patient home PBX is setup on a Plug computer for patient home use;
- a PBX module will be developed to handle a Speech to Text tool. This will allow when needed a direct transcription of calls for medical use including voice order handling. Such a centralized ASR (a great exploitation benefit) will handle multi-language tools and avoid unitary installation.

E. Performances

This unified platform is used for communication and videoconferencing purposes but also for robot remote control. Two different VoIP clients are used for performances and codecs compatibility : ekiga (www.ekiga.org) for PC platform and linphone (www.linphone.org) for PC and Android platform. These two clients are customized for HD and low delays communication. We use wideband Speex audio codec and H264 video codecs with a specific bandwidth adaptation module that reduces videoresolution in case of bandwidth congestion, keeping instant messaging and voice delays as low as possible. This solution allows low delays communication over internet (less than 100ms for a PC to PC communication over internet, less than 200ms for a PC to smartphone communication using WiFi). Tests with other standard VoIP clients and skype gave delays between 200ms and 500ms for long term communication (more than 3 hours long) and far less reliability (unexpected end of communication, freezing...). All these tests were done between 2 private networks with their own Asterisk IPBX, using trunking facilities to go through internet to demonstrate that it is possible to remote control a robot using standard Internet infrastructure.

IV. VOICE-BASED SYSTEM INTERFACE

A. ASR and VoIP

With such a centralized platform, ASR can be accessed using phone technologies facilities. Asterisk provides various speech tools but no embedded ASR tool. A proper commercial model, trained and annotaded by professionals costs. Open Source project Julius offers the services we need. Asterisk offers a generic speech API that is used with Julius. Julius can have input and output redirected to any socket. The aim was to have an Asterisk module

that can manage Asterisk speech functions, usable in a dialplan. The dialplan API is based around a single speech utilities application file, which exports many applications to be used for speech recognition. These include an application to prepare the ASR system, to activate the relevant grammar according to the context application context or menu and to play back a sound file while waiting for the person to speak.

We use app_julius module [15] developed by Danijel Korzinek and Dikshit Thapar. Dialplan Flow:

1) Create an ASR object using SpeechCreate()
2) Activate your grammars using SpeechActivateGrammar(Grammar Name)
3) Call SpeechStart() to indicate you are going to do recognize speech immediately
4) Play back your audio and wait for recognition using SpeechBackground(Sound File|Timeout)
5) Check the results and do things based on them
6) Deactivate your grammars using SpeechDeactivateGrammar(Grammar Name)
7) Destroy your speech recognition object using SpeechDestroy()

A simple macro is used in the dialplan to confirm word recognition. ARG1 is equal to the file to play back after "I heard..." is played.

```
[macro-speech-confirm]
exten => s,1,SpeechActivateGrammar(yes_no)
exten => s,2,Set(OLDTEXT0= ${SPEECH_TEXT(0)})
exten => s,3,Playback(heard)
exten => s,4,Playback(${ARG1})
exten => s,5,SpeechStart()
exten => s,6,SpeechBackground(correct)
exten => s,7,Set(CONFIRM=${SPEECH_TEXT(0)})
exten => s,8,GotoIf($["${SPEECH_TEXT(0)}" = "1"]?9:10)
exten => s,9,Set(CONFIRM=yes)
exten => s,10,Set(CONFIRMED=${OLDTEXT0})
exten => s,11,SpeechDeactivateGrammar(yes_no)
```

The voice-based MMI (Maximum Mutual Information) functionality uses a voice recognition module based on Julius and HTK softwares (Julius for recognition, HTK for training) with adaptation facilities to customize the system to the Elderly person use constraints and needs.

B. Julius, HTK

The voice recognition module is based on the use of conventional Hidden Markov Models (HMM) to model statistically the acoustic models of phonemes and / or words in the vocabulary. We use software tools such as HTK [21] and Julius [16]. Language models (linguistic probabilities, which are complementary to acoustic probabilities) are implicitly addressed in the use of such models to make robust word recognition in a given sentence (use of statistical N-grams and rules of grammar).

Sentence (repeated 10x by the same speaker)	Sentence totally correct(%)	Semantically correct(%)
hellep	100	100
help me	50	90
kom naar de keuken	100	100
kom eens naar de keuken	100	100
wil je naar de keuken komen	60	60
...	...	...
hector ik ga lunchen met kennisen	100	100
hector ik ga lunchen met buren	100	100
wolly ik ga uit eten	100	100
wolly ik ga uit eten met vrienden	100	100
wolly ik ga uit eten met kennisen	100	100
ja graag	40	80
Average percentage	86.39	94.44
Lowest percentage	40	60
Highest percentage	100	100

Table II
RECOGNITION RATES

C. Data, Adaptation

Adaptation strategies are implemented, especially those based on crossed multilingual adaptation between languages with rich phonetic materials.

Research of Tania Schultz [18], Rania Bayeh [3] and Gerard Chollet [6] on language independent and multilingual speech recognition, serve as a starting point.

V. EVALUATION

A first validation of the automatic speech recognition system was performed on data recorded in the European CompanionAble project (www.companionable.net). 22 Elderly Dutch speakers were recorded for one hour each in an experimental house (SmartHome) in Eindhoven. They repeated the phrases uttered by a "prompter', person reading at slow acoustic level the Prompt texts. Table II (containing 37 different phrases) gives results for one of these speakers, after adaptation of acoustic models by MLLR (classical technique of adaptation also studied in [4]) to its voice.

The rate of " semantically correct" is a way to describe that two sentences are at the same level of meaning (e.g., "help me" and "hellep"), so if "help me" is recognized instead of "hellep" the sentence is 100% correct (semantically) and voice dialogue can take place without problems.

The second software validation was conducted on 20 speakers (all seniors) without repetition of phrases, and the measurement is shown at the rate of word recognition. A classical MAP adaptation technique (also studied in [4]) was applied from a set of 10 adaptation sentences for each speaker. Figure 4 shows the results by language models and n-grams (2-gram and 6-gram), with and without adaptation of acoustic models.

Improvements are achieved through the hidden Markov models adaptation and the language model precision. We

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

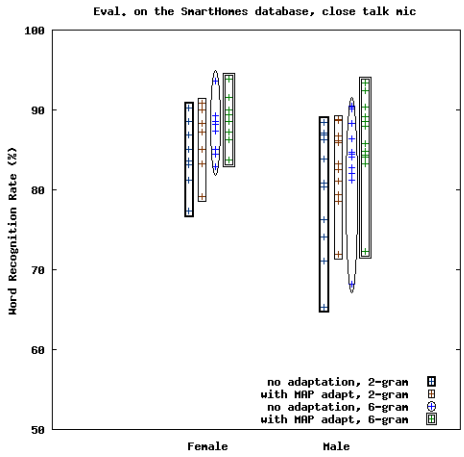


Figure 4. Evaluations of 20 elderly speakers (lavalier microphone).

also notice that not all users test the recognition system with the same success rate.

VI. Conclusion

The infrastructure for testing a mobile butler is in place. It uses both free software components for telecommunications (PBX - Asterisk) and for automatic processing of speech (Julius). Experimental results were obtained by automatic speech recognition of recorded data in the project CompanionAble. Under vAssist, a smartphone (Android) will be used [1]. The Asterisk server is ready for testing services related to usage scenarios listed in Section 2.

Today, telephony speech signals are generally sampled at 8kHz. First experimentations gave us good results but we need to work with higher-rates codecs (e.g., Speex 16kHz), better acoustic models and then finally to improve the platform from Narrowband to Wideband.

It is definitely interesting to achieve such a flexible level of communication using open source softwares. Although the need to work towards the modelization of more robust acoustic models for ASR (in order to increase the recognition rates), all the needed infra-structure is currently available and ready to make progress towards multiple kinds of applications, in many types of contexts (e.g., telemedicine, security, vocal commands, etc) that it is capable to handle.

Acknowledgment

Some parts of this research leading to these results has received funding from the European Commission Seventh Framework Programme (FP7/2007-2013) under grant agreement number 216487.

References

[1] Armstrong N., Nugent C., Moore G., and Finlay D., Using smartphones to address the needs of persons with Alzheimer's disease, Annales des Télécommunications, vol. 65, pp. 485-495 (2010);

[2] Baldinger, J.L. et al., Tele-surveillance System for Patient at Home: the MEDIVILLE system, 9th International Conference, ICCHP 2004, Paris France, Series : Lecture Notes in Computer Science, Ed. Springer, 2006. ;

[3] Bayeh, R. Reconnaissance de la Parole Multilingue: Adaptation de Modeles Acoustiques Multilingues vers une langue cible. Thèse (Doctorat) TELECOM Paristech, (2009);

[4] Caon, D.R.S. et al. Experiments on acoustic model supervised adaptation and evaluation by k-fold cross validation technique. In: ISIVC. 5th International Symposium on I/V Communications and Mobile Networks. Rabat, Morocco: IEEE, (2010);

[5] Clement, N., Tennant, C., and Muwanga, C.: Polytrauma in the elderly: predictors of the cause and time of death. Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine, v. 18, n. 1, p. 26, 2010. ISSN 1757-7241,http://www.sjtrem.com/content/18/1/26;

[6] Constantinescu, A. and Chollet, G. On cross-language experiments and data-driven units for alisp (automatic language independent speech processing). In: IEEE Workshop on Automatic Speech Recognition and Understanding. Santa Barbara, CA, USA: p. 606-613, (1997);

[7] Digium, The Open Source PBX & Telephony Platform, http://www.asterisk.org/.

[8] Gemmell, J., Bell, G. and Lueder, R., MyLifeBits: a personal database for everything, Communications of the ACM, vol. 49, Issue 1 (Jan 2006), pp. 88.95. http://research.microsoft.com/en-us/projects/mylifebits/

[9] Gitlin LN and Vause Earland T.. Améliorer la qualité de vie des personnes atteintes de démence: le rôle de l'approche non pharmacologique en réadaptation. In: JH Stone, M Blouin, editors. International Encyclopedia of Rehabilitation, (2011). Available online: http://cirrie.buffalo.edu/encyclopedia/fr/article/28/;

[10] Herlein G. et al., RTP Payload Format for the Speex Codec, draft-ietf-avt-rtp-speex-07, http://tools.ietf.org/html/draft-ietf-avt-rtp-speex-07, 2009.

[11] International Telecommunication Union, "H.261: Video codec for audiovisual services at p x 64 kbit/s", Line Transmission of Non-Telephone Signals, 1993.

[12] International Telecommunication Union, "H263: Video coding for low bit rate communication", SERIES H: Audiovisual and Multimedia Systems Infrastructure of audiovisual services, Coding of moving Video, 2005.

[13] International Telecommunication Union, "H264: Advanced video coding for generic audiovisual services", SERIES H: Audiovisual and Multimedia Systems Infrastructure of audiovisual services, Coding of moving Video, 2003.

[14] Julius ASR, http://julius.sourceforge.jp/en_index.php

[15] Korzinek, D, module app_julius, http://forge.asterisk.org/gf/project/julius/;

[16] Lee, A., Kawahara, T., and Shikano, K. In: EUROSPEECH. Julius - an open source real-time large vocabulary recognition engine. p. 1691-1694, (2001);

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

[17] Rigaud, A.S. et al. “Un exemple d’aide informatisé á domicile pour l’accompagnement de la maladie d’Alzheimer : le projet TANDEM”, NPG Neurologie - Psychiatrie - Gériatrie. N6, Vol.10, pp. 71-76, ISSN :1627-4830, LDAM édition/Elsevier, ScienceDirect, (April 2010).

[18] Schultz, T. and Katrin, K. Multilingual Speech Processing. Elsevier, (2006);

[19] Wiegand, T., Sullivan, G.J., Bjontegaard, G., and Luthra, A., Overview of the H.264/AVC Video Coding Standard, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, 2003.

[20] World Health Organization, 2002, The European Health Report, European Series, #97. ;

[21] YOUNG, S. J. et al. The HTK Book, version 3.4. Cambridge, UK: Cambridge University Engineering Department, (2006).

[22] Xiph.Org Foundation, ”Speex: A Free Codec For Free Speech”, http://speex.org/.

[23] SIP protocol, RFC 3261, http://www.ietf.org/rfc/rfc3261.txt

[24] RTP , RFC 3550, http://www.ietf.org/rfc/rfc3550.txt

[25] UDP, RFC 0768, http://www.ietf.org/rfc/rfc0768.txt

[26] UDP, RFC 3428, http://www.ietf.org/rfc/rfc3428.txt

[27] T.140, RFC 4103, http://www.ietf.org/rfc/rfc4103.txt

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

Automatic Conversion from CCM to HCM in State Transition Model

Akira Takura
Department of Career Planning and Information Studies
Jumonji University
Niiza, Japan
e-mail: takura@jumonji-u.ac.jp

Tadashi Ohta
IEICE Fellow

Tokorozawa, Japan
e-mail: myupapa@jcom.home.ne.jp

***Abstract*—As a program processing model for network software systems, the state transition model is well known. In the state transition model, there are two models: a central call model and a half call model. Each model has merits and demerits. So, if automatic conversion between the two models can be used, it would be convenient for system developers to design the system by adopting the appropriate model for the purpose of the design. A method of automatically converting from a half call model to a central call model has already been proposed. This paper proposes a method of automatically converting from a central call model to a half call model. The proposed method was applied to a three-way call service and it was confirmed that the conversion was correctly carried out.**

Keywords-state transition model; central call model; half call model; automatic conversion

I. INTRODUCTION

As a program processing model for network software systems, the state transition model has well been used [1][2]. The state transition model contains two models. One is called a central call model (abbreviated as CCM) where states of all terminals receiving a service are described in one state. The other model is called a half call model (abbreviated as HCM) where states of different terminals are described in different states, respectively. As states of all terminals receiving the service are described as one state in the CCM, service is easily understood. However, when the number of combinations for states of each terminal increases, the number of states in the CCM increases, resulting in increasing the size of programs implemented based on the CCM.

If an automatic conversion between CCM and HCM is established, the appropriate model can be applied for the purpose of the design. A method for automatic conversion from HCM to CCM has been proposed in [3]. In this paper, a method for automatically converting from CCM to HCM is proposed. The proposed method was applied to a three-way call service, call transfer service, call waiting service, and VoIP service, and it was confirmed that all state transition diagrams were correctly converted from CCM to HCM.

N. D. Grifeth proposed the method for creating a state transition diagram based on signals input to and output from the system whose function is unknown [4]. The created state transition diagram is based on CCM. S. K. Chakrabarti proposed the method for creating a state transition diagram to analyze functions of API [5]. Shimokura proposed the program described in a rule-based language , which is based on state transitions, for controlling networked robots [6]. The embedded program in AIBO, which is a robot made by Sony and was used in Shimokura's system, was designed based on state transition diagrams. Other researches of using a rule-based language are proposed by S. Sen, X. Fei, and L. Dongliang [7-9]. M. Ohba proposed the method for creating state transition diagrams based on CCM from programs described in a rule-based language [10].

In Section II, formal description of states is described so that computer processing is possible. In Section III, problems in converting from CCM to HCM such as conversion of states, state synchronization and reduction of states are described. In Section IV, solutions for the problems described in Section III are proposed. In Section V, an example of applying the proposed method to a three-way call service, main part, is described.

II. A FORMAL EXPRESSION OF STATE

For a computer to handle a state, the state of a state transition diagram has to be described formally. This is how it is done: Each state of the state transition diagram is described as a set of state description primitives (called primitive) which represent states of terminals which are receiving a service at that time [11][12]. A primitive consists of a primitive name which represents a state of a terminal and arguments which represent the terminals related to the primitive name. For example, when terminal A is in an idle state, it is described as idle(A). When users of terminals A and B are talking, it is described as talk(A,B). 'idle' and 'talk' are primitive names, and A and B are arguments, respectively. When terminals A and B are arguments of the same primitive, it is said that terminals A and B are connected. If terminals A and B are connected and terminals B and C are connected, then terminals A and C are connected. Thus, all terminals described in the same state of a state transition diagram are connected.

In a state transition model, where a state transition is decided based only on a current state and an event, the number of states becomes large. So, in the state transition model generally used, the state transition is decided not only on the current state and the event but also on state transition

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

conditions. State transition conditions are described, in the state transition diagram, in an analysis block below the current state. In this paper, state transition conditions are also described as a set of primitives. A primitive which is used only as a state transition condition and never used as an element of any state in the state transition diagram, is called, in this paper, a 'condition primitive'. On the other hand, a primitive which is used as an element of a state is called, in this paper, a 'state primitive'.

An example of a state transition diagram with two analysis blocks for a call waiting service is shown in Fig. 1. Four states and two analysis blocks are involved in Fig. 1. States are described as ovals, each of which has a set of primitives: dialtone(A), calling(A,B) {cw-calling(A,B), talk(B,C)}, and busy(A), respectively. Analysis blocks are described as triangles which have conditions: idle(B) and {m-cw(B), talk(B,C)}, respectively. When event dial(A,B) occurs at state dialtone(A), the next state is decided according to condition idle(B) described in the first analysis block. If idle(B) holds, the next state is calling(A,B). If idle(B) does not hold, the next state is decided according to condition {m-cw(B), talk(B,C)} described in the second analysis block.

III. PROBLEMS IN CONVERSION FROM CCM TO HCM

A. Conversion of States

In CCM, states of all terminals receiving a service are described in one state. But, in HCM, states of different terminals are described in different states, respectively. So, states in CCM should be converted so that states of different terminals are described in different states.

B. State Synchronization

In HCM, an event occurred in the system is independently accepted in individual state of state transition diagram related to a terminal which creates the event. More than one event might occur at one time. So, when the event is accepted in the state transition diagram, there is no guarantee that all terminals are at states which correspond to the same state of CCM. Thus, a mechanism which guarantees that all terminals are at states which correspond to the same state of CCM is needed.

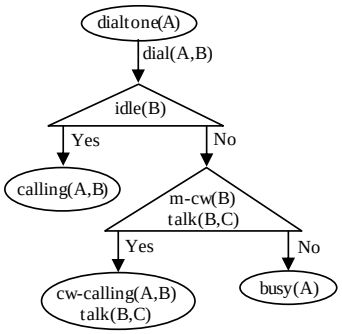


Figure 1. An Example of Two Analysis Blocks

C. Reduction of States

There is a case where, while a state of CCM is changed but a corresponding state of an HCM is not changed; there is a reduction of states. For example, in a three-way call service, when a user of terminal A, who is talking with a user of terminal B, wants to let a user of terminal C join the conversation, the user of terminal A pushes a flash button. Then terminal B is in hold state and is kept in the hold state until the three-way call is established. The state of CCM is changed but a state of terminal B in an HCM is not changed. Then, one state of terminal B in HCM corresponds to several states in CCM.

When the terminal, whose state corresponds to several states in CCM, creates an event, in HCM for terminal B, it has to ask all terminals which are receiving the service to distinguish to which state of CCM the state of terminal B in HCM corresponds.

IV. SOLUTIONS

A. Conversion of States

Each state in CCM is described as a set of primitives as mentioned in Section II. So, to convert a state in CCM to a corresponding state of terminal X in HCM, select primitives which have terminal X as an argument from the state in CCM. Thus the corresponding state of terminal X in HCM is gained as a set of the selected primitives. Primitives, that have more than one argument, appear in states of each argument in HCM. See Fig. 2.

B. Synchronization of States

First, a signal sequence for asking states of each terminal is proposed. Then, methods to synchronize states of each terminal are proposed based on the proposed signal sequence, in two cases; the state in the HCM where an event occurs is decided, and the state is not decided because of state reduction. The solution in the case that the state in HCM is reduced is described in *C*.

1) Signal sequence

In an HCM where an event occurs it has to ask current states of all terminals which are receiving the service by sending a signal to all terminals without a repetition. But, terminals to which the signal is sent are limited to terminals which are described in the current state of the HCM. In other word, the signal can be sent between terminals which are arguments of the same primitive described in the current state in the HCM.

First, based on a set of primitives in the current state of CCM, make a directed graph, by drawing arrows originating

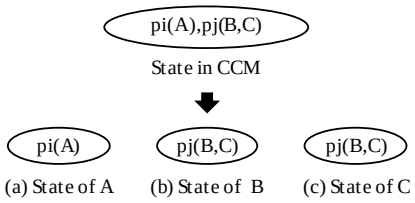


Figure 2. Converted States in HCM

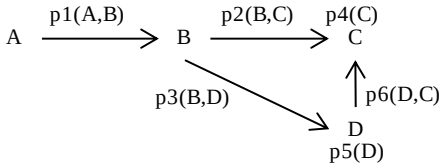


Figure 3. Graph Corresponding to a CCM State

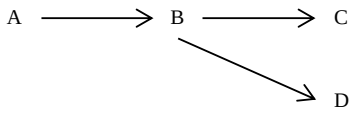


Figure 4. Extracted Tree from the Graph

from the terminal, which creates the event, to terminals which are described as arguments in the primitives where the originating terminal is also described as an argument [13]. As mentioned in Section II, all terminals described in the current state of CCM are connected. So, all terminals can be described as a node of the graph. Suppose the current state of CCM is {p1(A,B),p2(B,C),p3(B,D),p4(C), p5(D), p6(D,C)} and terminal A creates an event. In this case, the graph is shown in Fig. 3. A primitive, which has one argument, is described above or under the argument. A primitive, which has two arguments, are described above the arrow drawn between the two arguments.

Next, as the signal should be sent to each terminal only once, delete unnecessary arrows from the graph. Then delete primitives. Now, a tree which shows a signal sequence is obtained as shown in Fig. 4.

By sending the signal from the terminal which creates an event based on this tree, the signal can be sent to all terminals with certainty and without repetitions.

2) State of CCM is decided

In the case that the current state of CCM is decided, the HCM, where an event occurs, checks that all other terminals are in the states which correspond to the decided state, and synchronize states of all terminals. To realize the synchronization, the HCM corresponding to the root node in the tree shown in Fig. 4 sends an inquiring signal (REQ) to the next node and transits to a wait state. When the node, which receives REQ, is not in the corresponding state it sends back a signal NG to the former node which sent REQ. In this case, since there are no states which receive REQ, NG is sent from state analysis program. When the node which receives REQ is in the corresponding state and is not a leaf node, called an inner node in the tree, it transmits REQ to the next node and transits to a wait state to keep synchronization. If the node which receives REQ and is in the corresponding state is a leaf node, it sends back a signal OK to the former node which sends REQ, and transits to a wait state to keep synchronization.

The node, which receives NG, sends a reset signal to all the next nodes that have sent OK, sends NG to the former node, and transits to the current state. The node, which receives the reset signal and is an inner node, sends the reset signal to all the next nodes and transits to the current state. If the node is a leaf node, it transits to the current state.

When a node receives OK from the next node, if there is the next node to which REQ has not been sent, the node sends REQ to the next node. If there is not the next node to which REQ has not been sent there are two cases: the node is an inner node or the node is a root node. If the node is an inner node it sends OK to the former node. When the node is a root node, it sends a signal ST, which represents an instruction of state transition, to all the next nodes in parallel and transits to the next state. When an inner node receives ST it retransmits ST to all the next nodes in parallel and transits to the next state instructed by ST. When a leaf node receives ST it transits to the next state instructed by ST.

Note: The signal which instructs a state transition is not sent to the terminal which does not make a state transition nor retransmit the signal to another terminal.

C. State in HCM is Reduced

1) Process for synchronization

Information that each state in HCM corresponds to which state in CCM is saved in the process for converting a current state in CCM to individual states in HCM. For all terminals described in a current state of CCM, referring to the saved information, make an AND set of states in CCM to which a current state corresponds. Thus, the state in CCM, to which each state of individual terminals corresponds, can uniquely be decided. The signal to decide the unique state is sent based on the tree described in *B 1)*.

More precisely, the process is explained as follows: The root node sends REQ, which is a set of states in CCM corresponding to the current state of the node, to the next node as an inner event, and transits to wait state. An inner node which receives REQ, makes an AND set of received REQ and a set of states in CCM corresponding to the current state of the inner node to make a new REQ. Then, the inner node sends the REQ to the next node as a new inner event, and transits to wait state. If the AND set is null, a vacant set, the inner node sends back NG to the former node and transits to the current state. A leaf node which receives REQ, makes an AND set in the same way as the inner node. If the AND set is not null, the leaf node sends back the AND set as ANS to the former node, and transits to wait state. If the AND set is null, the leaf node sends back NG to the former node, and transits to the current state. The process hereafter is the same as one described in *B 2)* in this section under the condition of replacing signal OK with ANS and replacing "sending REQ" with "making new REQ and sending it". Thus, finally the root node can decide the corresponding state in CCM and can send ST to all the next nodes.

2) Example of signal sequence

Concrete examples of signal sequence are explained in the case where the current states of CCM are as follows:

S1: {p1(A,B),p2(B,C),p3(B,D),p4(C),p5(D),p6(D,C)}
S2: {p1(A,B),p2(B,C),p4(C)}
S3: {p1(A,B),p3(B,D),p5(D)}
S4: {p1(A,B),p2(B,C),p7(C)}

Terminals A, B, C and D are in states corresponding to S1, S2, S3, or S4 in CCM. But, for certain terminals, some

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

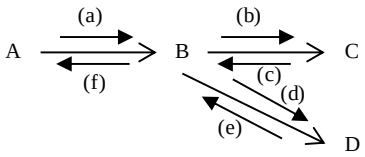


Figure 5. Tree Expression for S1

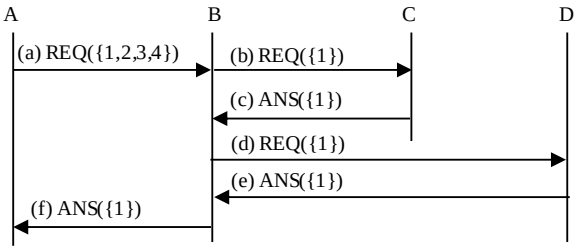


Figure 6. The Signal Sequence for S1

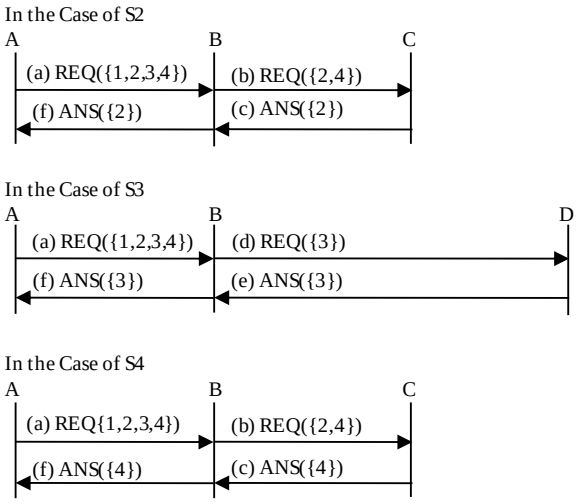


Figure 7. The Signal Sequence for S2, S3 and S4

states in HCM might be the same state. In this case, the terminal may not be able to decide to which state in CCM the current state of the terminal corresponds. In this example, terminal A is in the same state in HCM for S1, S2, S3, and S4 in CCM. Terminal B is in the same state in HCM for S2 and S4 in CCM. Terminal C is in the same state in HCM for S1 and S2 in CCM. So, when terminal A creates an event, in HCM terminal A sends {S1,S2,S3,S4} to the next node based on the tree described in *B 1)* in this section. Terminals A, B and C behave as described in the second paragraph in *C 1)* in this section. Thus, terminal A can distinguish to which state in CCM the current state of terminal A corresponds.

The tree expression for S1 is shown in Fig. 5. Signals are sent based on this tree. The signal sequence is shown in Fig. 6. (a): Since terminal A is in the state which corresponds to S1, S2, S3 and S4, terminal A sends {1,2,3,4} as REQ to terminal B. (b): As terminal B is in the state corresponding to S1, an AND set of REQ sent from terminal A and {1} is made, resulting in gaining {1} as new REQ. Terminal B sent {1} as REQ to terminal C. (c): Since terminal C is in S1, an AND set of REQ sent from terminal B and {1} is made, resulting in gaining {1}. As terminal C is a leaf node, it sends {1} as ANS to terminal B. (d): Since the node of terminal B has a branch to terminal D, terminal B sends {1} received as ANS from terminal C to terminal D as REQ. (e): By making an AND set of received REQ and its state which corresponds to S1, terminal D gains {1}. Terminal D is a leaf node, so it sends {1} as ANS to terminal B. (f): Terminal B sends {1} received from terminal D to terminal A as ANS. Thus, terminal A can recognize its state as S1.

Signal sequences for S2, S3 and S4 are shown in Fig. 7. Consequently, each state can be recognized.

D. Conversion for an Analysis Block

Primitives for plural terminals can be described in the analysis block. As far as the authors know, transition conditions for at most one terminal, other than a terminal which creates an event, is described in most of conventional telephone services. So, in this paper, the way of converting the analysis block is described in the case where transition conditions for at most one terminal, other than a terminal which creates an event, is described in the analysis block. The proposed method can be easily extended to the case where the analysis block has transition conditions for more than two terminals.

Analysis blocks of CCM are converted after conversion described in B and C in this section. In an analysis block conversion, if arguments of primitives represent the terminal where the event occurred, the analysis block is described at just below the current state in the HCM corresponding to the terminal. Otherwise, i.e. arguments of primitives in the analysis block are not the terminal, the way of converting the analysis block depends on whether primitives described in the analysis block are state primitives or condition ones as described below.

1) In the case of a state primitive

When the primitives described in the analysis block are state primitives, there is an argument which appears in arguments of the event. So, in the HCM where the event occurs, inquiry signal, REQ, to the HCM designated by the argument of the event is described below the event. In the HCM where REQ is received, REQ as an input signal and OK as a reply signal are described just below the state containing the primitive described in the analysis block of the CCM. And then, transition to wait state is described. Note that NG signal is not described in the HCM. This is because when the current state of the HCM cannot accept REQ, NG is sent by a state analysis program as mentioned in Section IV *B 2)*. In the HCM which receives OK, the OK signal is described as an input signal just below the wait state, and an instruction of state transition to the HCM, which sent the OK signal, is described.

Fig. 8 shows an example of converting a state transition diagram, where a state primitive p2(B) is described in the analysis block, from CCM to HCM by the proposed method.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

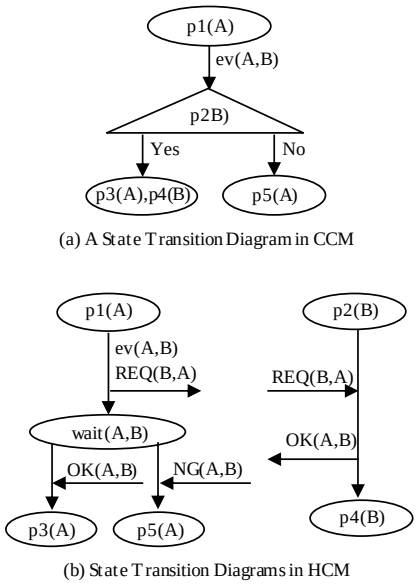


Figure 8. When a State Primitive is Described

2) In the case of a condition primitive

The analysis block is described just below the current state in the HCM corresponding to the first argument of the condition primitives. This current state of the HCM corresponds to that of the CCM. When the condition described in the analysis block is satisfied, OK is described as a reply signal to the HCM that sent REQ. When the condition described in the analysis block is not satisfied, the conversion is done in the following two cases: If, in CCM, the state of the terminal which received REQ does not change, in HCM, a transition to the current state is described. Otherwise, in HCM, a transition to wait state is described.

Fig. 9 shows an example of converting a state transition diagram, where a condition primitive pc(B) is described in the analysis block, from CCM to HCM. When terminal B is not in the state of {p2(B)}, NG1 is sent to terminal A. But, in HCM of terminal B, since NG1 is sent from the state analysis program NG1 does not appear in HCM of terminal B in Fig. 9.

3) In the case of both primitives

When both primitives, state primitives and condition ones, are described in the analysis block, the analysis block is converted in different ways depending on whether the both primitives have the same argument or not. If the both primitives do not have the same argument, the analysis block is described in the same way as mentioned in *1)* and *2)* above. If both primitives have the same argument, the conversion is carried out in two stages. First, REQ as an input signal is described just below the state consisting of the same state primitives described in the analysis block. Second, the analysis block is described just below the input signal, and the condition primitives, which have the argument corresponding to the HCM, are described in the analysis block. Input and output signals are described as described in *1)* and *2)*.

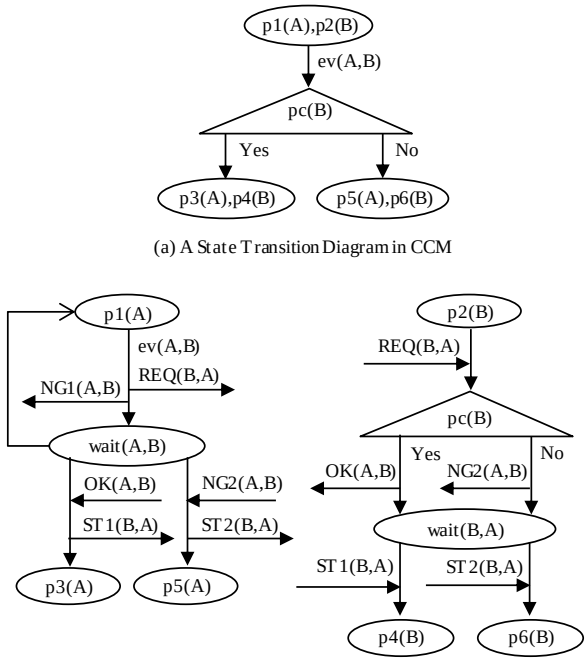


Figure 9. When a Condition Primitive is Described

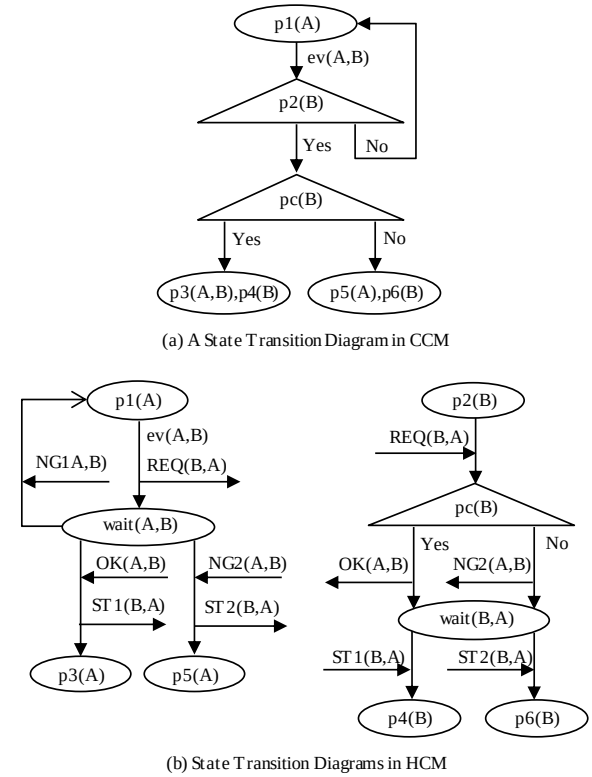


Figure 10. When State and Condition Primitives are Described

Fig. 10 shows an example of converting a state transition diagram, where a state primitive p2(B) and a condition

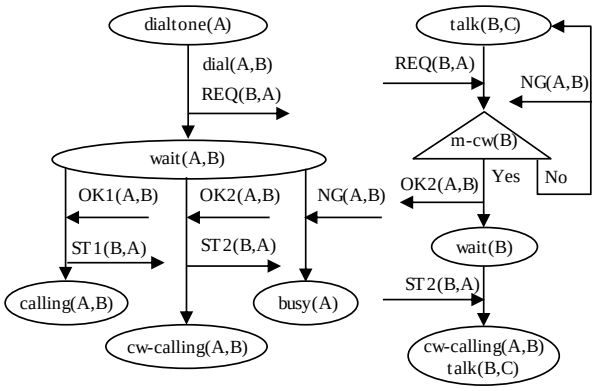


Figure 11. State Transition Diagrams in HCM (Call-Waiting Service)

primitive pc(B) are described in the analysis block, from CCM to HCM. When terminal B is not in the state of {p2(B)}, NG1 is sent to terminal A. But, in HCM of terminal B, since NG1 is sent from the state analysis program NG1 does not appear in HCM of terminal B in Fig. 10.

Fig. 11 shows the HCMs converted from the CCM of a call waiting service described in Fig. 1. Since, in Fig. 1, only the state of terminal A is described in the current state of CCM, the synchronization process described in *B* and *C* in this section is not described. OK1 is sent to terminal A when terminal B is in idle state. Namely this is a state transition for POTS. So, OK1 does not appear in HCM of terminal B in Fig. 11.

V. An Example of Conversion of State Transitions

An example of conversion for the three-way call service based on the method described in Section IV is shown.

A. State Transitions for Three-Way Call Service

The main part of state transitions for the three-way call service is described. Fig. 12 shows state transitions from two-party talk state (talk(A,B)) to three-way talk state (twctalk(A,B,C)) in the form of text.

Condition primitive m-twc(A) represents that terminal A subscribes to the three-way call service, flash(A) represents that flash button is pressed, hold(A,B) represents that terminal A holds terminal B, dialtone(A) represents that terminal A receives dial-tone, idle(C) represents that terminal C is in idle state, calling(A,C) represents that terminal A is calling terminal C, not[idle(A)] represents that terminal C is not in idle state, busy(A) represents that terminal A receives busy tone, dial(A,C) represents that a user of terminal A dials the number of terminal C, offhook(C) represents tresents that a user of terminal C hangs up the receiver.

```
S1 -> S2: {talk(A,B)} flash(A): {m-twc(A)} {hold(A,B),dialtone(A)}
S2 -> S3: {hold(A,B),dialtone(A)} dial(A,C): {idle(C)}
                    {hold(A,B),calling(A,C)}
S2 -> S5: {hold(A,B),dialtone(A)} dial(A,C): {not[idle(C)]}
                    {hold(A,B),busy(A)}
S3 -> S4: {hold(A,B),calling(A,C)} offhook(C): {} hold(A,B),talk(A,C)}
S4 -> S5: {hold(A,B),talk(A,C)} onhook(C): {} {hold(A,B),busy(A)}
S5 -> S1: {hold(A,B),busy(A)} flash(A): {} {talk(A,B)}
S4 -> S6: {hold(A,B),talk(A,C)} flash(A): {} {twctalk(A,B,C)}
          Syntax: {current state} event: {analysis block} {next state}
```

Figure 12. State Transitions of a Three-Way Call Service in CCM

The states of CCM described in Fig. 12 are the following six states. Note that S1 is a state of POTS (plain old telephone service).

S1 = {talk(A,B)},
S2 = {hold(A,B),dialtone(A)},
S3 = {hold(A,B),calling(A,C)},
S4 = {hold(A,B),talk(A,C)},
S5 = {hold(A,B),busy(A)},
S6 = {twctalk(A,B,C)}.

B. Conversion of Current States in HCM

Fig. 13 shows the converted current states in the HCM for terminal A, B and C from the current states S1 … S5 described in the CCM as mentioned in Section IV *B.*

C. Conversion to Tree Representation

Tree representations obtained from the current states S1 … S5 described in Section IV *B* are shown in Fig. 14. In the representation, the root represents the terminal where an event occurs. Terminal A, B, and C are represented as nodes.

D. Conversion of State Transitions

The state transition diagrams in HCM converted from the state transition diagram in CCM described in Section V *B* are shown in Fig. 15. State transitions for synchronization are omitted.

Since there are various types of state transitions in Fig. 12, in addition to processing of conversion of states and synchronization of states, some conversion methods described in Section V are applied.

a) In the state transition S1 -> S2: Since a condition primitive m-twc(A) is described in the analysis block, the conversion method described in IV *D 2)* is applied.

	A	B	C
S1:	talk(A,B)	talk(A,B)	-
S2:	dialtone(A),hold(A,B)	hold(A,B)	-
S3:	hold(A,B),calling(A,C)	hold(A,B)	calling(A,C)
S4:	hold(A,B),talk(A,C)		talk(A,C)
S5:	hold(A,B),busy(A)	hold(A,B)	-

Figure 13. Converted States in the HCM for Terminal A, B and C

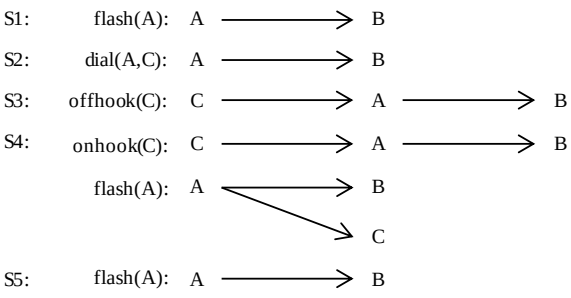


Figure 14. Tree Representation Obtained from States S1 … S5

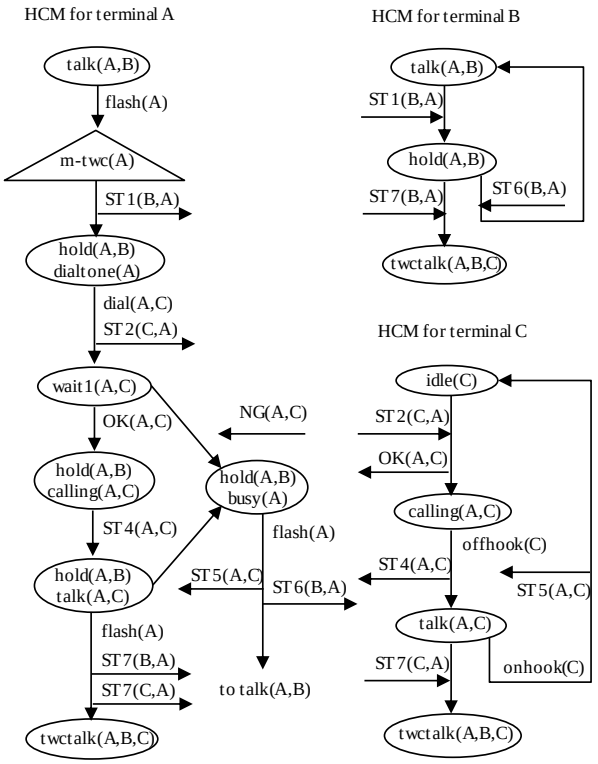


Figure 15. State Transition Diagrams in HCM Converted from CCM

b) In the state transition S2 -> S3: Since a state primitive idle(C) is described in the analysis block, the conversion method for the signal OK described in IV *D 1)* is applied.

c) In the state transition S2 -> S5: A state primitive not[idle(C)] which means the condition idle(C) is not satisfied is described in the analysis block. So, the conversion method for NG described in IV *D 1)* is applied.

d) In the state transition S4 -> S5: A branch appears in the tree expression for S4. So, the method of processing signals at the branch described in IV *B 2)* is applied.

Each state transition in Fig. 12 is converted to state transitions in HCM's in Fig. 15. State transition S1 -> S2 in Fig. 12 is converted to the state transition from talk(A,B) to {hold(A,B), dialtone(A)} in HCM for terminal A, and the state transition from talk(A,B) to hold(A,B) in HCM for terminal B, respectively. Suppose terminal A is in talk(A,B) state. At this moment, when an event flash(A) occurs and terminal A subscribes three-way call service, m-twc(A), signal ST1(B,A), which is an instruction to transit to hold(A,B), is sent to HCM for terminal B. In the HCM for terminal B, the state of terminal B transits to hold(A,B) if and only if ST1(B,A) is received when the state of terminal B is in at talk(A,B). Other six state transitions in Fig. 12 are also converted to state transition diagrams described in Fig. 15.

Thus, the proposed method can convert all types of state transitions in the three-way call service. Consequently, the proposed methods can give a good prospect for converting state transition diagrams of various services from CCM to HCM.

VI. Conclusion and Future Work

A conversion method from CCM to HCM was proposed, and by applying the method to the three-way call service it was confirmed that the proposed method can properly convert the state transition diagrams. Consequently, the proposed methods can give a good prospect for converting state transition diagrams of various services from CCM to HCM. Future work is to verify the validity of the proposed method by applying the method to various services including non-telephone services.

References

[1] H. Kawashima, K. Futami, and S. Kano, "Functional specification of call processing by state transition diagram," IEEE Trans. on Com., vol. COM-19, 1971.

[2] N. D. Grifeth, R. Blumenthal, J. C. Gregoire, and T. Ohta, "Feature interaction detection contest of the fifth international workshop on feature interactions," Computer Networks, vol. 32, pp. 487-510, Apr. 2000.

[3] A. Nakashima and T. Ohta, "Automatic conversion from HCM to CCM in telecommunication service simulation," Proc. ATS04, pp. 29-34, Apr. 2004.

[4] N. D. Griffeth, Y. Cantor, and C. Djouvas, "Testing a Network by Inferring Representative State Machines from Network Traces," ICSEA2006, Nov. 2006.

[5] S. K. Chakrabarti, and Y. N. Srikant, "Specification Based Regression Testing Using Explicit State Space Enumeration," ICSEA2006, Nov. 2006.

[6] M. Shimokura, S. Nakanishi and T. Ohta, "Networks for a symbiotic Human Life with Robots," Proc. of ROBOCOM2007, Oct. 2007.

[7] S. Sen and R. Cardell-Oliver, "A Rule-Based Language for Programming Wireless Sensor Actuator Networks using Frequency and Communication," Third Workshop on Embedded Networked Sensors (EmNets 2006), 2006.

[8] X. Fei and E. Magill, "Rule Execution and Event Distribution Middleware for PROSEN-WSN," 2008 Second International Conference on Sensor Technologies and Applications, pp. 580-585, Aug. 2008.

[9] L. Dongliang, Z. Kanyu, and L. Xiaojing, "ECA Rule-based IO Agent Framework for Greenhouse Control System," ISCID, 2008 International Symposium on Computational Intelligence and Design, vol. 1, pp. 482-485, Oct. 2008.

[10] M. Ohba, K. Matsuoka, and T. Ohta, "Eliciting State Transition Diagrams from Programs described in a Rule-based Language," ISAST Transaction on Computers and Intelligent Systems, No. 2, Vol. 1, pp. 58-66, Jan. 2009.

[11] Y. Hirakawa and T. Takenaka., "Telecommunication service description using state transition rule," Int. Workshop on Software Specification and Design, Oct. 1991.

[12] T. Yoneda and T. Ohta, "The declarative language STR," Proc. of FIREworks workshop, pp. 197-212, May 2000.

[13] A. Takura, T. Ohta, and K. Kawata, "Process specification generation from communications service specification," Automated Software Eng., No. 2, pp. 167-182,1995.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

From UML to SRN: A Performability Modeling Framework Considering Service Components Deployment

Razib Hayat Khan
Department of Telematics
NTNU, Norway
rkhan@item.ntnu.no

Fumio Machida
Service Platform Research
NEC, Japan
h-machida@ab.jp.nec.com

Poul E. Heegaard
Department of Telematics
NTNU, Norway
poul.heegaard@item.ntnu.no

Kishor S. Trivedi
Department of ECE
Duke University, NC, USA
kst@ee.duke.edu

***Abstract*-Conducting the performance modeling of distributed system separately from the dependability modeling fails to asses the anticipated system performance in the presence of system components failure and recovery. System dynamics is affected by any state changes of the system components due to failure and recovery. This introduces the concept of performability that considers the behavioral change of the system components due to failures and also reveals how this behavioral change affects the system performance. But, to design a composite model for distributed system, perfect modeling of the overall system behavior is crucial and sometimes very cumbersome. Additionally evaluation of the required measures by solving the composite model are also intricate and error prone. Bearing this concept in mind, we delineate a performability modeling framework for a distributed system that proposes an automated transformation process from high level UML notation to SRN model and solves the model to generate various numerical results. To capture system dynamics through our proposed framework, we outline a specification style that focuses on UML collaboration and activity as reusable specification building blocks, while deployment diagram identifies the physical components of the system and the assignment of software artifacts to the identified system components. Optimal deployment mapping of software artifacts on the available physical resources of the system is investigated by deriving the cost function. State machine diagram is utilized to capture state changes of system components such as failure and recovery. Later on, model composition is achieved by assigning guard function.**

Keywords: UML, SRN, Performability, Deployment

I. INTRODUCTION

The analysis of the system behavior from the pure performance viewpoint tends to be optimistic since it ignores the failure and repair behavior of the system components. On the other hand, pure dependability analysis tends to be too conservative since performance considerations are not taken into account [3]. When the service is deployed it might be the case that something goes wrong in the system because of performance or dependability bottlenecks of the resources and that might adversely affects the service request completion. This bottleneck is an impediment to assure the effectiveness and efficiency requirements to achieve the purpose of system to deliver services proficiently and in timely manner [2]. Therefore, in real systems, availability, reliability and performance are important QoS indices which should be investigated in a combined manner that introduces the concept performability. Performability considers the effect of state changes because of failure and recovery of the system components and their impact on the overall performance of the system [1]. Bearing the above concept we therefore introduce a performability modeling framework for distributed system to allow modeling of the performance and dependability related behaviors in a combined way not only to model functional attributes of the service provided by the system but also to investigate dependability attributes to reflect how the changes in the dependability attributes affect the system performance. For ease of understanding the complexity behind the modeling of performability attributes the proposed modeling framework works in two different layers such as performance modeling layer and dependability modeling layer. The proposed framework achieves its objective by maintaining harmonization between performance and dependability modeling layer with the assist of model synchronization.

However in a distributed system, system behavior is normally distributed among several objects. The overall behavior of the system is composed of the partial behavior of the distributed objects of the system. So it is obvious to model the behavior of the distributed objects perfectly for appropriate demonstration of the system dynamics. Hence we adopt UML (Unified Modeling Language) collaboration, state machine and activity oriented approach as UML is the most widely used modeling language which models both the system requirements and qualitative behaviors through different notations [4]. Collaboration and activity diagram are utilized in the performance modeling layer to demonstrate the overall system behavior by defining both the structure of the partial object behaviors as well as the interaction between them. State machine is employed in the dependability modeling layer to capture system component behavior with respect to failure and repair events. Later the UML specification styles are applied to generate the SRN (Stochastic Reward Net) model automatically by our proposed framework. SRN models generated in both performance and dependability modeling layer are synchronized by the model synchronization role by designing guard functions (a special property of the SRN model [5]) to properly model the system performance behavior with respect to any state changes in the system due to component failure [1]. The proposed modeling framework considers system architecture to realize the deployment of the service components. Abstract view of the system architecture is captured by the UML deployment diagram,

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

which defines the execution architecture of the system by identifying the system components and the assignment of software artifacts to those identified system components [4]. Considering the system architecture to design the proposed framework resolves the bottleneck of system performance by finding a better allocation of service components to the physical nodes. This needs for an efficient approach to deploy the service components on the available hosts of distributed environment to achieve preferably high performance and low cost levels. Moreover, UML models are annotated according to the *UML profile for MARTE* [7] and *UML profile for Modeling Quality of Service and Fault Tolerance Characteristics & Mechanisms* to include quantitative system parameters [12].

Markov model, SPN (Stochastic Petri Nets) and SRN are probably the best studied performability modeling techniques [3]. Among all of them, we will focus on the SRN model generated by our proposed framework due to its some prominent and interesting properties such as priorities assignment in transitions, presence of guard functions for enabling transitions that can use entire state of the net rather than a particular state, marking dependent arc multiplicity that can change the structure of the net, marking-dependent firing rates, and reward rates defined at the net level [5].

Several approaches have been followed to conduct the performability analysis model from system design specification [8] [9] [10] [11]. However, most existing approaches do not highlight more on the issues that how to optimally conduct the system modeling to capture system dynamics and to conduct performability evaluation. The framework presented here is the first known approach that introduces a new specification style utilizing UML behavioral diagrams as reusable specification building block to characterize system dynamics. Building blocks describe the local behavior of several components and the interaction between them. This provides the advantage of reusability of building blocks, since solution that requires the cooperation of several components may be reused within one self-contained, encapsulated building block. This reusability provides the opportunity to design new system's behavior rapidly utilizing the existing building blocks according to the specification rather than starting the design process from the scratch. In addition the resulting deployment mapping provided by our framework has greater impact with respect to QoS provided by the system. Our aim here is to deal with vector of QoS properties rather than restricting in one dimension. Our presented deployment logic is surely able to handle any properties of the service, as long as we can provide a cost function for the specific property. The cost function defined here is flexible enough to keep pace with the changing size of search space of available hosts in the execution environment to ensure an efficient deployment of service components. Furthermore we aim to be able to aid the deployment of several different services at the same time using the same proposed framework. Moreover the introduction of model synchronization activity relinquishes the complexity and unwieldy affects in modeling and evaluation task of large and multifaceted systems. Model synchronization hides the intricacy behind demonstration of composite model behavior by designing guard functions [5]. Guard functions take charge of the proper functioning of the composite model by considering any changes either in the performance model or in the dependability model.

The paper is organized as follows: Section II introduces our proposed modeling framework, Section III depicts UML based model description, Section IV explains service component deployment issue, Section V clarifies model annotation, Section VI delineates model translation rules, Section VII introduces the model synchronization mechanism, Section VIII describes the fault tree model, Section IX demonstrates the application example to show the applicability of our modeling framework and Section X delineates the conclusion with future directions.

II. OVERVIEW OF PROPOSED FRAMEWORK

Our proposed performability framework is composed of 2 layers: performance modeling layer and dependability modeling layer. The performance modeling layer mainly focuses on capturing the system's dynamics to deliver certain services deployed on a distributed system. The performance modeling layer is divided into 5 steps shown in Fig.1 where the first 2 steps are the parts of Arctis tool suite which is integrated as plug-ins into the eclipse IDE [14]. Arctis focuses on the abstract, reusable service specifications that are composed form UML 2.2 collaborations and activities [14]. It uses collaborative building blocks to create comprehensive services through composition. To support the construction of building block consisting of collaborations and activities, Arctis offers special actions and wizards.

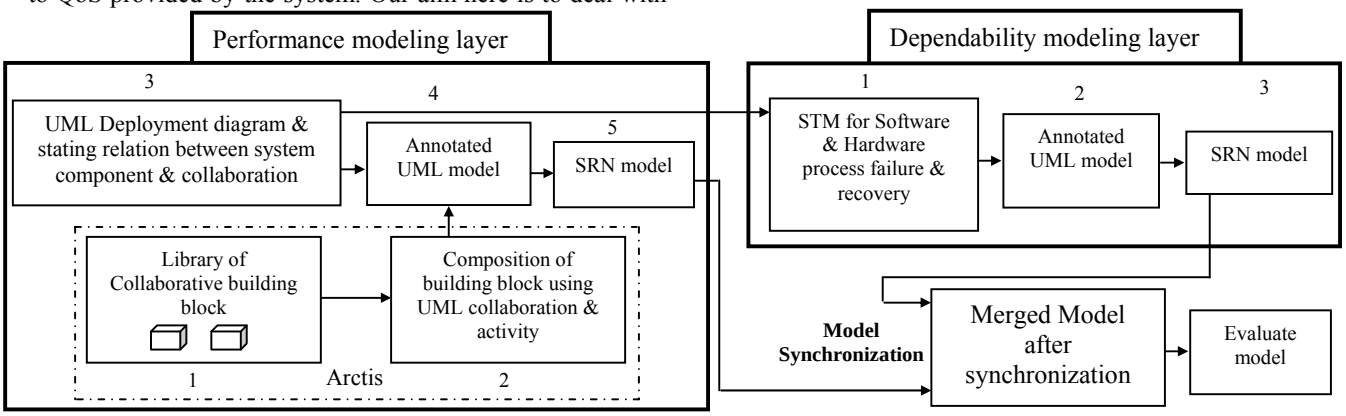


Figure 1. Proposed performability modeling framework

In the first step of performance modeling layer, a developer consults a library to check if an already existing basic collaboration role block or collaboration between several blocks solve a certain task. Missing blocks can also be created from existing building blocks and stored in the library for later reuse. The building blocks are expressed as UML models. The structural aspect, for example the service component and their multiplicity, is expressed by means of UML 2.2 collaborations. For the detailed internal behavior, UML 2.2 activities have been used. In the second step, the building blocks are combined into more comprehensive service by composition to specify the detailed behavior of how the different events of collaborations are composed so that the desired overall system behavior can be obtained. For this composition, UML collaborations and activities are used complementary to each other [14]. In the third step, the deployment diagram of our proposed system is delineated and the relationship between system component and collaboration is outlined to describe how the service is delivered by the joint behavior of the system components. In the fourth step, performance information is incorporated into the UML activity diagram and deployment diagram according to *UML profile for MARTE* [7]. The next step is devoted to automate generation of SRN model following the transformation rules. The SRN model generated in this layer is called performance SRN.

The dependability modeling layer is responsible for capturing any state changes in the system because of failure and recovery behaviors of system components. The dependability modeling layer is composed of three steps shown in Figure. 1. In the first step, UML state machine diagram (STM) is used to describe the state transitions of software and hardware components of the system to capture the failure and recovery behaviors. In the next step, dependability parameter is incorporated into the STM diagram according to *UML profile for Modeling Quality of Service and Fault Tolerance Characteristics & Mechanisms Specification* [12]. The last step reflects the automated generation of the SRN model from the STM diagram following the defined transformation rules. The SRN model generated in this layer is called dependability SRN.

The model synchronization is used as glue between performance SRN and dependability SRN. The synchronization task guides performance SRN to synchronize with the dependability SRN by identifying the transitions in the dependability SRN. The synchronization between performance and dependability SRN is achieved by defining the guard functions. Once the performance SRN model synchronized with dependability SRN model a merged SRN model will be obtained and various performability measures can be evaluated from the merged model using the software package such as SHARPE [15].

III. UML BASED SYSTEM DESCRIPTION

Construction of collaborative building blocks: The proposed framework utilizes collaboration as main entity. Collaboration is an illustration of the relationship and interaction among software objects in the UML. Objects are shown as rectangles with naming label inside. The relationships between the objects are shown as line connecting the rectangles [4]. The specifications for collaborations here are given as coherent, self-contained reusable building blocks. The structure of the building block is described by UML 2.2 collaboration. The building block declares the participants (as collaboration roles) and connection between them. The internal behavior of building block is described by UML activity. It is declared as the classifier behavior of the collaboration and has one activity partition for each collaboration role in the structural description. For each collaboration, the activity declares a corresponding call behavior action refereeing to the activities of the employed building blocks. For example, the general structure of the building block *t* is given in Fig. 2 where it only declares the participants *A* and *B* as collaboration roles and the connection between them is defined as collaboration t_x $(x=1\ldots n_{AB}$ (number of collaborations between collaboration roles A & B)). The internal behavior of the same building block is shown in Fig. 3(b). The activity $transfer_{ij}$ (where *ij* = AB) describes the behavior of the corresponding collaboration. It has one activity partition for each collaboration role: *A* and *B*. Activities base their semantics on token flow [1]. The activity starts by forwarding a token when there is a response (indicated by the streaming pin *res*) to transfer from the participant *A* to *B*. The token is then transferred by the participant *A* to participant *B* represented by the call operation action *forward* after completion of the processing by the collaboration role *A*. After getting the response of the participant *A* the participant *B* starts the processing of the request (indicated by the streaming pin *req*).

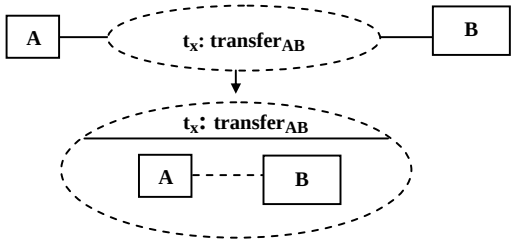


Figure 2. Structure of the Building block

Composition of building block using UML collaboration & activity: To generate the performance model, the structural information about how the collaborations are composed is not sufficient. It is necessary to specify the detailed behavior of how the different events of collaborations are composed so that the desired overall system behavior can be obtained. For the composition, UML collaborations and activities are used complementary to each other. UML collaborations focus on the role binding and structural aspect, while UML activities complement this by covering also the behavioral aspect for composition. Therefore, the activity contains a separate call behavior action for all collaboration of the system. Collaboration is represented by connecting their input and output pins. Arbitrary logic between pins may be used to synchronize the building block events and transfer data between them.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

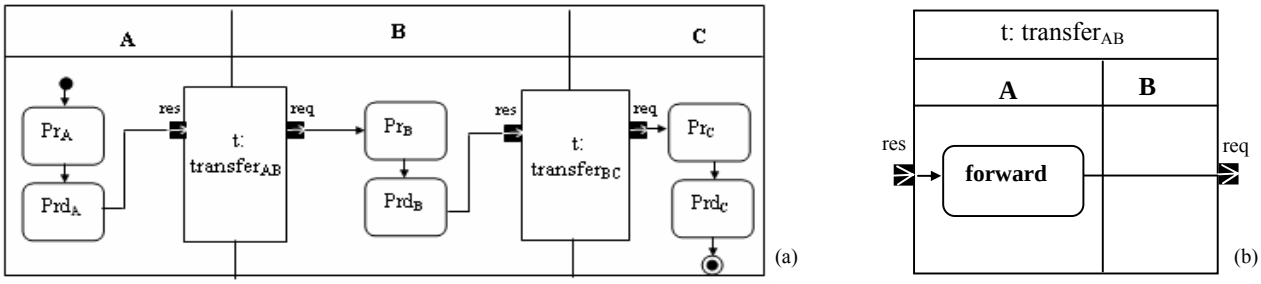


Figure 3. (a) Detail behavior of the event of the collaboration using activity (b) internal behavior of the collaboration

By connecting the individual input and output pins of the call behavior actions, the events occurring in different collaborations can be coupled with each other. Semantics of the different kinds of pins are given in more detailed in [14]. For example the detailed behavior and composition of the collaboration is given in following Fig. 3(a). The initial node (●) indicates the starting of the activity. The activity is started from the participant *A*. After being activated, each participant starts its processing of request which is mentioned by call operation action Pr_i *(Processing$_i$, where i = A, B & C)*. Completion of the processing by the participants are mentioned by the call operation action *Prdi (Processing_done$_i$, where i = A, B & C).* After completion of the processing, the response is delivered to the corresponding participant. When the processing of the task by the participant *A* completes, the response (indicated by streaming pin *res)* is transferred to the participant *B* mentioned by collaboration *t: transfer$_{ij}$* (*where ij = AB)* and participant *B* starts the processing of the request (indicated by streaming pin *req).* After completion of processing participant B transfers the response to the participant *C* mentioned by collaboration *t: transfer$_{ij}$* (*where ij = BC).* Participant *C* starts the processing after getting the response form *B* and activity is terminated after completion of the processing which is mentioned by the terminating node (◉).

Modeling failure & repair behavior of software & hardware component using STM: State transitions of a system element are described using STM diagram. In an STM, a state is depicted as a rounded rectangle and a transition from one state to another is represented by an arrow. Here STM is used to describe the failure and recovery behavior of software and hardware component. The STM of software process is shown in Fig. 4(a). The initial node (●) indicates the starting of the operation of software process. Then the process enters Running state. Running is the only available state in the STM. If the software process fails during the operation, the process enters Failed state. When the failure is detected by the external monitoring service the software process enters Recovery state and the repair operation will be started. When the failure of the process is recovered the software process returns to Running state. The STM of hardware node is shown in Fig. 4 (b). States of the hardware node start from the Stop state. The hardware node starts the operation when the on command is invoked and the node enters Running state. Running is the only available state here. If the node fails during the operation, the node enters Failed state. When the failure is detected the repair operation of the hardware node is started. When the failure of the node is repaired the node returns to Running state. The hardware node operation is terminated by the off operation and enters Stop state.

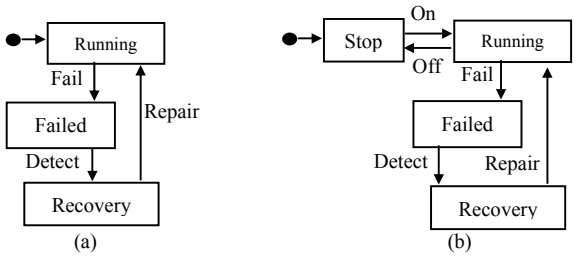


Figure 4. (a) STM of Software Process (b) STM of Hardware component

IV. DEPLOYMENT DIAGRAM & STATING RELATION BETWEEN SYSTEM & SERVICE COMPONENT

We model the system as collection of N interconnected nodes. Our objective is to find a deployment mapping for this execution environment for a set of service components C available for deployment that comprises the service. Deployment mapping can be defined as M: C $\rightarrow$ N between a numbers of service components instances C, onto nodes N. We consider four types of requirements in the deployment problem. (1) Components have execution costs, (2) collaborations have communication costs and (3) costs for running of background process known as overhead cost and (4) some of the components can be restricted in the deployment mapping to specific nodes which are called bound components. We observe the processing cost that nodes impose while host the components and also the target balancing of cost among the nodes available in the network. Communication costs are considered if collaboration between two components happens remotely, i.e., it happens between two nodes [6]. In other words, if two components are placed onto the same node the communication cost between them will not be considered. The cost for executing the background process for conducting the communication between the components is always considerable no matter whether the components deploy on the same or different nodes. Using the above specified input, the deployment logic provides an optimal deployment architecture taking into account the QoS requirements for the components providing the specified service. We then define the objective of the deployment logic as obtaining an efficient (low-cost, if possible optimum) mapping of component onto the nodes

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

that satisfies the requirements in reasonable time. The deployment logic providing optimal deployment architecture is guided by the cost function F (M). The cost function is designed here to reflect the goal of balancing the execution cost and minimizing the communications cost [6]. This is in turn utilized to achieve reduced task turnaround time by maximizing the utilization of resources while minimizing any communication between processing node. That will offer a high system throughput, taking into account the expected execution and inter-node communication requirements of the service components on the given hardware architectures which is already highlighted in [13]. The evaluation of cost function F (M) is mainly influenced by our way of service definition. Service is defined in our approach as a collaboration of total E components labeled as c_i (where i = 1…. E) to be deployed and total K collaboration between them labeled as k_j, (where j = 1 … K). The execution cost of each service component can be labeled as fc_i; the communication cost between the service components is labeled as fk_j and the cost for executing the background process for conducting the communication between the service components is labeled as f_{Bj}. Accordingly we only observe the total cost ($\widehat{l}_n$, n = 1…N) of a given deployment mapping at every node. We will strive for an optimal solution of equally distributed cost among the processing nodes and the lowest cost possible, while taking into account the execution cost fc_i, i = 1….E, communication cost fkj, j = 1….K and cost for executing the background process f_{Bj}, j = 1….k. fc_i, f_{kj} and f_{Bj} are derived from the service specification, thus the offered execution cost can be calculated as $\sum_{i=1}^{|E|} fc_i$. This way, the logic can be aware of the target cost T [6]:

$$T = \frac{\sum_{i=1}^{|E|} fc_i}{|N|} \quad (1)$$

To cater for the communication cost f_{kj}, of the collaboration k_j in the service, the function q_0 (M, c) is defined first [16]:

$$q_0 (M, c) = \{n \in N \mid \exists (c \rightarrow n) \in M\} \quad (2)$$

This means that q0 (M, c) returns the node n that host component in the list mapping M. Let collaboration $k_j = (c_1, c_2)$. The communication cost of k_j is 0 if components c_1 and c_2 are collocated, i.e. $q_0 (M, c_1) = q_0 (M, c_2)$, and the cost is f_{kj} if components are otherwise (i.e. the collaboration is remote). Using an indicator function I(x), which is 1 if x is true and 0 otherwise, this expressed as $I (q_0 (M, c_1) \neq q_0 (M, c_2)) = 1$, if the collaboration is remote and 0 otherwise. To determine which collaboration k_j is remote, the set of mapping M is used. Given the indicator function, the overall communication cost of service, F_k (M), is the sum [16]

$$F_k (M) = \sum_{j=1}^{|k|} I (q_0 (M, K_{j,1}) \neq q0 (M, K_{j,2})). f_{kj} \quad (3)$$

Given a mapping M = {m_n} (where m_n is the set of components at node n & $n \in N$) the total cost can be obtained as $\widehat{l}_n = \sum_{c_i \in m_n} fc_i$. Furthermore the overall cost function F (M) becomes [16]:

$$F (M) = \sum_{n=1}^{|N|} | \widehat{l}_n - T | + F_k (M) + \sum_{j=1}^{|K|} f_{Bj} \quad (4)$$

V. ANNOTATION

To annotate the UML diagram the stereotype *saStep, computingResource, scheduler, QoSDimension* and the tag value *execTime, deadline, mean-time-to-repair, mean-time-between-failures* and *schedPolicy* are used according to the *UML profile for MARTE* and *UML Profile for Modeling Quality of Service & Fault Tolerance Characteristics* [7],[12]. *saStep* is a kind of step that begins and ends when decisions about the allocation of system resources are made. The duration of the execution time is mentioned by the tag value *execTime* which is the average time in our case. *deadline* defines the maximum time bound on the completion of the particular execution segment that must be met. A *ComputingResource* represents either virtual or physical processing devices capable of storing and executing program code. Hence its fundamental service is to compute. A *Scheduler* is defined as a kind of ResourceBroker that brings access to its brokered ProcessingResource or resources following a certain scheduling policy tagged by *schedPolicy*. The ResourceBroker is a kind of resource that is responsible for allocation and de-allocation of a set of resource instances (or their services) to clients according to a specific access control policy [7]. *QoSDimension* provides support for the quantification of QoS characteristics and attributes *mean-time-to-repair* and *mean-time-between-failures* [12]. We also introduce a new stereotype <<*transition*>> and three tag values *mean-time-to-stop, mean-time-to-start* and *mean-time-to-failure-detect*. <<*transition*>> induces a state transition of a scenario. *mean-time-to-stop* defines the mean time required to stop working of a hardware instance, *mean-time-to-start* states the time required to start working of a hardware instance, *mean- time-to-failure-detect* defines the mean time required to detect failures in the system.

VI. MODEL TRANSLATION

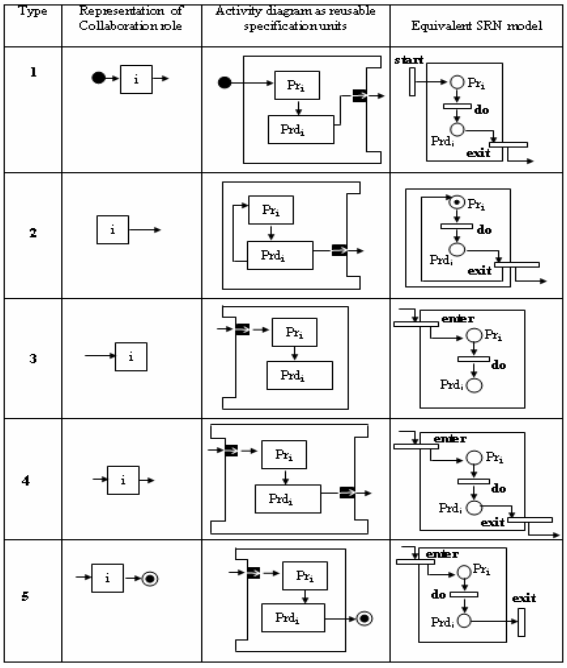
This section highlights the rules for the model translation from various UML models to SRN model. Since all the models will be translated into SRN we will give a brief introduction about SRN model. SRN is based on the Generalized Stochastic Petri net (GSPN) [3] and extends them further by introducing prominent extensions such as guard function, reward function and marking dependent firing rate [5]. A guard function is assigned to a transition. It specifies the condition to enable or disable the transition and can use the entire state of the net rather than just the number of tokens in places [5]. Reward function defines the reward rate for each tangible marking of Petri Net based on which various quantitative measures can be done in the Net level. Marking dependent firing rate allows using the number of token in a chosen place multiplying the basic rate of the transition. SRN model has the following elements: Finite set

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

of the places (drawn as circles), Finite set of the transitions defined as either timed transition (drawn as thick transparent bar) or immediate transition (drawn as thick black bar), set of arcs connecting places and transition, multiplicity associated with the arcs, marking that denotes the number of token in each place.

Before introducing the translation rules different types of collaboration roles as reusable basic building block are demonstrated with the corresponding SRN model in Table I that can be utilized to form the collaborative building blocks.

TABLE I. SPECIFICATION OF REUSABLE UNITES AND EQUIVALENT SRN MODEL



The rules are the following:

Rule1: The SRN model of a collaboration (Fig. 5), where collaboration connects only two collaboration roles, is formed by combining the basic building blocks type 2 and type 3 from Table I. Transition **t** in the SRN model is only realized by the overhead cost if service components A & B deploy on the same physical node as in this case communication cost = 0, otherwise **t** is realized by both the communication & overhead cost.

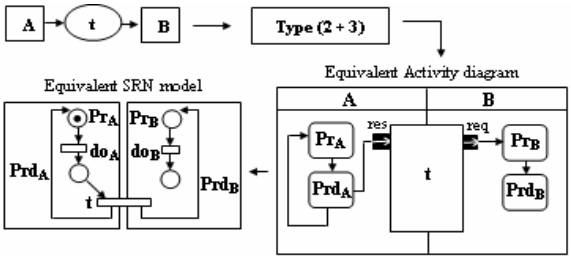


Figure 5. Graphical representation of rule 1

In the same way, SRN model of the collaboration can be demonstrated where the starting of the execution of the SRN model of collaboration role *A* depends on the receiving token from external source.

Rule 2: For a composite structure, when a collaboration role *A* connects with *n* collaboration roles by *n* collaborations like a star graph (where n=2, 3, 4, …..) where each collaboration connects only two collaboration roles, the SRN model is formed by the utilizing the basic building block of Table I which is shown in Fig. 6.In the first diagram in Fig. 6, if component A contains its own token equivalent SRN model of the collaboration role A will be formed using basic building block type 1 from Table I. The same applies to the component B and C in the second diagram in Fig. 6.

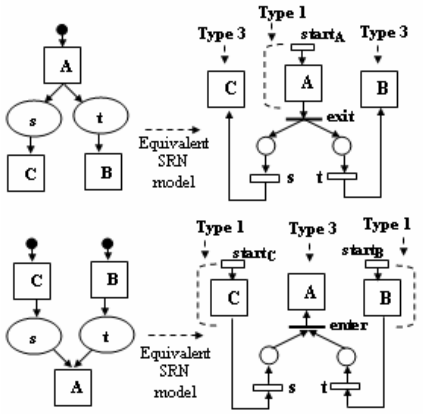


Figure 6. Graphical representation of rule 2

STM can be translated into a SRN model by converting each state into place and each transition into a timed transition with input/output arcs which is reflected in the transformation Rules 3.

Rule 3: Rule 3 demonstrates the equivalent SRN model of the STM of hardware and software components which are shown in the Fig. 7.

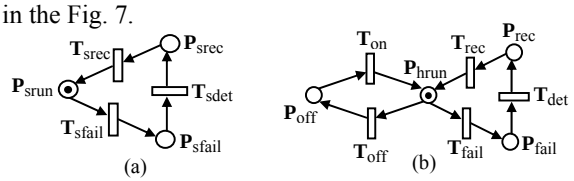


Figure 7. (a) SRN of Software process (b) SRN of hardware component

VII. MODEL SYNCHRONIZATION

The model synchronization is achieved hierarchically. Performance SRN is dependent on the Dependability SRN. Transitions in dependability SRN may change the behavior of the performance SRN. Moreover transitions in the SRN model for the software process also depend on the transitions in the SRN model of the hardware component. These dependencies in the SRN models are handled by the model synchronization by incorporating the guard functions [5].

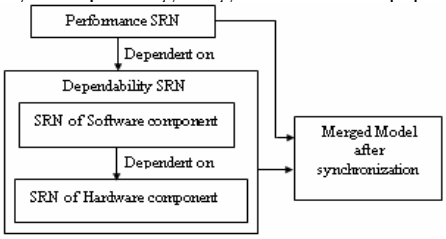


Figure 8: Model synchronization hierarchy

The model synchronization is focused in details here:

Synchronization between the Dependability SRN models in the dependability modeling layer: SRN model for the software process (Fig. 7(a)) is expanded by incorporating one additional place P_{hf}, three immediate transitions t_{hf}, t_{hfl}, t_{hfr} and one timed transition T_{recv} to synchronize the transitions in the SRN model for the software process with the SRN model for the hardware component. The expanded SRN model (Fig. 9(a)) is

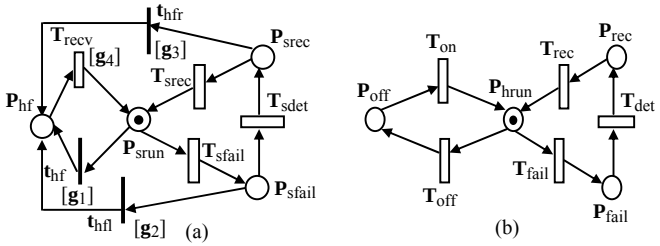


Figure 9. (a) Synchronized transition in the SRN model of the software process with the (b) SRN model of the hardware component

associated with four additional arcs such as ($P_{sfail} \times t_{hfl}$) $\cup$ ($t_{hfl} \times P_{hf}$), ($P_{srec} \times t_{hfr}$) $\cup$ ($t_{hfr} \times P_{hf}$), ($P_{srun} \times t_{hf}$) $\cup$ ($t_{hf} \times P_{hf}$) and ($P_{hf} \times T_{recv}$) $\cup$ ($T_{recv} \times P_{srun}$). The immediate transitions t_{hf}, t_{hfl}, t_{hfr} will be enabled only when the hardware node (in Fig. 9 (b)) fails as failure of hardware node will stop the operation of software process. The timed transition T_{recv} will be enabled only when the hardware node will again start working after being recovered from failure. Four guard functions g_1, g_2, g_3, g_4 allow the four additional transitions t_{hf}, t_{hfl}, t_{hfr} and T_{recv} of software process to work consistently with the change of states of the hardware node. The guard functions definitions are given in the Table III.

Synchronization between the dependability SRN & performance SRN: To synchronize the collaboration role activity, performance SRN model is expanded by incorporating one additional place P_{fl} and one immediate transition f_A shown in Fig. 10. After being deployed when collaboration role "A" starts execution a checking will be performed to examine whether both software and hardware components are running or not. If both the components work the timed transition do_A will fire which represents the continuation of the execution of the collaboration role "A". But if software resp. hardware components fail the immediate transition f_A will be fired which represents the quitting of the operation of collaboration role "A". Guard function gr_A allows the immediate transition f_A to work consistently with the change of states of the software and hardware components.

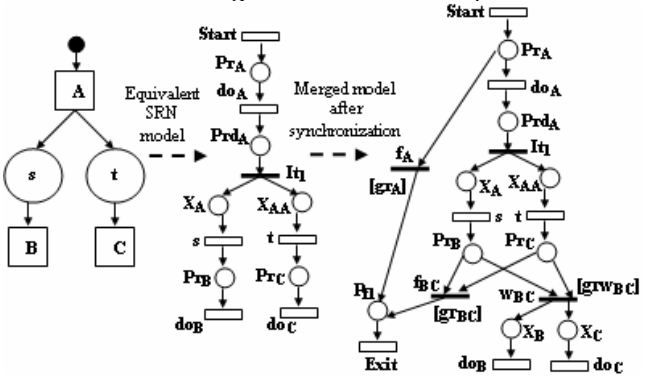


Figure 10. Synchronize the performance SRN model with dependability SRN

Performance SRN model of parallel execution of collaboration roles are expanded by incorporating one additional place P_{fl} and immediate transitions f_{BC}, W_{BC} shown in Fig. 10. In our discussion, during the synchronization of the parallel processes it needs to ensure that failure of one process eventually stop providing service to the users. This could be achieved by immediate transition f_{BC}. If software resp. hardware components (Fig. 9) fail immediate transition f_A will be fired which symbolizes the quitting of the operation of both parallel processes "B" and "C" rather than stopping either process "B" or "C", thus postponing the execution of the service. Stopping only either the process "B" or "C" will result inconsistent execution of the whole SRN and produce erroneous result. If both the software and hardware components work fine the timed transition W_{BC} will fire to continue the execution of parallel processes "B" and "C". Guard functions gr_{BC}, grw_{BC} allow the immediate transition f_{BC}, W_{BC} to work consistently with the change of the states of the software and hardware components. The guard function definitions are shown in the Table III.

VIII. HIERARCHICAL MODEL FOR MTTF CALCULATION

It is very demanding and not efficient with respect to execution time to consider behavior of all the hardware components during the SRN model generation. SRN model becomes very cumbersome and inefficient to execute. To solve the problem, we evaluate the MTTF (Mean time to failure) of system using the hierarchical model in which a fault tree is used to represent the MTTF of the system by considering MTTF of every hardware component in the system. Later we consider this MTTF of the system in our dependability SRN model for hardware components (Fig. 7(b)) rather than considering failure behavior of all the hardware components individually. The below Fig. 11 introduces one example scenario of capturing failure behavior of the hardware components using fault tree where system is composed of different hardware devices such as one CPU, two memory interfaces, one storage device and one cooler. The system will work when CPU, one of the memory interfaces, storage device and cooler will run. Failure of both memory interfaces or failure of either CPU or storage device or cooler will result the system unavailability.

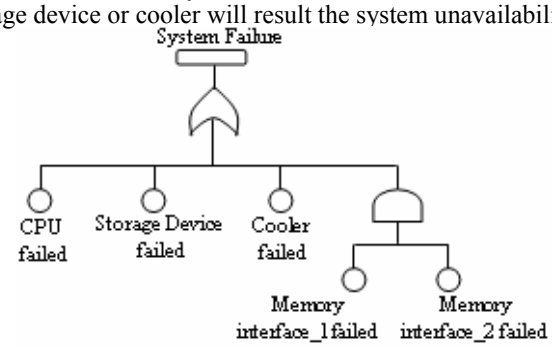


Figure 11. Fault tree model of System Failure

IX. CASE STUDY

As a representative example, we consider the scenario dealing with heuristically clustering of modules and assignment of clusters to nodes [16]. This scenario is sufficiently complex to show the applicability of our proposed framework. The problem is defined in our approach as collaboration of $E = 10$ service components or collaboration role (labeled $C_1 \ldots C_{10}$) to be deployed and $K = 14$ collaborations between them depicted in Fig. 12. We consider four types of requirements in this specification. Besides the execution cost, communication costs and cost for running background process, we have a restriction on components C_2, C_7, C_9 regarding their location. They must be bound to nodes $\boldsymbol{n_2}$, $\boldsymbol{n_1}$, $\boldsymbol{n_3}$ respectively. In this scenario, new service is generated by integrating and combining the existing service components that will be delivered conveniently by the system. For example, one new service is composed by combining the service components C_1, C_7, C_6, C_8, C_9 shown in Fig. 12 as thick dashed line. The internal behavior of the collaboration K_i is realized by the call behavior actions through UML activity like structure already demonstrated in Fig. 3(b). The composition of the collaboration role C_i of the delivered service by the system is demonstrated in Fig. 14. The initial node (●) indicates the starting of the activity. After being activated, each participant starts its processing of request which is mentioned by call behavior action Pr_i (Processing of the ith service component). Completions of the processing by the participants are mentioned by the call behavior action Prd_i (Processing done of the ith service component). The activity is started from the component C_1 where the semantics of the activity is realized by the token flow. After completion of the processing of the component C_1 the response is divided into two flows which are shown by the fork node f_1. The flows are activated towards component C_7 and C_6. After getting the response from the component C_1, processing of the components C_7 and C_6 will be started. The response and request are mentioned by the streaming pin *res* and *req*. The processing of the Component C_8 will be started after getting the responses from both component C_7 and C_6 which is realized by the join node j_8. After completion of the processing of component C_8 component C_9 starts its processing and later on activity is terminated which is mentioned by the end node (◉). In this example, the target environment consists of $N = 3$ identical, interconnected nodes with no failure of network link, with a single provided property, namely processing power, and with infinite communication capacities depicted in Fig. 13. The optimal deployment mapping can be observed in Table II. The lowest possible deployment cost, according to equation (4) is: 17 + 100 + 70 = 187.

To annotate the UML diagrams in Fig. 13 & 15 we use the stereotypes *<<saStep>>* *<<computingResource>>*, *<<scheduler>>* and the tag values *execTime*, *deadline* and *schedPolicy* which are already explained in section 5. Collaboration K_i (Fig. 15) is associated with two instances of *deadline* as collaborations in example scenario are associated with two kinds of cost: communication cost & cost for running background process (BP). To annotate the STM UML diagram of software process (shown in Fig. 14) we use the stereotype *<<QoSDimension>>*, *<<transition>>* and attributes *mean-time-between-failures*, *mean-time-to-failure detect* and *mean-time-to-repair* already mentioned in section 5. Annotation of the STM of hardware component can be demonstrated in the same way as STM of software process.

By considering the deployment mapping and the transformation rules the analogous SRN model of our example service (in Fig. 15) is depicted in Fig. 16. In our discussion, we consider M/M/1/n queuing system so that at most n jobs can be in the system at a time [3]. For generating the SRN model, firstly we will consider the starting node (●). According to rule 1, it is represented by timed transition

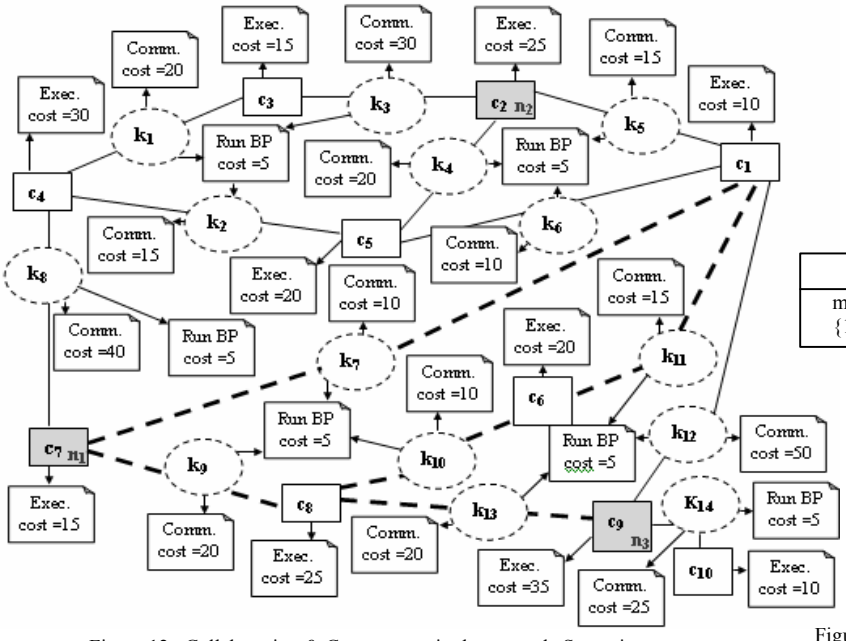


Figure 12. Collaboration & Components in the example Scenario

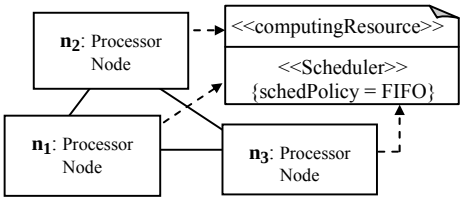


Figure 13. The target network of hosts

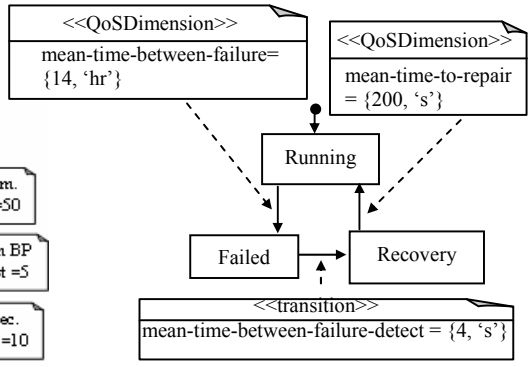


Figure 14. Annotated STM diagram of software component

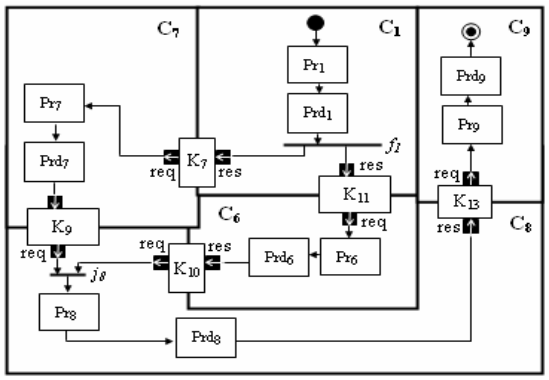


Figure 15. Service composition & Detail behavior of the event of the Collaboration using activity

(denoted as start) and the arc connected to place Pr_1 (states of component C_1). When a token is deposited in place Pr_1, immediately a checking is done about the availability of both software and hardware components by inspecting the corresponding SRN models (Fig. 9). The availability of software and hardware component allow the firing of timed transition t_1 mentioning the continuation of the further execution. Otherwise immediate transition f_1 will be fired mentioning the ending of the further execution because of software resp. hardware component failure. The enabling of immediate transition f_1 is realized by the guard function $\boldsymbol{gr_1}$. After the completion of the state transition from Pr_1 to Prd_1 (states of component C_1) the flow is divided into two branches (denoted by the immediate transition It_1) according to rule 2. The token will be deposited to place Pr_7 (states of component C_7) and Pr_6 (states of component C_6) after the firing of transitions K_7 and K_{11}. The collaboration K_7 is realized both by the communication cost and cost for running background process as C_1 and C_7 deploy on the two different nodes $\boldsymbol{n_3}$ and $\boldsymbol{n_1}$. According to rule 1, collaboration K_{11} is realized only by the cost for running background process as C_1 and C_6 deploy on the same processor node $\boldsymbol{n_3}$. When a token is deposited into place Pr_7 and Pr_6, immediately a checking is done about the availability of both software and hardware components by inspecting the corresponding dependability SRN model (Fig. 9). The availability of software and hardware components allow the firing of immediate transition w_{76} which eventually enables the firing of timed transition t_7 and t_6 mentioning the continuation of

TABLE II. OPTIMAL DEPLOYMENT MAPPING

Node	Components	$\widehat{l}_n$	$\lvert \widehat{l}_n - T \rvert$	Internal collaborations
n_1	c_4, c_7, c_8	70	2	k_8, k_9
n_2	c_2, c_3, c_5	60	8	k_3, k_4
n_3	c_1, c_6, c_9, c_{10}	75	7	k_{11}, k_{12}, k_{14}
$\sum$ cost			17	100

the further execution. The enabling of immediate transition w_{76} is realized by the guard function $\boldsymbol{grw_{76}}$. Otherwise immediate transition f_{76} will be fired mentioning the ending of the further execution because of failure of software resp. hardware component. The enabling of immediate transition f_{76} is realized by the guard function $\boldsymbol{gr_{76}}$. After the completion of the state transition from Pr_7 to Prd_7 (states of component C_7) and from Pr_6 to Prd_6 (states of component C_6) component C_8 starts processing. The merging of result is realized by the immediate transition It_2 after the firing of transitions K_9 and K_{10}. Collaboration K_9 is realized only by the cost for running background process as C_7 and C_8 deploy on the same processor node $\boldsymbol{n_1}$. K_{10} is translated by the timed transition which is realized both by the communication cost and cost for running background process as C_6 and C_8 deploy on the two different nodes $\boldsymbol{n_3}$ and $\boldsymbol{n_1}$. When a token is deposited in place Pr_8, immediately a checking is done about the availability of both software and hardware components by inspecting the corresponding SRN model (Fig. 9). The availability of software and hardware components allow the firing of timed transition t_8 mentioning the continuation of the further execution. Otherwise immediate transition f_8 will be fired mentioning the ending of the further execution because of software resp. hardware component failure. The enabling of immediate transition f_8 is realized by the guard function $\boldsymbol{gr_8}$. After the completion of the state transition from Pr_8 to Prd_8 (states of component C_8) the token is passed to place Pr_9 by firing of timed transition K_{13}. K_{13} is realized by both communication cost and cost for running background process as C_8 and C_9 deploy on the two different nodes $\boldsymbol{n_1}$ and $\boldsymbol{n_3}$. When a token is deposited in place Pr_9, immediately a checking is done about the availability of both software and hardware components by inspecting the corresponding SRN model (Fig. 9). The availability of software and hardware component allow the firing of timed transition t_9 mentioning the continuation of the further execution. Otherwise immediate transition f_9 will be fired mentioning the ending of the further execution because of software resp. hardware component failure and the ending of the execution of the SRN model is realized by the timed transition $Exit_2$. The enabling of immediate transition f_9 is realized by the guard function $\boldsymbol{gr_9}$. After the completion of the state transition from Pr_9 to Prd_9 (states of component C_9) the ending of the execution of the SRN model is realized by the timed transition $Exit_1$. The definition of guard functions are shown in Table III (Phrun & Psrun are shown in Fig. 9).

TABLE III. GUARD FUNTIONS DEFINITION

Function	Definition
g_1, g_2, g_3	if (# Phrun = = 0) 1 else 0
g_4	if (# Phrun = = 1) 1 else 0
gr_A, gr_{BC}, gr_1, gr_{76}, gr_8, gr_9	if (# Psrun = = 0) 1 else 0
grw_{BC},grw_{76}	if (# Psrun = = 1) 1 else 0

We use SHARPE [15] to execute the obtained model and calculate the system's throughput. The throughput of successful jobs can be computed by checking the throughput of the transition $Exit_1$ by SHARPE [15]. The throughput result is summarized in Tab. IV and graph in Fig. 17 shows throughput variation of the system against the change of failure rate of both hardware and software components.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

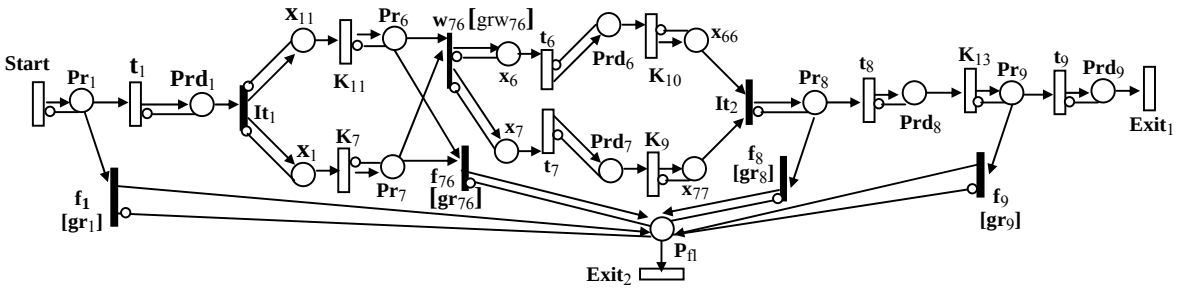


Figure 16. Equivalent SRN model of the example service

X. CONCLUSION AND FUTURE WORK

We presented a novel approach for model based performability evaluation of a distributed system which spans from system's dynamics demonstration and capturing behavior of system components through UML diagram as reusable building blocks to efficient deployment of service components in a distributed manner by focusing the QoS requirements. We put emphasis to establish some important concerns relating specification and solution of performability models emphasizing the analysis of the system's dynamics. We design the framework in a hierarchical and modular way which has the advantage to introduce any modification or adjustment at a specific layer in a particular submodel rather than in the combined model according to any change in the specification. Among the important issues that come up in our development is flexibility of capturing the system's dynamics using our new reusable specification of building blocks and ease of understanding the intricacy of combined model generation and evaluation from that specification by proposing transformation from UML diagram to corresponding SRN elements like states, different pseudostates and transitions. However, our eventual goal is to develop support for runtime redeployment of components, this way keeping the service within an allowed region of parameters defined by the requirements. As a result, with our proposed framework we can show that our logic will be a prominent candidate for a robust and adaptive service execution platform. However, the size of the underlying reachability set to generate SRN model is major limitation for large and complex systems. Further work includes tackling the state explosion problems of reachability marking of large distributed systems.

TABLE IV. THROUGHPUT CALCULATION

	Throughput
Performability model	0.0095
Pure performance model	0.01385

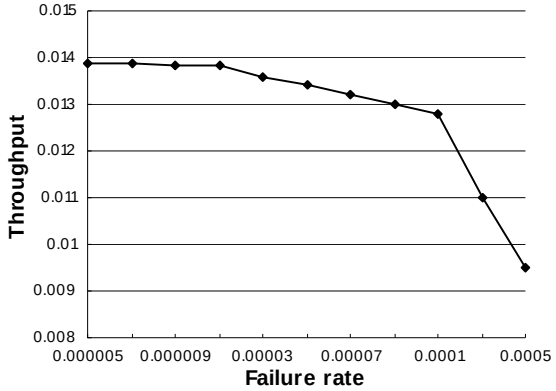


Figure 17. Numerical result of our example scenario

REFERENCES

[1] F. A. Jawad and E. Johnsen, "Performability: the vital evaluation method for degradable systems and its most commonly used modeling method, Markov reward modeling", http://www.doc.ic.ac.uk/~nd/surprise_95/journal/vol4/eaj2/report.html, <retrieved May 2011>

[2] E. de Souza e. Silva and H. R. Gali,"Performability analysis of computer systems: from model specification to solution", Performance evaluation 14, pp. 157-196, 1992

[3] K. S. Trivedi, "Probability and Statistics with Reliability, Queuing and Computer Science application", Wiley-Interscience publication, ISBN 0-471-33341-7, 2001

[4] OMG UML Superstructure, Version-2.2

[5] G. Ciardo, J. Muppala, and K. S. Trivedi, "Analyzing concurrent and fault-tolerant software using stochastic reward nets", Journal of Parallel and Distributed Computing, Vol. 15, 1992

[6] M. Csorba, P. Heegaard, and P. Herrmann, "Cost-Efficient Deployment of Collaborating Components", DAIS, pp. 253–268, Springer, 2008

[7] OMG 2009, "UML Profile for MARTE: Modeling & Analysis of Real-Time Embedded Systems", V – 1.0

[8] N. Sato and Trivedi, "Stochastic Modeling of Composite Web Services for Closed-Form Analysis of Their Performance and Reliability Bottlenecks", ICSOC, pp. 107-118, Springer, 2007

[9] P. Bracchi, B. Cukic, and Cortellesa, "Performability modeling of mobile software systems", ISSRE, pp. 77-84, 2004

[10] N. D. Wet and P. Kritzinger, "Towards Model-Based Communication Protocol Performability Analysis with UML 2.0", http://pubs.cs.uct.ac.za/archive/00000150/01/No_10, <retrieved May 2011>

[11] Gonczy, Deri, and Varro, "Model Driven Performability Analysis of Service Configurations with Reliable Messaging", MDWE, 2008

[12] OMG 2009, "UML Profile for Modeling Quality of Service & Fault Tolerance Characteristics Specification", V-1.1

[13] R. H. Khan and P. Heegaard, "A Performance modeling framework incorporating cost efficient deployment of multiple collaborating components" ICSECS, pp. 31-45, Springer, 2011

[14] F. A. Kramer, "ARCTIS", Department of Telematics, NTNU, http://arctis.item.ntnu.no, <retrieved May 2011>

[15] K. S. Trivedi and R. Sahner, "Symbolic Hierarchical Automated Reliability / Performance Evaluator (SHARPE)", Duke University, NC, 2002

[16] Mate J. Csorba, "Cost efficient deployment of distributed software services", PhD Thesis, NTNU, Norway, 2011

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

Content-based Clustering in Flooding-based Routing: The case of Decentralized Control Systems

Soroush Afkhami Meybodi, Jan Bendtsen, Jens Dalsgaard Nielsen
Department of Electronic Systems
Aalborg University
Aalborg, Denmark
Emails: {sam, dimon, jdn}@es.aau.dk

***Abstract*—This paper investigates a problem that is usually studied in communication theory, namely *routing* in wireless networks, but it offers a control oriented solution – particularly for decentralized control systems – by introducing a new routing metric. Routing algorithms in wireless networks have a strong impact on the performance of networked control systems which are built upon them, by imposing latency, jitter, and packet drop out. Here, we have gone one step further from only investigating the effect of such communication constraints, and have directly intervened in the design of the routing algorithm for control systems in order to: 1) realize our preferred network topology and data traffic pattern, and 2) making it feasible to add and remove sensors, actuators, and controllers without having to decommission and/or re-design the system. Moreover, the end-to-end latency and jitter in our system tend to be minimal as a result of robustness of the algorithm to topology modifications. The proposed routing solution combines traditional flooding-based routing scheme with a novel method of clustering nodes based on correlation analysis between existing and emergent sensors and actuators of a control system.**

***Keywords*-routing; flooding; clustering; system identification**

I. INTRODUCTION

Chronologically, flooding is the first type of routing solutions that appeared [1]. The name describes well how it works. In pure flooding – that is flooding without any network structure – when a node receives a packet, it checks whether it is the final recipient of the packet or it has received the same packet before. The latter may happen if a packet reaches the same node from different routes. A negative answer to both questions results in retransmission of the packet. Assuming no congestion at the MAC layer, flooding is unquestionably the fastest routing method with the minimum latency which offers a true peer-to-peer (P2P) traffic support. It is not sensitive to modifications of the physical topology because it does not rely on identifying and keeping track of optimal routes. Actually, no topology maintenance is required at all.

All of these advantages come at the expense of a serious drawback. Flooding imposes a heavy overhead, hence is painfully resource consuming. Too many healthy retransmissions happen before a packet is faded from the network. This will not only consume much energy, which is especially valuable for battery-operated nodes, but also causes congestion and increases collision chance at the MAC layer. This diminishes the main benefits of flooding and increases energy consumption both because of the high number of transmissions and the high number of re-transmissions after collisions at the MAC layer happen. That is why pure flooding works well only for small networks with a few number of nodes [1].

Several remedies have been proposed to reduce the number of unnecessary retransmissions in flooding-based routing. Some assume that a limited number of retransmissions are enough to reach the destination and do not propagate the packet any further. Such a number is derived either probabilistically or deterministically by considering the worst case [1]. All of the other methods impose a structure on the network. In the context of flooding-based routing, structuring a network is equivalent to partitioning it into separate or overlapping clusters.

Clustering confines the domain of flooding each packet and can be done by any of the following methods:

- *Coordinates-based Clustering*: Clustering nodes based on their geographical, relative, or virtual coordinates is a simple task provided that the nodes are aware of their geographical location, or can infer their relative or virtual locations. Any kind of distance measure between a node and a cluster center might be used as the membership criterion. If a node is located close enough to the cluster center, it will be a member of that cluster.
- *Metric-based Clustering*: Communication metrics might also be exploited in setting a structure for the network, e.g. by omitting the nodes that have little residual energy or blacklisting the links which are not reliable enough. This category represents a number of popular protocols. Here is how they generally work: At pre-scheduled time intervals, several nodes elect themselves as cluster heads. This could be done by either a pre-defined probability value in each node as in Low-Energy Adaptive Clustering Hierarchy (LEACH) [2], or based on the remaining energy of battery operated nodes as in Hybrid Energy-efficient,

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

Distributed clustering (HEED) [3]. Then the remaining nodes attach to the *nearest* cluster head. "Near" could be interpreted by any kind of distance measure which is derived from a routing metric, e.g. strength of the received signal, expected transmission count (ETX), hop count, etc. [4].

- *Content-based Clustering*: This is a data-centric approach in which the data content of the packet is used to alleviate flooding overhead. Current solutions are either based on tailoring redundant data or aggregating correlated or similar data [5].

In this paper, we are going to introduce another approach towards content-based clustering which is based on the specific characteristics and requirements of the top layer application, i.e. a decentralized control system. Our solution utilizes control oriented metrics to form clusters [6], instead of typically used communication based metrics. Although clustered flooding-based routing is well known, how to create and maintain these clusters makes our routing solution novel.

The remaining of the paper fulfills the above mentioned objectives by coping with the following structure: In Section II, some preliminaries regarding the traffic pattern of our application, which we will call decentralized Wireless Networked Control Systems (WNCS), are stated. They are followed by introducing the assumed networking topology. The main result is given in Section III by proposing a routing algorithm for decentralized WNCS followed by implementation details and step-by-step procedures on cluster formation and maintenance. Some performance related remarks are presented in Section IV. The paper is concluded in Section V.

II. PRELIMINARIES

A. Dominant Traffic Pattern

In a WNCS, there are three kinds of nodes: actuators, controllers, and sensors. All of them should have the capability to act both as a data *sink* and as a data *source*, described as follows.

- A Sensor is regularly a data source to send sensory data towards relevant controllers, typically once per control *cycle time* interval, equivalent to the control loop sampling time, e.g. 100 ms.
- A sensor sporadically acts as a data sink to receive configuration data from its associated controllers.
- An actuator is regularly a data sink which receives commands from the associated controller at each control cycle time and implements them.
- An actuator sporadically acts as a data source to report failures.
- Controllers should send and receive data in each control cycle time interval. They gather data from sensors at the beginning of a typical cycle time interval, and send commands to the actuators at the end of the interval. This is exactly what happens in a Programmable Logic Controller (PLC), supplied either with single cable inputs/outputs or aggregated bus communication modules.

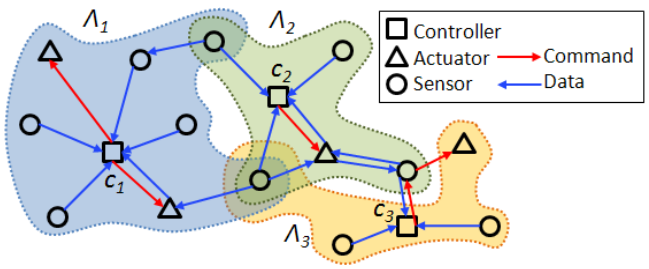


Figure 1. Logical network topology of a simple decentralized wireless networked control system

Unlike newer routing protocols for WNCSs, which assume a Multi-point to Point (MP2P) traffic with controllers as the sink nodes [7], the above list suggests a P2P traffic pattern. Especially the third item, which has the same importance as the first one and happens at the same frequency, makes WNCSs incompatible with MP2P architecture. Note that, this would not be the case if only monitoring and open loop control were of interest as in [7].

B. Network Topology

Fig. 1 shows the logical network topology of a simple decentralized WNCS. Arrows represent data direction in its regular functioning mode, but information flow in the reverse direction is also required as explained earlier in II.*A*.

Fig. 1 depicts the key assumptions that we have considered in topology design, described in the following:

1) There could be a large number of controllers ($C_i, i = 1, ..., n$).
2) Each controller and its associated sensors and actuators form a set, called a *cluster* henceforth. Clusters are identified by their unique tag ($\Lambda_i, i = 1...n$).
3) There are as many clusters as controller nodes which are called *cluster heads*.
4) Each packet contains a cluster *association* field. In general, a packet might be tied to one or more clusters. It is also possible that a packet is not associated with any cluster.
5) A sensor might be a *member* of multiple clusters, meaning that its generated data could be associated to more than one cluster head. In other words, a packet that is generated at a sensor node, might be reported to more than one controller.
6) An actuator could be a member of at most one cluster, meaning that it may not receive commands from more than one controller.
7) Data packets to/from members of a cluster should be sent from/to the cluster head. In other words,

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

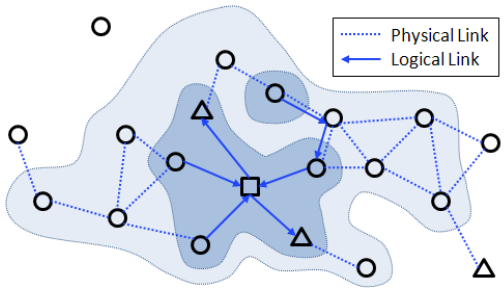


Figure 2. Preserving connectivity of cluster members by a surrounding cloud whose thickness is adjusted by *relay credit*, here defined as hop-count ≤ 2

a controller node is either the source or the final destination in every transmission path.

8) Relaying data packets associated to the cluster (Λ_j) could be done via all of the nodes in the entire network, irrespective of their membership status in Λ_j, provided that the relayed packet has enough remaining *relay credit*.

Relay credit can be defined in terms of any scalar node or link routing metric, e.g. hop count as a node metric or expected transmission count (ETX) as an accumulative link metric [4].

The last assumption clarifies that cluster membership is not necessary when relaying a packet. The purpose of assigning relay credit to a packet is to preserve connectivity in a cluster while constraining the number of retransmissions that might occur to a packet among non-member nodes. Each data packet is given an initial relay credit besides cluster membership tags, when generated. While it is roaming inside its own cluster, it does not spend any relay credit. However, when it is being relayed among non-member nodes, the initial relay credit is decreased at each non-member node until the remaining credit is not enough for more retransmissions among non-member nodes. Fig. 2 illustrates an example.

In Fig. 2, the dark area shows the members of a cluster Λ, and the lighter surrounding area shows the maximum penetration depth of the packets of Λ if hop-count ≤ 2 is considered as the relay credit criterion. It helps preserving connectivity of Λ members, when there is no direct physical link between them. This is shown in Fig. 1 too, where a sensor node which is a member of Λ_3 connects to its own cluster via an actuator and a sensor of Λ_2.

III. The Routing Protocol

A. Control Oriented Clustering of Nodes

To propose a method to form clusters based on requirements of the control system, we rely on the results of [6]. It offers three stochastic correlation-based measures that indicate *usefulness* of incorporating a new sensor/actuator in a present system model. We will use these scalar measures as application-based routing metrics in order to extend the concept of *distance*. Consequently, membership in a specific cluster Λ_j is granted to a node if that node is closer to the cluster head C_j than a specific threshold value d_j.

1) Addition of a new sensor: For a newly added sensor node, the following two complementary measures are proposed.

$$d^2_{U_p,y_a} = \frac{E[y_a(k) - \hat{y}_a(k|U_p^{k-1})]^2}{E[y_a(k) - E(y_a(k))]^2} \quad (1)$$

$$d^2_{U_pY_pY_a,y_a} = \frac{E[y_a(k) - \hat{y}_a(k|U_p^{k-1}, Y_p^{k-1}, Y_a^{k-1})]^2}{E[y_a(k) - E(y_a(k))]^2} \quad (2)$$

in which d^2 represents the correlation based distance and varies between 0 and 1. Subscripts $(.)_p$ and $(.)_a$ refer to present model and added device, respectively. $y(k)$ and $u(k)$ mean individual samples of a sensor's data and an actuator's command at time k, while Y^k and U^k indicate the set of all samples from the beginning up to and including time k. $E(.)$ stands for expected value operator over a finite number of data samples, N, which is pre-defined in the sensor node. Superscript $\hat{(.)}$ stands for the least-squares estimation based on the available model.

With respect to the above mentioned definitions, interpretation of (1) and (2) is given in the following paragraph, assuming that: 1) the model of the present system is discrete-time linear time-invariant and 2) the new sensor provides sufficiently exciting data, and 3) a consistent un-biased least-squares estimation is given when $N \to \infty$.

The denominator in (1) and (2) is the variance of the data gathered by the new sensor, i.e. y_a. The numerator in (1) indicates how predictable the current $y_a(k)$ is if the commands of all actuators are known in the previous samples. If, according to the present model, none of the $k-1$ samples of all of the actuators have any tangible effect on the k^{th} sample of y_a, the following equation holds true.

$$E[y_a(k)|U_p^{k-1}] = E(y_a(k)) \quad (3)$$

Furthermore, if an unbiased estimation is assumed, we have $\hat{y}_a(k|U_p^{k-1}) = E[y_a(k)|U_p^{k-1}]$ which in combination with (3) results in the following expression:

$$\hat{y}_a(k|U_p^{k-1}) = E(y_a(k)) \quad (4)$$

Equation (4) means that the conditional least squares estimation of y_a is equal to its actual expected value. Therefore, the present model is good enough and the new measurement does not add any value to it. In this situation, $d^2_{U_p,y_a} = 1$, which should be read as: the new node is too far from the cluster head and cannot become a member, that is it is irrelevant to the control loop in question.

Equation (1) measures how much the additional sensor is affected by the present actuators in open loop. Nevertheless, this measure only reveals linear correlation. To look for nonlinear correlations, (1) should be modified according to the specific nonlinearity we are looking for. This is the easy

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

step, but the difficult part is to perform nonlinear online incremental system identification to find $\hat{y}_a$. We do not consider this case in this paper.

The numerator in (2) measures how much the additional output could be controlled by the present actuators in closed loop. The interpretation is similar to (1), but this time the data from all of the sensors, including the new one, are also used in the least squares estimation, hence making it a more computationally intensive problem. Either (1) or (2) could be used in a given setting. Exploiting (1) is recommended in cases where y_a cannot be controlled independently of y_p [6].

2) Addition of a new actuator: When a new actuator is added, the following measure is proposed.

$$d^2_{U_a y_p | U_p Y_p} = \frac{E[y_p(k) - \hat{y}_p(k|U_p^{k-1}, Y_p^{k-1}, U_a^{k-1})]^2}{E[y_p(k) - \hat{y}_p(k|U_p^{k-1}, Y_p^{k-1})]^2} \quad (5)$$

Equation (5) measure how much influence the additional actuator has on the present sensors in closed loop. If U_a^{k-1} does not have any effect on improving prediction of $\hat{y}_p$, then the prediction errors in the numerator and denominator will look alike and d will get its maximum value ≈ 1 On the other hand, if U_a^{k-1} is useful such that the prediction error in numerator is much less than that in denominator, then we have: $d \to 0$. Equation (5) should be interpreted similar to the previous measures with similar concerns. The same assumptions hold for a consistent estimation of $\hat{y}_p$. In practice, to provide a sufficiently exciting control signal, the actuator has to be driven by an external signal.

Remark 1: Latency in the communication network has a considerable impact on all of the introduced measures. But at the same time, it influences control performance too. Therefore, it is reasonable to consider this effect on evaluating usefulness of adding new sensors and actuators.

B. Network Layer Packet Format

To illustrate details of the protocol, the packet format shown in Fig. 3, is chosen for the Network layer.

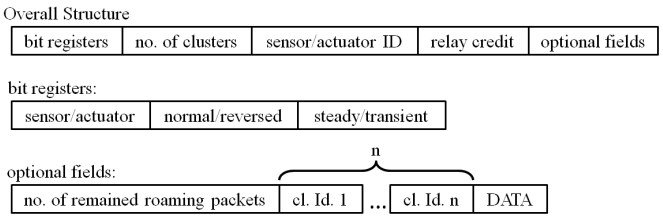


Figure 3. Structure of data packets at Network layer

The first field in the packet is a set of bit registers. The first bit flag determines whether the packet is sensor or actuator related. A sensor packet is generated in one of the sensors and should be routed to one or more controller nodes. An actuator packet is generated in a controller node and should be routed to an actuator node. The second bit indicates the direction of data. For sensor packets, "normal" means "from sensor to controller" and "reverse" means the other way around. The converse is true for actuator packets. The third bit register shows if the packet is sent in *steady* or *transient* operating mode. In transient mode, the packet contains an additional field, namely *number of remained roaming packets*, which is listed among optional fields. The difference between these two modes and the function of roaming packets is explained later in Section III.C.

The second field stores the number of associated clusters. "Zero" in this field means that the packet is not associated with any cluster. If the packet is linked to $n > 0$ clusters, n additional fields are included in the packet, each of which containing the address of one of the associated clusters.

The next field is *Sensor/Actuator ID*. In sensor packets, it contains address of the sensor node that has generated the packet. In actuator packets, it contains address of the actuator node that the packet is destined to. Assignment of unique addresses to all of the nodes in the entire network is a prerequisite to our routing solution. The same demanding requirement exists in emerging standards, e.g. IETF ROLL, by incorporating IPv6 as a worldwide addressing standard [7].

The packet also contains a *relay credit* ≥ 0 which is explained in details, earlier in Section II.B.

The last optional field is DATA which actually contains the application layer packet.

C. Cluster Formation

Here, we give a high level description of cluster formation. Assume that the controller nodes $(C_i, i = 1, ..., n)$ are deployed as cluster heads. In a realistic scenario, each cluster head has a built-in model of the subsystem it is supposed to control. All of the initially deployed sensor and actuator nodes are already bound to their controllers. In other words, in the network setup phase, all of the nodes are aware of their cluster membership. As a result, sensor nodes immediately start to function in their normal operating mode. See Fig. 4.

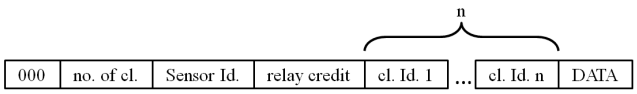


Figure 4. Structure of sensor packets at steady operation

Actuator nodes should start working in a safe mode and wait until they receive commands from the cluster head, i.e. the controller unit. The cluster head starts sending commands to the actuator as soon as it can devise the commands based on received sensor packets and the pre-programmed plant model. Command carrying packets are illustrated in Fig. 5.

In both above cases, the packets flood their pertinent clusters. Moreover, they propagate among the nodes of

100	1	Actuator Id.	relay credit	cl. Id	DATA

Figure 5. Structure of actuator packets at steady operation

neighboring clusters into a certain depth defined by their relay credit.

Later on, when a new sensor pops up, it does not initially belong to any cluster and it is in transient operating mode. Thus, it publishes data in *roaming packets* as shown in Fig. 6.

001	0	Sensor Id.	relay credit	no. of remained roaming packets	DATA

Figure 6. Structure of packets generated by a sensor when it is just turned on

When a roaming packet arrives at a neighboring node that is operating in *Steady* mode, it inherits all of the cluster tags of that node – meaning that the $cl.Id.$ fields of the packet are refreshed. This action is performed only if *relay credit* > 0. If either the neighboring node is in *Transient* mode or *relay credit* $= 0$, cluster tags of the packet remain unchanged.

In steady mode, embedding a cluster tag into a packet gives it the right to freely flood in that cluster without spending relay credit, but it is not the case in transient mode in which relay credit is constantly spent for every transmission. Therefore, embedding cluster tags into roaming packets does not give them a free pass. On the other hand, running out of relay credit is not the stop criterion when retransmitting a packet in transient mode. It just kills its ability to inherit new cluster tags. At the end, a roaming packet floods into the clusters that it managed to enter before running out of relay credit.

Example 1: Fig. 7 depicts an example when a new sensor is placed amongst nodes of Λ_1. However, some of its roaming packets could also reach the borders of Λ_2 before consuming all of their relay credit. Thus, presence of the new sensor is advertised through the union of nodes of Λ_1, Λ_2, and in the *relay credit* > 0 zone. Based on the above setting, C_1 and C_2 start calculating (1), or (2), or both. Note that, C_3 might also receive the roaming packets of the new sensor if it is placed in $\Lambda_3 \bigcap \Lambda_1$ or $\Lambda_3 \bigcap \Lambda_2$, but it will not calculate (1) or (2) because Λ_3 is not listed among clusters in the packets.

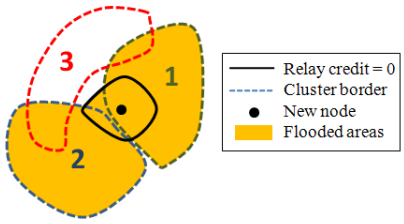


Figure 7. Effect of *relay credit* when a new node is joined

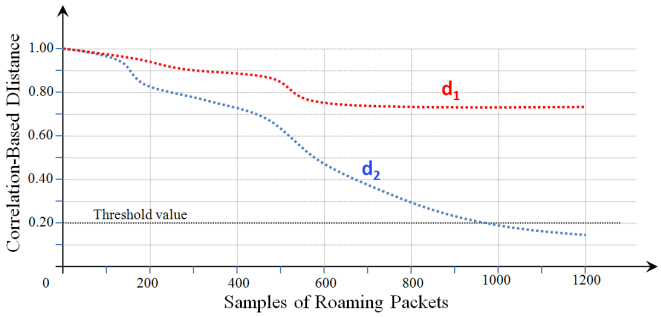


Figure 8. Development of usefulness measure (1) in C_1 and C_2 during identification process

Fig. 8 shows how the correlation-based distance measures (1) or (2) might develop over time in C_1 and C_2. It is assumed that the new sensor generates 1200 roaming packets.

After collecting sufficient samples at the cluster heads, each C_i decides whether the new sensor should become a member of their cluster or not. In Fig. 8, C_2 concludes that the new sensor is relevant well before the roaming packets are discontinued, but C_1 finds the new sensor irrelevant to the control performance of its internal model. C_2 continues the joining process by sending a *join* request to the new sensor. The data packet which contains the join request is in the following form, shown in Fig. 9.

010	1	Sensor Id.	relay credit	cl. Id. 1

Figure 9. Join request from cluster head to a sensor

Note that, this packet is in "steady operating mode", which means no relay credit is deducted unless it leaves the source cluster, Λ_2 in our example. This mechanism is useful when the new node is only reachable via nodes of other clusters.

After sending all roaming packets, equal to 1200 in our example, the new sensor applies received join requests. Join requests arrive at the sensor node asynchronously. Therefore, the node should continually accept join requests, at least until a pre-defined time. If no join request is received, the sensor starts another round of generating roaming packets.

If a group of sensors are deployed simultaneously, the ones which have other cluster members in their vicinity will find their clusters earlier. The others that do not have any neighbor operating in steady mode, may not inherit cluster tags, hence their flooding domain is limited to the *relay credit* > 0 zone.

The above procedure is slightly different for a new actuator. Each newly turned on actuator applies a pre-specified control sequence for the purpose of sufficiently exciting the plant and creating measureable outcomes. Simultaneously, it publishes roaming packets as shown in Fig. 10, which contain current value of the actuator output.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

111	0	Actuator Id.	relay credit	no. of remained roaming packets	DATA

Figure 10. Packets generated by an actuator when it is just turned on

The cluster heads which receive these packets, use the DATA field in evaluating (5) similar to what was shown in Fig. 8. When the roaming packets are discontinued – meaning that the actuator is waiting for the decision – each cluster head returns the calculated *usefulness measure* to the actuator by packets shown in Fig. 11.

100	1	Actuator Id.	relay credit	cl. Id. 1	DATA

Figure 11. Packets that return usefulness of utilizing an actuator in a cluster

The actuator waits for a certain time to receive evaluation results from all involved cluster heads. Then it compares the received usefulness values, which are embedded in DATA fields, and selects the cluster that has returned the highest value. If at least one evaluation result is received, the actuator chooses its own cluster and sends a join request to that cluster as illustrated in Fig. 12. If no evaluated usefulness measure is received, the above procedure starts from the beginning.

110	1	Actuator Id.	relay credit	cl. Id. 1

Figure 12. Join request from an actuator to a cluster head

Note that, when a new sensor is added, it receives individual join requests from controllers. But when a new actuator is added, it sends the join request to a single controller.

IV. Performance Related Remarks

Remark 1: Unlike other clustered flooding-based routing protocols whose acceptable performance depend heavily on the optimal choice of the number of clusters and the thoughtful selection of cluster heads [2], [3], these parameters are pre-defined in our protocol because all of the controller units $(C_j, j = 1...n)$ and only the controller units are cluster heads. Moreover, the controller nodes are fixed in the whole lifetime of the network.

Remark 2: Another issue is the influence of lower layer protocols on performance of the routing protocol. Unlike [2], that has utilized a TDMA-based MAC in sake of energy-efficient collision-free transmissions, the MAC layer in our system cannot accommodate a deterministic reservation-based protocol. It is mainly due to the constrained coverage range of nodes which does not guarantee existence of direct links between cluster heads and every member of the cluster to schedule a frame-based MAC. After all, it is a prerequisite for framed MACs that all of nodes can be accessed from a single base station for scheduling purposes. Otherwise, many time frames must be kept unused and reserved for future extension, as in Time Synchronized Mesh Protocol (TSMP) [8].

Our routing solution may be built either on a contention-based or a preamble-sampling MAC which are inferior to deterministic MACs in terms of energy efficiency and end-to-end latency if data transmission among nodes is frequent.

Therefore, the main benefit of our routing solution is to pick the members of each cluster so prudently such that it results in the minimum number of nodes in a cluster, and more efficient flooding in clusters.

V. Conclusion

In this paper, we have proposed a method to form clusters of nodes to be utilized by a flooding-based routing algorithm in decentralized wireless networked control systems. Our cluster formation method is data-centric and originates from the requirements of the control application. It makes use of model-based correlation estimation between new nodes and the existing model of the system. Furthermore, we have proposed to exploit non-member nodes in providing connectivity among nodes of an individual cluster. To this end, we have suggested to use an arbitrary routing metric, e.g. hop count. Operation of the proposed clustering mechanism is described in details.

References

[1] T. Watteyne, A. Molinaro, M. G. Richichi, and M. Dohler, "From MANET to IETF ROLL standardization: A paradigm shift in WSN routing protocols," to appear in IEEE Communications Surveys & Tutorials.

[2] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 660–670, October 2002.

[3] O. Younis and S. Fahmy, "Heed: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," *IEEE Transactions on Mobile Computing*, vol. 3, no. 4, pp. 366–379, October-December 2004.

[4] IETF ROLL: routing metrics used for path calculation in low power and lossy networks. [Online]. Available: http://tools.ietf.org/wg/roll/draft-ietf-roll-routing-metrics/

[5] J. Kulik, W. Heinzelman, and H. Balakrishnan, "Negotiation-based protocols for disseminating information in wireless sensor networks," *Wireless Networks*, vol. 8, no. 2, pp. 169–185, March-May 2002.

[6] T. Knudsen, J. Bendtsen, and K. Trangbaek, "Awareness and its use in incremental data driven modelling for Plug and Play Process Control," to appear in European Journal of Control, 2011.

[7] IETF ROLL: IPv6 Routing Protocol, draft standard. [Online]. Available: http://tools.ietf.org/wg/roll/draft-ietf-roll-rpl/

[8] K. S. J. Pister and L. Doherty, "TSMP: time synchronized mesh protocol," in *Proceedings of the IASTED International Symposium on Distributed Sensor Networks*, November 2008.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

Utilizing a Risk-Driven Operational Security Assurance Methodology and Measurement Architecture – Experiences from a Case Study

Reijo M. Savola, Teemu Kanstrén,
Heimo Pentikäinen, Petri Jurmu
VTT Technical Research Centre of Finland
Oulu, Finland
e-mail: {Reijo.Savola, Teemu.Kanstren,
Heimo.Pentikainen, Petri.Jurmu}@vtt.fi

Mauri Myllyaho
EXFO NetHawk
Oulu, Finland
e-mail: Mauri.Myllyaho@exfo.com

Kimmo Hätönen
Nokia Siemens Networks
Espoo, Finland
e-mail: Kimmo.Hatonen@nsn.com

***Abstract*—Practical measurement of information security of telecoms services is a remarkable challenge because of the lack of applicable generic tools and methods, the difficult-to-predict nature of security risks, the complexity of the systems, and the low observability of security issues in them. We discuss our experiences in utilizing a risk-driven methodology and associated measurement architecture in a practical case study. Effectiveness and efficiency are of main interest to stakeholders responsible for security. We note, however, that security configuration correctness and compliance with requirements are, in practice, the core objectives from an operational perspective. For these objectives there is more evidence available and it is easier to attain it. Our findings in this case study show a need for a wide range of security metrics to offer sufficient evidence of the design, implementation, and deployment of security controls. The case study also shows how visualization tools can be used efficiently to support the management of collections of these metrics.**

Keywords-Security; metrics; monitoring; risk analysis

I. INTRODUCTION

In the modern world, telecoms services are becoming more and more exposed to security threats. Sufficient and effective operational security is a result of adequate solutions and different stakeholders' activities at various levels, from the overall system administration to end-user applications and their behavior. Systematically obtained and managed evidence of the performance of these systems' security solutions benefits system development and maintenance.

The term *(information) security metrics* has become standard when referring to the security level or performance of a System under Investigation (SuI). A more appropriate term is *security indicators*, given the unpredictability of security risks, the complexity of systems, and their low observability in the absence of suitable measurement architectures. However, the former term is used in this study, to follow the most widely used terminology. Examples of security metrics application areas include risk management, comparison of different solutions, security assurance, testing, and monitoring [1]. This study focuses on operational security assurance.

In this study, we assume that there are three fundamental objectives of security measurement: effectiveness, efficiency and correctness. *Security effectiveness* means assurance that the stated security objectives are met in the SuI and the expectations for resilience in the use environment are satisfied, while the SuI does not behave in any way other than what is intended [2, 3, 4]. It is very difficult to measure security effectiveness directly; though activities such as long-term system use and penetration testing give some confidence. The quality of knowledge of risks is vital for security effectiveness. *Security efficiency* is assurance that adequate security effectiveness has been achieved in the SuI, in view of the resource, time, and cost constraints [2]. *Security correctness* is assurance that the security controls defined have been correctly implemented in the SuI, and the system, its components, the interfaces, and the processed data meet the security requirements [2, 3, 4]. Specific requirements, standards and best practices are used as references for security correctness assurance. While most experts agree that 100% secure systems are not possible, security correctness, including legal and standards compliance, is an important and achievable objective in practical security work.

From a security measurement perspective, the optimal ratio of security effectiveness and efficiency is of great interest. The goal of all security work is to ensure adequate security performance with respect to capability of mitigating and/or eliminating actual risks (effectiveness) using resources (e.g., time, money and functional performance) efficiently. We define operational security assurance, in line with [5], as grounds for confidence that security control realization is as expected in the operational system. This definition clearly emphasizes the security controls'

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

correctness, although it indirectly addresses effectiveness and efficiency too.

In this study, we discuss the role of operational security assurance in aiming at an optimal ratio of security effectiveness to efficiency in a Push E-mail system case study. We describe the security model components and discuss issues that we encountered when we implemented the model in an operational system.

Today smart phones are used in increasing amounts for various Web-based social services, such as Facebook and a variety of email applications. Additionally, the purchase of new music, goods, or software by means of mobile devices has become more common. The phones utilize several types of network connection at the same time. The shift in mobile devices' usage to other than only voice calls or Short Message Service (SMS) messaging has created a need for network and Internet operators to offer these services securely for the mobile devices' users. The demand for "always-on" functionality, especially in hand-held devices, has resulted in Push E-mail systems.

The main contribution of this study is in the benefit and challenge analysis of utilizing risk-driven security metrics and associated measurement architecture in a practical telecoms service case study. The metrics and measurement approach used are introduced in our previous work in [1, 2, 6–12]. The approach enables systematic and practical gathering and management of security evidence for different security related decision-making purposes.

The rest of the paper is structured as follows. Section II discusses the background and summarizes our previous work on this topic. Section III presents related work. Section IV presents the case study, with example metrics, and discusses our experiences of it. Section V addresses the benefits and challenges of utilizing our approach, before Section VI offers conclusions and poses future research questions.

II. BACKGROUND AND PREVIOUS WORK

In the discussion that follows, we offer a brief presentation of our previous work with security metrics and measurements and in the development of measurement architectures for them.

In [6], we introduced an iterative hierarchical security metrics development methodology, shown in simplified form in Fig. 1. This methodology is aimed at producing a balanced and detailed collection of security metrics, along with an associated measurement architecture. The measurement architecture includes the technical, administrative, legal and other means for gathering evidence. Note that a measurement architecture can be considered to be *risk-driven* if it is designed on the basis of risk-driven Security Objectives (SOs). The figure is identical to the description given in [5] apart from removal of the Quality-of-Service (QoS) metrics branch and replacement of the term "threat and vulnerability analysis" by "Risk Analysis" (RA). Term "threat and vulnerability analysis" has been used in the industry in referring to technical-level (or "architectural-level") RA. The term "RA" better represents the starting point; RA (either company-level or technical-level) as a holistic activity is the best choice as a foundation for security measurement goals. In [7], we integrated the above process into an industrial pilot study to match an iterative RA process and Agile software development. Experiences from the pilot showed the potential of security metrics in offering early visibility of security effectiveness and efficiency during security-critical phases of R&D. It became evident also that individual security metrics do not offer enough benefits; instead, collections of them are needed. Not much security effectiveness evidence is available during the first iterations of RA, when the need for it is at its highest.

In References [8] and [9], we discussed *Base Measures* (BM), *Derived Measures* (DMs), *measurement probes* and *measurement points*. BMs are abstract measurable properties of the SuI. Basic Measurable Component (BMC), discussed in [5], is a similar concept to BM. The difference between BMs and BMCs is that the latter represent the measurable properties *that are components of the decomposition* of SOs, whereas BMs can be standalone measures. It is possible that a property described by a BMC cannot be fully measured (through unavailability or unattainability of the evidence needed). DMs are interpretations of the BMs. In practice, one or multiple DMs can represent each BM. In generic models, the DMs that will be available in the future are not known, so only BMs can be presented. Development of detailed metrics, or DMs, for both of them may mean utilization of different measurement architectures. A measurement probe is a tool for performing checks of infrastructure objects in order to provide the information needed for purposes of measurements as defined by metrics. A measurement point is a point in the SuI, where one or more measurement probes are deployed.

In [8] and [9], we introduced a reference architecture for building a general monitoring framework, which can be utilized in Stage 5 in Fig. 1 for obtaining automated technical evidence for the purposes of continuous operational security assurance. This approach consists of four layers: (i) at the bottom, the Base Measure Layer, (ii) next, above it, the Data Collection layer, (iii) the Measurement Control and Processing Layer, and, at the top, (iv) the Presentation, Evaluation, and Management Layer.

As collections of security metrics can grow rather large, their management is a challenge. Moreover, aggregation of measurement values has pitfalls: relying blindly on an aggregated value can result in loss of important information and can lead to a false sense of security. There is no optimal weighting among branches, since many security problems arise from weakest links, which can be present in any sub-hierarchy. The benefits of visualization for human cognition can be utilized to increase the manageability of security metrics collections. In [10], we introduced a modeling and visualization tool called the Metrics Visualization System, or MVS, for the management of hierarchical security metrics and measurements. In the MVS *security metrics model* (SMM), the basic building block is a *security metrics node* (SMN). In an SMM, SMNs form a hierarchy. Same (or slightly customized) sub-hierarchies can be attached to different security controls at the higher level, because similar Security Controls (SCs) often mitigate or remove different security risks. All SMNs in the SMM have the same default

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

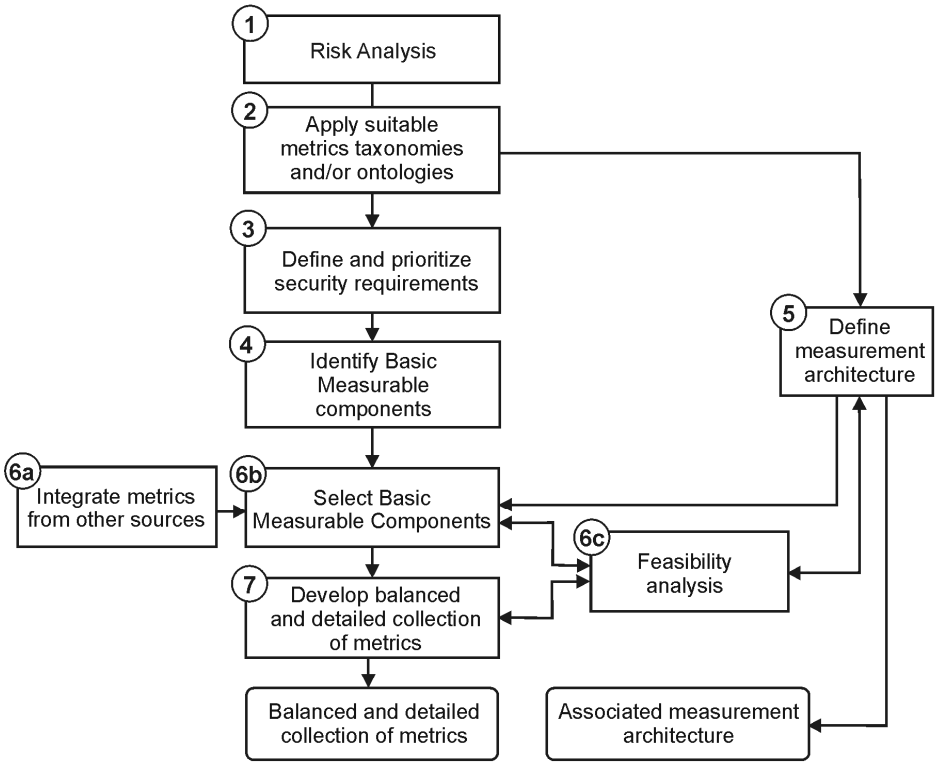


Figure 1. A security metrics development approach based on [5].

property fields: a distinctive name, metric confidence value (range 0…1), operation specification (logical expression), threshold criteria and associated visualization, polling frequency field for automated measurements, and enable/disable flag for operation value evaluation. The metrics in SMNs can be defined in terms of logical operations. All nodes can be colored or left blank. The default coloring scheme of the MVS imitates traffic lights: red stands for insufficient level, yellow for intermediate level, and green for sufficient level [10].

An important challenge is that many objects of the SuI system architecture are *unmanaged* [13]: they are not within the Administration Domain (AD) of the stakeholder carrying out security management and/or measurement of the SuI. Direct security measurements are not possible for an unmanaged object. However, a *trust value*, a certain value representing the amount of trust that the security of the object is adequate, can be associated with the object [11]. In practice, in many cases direct measurement of the SuI is not possible. Assessment of the properties essential to the security level can be used to replace direct measurements to achieve enough *indication* of the security level. In [12], a taxonomy of quality metrics for assessment of security correctness was introduced. The taxonomy uses a presentation inspired by the Common Criteria (CC) [14]. The quality metric families comprise (i) coverage, (ii) rigor, (iii) depth, and (iv) independence of verification. Any of six different quality levels discussed in [12] can be assigned to each family.

III. Related Work

Security quantification has been studied in the research already for several years now. Comprehensive overviews of security metrics approaches and objectives are found in, for example, [15–17]. Critical discussions and surveys are available in [18–20]. Skeptics often consider the current state of the art of security so low that any attempt to measure it would not be as success [17]. Evidently, the poor level is a result of the lack of usable tools and methods capable of systematizing security work and obtaining evidence of it. Furthermore, systematic security methods have not been emphasized enough in software engineering. The problem in the many research efforts aimed at security quantification has been that there is a lack of their validation in real or realistic case studies. Although no exhaustive validation has been carried out in this study, it offers many practical insights towards defining an industrial-strength security measurement framework. Among the major attempts to standardize security evaluation are the ISO/IEC 15408 Standard (CC) [14], the ISO/IEC 27000 series of standards [21], and many similar standards preceding them. A severe shortcoming in these efforts is that they are generic in nature and do not focus enough on security risk. Risk-driven and practical frameworks such as the one discussed in this paper offer new potential also for standardization.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

IV. CASE STUDY: A PUSH E-MAIL SYSTEM

In this section, we briefly present the SuI of our case study, a company's Push E-mail service. We also present its identified security risks at high level, SOs and examples of metrics and measurements. The aim is three-fold: (i) to give an implicit example of the application of the approach discussed in Section II, (ii) to gather findings addressing the potential of security metrics and measurements, and (iii) to investigate shortcomings in the approach.

During the case study, different components of the SuI were integrated in a laboratory environment and their security risks and associated controls investigated, along with the metrics modeling and development. These activities were carried out in co-operation between the research and industrial partners of the BUGYO Beyond Eureka CELTIC cluster project. Project's main advances are summarized in [5].

A. The System under Investigation

The Push E-mail [22] functionality is situated at the last hop of the e-mail system, from the Receiver's E-mail Server to the Receiver's Client, which is today often a smartphone. Assume that a Sender would like to send an e-mail message to the Receiver at address name@a-company.com. The sequence of e-mail transfer consists of the following steps [11]:

1. The Sender asks from an E-mail Client called a Mail User Agent (MUA) to send an e-mail message to a Mail Transfer Agent (MTA) on the E-mail Server run by the Sender's Internet Service Provider (ISP).
2. The MTA requests the IP address corresponding to the "to"-address of the e-mail message from the Domain Name System (DNS).
3. The DNS responds with the address resolution information.
4. The Sender's MTA sends the message to the Receiver's MTA using the Simple Mail Transfer Protocol (SMTP).
5. The Receiver's MTA sends the message to his MUA using Post Office Protocol version 3 (POP3) or the Internet Message Access Protocol (IMAP).
6. In the case of an e-mail address managed by a local server, the message is passed to the Mail Delivery Agent (MDA) of the server instead of the MTA.

B. Risk Analysis and Security Objectives

Prioritized SOs for the SuI are agreed upon according to the RA. As concluded in [7], RA should be iterative throughout the system lifecycle. The major categories of risk identified are listed in Table I. Note that the risk categories in the table are not quantified and prioritized. A systematic prioritization effort by a group of core stakeholders is needed. In the table, "C" represents for confidentiality, "I" integrity, "A" availability, and "P" privacy risk.

R1 can result from exploitation strategies of many types – example cases where include an attacker using social engineering, or malicious insiders' knowledge, discovering critical vulnerabilities (e.g., weaknesses in a core configuration file), utilizing knowledge otherwise acquired, using malware, and exploiting a situation in which authentication is not strong enough and there are problems with security patches. R2 can stem from unintentional configuration problems (low-quality configuration management, security patch problems, and human error) or can be a result of attacker activity. R2 has potential to contribute to R1 too. R3 can be realized via brute-force (e.g., dictionary) attacks, or through network eavesdropping and exploitation of default e-mail user passwords. Loss of availability (R4) can be caused by Denial of Service (DoS) attacks, including Distributed Denial of Service (DDoS). Attack strategies for R5 exploit, first and foremost, low end-user security awareness or too great trust.

TABLE I. MAJOR RISK CATEGORIES FOR PUSH E-MAIL SERVICE

#	Risk	C/I/A/P
R1	Attacker gaining unauthorized access to the e-mail system as an administrator and potentially seizing it or even a larger system within or outside the AD	C/I/A/P
R2	Unintentional or deliberate misconfiguration of the system, making it vulnerable to attack	C/I/A/P
R3	Attacker gaining unauthorized access to e-mail messages and their content	C/I/P
R4	Attacker causing the e-mail service to crash or causing delays in it	A
R5	Phishing and spam causing indirect losses to the e-mail user	C/I/A/P

Nowadays, the environment in which Push E-mail services are used is vulnerable to the risks discussed above. In comparison to an e-mail service run on personal computers, typical Push E-mail clients run in a more challenging environment, on various types of mobile devices. Nowadays, keeping smart-phones up to date from a security perspective is not a trivial task. These problems seriously affect the operational security level of the SuI. Examples of these problems are the following:

- There are often changes in application and platform SW and in the service concept which the smart phone uses. Consequently, it is difficult to maintain a trusted and consistent up-to-date system configuration.
- Administration responsibilities are often unclear. The end-user might not have the sufficient rights to keep the configuration up-dated, and the administrator possessing those rights might not be able to maintain an up-to-date configuration. This is because the smart phones under any given company's administration may feature a myriad of network protocols with varying security levels.
- Some smart-phone models have advanced automated functionality, and the case of the end-user having enabled the wrong mode, these functions can cause security risks.

Note that typical Service Level Agreements (SLAs), which can be used to set requirements for companies' international services also, emphasize availability (R4). Other security risks are typically not addressed. In practice, this challenge contributes to difficulties in communicating the risks throughout the development, implementation and operation of services.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

TABLE II. EXAMPLES OF SECURITY OBJECTIVES

#	SO on which an associated high-level SC is based	Risk
1	Authenticity and authorization of administration users and e-mail service users • End-user authentication (e-mail service, end-user role within the AD, and device) • Authentication of administration users • Client/server authentication • Access control in the AD	R1, R2, R3, R4
2	Up-to-date and secure configuration and SW versions for all relevant infrastructure objects • Clear responsibilities • Client: operating system, anti-virus and E-mail SW Client/server authentication • Authentication, Authorization and Accounting (AAA) Server SW within the AD • SW outside the AD (E-mail Server of ISP, MTA, and DNS)	R2, R3
3	Confidentiality and integrity of traffic and messages • Server/client traffic • Secure Sockets Layers/Transport Layer Security (TSL/SSL) channel • Authentication and authorization traffic • Wireless Local Areas Network (WLAN) channel	R3, R4
4	Up-to-date and effective anti-spam, anti-phishing and malicious attachment removal solutions	R5

In addition to the results of RA, the high-level objectives contributing to SOs can be based on suitable best practice, such as the ISO/IEC 27000 series of standards [21], which defines generic confidentiality, integrity, availability and privacy goals. Moreover, company-level security requirements and guidelines can be used. Table II presents examples of specific SOs of the SuI, and their connection to the risks specified in Table I. The actual security solutions of the system, SCs, are based on the SOs. Note that the list presented here is not complete. In the table, "AD" refers to the e-mail service AD of a company.

C. Modeling Metrics' Relationships to Security Objectives

Following the process of Fig. 1, security metrics models are constructed on the basis of the RA results and identified SOs in a prioritized manner. Fig. 2 shows a screenshot from the MVS tool depicting the highest levels of an SMM for an Authorization SC (SC1). The SMM includes the relevant SCs as SMNs immediately below the highest level entity, SuI node. Four other controls are shown in the figure, but suppressed for space reasons: SC2: Secure configuration and versioning, SC3: Confidentiality management, SC4: Spam filtering and malicious-attachment removal, and SC5: Service availability. Suppression of lower-level details is denoted by "+" at the bottom of a metrics node. It is important to note that alternative hierarchical classifications of SCs and, consequently, sub-hierarchies of the SMM, are possible. The choice of risks to be shown at the highest-level depends on the priority of them [7]. For example, secure configuration can be incorporated into separate sub-hierarchies. In the example SMM, we chose to emphasize it as a separate SC and associated SMM branch because of its importance in a typical Push E-mail use environment. In addition to the general confidentiality management sub-hierarchy (SC3), confidentiality concerns of other SCs are emphasized under the relevant sub-hierarchy.

In the example SMM, authorization is divided into two main branches, authentication and access control. The Fig. 5 in the Appendix 1 shows the MVS sub-hierarchy SMM for authentication. The leaves in the SMM represent BMCs with no expansion possibility or components for which a further breakdown is possible. Authentication is further divided into end-user authentication, administration personnel authentication and client/server authentication branches.

Because of the serious consequences of attacks could have for the administration of the e-mail service, a separate authentication metrics collection is used, with requirements stricter than the end-user requirements. Note that, depending on the smart phone device, the device authentication solutions can differ. Furthermore, the administration domain and e-mail service require dedicated authentication solutions

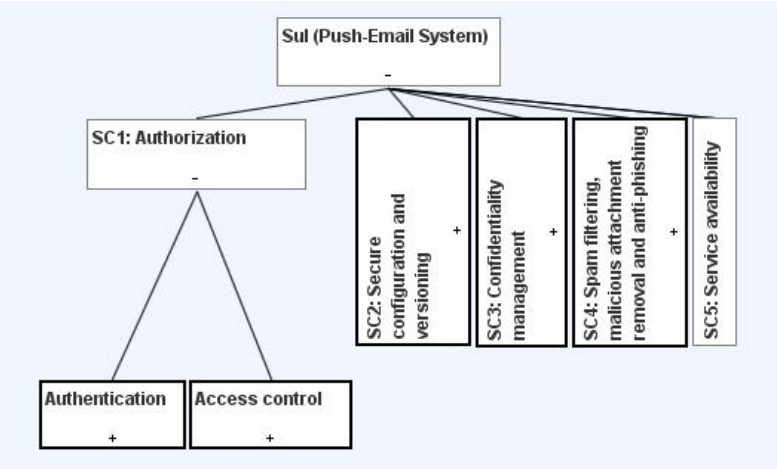


Figure 2. An example high level SMM for the Push E-mail case system. A screenshot from the MVS tool.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

(that can be federated). All authentication branches mentioned incorporate ID Strength and Mechanism Strength sub-hierarchies, following the taxonomy shown in [6]. The ID Strength branch is opened under "End-user Authentication," and Mechanism Strength is shown under "Administration Personnel Authentication," to BMC level.

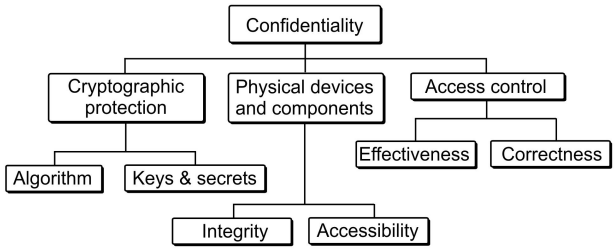


Figure 3. Confidentiality decomposition [6].

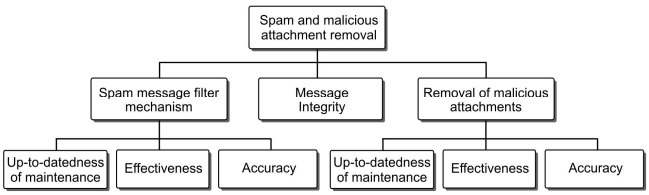


Figure 4. Spam and malicious attachment removal breakdown [11].

Similar sub-hierarchies can be constructed for other SCs via the MVS tool. The SMM sub-hierarchy for SC2 includes configuration, version control and testing and monitoring results, with different infrastructure objects relevant to the security solutions of the SuI forming sub-hierarchies. The infrastructure objects important for the Push E-mail service include the Mail Server, AAA Server, mobile device(s), Push E-mail Client SW, Spam Filter, and firewalls. In addition, several important infrastructure objects reside outside the AD (relevant for such purposes as identity management). Fig. 3 shows a decomposition from [6], which can be applied for metrics modeling associated with SC3. An example metrics hierarchy for spam filtering and malicious attachment removal (SC4) is shown in Fig. 4. SC4 includes security awareness metrics also, to reflect the level of the end-users' capability to withstand phishing attacks. Use of the availability sub-hierarchy (SC5) emphasizes the system's effectiveness in combating DoS and DDoS attacks, and evidence from robustness testing.

The example SMM is risk-oriented. However, it can be arranged in other ways, so as to match the needs of the users of the metrics better. For example, server administration personnel utilize various server programs. A metrics view showing these programs and their main configurations at high level would be beneficial for them.

D. Example Operational Metrics for Configuration Correctness and Deployment of Security Controls

Below, we discuss the difference between BMs and DMs by using some metrics examples from the Push E-mail system. Table III shows how some security metrics of operational BMs for the Push-Email service studied were defined. The examples listed here emphasize configuration correctness and adequate deployment of security controls. Examples of metrics for effectiveness of authentication, authorization, integrity, confidentiality, and availability are given in [6], and for spam filtering and malicious attachment removal in [11].

TABLE III. EXAMPLES OF BASE MEASURES IN PUSH-EMAIL SERVICE

BM	**SC**	**DM**
Mail Server: User authentication mode	1	1.1
Mail Server: Denial of plaintext authentication without encryption	1	1.2
Mobile device: AD password strength	1	
Mobile device: Device password strength	1	
Push E-mail Client SW: User authentication mode	1	1.3
Mail Server: Operating System UTD	2	2.1
Mobile device: Operating System UTD	2	2.2
Push E-mail Client SW: Service SW UTD	2	
Mail Server: IMAP server encryption activated mode	3	3.1
Mail Server: IMAP minimum encryption key length	3	3.2
Mobile device: WLAN encryption configuration	3	
Push-Email Client SW: Encryption algorithm mode	3	3.3
Mail Server: Spam filter SW UTD	4	
Mail Server: Malicious attachment removal SW UTD	4	
Mail Server: Mail backup UTD	5	5.1

TABLE IV. EXAMPLES OF DERIVED MEASURES

DM	**Example expression**
1.1	Configuration command check: `auth_mechanisms = plain login cram-md5`
1.2	Configuration command check in Dovecot configuration file: `disable_plaintext_auth`
1.3	Configuration command check: `auth_mechanisms = plain login cram-md5`
2.1	Version information query – example reply from the system: *Linux webrouter 2.6.32-24-generic #43-Ubuntu SMP Thu Sep 16 14:17:33 UTC 2010 i686 GNU/Linux*
2.2	Version information query – example reply from the system: *Linux Nokia-N900 2.6.28-omap1 1 PREEMPT Fri Aug 6 11:50:00 EEST 2010 armv7l unknown*
3.1	Configuration command check: `ssl = required`
3.2	Configuration command check: `ssl_cipher_list = ALL:!LOW:!SSLv2`
3.3	Configuration directive check: `SSLCipherSuite AES256-SHA:AES128-SHA`
5.1	Checking appropriate use of the `rsync` application, example: `rsync –a /home/user/Maildir /media/backupdrive/mail`

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

The BM examples listed in Table III originate from different abstraction levels in the metrics hierarchy. For example, "Denial of plaintext authentication without encryption" is a more detailed BM than "AD password strength." The SC number associated with the BM is shown. Furthermore, the table gives a reference to a DM derived from the BM, utilizing OpenSSL [23] and Dovecot Secure IMAP Server [24] commands, listed in Table IV. "UTD" refers to *Up-To-Datedness*. Mapping DMs from BMs is a 1-to-*N* process: one or several DMs represent each BM. In Table IV, only one example DM is given for each BM.

It is evident that the *SO representativeness* of scattered DMs is not enough for sufficient security evidence at the SO level. The situation is due to the information being missing, or the evidence needed being either unavailable or unattainable. Comparison of the examples in Table III and IV shows that, especially during the process of interpretation of BMs, a lot of information is lost. First of all, it is important that this kind of measurability challenges be duly kept on track in the metrics hierarchy properties.

Certain evidence requires other types of information-gathering than technical measurement architectures. For example, anti-phishing (SC4) metrics require measurement of end-users' security awareness. Puhakainen [25] investigates factors contributing to this awareness. He introduces theories based on training, awareness campaigns and punishment/reward.

The problem of missing information gaps can be mitigated by means of suitable assessment methods to give evidence of the situation. The quality metrics from [12] can be utilized for this purpose. The level of investigation preferred for assessment is the BMC level, but, depending on the granularity of metrics, the level may be higher or lower. The following questions form the basis for assessment:

1. Coverage: How widely has the BMC been investigated?
2. Rigor: Has there been enough rigor in the investigation?
3. Depth: In what depth has the BMC been investigated?
4. Independence: Has the investigation been carried out independently of system development and/or operation?

Measurement intervals for security-related measurements depend on various factors. The most important ones are the type of evidence needed, critical changes in the SuI, its availability, its attainability, and the efficiency constraints affecting measurements. Measurement needs can change in response to any combination of these factors.

V. Discussion of Benefits and Challenges

This section discusses the results from the application of our approach in the case study in a more general context.

A. Benefits

Today, state-of-practice activities in operational security assurance for telecoms services are largely based on *ad hoc* practices. Obviously, systematic evidence-driven security approaches bring several advantages.

By utilizing metrics, one can make more evident the potential bias between the security implementation and its specification [7]. This enables decision-makers to make informed decisions about investments in security countermeasures and risk mitigation. Visibility and constant evaluation of the status of operational security assurance highlight areas of potential problems and allow addressing them before risks are actualized. Security level in new R&D efforts and system operation will improve if factors contributing to security effectiveness and efficiency, along with their relationships, can be analyzed and documented.

The traceability of the objective requirement chain from the outcome of the first iterations of RA to SOs, and further to design and operational requirements, is systematized and better managed via the collection of metrics. Systematic risk-based thinking throughout the system lifecycle supports more effective and efficient security solutions. Feedback to R&D activities from the system's operation, utilizing metrics and measurements, is a powerful tool in assisting the future R&D efforts to focus on relevant security issues.

B. Challenges

As can be seen from the SMM of Figs. 2 and 3, obtaining sufficient evidence of security issues in a realistic system requires a wide collection of metrics, measurements and assessments. The need for wide metrics collections can result in a burden for practical service administration if there are no usable tools offering the right type of information.

Despite advances, such as the MVS tool, there are still many question marks in efficient metrics management. Simple measurement result aggregation, in combination with poor representativeness of the metrics used, results in the problem that the model does not express security phenomena in a full enough and credible way. Moreover, ensuring the correctness of metrics, i.e., that they represent the correct aspects relevant to SOs – still remains a challenge.

For many security issues, automated measurement is not possible; some of the required information is simply not available or attainable. Therefore, to increase the representativeness of metrics (i.e., fill the gaps between RA results, SOs, SCs, BMCs, BMs and DMs), one should use assessments. Credible assessment techniques still require advances. Moreover, common agreements on *trust value* management are needed.

The cost and effort in creation, maintenance and evolution of metrics and measurement architectures is a challenge. The advantages of using metrics and measurements should be compared with the added burden. Since the present study was a laboratory research effort, cost-effectiveness was not investigated. Today, proper administration of complex servers connected to the Internet requires personnel to follow their status constantly. In practice, resourcing can be troublesome. Accordingly, existing infrastructure, functionality and processes should be exploited as much as possible, to incur minimal overhead. Daily manual follow-up of logs is not feasible for visualization of every security aspect, or even every relevant one. The timing of responses to security problems is an issue too. For example, if a security risk related to a server configuration is detected in time, the question remains of whether it can be dealt with right away or only during off-peak usage hours. Live server setups require frequent updates to address security concerns and integration with

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

management tool updates. Otherwise, the tools would only provide snapshots of certain situations. If information-gathering is too detailed or frequent, the measurement approach can affect performance of the actual SuI. Moreover, the logs can grow so big that they can be stored for only short measurement periods.

VI. CONCLUSIONS AND FUTURE WORK

We have discussed our experiences from development of security metrics and corresponding measurements in a Push E-mail service. The approach used is based on risk-driven hierarchical security metrics development, and utilization of a visualization tool and associated measurement and assessment approaches. Through the use of security metrics and measurements, the differences between security design and its implementation can be made evident, enabling informed decision-making. Moreover, the security objectives and requirements can be managed and traced throughout the system lifecycle.

Our experiences from the modeling of the case system showed that sufficient and credible security evidence consists of a wide collection of metrics, which should be managed in such a way that the relationships extending from high-level risk-driven security objectives and detailed measurements can be traced. In practice, the detailed measurements' correspondence with security objectives is often poor. Consequently, assessment and careful utilization of the available evidence is needed if we are to be able to fill the information gaps.

The cost-effectiveness of metrics and measurements was not addressed in this research effort. Our future work will include cost-effectiveness analysis of the proposed approach in real-world scenarios. Further evolution of the approach is planned in connection with this work.

ACKNOWLEDGEMENTS

The work presented here has been carried out in three European and Finnish national research projects: the BUGYO Beyond Eureka CELTIC cluster project (2008–2011), GEMOM EU FP7 ICT project (2008–2010), and Cloud Software Program (2008–2013) launched by the Finnish Strategic Centre for Science, Technology and Innovation TIVIT Plc. We wish to thank our colleagues involved in these projects for their related work and helpful discussions that made this study possible.

REFERENCES

[1] R. Savola, "A taxonomical approach for information security metrics development," NORDSEC '07, 2007.

[2] R. Savola, "A security metrics taxonomization model for software-intensive systems," Journal of Information Processing Systems, Vol. 5, No. 4 (Dec. 2009), pp. 197–206.

[3] W. Jansen, "Directions in security metrics research," U.S. National Institute of Standards and Technology, NISTIR 7564, Apr. 2009, 21 p.

[4] Information Technology Security Evaluation Criteria (ITSEC), Version 1.2, Commission for the European Communities, 1991.

[5] S. Haddad, S. Dubus, A. Hecker, T. Kanstrén, B. Marquet and R. Savola, "Operational security assurance evaluation in open infrastructures," Proceedings of CRiSIS 2010, Sept. 26-28, 2011, Timisoara, Romania, pp. 100–105.

[6] R. Savola and H. Abie, "Development of measurable security for a distributed messaging system," Int. Journal on Advances in Security, Vol. 2, No. 4, pp. 358–380.

[7] R. Savola, C. Frühwirth and A. Pietikäinen, "Risk-driven security metrics in agile software development – an industrial pilot study," Submitted (2012).

[8] T. Kanstrén, R. Savola, A. Evesti, H. Pentikäinen, A. Hecker, M. Ouedraogo, K. Hätönen, P. Halonen, C. Blad, O. López and S. Ros, "Towards an abstraction layer for security assurance measurements (invited paper)," Proceedings of ECSA: Companion Volume, pp. 189–196.

[9] T. Kanstrén, R. Savola, S. Haddad and A. Hecker, "An adaptive and dependable distributed monitoring framework," Int. Journal on Advances in Security, Vol. 4, Nos. 1 & 2, 2011, pp. 80–94.

[10] R. Savola and P. Heinonen, "A visualization and modeling tool for security metrics and measurements management," Proceedings of ISSA 2011, Johannesburg, South Africa, 8 p.

[11] R. Savola, H. Pentikäinen, and M. Ouedraogo, "Towards security effectiveness measurement utilizing risk-based security assurance," Proceedings of ISSA 2010, Aug. 2–4, 2010, Sandton, South Africa, 8 p.

[12] M. Ouedraogo, R. Savola, H. Mouratidis, D. Preston, D. Khadraoui, and E. Dubois, "Taxonomy of quality metrics for assessing assurance of security correctness," Software Quality Journal, Online First, Nov. 30, 2011, 30 p.

[13] M. Ouedraogo, D. Khadraoui, B. de Rémont, E. Dubois, and H. Mouratidis, "Deployment of a security assurance monitoring framework for telecommunication service infrastructure on a VoIP system," Proceedings of NTMS '98.

[14] ISO/IEC 15408-1:2005: "Common Criteria for information technology security evaluation – Part 1: Introduction and general model," ISO/IEC, 2005.

[15] D. S. Hermann, "Complete guide to security and privacy metrics – measuring regulatory compliance, operational resilience and ROI," Auerbach Publications, 2007, 824 p.

[16] A. Jaquith, "Security metrics: Replacing fear, uncertainty and doubt," Addison-Wesley, 2007.

[17] N. Bartol, B. Bates, K.M. Goertzel and T. Winograd," Measuring cyber security and information assurance: A state-of-the-art report," Information Assurance Technology Analysis Center (IATAC), May 2009.

[18] J. McHugh, "Quantitative measures of assurance: Prophecy, process or pipedream?" Workshop on Information Security System Scoring and Ranking (WISSSR), ACSA and MITRE, Williamsburg, Virginia, May, 2001 (2002).

[19] D. McCallam, "The case against numerical measures of information assurance," Workshop on Information Security System Scoring and Ranking (WISSSR), ACSA and MITRE, Williamsburg, Virginia, May, 2001 (2002).

[20] V. Verendel, "Quantified security is a weak hypothesis: A critical survey of results and assumptions," New Security Paradigms Workshop, Oxford, U.K., 2009, pp. 37–50.

[21] ISO/IEC 27000:2009: "Information technology – Security techniques – Information security management systems – Overview and vocabulary," ISO/IEC, 2009

[22] R. W. Smith, "LPIC-2: Linux Professional Institute Certification, study guide," Sybex, 2011, 694 p.

[23] ——, "OpenSSL project – Cryptography and SSL/TSL toolkit," Website: www.openssl.org [Accessed Jan. 15, 2012].

[24] ——, "Dovecot – Secure IMAP Server," Website: www.dovecot.org [Accessed Jan. 15, 2012].

[25] P. Puhakainen, "A design theory for information security awareness," PhD thesis, University of Oulu, Finland, 2006.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

APPENDIX 1

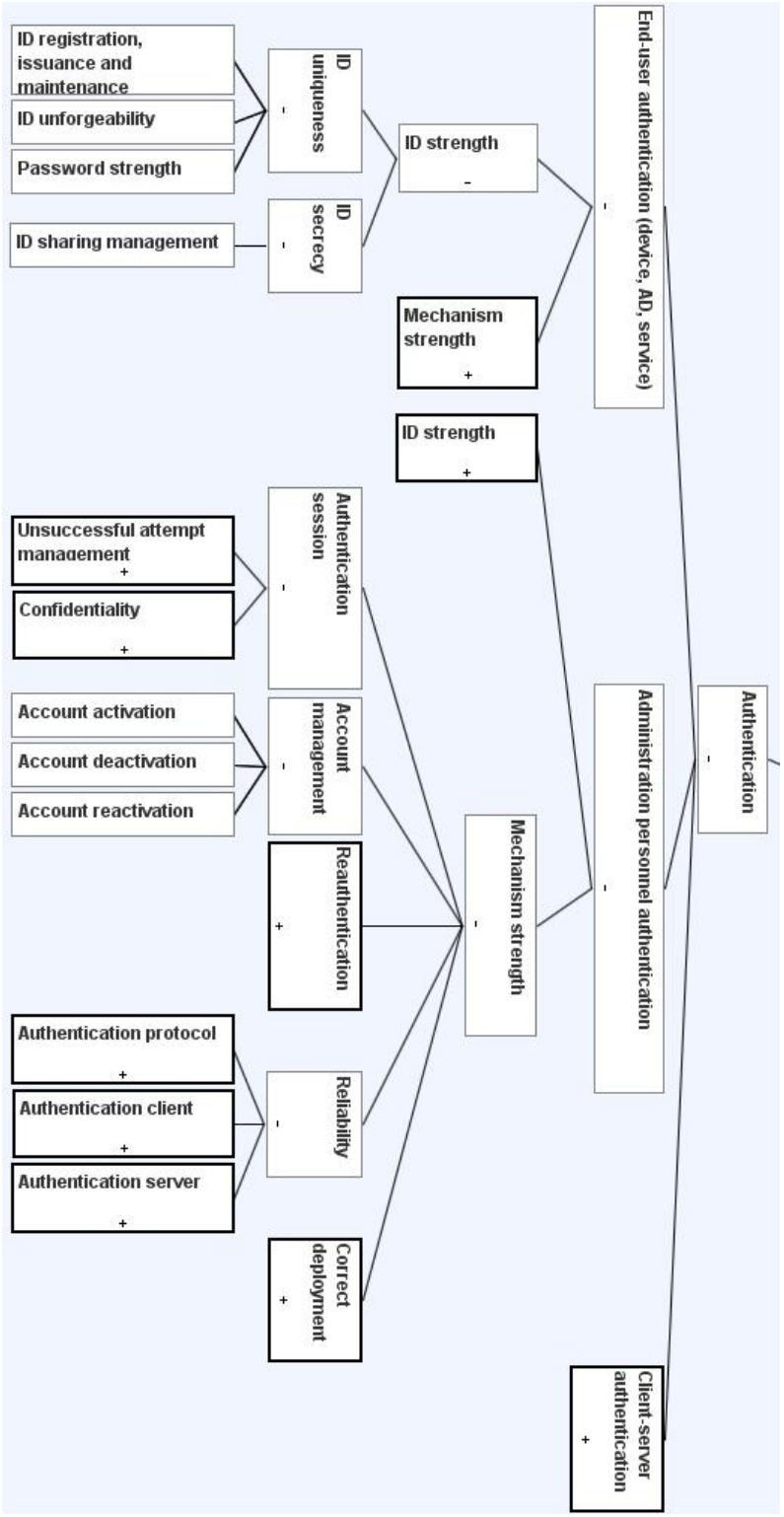


Figure 5. A screenshot of the authentication SMM branch.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

Impact of Gaming during Channel Zapping on Quality of Experience

Robert E. Kooij[1,2]
[1]TNO (Netherlands Organization for Applied Scientific Research)
Delft, the Netherlands
robert.kooij@ tno.nl
[2]Fac. of Electrical Engineering, Mathematics and Computer Science
University of Technology Delft
Delft, the Netherlands

Michiel Geijer
Earth Sciences, Faculty of Geosciences
Utrecht University
Utrecht, the Netherlands
michielgeijer@gmail.com

Abstract—**This paper belongs to a research program started in 2006, dealing with Quality of Experience (QoE) aspects of channel zapping. The program, which has relevance for the broader topics Quality of Service and Next Generation Networks, started with quantifying the QoE expressed as a so-called Mean Opinion Score (MOS) for situations when a black screen is visible during channel zapping. Based upon the observation that the QoE is (possibly) increased by showing information while the user waits for the target channel to appear, the program continued with assessing the QoE in case advertisements are shown during channel switching. In this paper, we quantify the impact on QoE when offering users an interactive game during channel zapping. Our subjective experiment shows that for zapping times greater than 2.25 seconds, offering games during zapping instead of the usual black screen, leads to a better QoE. For zapping times larger then about 3 seconds, the MOS for the 'game' scenario is larger than 3.5, indicating the onset to acceptable quality. For zapping times below 1 second, the MOS for the 'game' scenario is very low, indicating that it is a bad idea to show games in case of such short zapping times. The scenario where during zapping advertisements are shown is always outperformed by either the 'black screen' or 'game' scenario.**

For small zapping times, the ease of play is much too low. For instance, for a zapping time of 1 second, only 18% of the test subjects managed to play the game in time. For zapping times of 3 seconds the ease of play becomes higher than 80%. At zapping times of 5 seconds, the game score is as high as 82%.

Keywords - Channel Zapping; Quality of Experience; Mean Opinion Score; IPTV; gaming.

I. INTRODUCTION

This paper belongs to a series of papers, the first of which was published in 2006 [1], that all deal with Quality of Experience (QoE) aspects of channel zapping. In the highly competitive market of Triple Play (i.e., the commercial bundling of voice, video and data on a common IP based network infrastructure), Service Providers, which are offering high quality IPTV services, need to address the QoE requirements of IPTV. QoE takes into account how well a service meets customers goals and expectations rather than focusing only on the network performance.

One of the key elements of QoE of IPTV is how quickly users can change between TV channels, which is called channel zapping. The zapping time is the total duration from the time that a viewer presses the channel change button, to the point the picture of the new channel is displayed, along with the corresponding audio. Minimum quality requirements for many aspects related to IPTV have been specified by both the ITU [2] and the DSL Forum [3]. However, in the ITU document there are no recommendations at all related to zapping times, while in the DSL forum document it is recommended to limit zapping time to an arbitrary maximum of 2 seconds. Additionally it is noticed in the document that providers should strive for zapping times in the order of 1 second.

Because these quality requirements are rather vague Kooij et al. [1] started a research program in order to get insight in the relation between QoE and zapping time. In [1], a number of subjective tests was described, in which, during channel zapping, a black screen, which contained the number of the target channel, was visible. The QoE was expressed as a so-called Mean Opinion Score (MOS). The test subjects (21 in total) could select one of the following five opinion scores, motivated by the ITU-T ACR (Absolute Category Rating) scale, see [4]: 5: *Excellent zapping quality*, 4: *Good zapping quality*, 3: *Fair zapping quality*, 2: *Poor zapping quality*, 1: *Bad zapping quality*.

The main result of [1] is an explicit relation between the user perceived QoE and the zapping time. From this relation it was deduced that in order to guarantee a MOS of at least 3.5, which is considered the lower bound for acceptable quality of service (see [4]), we need to ascertain that the zapping time is less than 430 ms. Note that for MOS = 3.5 the average user will detect a slight degradation of the quality of the considered service. The requirement on the zapping time mentioned above is currently not met in any implementation of IPTV (see, for instance, [5] and [6]).

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

In [7], new subjective experiments were described where the zapping took place under different conditions. These experiments included 'lean backward' zapping, that is, zapping while sitting on a sofa with a remote control. The subjects are more forgiving in this case and the requirement for acceptable QoE could be relaxed to 670 ms. In addition, [7] reports on subjective experiments where the zapping times were varying. It is found that the MOS rating decreased if zapping delay times were varying.

The research program on QoE and zapping continued based on the observation that in order to increase the QoE of channel zapping, two approaches are possible. In the first approach, the actual zapping time is reduced. An example of this method is given by Degrande et al. [8]. They suggest to retain the most recent video part in a circular buffer and display this video until the incoming channel is ready.

In the second approach, the QoE is (possibly) increased by showing information while the user waits for the target channel to appear. The displayed information could be about the target channel, personalized content or advertisements (see also [9]). Subjective tests following the second approach have been described in [10]. The main conclusion of [10] was that by showing advertisements during channel zapping, instead of the usual black screen, users rated the experience higher, at least for realistic values of the zapping time.

Kooij et al. [5] extended the results of [10] in the following way: the number of persons that participated in the subjective experiments was increased from 12 to 30, the measurements were added that provide insight about zapping times for today's digital television services, and a completely new section was added about how the finding of this research could be used to design a system for optimal zapping experience. This system for optimal zapping experience has two patents pending.

The aim of this paper is to assess the QoE of channel zapping when, during zapping, the user is offered interactive content in the form of a game, instead of the usual black screen. To our knowledge, this paper is the first ever that deals with interactive content during zapping.

The rest of this paper is organized as follows. In Section II, the possible effect of gaming on IPTV perceived quality is analyzed and various factors that contribute to the results are listed. In Section III the experiment performed to quantify the user perception is described. In Section IV, the results obtained from the subjective tests are presented. Finally, conclusions are given in Section V.

II. QUALITY OF EXPERIENCE AND GAMING

Using content such as advertisements or gaming during IPTV channel zapping is an approach that tries to increase the QoE while the service quality or zapping time remains unchanged. Obviously, not all people would be happy to see advertisements during zapping. People are probably more open to gaming during zapping because through a game educational or entertaining content can be presented. Although the business driver for advertisements seems of high impoprtance we anticipate that gaming during zapping can also boost the QoE, in two different ways.

a) Users playing games during channel zapping, will not be bored with the longer zapping times. Hence the QoE of the users for the channels with games could increase with respect to the black screens. This is actually what we have measured in the conducted subjective experiment.

b) The second consequence is that the providers might earn money from these games, through a similar business case as apps for smart phones. Therefore, they can lower the price of the service. Obviously, a lower price is one of the factors that can boost the QoE.

It should be noted that the effect of games (if they were implemented) on QoE is not just straightforward. Rather, it depends on various factors, which could affect the QoE positively or negatively.

a) The type of game: A particular user could like some sort of game and dislike other type of games.

b) The difficulty of the game, in relation to the length of the zapping time. For example, if the zapping time is very short, then it is probably not even possible to play the game, hence a game in this scenario could be quite annoying. On the other hand, if the zapping time is very long and the game is too simple, users may not be challenged enough to appreciate the game.

c) Obviously, the game that is offered to the user should not be exactly the same all the time. Probably it is best to offer different variations of the same game to the users.

Some of the factors above could positively affect the user perception. However, the implementation complexity also increases if all these issues are to be properly addressed. The best approach to use these games is to select a game randomly from a set of pre-rendered games stored in the Set-top Box (STB) when the user zaps to a different channel. It is recommended to use games that the user likes. Using pre-rendered games is important because the zap screen can then be displayed immediately.

III. THE EXPERIMENT

A. *Design of the experiment*

For the TV channel zapping experiment, a HTML page containing two frames is implemented. Through the lower frame the user can switch between 5 different TV channels. The TV channels are implemented as Flash files in the upper frame of the HTML page. The TV channels contain the following content: a dancing girl, a cartoon, three men in a suit, a room service scene and a dancing man.

Figure 1 shows the HTML page with the dancing girl channel being on. Although the Flash files contain audio, sound is switched off during the experiments. Audio might be added to the experiment but the synchronization problem

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

will be another cause for quality degradation. So, to assess the quality experienced for zapping times, we felt it is better to make the experiments with no sound, because otherwise the test subjects opinions might be biased by the synchronization quality. The videos within the upper frame are displayed in a screen of size of 550x400 pixels.

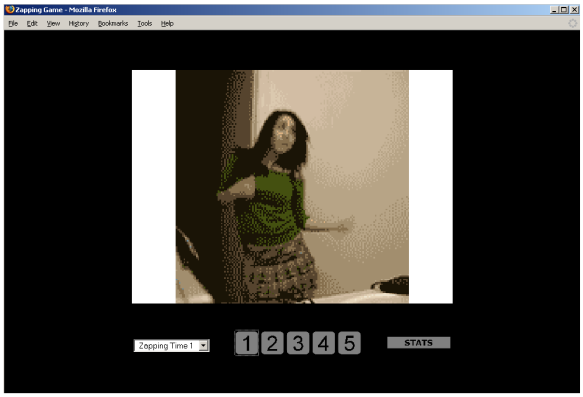


Figure 1: HTML page used in the experiment

In the experiments reported in [1] and [10], seven zapping times between 0 and 5 second were implemented in arrays in the javascript code. These zapping times were 0, 0.1, 0.2, 0.5, 1, 2 and 5 seconds. Moreover, a random ordering of these zapping times was implemented for each of the test subjects that participated in the subjective experiments. The number of test subjects in [1] was 21 while 12 test subjects participated in [10]. Additional subjective experiments with 18 test subjects, including also the zapping times 3 and 4 seconds, were reported in [5].

When the user zaps to a new channel, the page sleeps for a time corresponding to the implemented zapping time before the requested channel is displayed. During this time, either a black screen is shown [1], [7], or an advertisement [5], [10]. This paper assesses the impact of having an interactive game on display during channel zapping. The game consists of a traffic situation where the user has to determine whether or not it is allowed to follow the direction that is indicated by the arrow. The perspective of the user is that of the driver of the car in the bottom of the picture (see Figure 2).

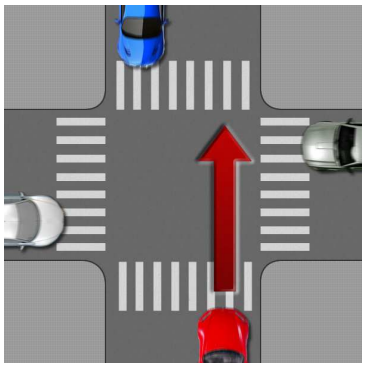


Figure 2: Example of a game situation

The chain of events is depicted in Figure 3. All games situations are depicted through pictures of size 400x400 pixels.

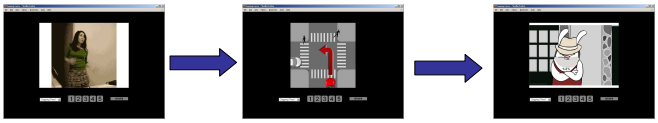


Figure 3. Showing a game during zapping

For the experiment the zapping times were 0.5, 1, 2, 3, 4 and 5 seconds. We left out some of the original zapping times used in [1] and [10], in order to keep the length of the test sufficiently short, thus preventing fatigue of the test objects, and because zapping times of 0.1 and 0.2 seconds are presumably too short for interactive gaming.

B. The actual experiment

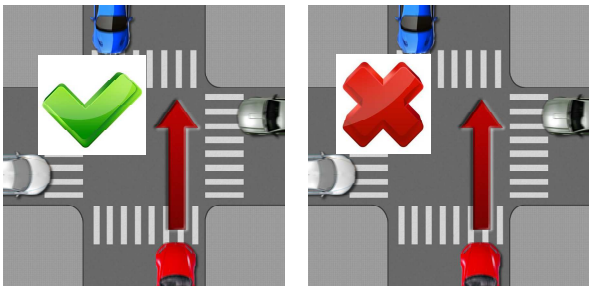
In this subsection, we describe the details of the subjective experiment. The test subjects consist of a total of 21 people at TNO in Delft, the Netherlands and at the second author's house in Breda, the Netherlands. The test subjects vary in age (17 – 57 years), gender (14 male, 7 female) and experience. According to [4] at least 15 observers should participate in subjective testing of multimedia services. They should not be directly involved in quality assessment as part of their work and should not be experienced assessors. These conditions are met for the group of test subjects.

To view the channels a laptop (Core i3, 4GB RAM, windows XP, 1600x900 pixels screen resolution) is used as a TV set. The experiment that we have conducted is of 'lean backward zapping' type. That means the user will sit back in a chair and use the remote control to zap between the channels. A Sony Ericsson w660i with the Bluetooth Human Interface Devices (HID) protocol implemented on it, is used as a remote control device. The mobile phone is depicted in Figure 4.

Figure 4: Sony Ericsson w660i used as remote control

Through the HID protocol, pressing the buttons 1 to 5 on the phone realize a channel switch to the corresponding channel on the HTML page. In order to play the game that is shown during zapping, the user needs to press the upper right button of the phone when he/she thinks it is allowable to

drive the car in the indicated direction. If he/she thinks it is NOT permitted to drive the car, the upper left button needs to be pressed (see also Figure 4). After the test subject has presses one of the two buttons, visual feedback is given by displaying either a “check symbol” (correct answer) or a “cross symbol” (incorrect answer) on top of the depicted traffic situation (see Figure 5).



(a) correct (b) incorrect

Figure 5: Feedback on user input

The experiment contains two parts, the training and the actual experiment. The training part is intended to familiarize the subject with the test environment. The training starts with a situation requiring immediate zapping, that is, a situation corresponding to MOS = 5. The test subject can switch between channels by pressing the buttons numbered 1 to 5 on the mobile phone.

In the second part of the training, the test subject is offered a situation with very long zapping time (i.e., 5 seconds), where during zapping a black screen is shown. This situation corresponds with MOS = 1.

In the third part of the training, the test subject is offered the traffic game during zapping. The aim of the game is to determine for each offered traffic situation whether the car can drive in the indicated direction or not.

After the test subject has switched channels a number of times (approximately ten times) he/she then presses the “stats” button (see Figure 1) to reveal the score of the games played by the test subject. An alert box then appears, containing the statistics of the gameplay (see Figure 6).

Note that it is possible that the number of zaps is larger than the sum of “number of correct answers” and “number of wrong answers”, namely when the test subject was too late to press one of the answer buttons during zapping.

At this point, the test subject is ready to start the actual experiment. Below we show the literal text that is presented to the test subjects, who first had to do the ‘black screen’ scenario, and then the ‘game’ scenario.

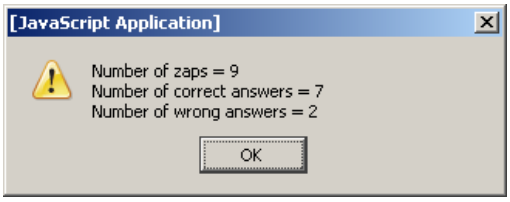


Figure 6: Statistics of the gameplay

Black screen

1. In this paper form, fill in your name, age, your experience with gaming and SMS and the date
2. Write down in this paper form the Scenario (for “black screen”) you are in. You can see this in the URL in the Firefox browser
3. Select “Zapping Time 1” in the drop-down list on the left-bottom of the screen. Experience the zapping time, by pressing on the buttons numbered from 1 to 5. Try to do a total of about 10 zaps.
4. Write down your MOS value on the table shown on the form.
5. Then, select “Zapping Time 2” in the drop-down list, repeat steps 3 and 4, until you have assessed all 6 Zapping Times.

Game

6. Write down in the paper form the Scenario (for the game) you are in. You can see this in the URL in the Firefox browser
7. Select “Zapping Time 1” in the drop-down list on the left-bottom of the screen. Experience the zapping time, by pressing on the buttons numbered from 1 to 5. Also try to play, during the zapping, the game. Try to do a total of about 10 zaps. When that is done press the “stats” button.
8. Write down your MOS value in the table shown on the form. Also write down the statistics from the alert-box.
9. Then, select “Zapping Time 2” in the drop-down list, repeat steps 3 and 4, until you have assessed all 6 Zapping Times.
10. If you want to, you can write general comments in the box on the bottom of the form.

The order of the six zapping times (0.5, 1, 2, 3, 4, 5 seconds) was randomized into four different orders (denoted by A, B, C and D). Each test subject was offered a different order for the ‘black screen’ scenario and the ‘game’ scenario.

IV. RESULTS

A. *MOS results*

The results obtained for each zapping time are analyzed and averaged over the number of test subjects to obtain the

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

MOS for each zapping time. This is done for both the 'black screen' and the 'game' scenario, that is, the case where, respectively, during zapping a black screen is shown and the case where the game is shown. The obtained MOS scores, together with their 95% confidence intervals, are shown in Figure 7.

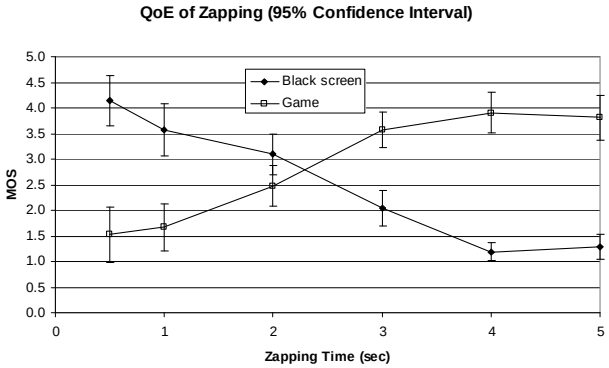


Figure 7: MOS for "black screen' and 'game'

The following important insights can be obtained from Figure 7:

- The MOS for 'game' exceeds the MOS for 'black screen' for zapping times greater than about 2.25 seconds. This implies that the users prefer 'game' only when the zapping time is sufficiently large.
- For zapping times larger then about 3 seconds the MOS for 'game' is larger than 3.5, indicating the onset to acceptable quality.
- For zapping times below 1 second the MOS for 'game' is very low, indicating that it is a bad idea to show games in case of such short zapping times.

B. Comparison with previous results

The 'black screen' experiment was conducted before, see [1]. The authors of [1] suggested the following model for the relation between zapping time (in seconds) and QoE (expressed in MOS), for the 'black screen' case:

$$MOS = \max\{1, \min\{-1.02 \cdot \ln(ZappingTime) + 2.65, 5\}\}. \quad (1)$$

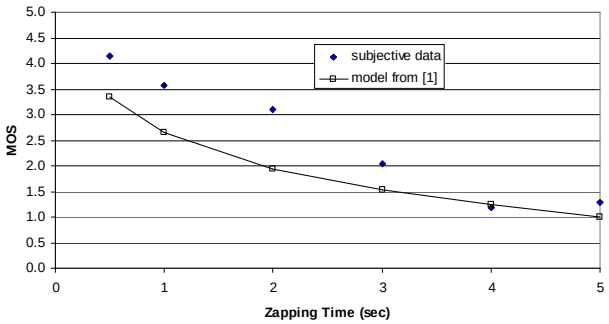


Figure 8. Comparing our 'black screen' results with the model from [1]

It is clear from Figure 8 that our new subjective data is much more optimistic than that predicted by model (1), especially for zapping times smaller than 4 seconds. We can think of three possible explanations for this. First, the experiment reported in [1] was 'lean forward', while our experiment was 'lean backward'. It is known from [7] that this leads to higher MOS scores. Secondly, the zapping times offered to the test subjects in [1] were on a logarithmic scale (0, 0.1, 0.2, 0.5, 1, 2, 5 seconds) while for our experiment we basically used a linear scale. This possibly might lead to a bias in the form of lower MOS scores for small zapping times. Finally, the tests described in [1] were conducted in 2006, while our tests took place in 2011. People may have become more accustomed to zapping times in the order of 1 to 3 seconds in the last five years.

In [5] and [10], we have also assessed the QoE for zapping in case advertisements were shown during channel switching. These experiments, using the same zapping times as in our experiment, took place in 2009, therefore we feel we can compare the results from [5] and [10] with our new results, see Figure 9. Note that the performance evaluation of the advertisement scenario did not take the reduction of user's cost into account.

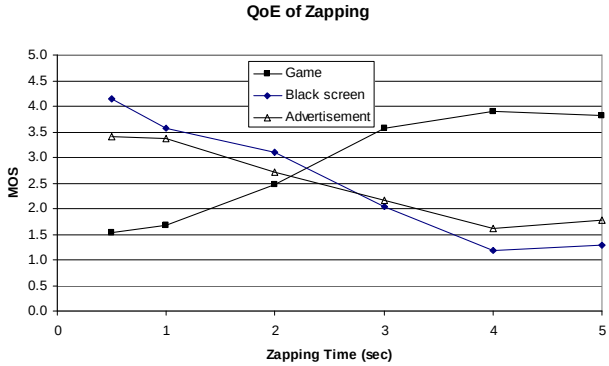


Figure 9: MOS for "black screen', 'game' and 'advertisement'

We conclude from Figure 9 that the 'advertisement' scenario never performs better than both the 'black screen' and the 'game' scenario at the same time. Note that it was anticipated in [5] that there could be a regime for which the advertisement scenario would be preferred but this is not reflected in Figure 9.

C. Results on ease of play and game score

As explained in Section III.B, we have also assessed the ease of play and the scores of the game played during channel zapping. The ease of play is quantified as follows. For each test subject we know, for a given zapping time, the total number of zaps (Z) and the number of correct (C) and the number of incorrect answers (I). The ease of play for this test subject and zapping time is defined as (C+I)/Z*100%. The overall ease of play for this zapping time is obtained by averaging this quantity over all 21 test subjects. In a similar way the overall game score is obtained by averaging the

quantity C/Z*100% over all 21 test subjects. The results are depicted in Figure 10.

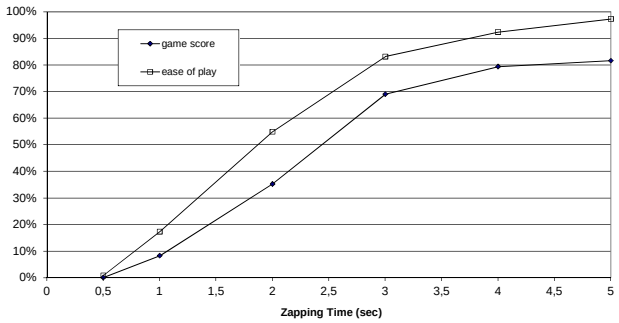


Figure 10: Ease of play and game score

It is clear that for small zapping times the ease of play is much too low. For instance, for a zapping time of 1 second, only 18% managed to play the game in time. Around zapping times of 3 seconds the ease of play becomes higher than 80%. The game score curve shows an s-shape with the inflection point around 2.5 seconds. At zapping times of 5 seconds the game score is as high as 82%. It is interesting to notice that the percentage of incorrect answers remains more or less constant at 18% for zapping times larger than 2 seconds.

D. Discussion on user comments

In addition to evaluating the MOS and keeping their gaming scores, users were asked to comment on the subjective experiments. The following are the main comments of the users.

a) More than 50% of the test subjects made comments on the positions of the "correct" and "incorrect" buttons on the mobile phone. As on most mobile phones the upper left button is green and the upper right is red, indicating some positive action (e.g., making a call) or negative action (e.g., terminating a call), they might have found it more intuitive if the "correct" and "incorrect" buttons would have been placed the other way around. A few test subjects remarked that the buttons on the mobile phone are too small to make the gameplay enjoyable.

b) A number of test subjects mentioned that for small zapping times it becomes impossible to play the game, which leads to irritation. Some suggest that it is better in these cases to show a black screen instead of the game. Of course this is completely in line with the system for optimal zapping experience we have suggested in [5].

V. CONCLUSIONS

From our conducted subjective experiment, it is found that for zapping times greater than 2.25 seconds, offering games during zapping instead of the usual black screen leads to a better QoE. For zapping times larger then about 3 seconds, the MOS for the 'game' scenario is larger than 3.5, indicating the onset to acceptable quality. For zapping times below 1 second the MOS for the 'game' scenario is very low, indicating that it is a bad idea to show games in case of such short zapping times. The scenario where during zapping advertisements are shown, is always outperformed by showing either the 'black screen' or 'game' scenario.

For small zapping times the ease of play is much too low. For instance for a zapping time of 1 second, only 18% of the test subjects managed to play the game in time. Around zapping times of 3 seconds the ease of play becomes higher than 80%. The game score curve shows an s-shape with the inflection point around 2.5 seconds. At zapping times of 5 seconds the game score is as high as 82%.

The results obtained in this paper are useful when implementing a system for optimal zapping experience, which is described in [5]. In the near future we want to implement the system in a field trial and conduct further subjective experiments with the system, taking into account longer time scales, i.e., we will assess QoE after a period of, for instance, 3 months. Finally, we suggest that the system might be adapted to function as a broader 'System for Optimal Waiting Experience'. Hopefully this inspires others to improve the quality of experience in our day-to-day activities.

REFERENCES

[1] R.E. Kooij, O.K. Ahmed and K. Brunnström, "Perceived Quality of Channel Zapping", Proc. of the fifth IASTED International Conference, Comm. Systems and Networks, Aug. 28-30, 2006, Palma de Mallorca, Spain, pp. 155-158.

[2] ITU Focus group on IPTV, "Quality of Experience requirements for IPTV", FG IPTV-DOC-0118, July 2007.

[3] DSL Forum, "Triple Play Services Quality of Experience (QoE) Requirements and Mechanisms", Technical Report TR-126, December 2006.

[4] ITU-T Rec. P.910, "Subjective video quality assessment methods for multimedia applications", April 2008.

[5] R.E. Kooij, V.B. Klos, B.E. Godana, F.P. Nicolai and O.K. Ahmed, "Optimising the Quality of Experience during Channel Zapping", International Journal on Advances in Systems and Measurements, vol 2, no. 2&3, pp. 204-213, December 2009.

[6] F.M.V. Ramos, J. Crowcroft, R. J. Gibbens, P. Rodriguez and I.H. White, "Reducing channel change delay in IPTV by predictive pre-joining of TV channels", Signal Processing: Image Communication, Volume 26, Issue 7, Pages 400-412, August 2011.

[7] R.E. Kooij, F. Nicolai, O.K. Ahmed and K. Brunnström , "Model validation of channel zapping quality", Proc. of Human Vision and Electronic Imaging Conf., Jan. 19-22, 2009.

[8] N. Degrande, K. Laevens, D. De Vleeschauwer and R. Sharpe, "Increasing the user perceived quality for IPTV services", IEEE Communications Magazine, Vol. 46, Issue 2, Feb. 2008, pp. 94-100.

[9] Bigband Networks, "Methods and apparatus for using delay time during switching events to display previously stored information elements", US Patents 7237251, March 2000.

[10] B.E. Godana, R.E. Kooij and O.K. Ahmed, "Impact of Advertisements during Channel Zapping on Quality of Experience", Proc. of The Fifth International Conference on Networking and Services, ICNS 2009, Valencia, Spain April 20-25, 2009.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

Improving Perceived Fairness and QoE for Adaptive Video Streams

Bjørn J. Villa
Department of Telematics
Norwegian Institute of Science and Technology
Trondheim, Norway
bjorn.villa@item.ntnu.no

Poul E. Heegaard
Department of Telematics
Norwegian Institute of Science and Technology
Trondheim, Norway
poul.heegaard@item.ntnu.no

Abstract - **This paper presents an enhancement to a category of Adaptive Video Streaming solutions aimed at improving both Quality of Service (QoS) and Quality of Experience (QoE). The specific solution used as baseline for the work is the Smooth Streaming framework from Microsoft. The presented enhancement relates to the rate adaption scheme used, and suggests applying a stochastic variable for the rate adjustment intervals rather than the fixed approach. The main novelty of the paper is the simultaneous study of both network oriented fairness in the QoS domain and perception based fairness from the QoE domain, when introducing the suggested mechanism. The method used for this study is by means of simulations and numerical optimization. Perception based fairness is suggested as an objective QoE metric which, requires no reference to original content. The results show that the suggested enhancement has great potential in improving QoE, while maintaining QoS.**

Keywords - Adaptive Video Streaming; Fairness; QoE.

I. INTRODUCTION

Solutions for Adaptive Video Streaming are part of the more general concept of ABR (Adaptive Bit Rate) streaming which, covers any content type. The implementation of ABR streaming for video varies between different vendors, and among the more successful one today is the Microsoft Smooth Streaming (SilverLight) framework [1]. In general, the different implementations use many undisclosed and proprietary functions, awaiting results from ongoing standardization.

The basic behavior of adaptive video streaming solutions is that the client continuously performs a measurement and estimation of available resources in order to decide which, quality level to request. The relevant resource from the network side is the available capacity along the path between the server and client. Based on this, at certain intervals the client decides to either go up or down in quality level or remain at the current level. The levels are predefined and communicated to the client by the server at session startup. The changes in quality levels are normally done in an incremental approach, rather than by larger jumps in rate level. The rationale behind this is the objective to provide a smooth watching experience for the user. However, it may also be related to the CPU monitoring done by the client, as this is a key resource required. It may be the case that even if the network can provide you with a much higher rate level, the CPU on the device being used would not be able to process it. During the initial phase of an adaptive streaming session the potential requests of change in rate level are more frequent than later on when operating in a more steady-state phase. To some extent this is a rather aggressive behavior from a single client which, may have undesirable inter-stream impacts. At the same time, in order to give the user a good first impression and make him want to continue using the service it is desirable to reach a high quality as soon as possible.

Among the strongest drivers for commercial use of ABR based services on the Internet are Over-The-Top content providers. These are providers which, rely on the best effort Internet service as transport towards their customers. Therefore, technologies aiming at making services survive almost any network state are of great interest. In addition to focus on the network based QoS dimensions of services and involved networks, there is also a growing interest in the QoE dimension [2]. The latter should be considered as not only a richer definition of quality, but also more focused towards who decides whether something is good or bad, i.e., the end user. The evolution of successful services on Internet indicates that the focus on QoE for Over-The-Top providers is a good strategy.

A. Problem Statement

The concept of Adaptive Video Streaming is without a doubt very promising. However, as more and more services are adopting this concept the success brings new challenges. The first challenge with effects visible to the end users is how well these services behave when they compete for a shared resource, such as the broadband access to a household. With a strong dominance of video based service on the Internet this issue is important to address. As each client operates independently of each other, it has no understanding of the traffic it competes with. Different clients consider each other as just background traffic. This leads to unpredictable and potentially oscillating behavior of each session, especially in a home environment this type of interference is likely to have a very negative impact on each user QoE.

B. Research Approach

The method investigated in this paper to address the problem at hand is to apply specific changes in the algorithm used by each ABR client controlling the adaptive behavior. The specific change suggested is related to the

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

rate adjustment interval used [1]. The effect of changing the duration of the rate adjustment interval from a fixed value T to some stochastic variable is presented and analyzed.

The ABR solution used as reference point for the work is the one from Microsoft (Smooth Streaming). However, the key principles would still apply to other solutions based on similar principles.

C. Paper Outline

The structure of this paper is as follows. Section II provides an overview of methodology and metrics; Section III describes the simulation model; Section IV presents simulation results; Section V gives an analysis of the results; Section VI provides the conclusions and an outline of future work.

II. RELATED WORK

It has been shown in [3] that competing adaptive streams can cause unpredictable performance for each stream, both in terms of oscillations and ability to achieve fairness in terms of bandwidth sharing. The experimental results presented give clear indication on that competing ABR clients cause degraded and unpredictable performance. Apart from this paper, the topic at hand does not seem to have been addressed by the academic research community to the extent it deserves.

In another paper [4], the authors have investigated how well adaptive streaming performs when being subject to variable available bandwidth in general. Their findings were that the adaptive streams are performing quite well in this type of scenario except for some transient behavior. These findings do not contradict the findings in [3] as the type of background traffic used do not have the adaptive behavior itself, but is rather controlled by the basic TCP mechanisms.

Rate-control algorithms for TCP streaming in general and selected bandwidth estimation algorithms are described in [5]. This work is relevant to any TCP based application delivering a video stream.

In some of our own previous work we have described and analyzed how competing adaptive streams can be controlled using a knowledge based bandwidth broker in the home gateway [6] [7].

III. METHODOLOGY AND METRICS

In this section, we introduce the relevant performance metrics and together with motivation for the chosen focus. Thereafter, some candidate methods on how to improve the performance metrics are given, and finally, the specific method subject for study is presented.

A. Flow Based Performance Metrics

For transport flows it is common [8] to focus on the following metrics in order to assess their performance: inter-flow fairness, stability and convergence time. This in addition to the general QoS metrics: bandwidth, packet loss, delay and jitter. The same metrics can be applied to adaptive video streams as they by definition also are flows with similar concerns. The analysis of these metrics can be done from a strict network oriented perspective (QoS), but to some extent also bridged over to a user perception domain (QoE). When focusing on the inter-flow fairness metric this is traditionally analyzed [9] using, e.g., the Jain's fairness index [10], the product measure [11] or Epsilon-fairness [12] for flows with equal resource requirements. For flows with different resource requirement, the Max-Min fairness [13], proportional fairness [14] or minimum potential delay fairness [15] approaches are commonly seen. Real life adaptive video streams would typically belong to the last category.

Max-Min fairness: The objective of max-min fairness is to maximize the smallest throughput rate among the flows. When this is met, the next-smallest throughput rate must be as large as possible, and so on. Max-min fairness can also be explained by considering it as a progressive filling algorithm, where all flows start at zero and grow at the same pace until the link is full. With this approach the max-min fairness gives priority to the smallest flows. The least demanding flows always have the best chance of getting access to all the resources it needs.

Proportional fairness: The original definition of proportional fairness comes from economic disciplines [14] for the purpose of charging. The original definition is used in the relevant RFC [9] but it does not come across as very constructive for the purpose of analyzing fairness in single resource (e.g., bandwidth) sharing among flows. In this context more recent definitions and interpretations are more suitable [16]. The principle of this would be that a resource allocation is considered proportional fair if it is made to the flow which, has the highest ratio between potential maximum resource consumption and its average resource consumption so far. A further simplification would be to use the current resource usage (if greater than 0) instead of the average in the ratio calculation. The same ratio numbers for each flow could then be used to give a view on the current system fairness by comparing them. If they are all equal the system could be stated as proportionally fair.

Minimum potential delay fairness: The idea behind minimum potential delay fairness is based on the assumption that the involved flows are generated by applications transferring files of certain sizes. A relevant bandwidth sharing objective would be to minimize the time needed to complete those transfers. However, this does not apply to an adaptive streaming scenario and is therefore not discussed any further.

B. Perception Based Performance Metrics

There is a wide range of metrics which, influence how satisfied an end user is with a service such as e.g., video streaming. Many of these are not related to network aspects, and therefore difficult to influence by means in this domain. However, one of the perceived performance metrics which, could be correlated with network aspect is the notion of

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

perceived fairness. It is then of great interest to try and find methods of influencing this in a positive manner.

Looking at fairness from an end user perception, research from the social science and psychology domain [17] states that this is closely related to what is called 'Social Justice'. In this context a queuing system or any other resource allocation mechanism would be considered as a 'Social System'. It has further been found that users react negatively to any system behavior which, gives better service to other user, unless justification is provided. Such system behavior is considered un-fair, i.e., in violation with the social justice of the system as the end users considers it as discrimination.

The end user notion of system discrimination has been suggested by [18] as an important measure of perceived service quality, and more specifically the perceived fairness is stated to be closely related to the discrimination frequency. It should be noted that analyzing this type of end user perceived discriminations has a challenge in terms of handling the false positive and false negative cases.

Applying the concept of discrimination to competing adaptive streams, it would be related to situations where end user expectations are not met during steady state periods and also negative changes in service delivery during more transient periods. In other words, whatever makes the end user think that he is being discriminated due to other users in the system, will lead to reduced perceived service quality.

In order to use this type of perceived end user discrimination as a measure for how well the algorithm which, controls the adaptive streams are performing, a clear definition regarding what end users are considering as discrimination is required. This could, e.g., be periods with session rate below some threshold, any change in session rate to a lower level or the session rate change frequency.

C. Methods for Improving Performance

There are several things that one could try to incorporate into the adaptive algorithms controlling the ABR service [1] in order to make them perform better in a multi-stream scenario.

The selected performance metrics to be studied are from the network side proportional fairness, and from the end user side the perceived fairness metric as earlier described. Whether it is possible to improve both these fairness metrics at the same time will be an important part of the results.

Randomization of time intervals: The fixed rate adjustment intervals (T) used by each adaptive stream while in steady-state may be a contributing factor to inaccurate estimations of available bandwidth and thereby oscillating behavior. An alternative to fixed intervals would be to randomize them by using a per-session stochastic parameter (within certain reasonable bounds). By doing so the available bandwidth estimation methods may become more accurate.

Back-off periods: Whenever a service is reducing its rate level due to observed congestion it may try to increase again after the same amount of time (T). In addition to the previous described randomization of this interval, one could also consider introducing a back-off period. This would imply that after a service has reduced its rate level, it enters a back-off period of a certain duration during which, no increase is allowed.

Threshold based behavior: Rather than using the same intervals of potential rate changes all the time, one could introduce a threshold for when it operates more or less aggressive. This threshold could be the mean available rate level for a specific session, or even a smoothed average value for the actual achieved level. This concept is applied with success in more recent TCP versions for the purpose of optimizing performance.

The method chosen for the simulations is according to the first approach described, i.e., a randomization of the intervals between each potential rate change as originally suggested in [3]. As baseline for the simulations, the fixed interval with $T=2s$ has been used. Then as stochastic alternatives, both a uniform distribution and a negexp distribution have been implemented. The uniform distribution gives values of T between [1.6, 2.4]s, while the negexp alternative gives values of T according to the distribution function with λ=0.5 and expected value ($1/\lambda$) = 2s.

IV. SIMULATION MODEL

As the adaptive streaming solutions of today are highly proprietary, the details concerning their implementation are not disclosed. Due to this, there will always be some degree of uncertainty concerning their internal functions.

A. Assumptions

One of the key functions of an ABR client is the method used for determining whether to go up or down in rate level during times of varying available bandwidth. From studying live traffic it does not seem as if the clients use additional network probing beyond the actual information obtained through download of video segments. Further on, in the likely absence of a per stream traffic shaper at the server side (for scalability and performance reasons), it will give a traffic pattern for each stream which, typically contains a sequence of busy and idle periods. The measured busy period rate is then higher than the actual stream rate level. Also, it is likely that there will be sub-periods within the busy periods where per packet rate is close to the total available bandwidth. As such, the client can probably obtain a rather accurate indication of maximum available bandwidth by just looking at minimum observed inter-arrival time of packets of known size belonging to the same stream.

However, not all streams will have interleaved busy periods so there is a good chance for each stream to overestimate the potential for additional bandwidth. There is a wide range of bandwidth estimation methods and a few of these are described in [19], but again - as the details of the

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

adaptive streaming solutions are not disclosed we will not discuss this part any further. Independent of which, method being used, there will be some degree of uncertainty which, contributes to variable performance. Further on, we assume the following to be true for the ABR sessions to be studied

- No stream coordination at server side
- No involvement from mechanisms in the network between the client and server
- All clients operate independently and do not communicate
- All clients are well behaved in the sense that they follow the same scheme
- At each defined stream rate level there are no variations due to i.e., picture dynamics

B. Session Type and Schedule

The ABR sessions used in the simulator are based on profiles observed in commercial services. The quality levels defined are {0, 250, 750, 1500, 2500, 3500, 5000} Kbps. All sessions are of the same type. The sessions are initiated by 10 different users and start time scheduling are done according to stochastic distributed parameters t_a – Uniform [0, 2000] ms and t_b – Uniform [0, 60] s. This gives that all sessions start during the first 60 seconds (t_b), but shifted by some milliseconds (t_a) in order to avoid synchronization of the rate adjustment intervals.

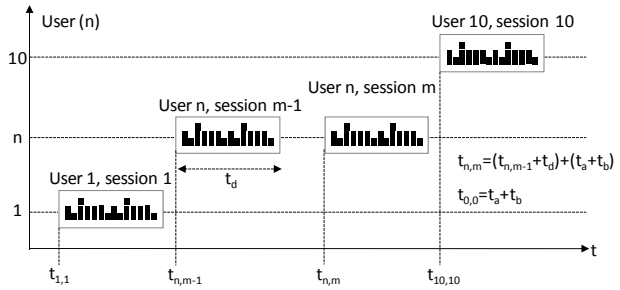


Figure 1. Session scheduling per user

During one simulation run, each user executes a total of 10 sessions sequentially. Time for starting the next session (m) for specific user (n) is noted $t_{n,m}$ (cf. Figure 1). The duration of each session t_d is deterministic and set to 40 minutes. A total of 10 simulation runs using different seeds are executed, corresponding to an aggregated session time of approximately 66 hours per user.

C. Rate Adaption Algorithm

The model for rate adaption per session is based on periodic estimation of available bandwidth $A_s(t)$ and calculation of a smoothed average $SA_s(t)$.

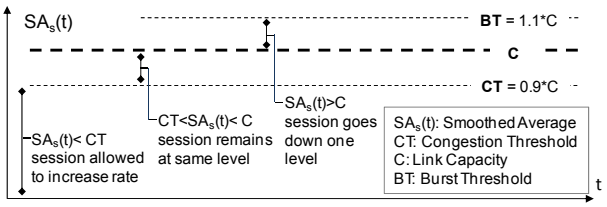


Figure 2. Thresholds for smoothed average

This smoothed average (cf. Figure 2) is compared to a congestion threshold (*CT*), the link capacity (*C*) and a burst threshold (*BT*) in order to trigger a rate adjustment.

Whenever the sum of requested rates from sessions is above the burst threshold (*BT*), the next session which, calculates $SA_s(t)$ will be forced down, independent of the value of $SA_s(t)$. This function is implemented in the simulator in order to incorporate the somewhat unpredictable behavior during times of heavy congestion.

The calculation of smoothed average $SA_s(t)$ is based on [3], and is expressed in (1). The parameter δ gives the weighting of the estimated available bandwidth for the two periods included in the calculation.

$$SA_s(t) = \delta\, A_s(t_{i-1}) + (1-\delta)A_s(t_i) \qquad (1)$$

The available bandwidth estimation function used in the simulations is based on the assumption that sessions running at high rates are able to make more accurate estimations than those running at lower rates. An abstraction of the function itself is made by a number of n bandwidth samples $C_{i,j}$ (cf. Figure 3)

A specific session is then given access to a number of these samples according to its current rate level, and then it will use this as basis for its estimation. A high rate gives a high number of samples available, and then, also, a higher degree of accuracy.

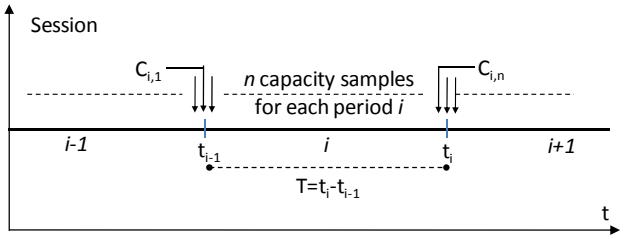


Figure 3. Capacity samples per period

The number of samples $x_{s,i}$ available to a specific session *s* for period *i* is given by its ratio between current rate $R_s(t_i)$ and max rate $R_s max$, multiplied by *n* as per (2).

$$x_{s,i} = n^{R_s(t_i)}/_{R_s max} \qquad (2)$$

In the simulations, the value of n was set to 20 and $R_s max$ was according to the session definition 5000Kbps. The available bandwidth estimated $A_s(t)$ for period i is then given by the following (3).

$$A_s(t_i) = \sum_{l=1}^{x_i} C_{i,l} / x_{s,i} \quad (3)$$

By combination with the expression for $SA_s(t)$ it gives the following expression (4).

$$SA_s(t) = \delta \sum_{l=1}^{x_{i-1}} C_{i,l} / x_{s,i-1} + (1-\delta) \sum_{l=1}^{x_i} C_{i,l} / x_{s,i} \quad (4)$$

The value of δ was set to 0.8 as per [3], thus giving most weight to the available bandwidth estimation from the previous period.

D. Simulation Tool

The simulator was built using the process oriented Simula [20] programming language and the Discrete Event Modeling On Simula (DEMOS) context class [21].

This programming language is considered as one of the first object oriented programming languages, and remains a strong tool for performing simulations.

V. RESULTS

The simulation results are presented for different congestion levels on the access link. The chosen capacities are 10, 20, 30 and 40Mbps. The lowest capacity would represent a highly congested scenario. The simulations were also run for all levels from 10 to 40 with increments of 200Kbps but for the sake of clarity these details are left out as they did not change the conclusions.

The studied fairness parameters (proportional and perceived), are compared for the 10 independent users sharing the access link. In order to present more information regarding variations in quality levels, a presentation of Coefficient of Variation (CV) is given. Values for CV below 1 is considered low-variance, while above 1 is considered high-variance.

The simulation results to be presented are based on that all users are accessing the same service, with identical session properties (i.e., quality levels). However, the simulations were also run for other service types and a mixture of services. These results are also left out, as they did not change the conclusion.

A. Proportional Fairness

Proportional fairness is measured as achieved session average rate per user, divided by session max – as per the definition earlier (cf. Figure 4, Figure 5, Figure 6). A high value is good and the maximum value is 1.

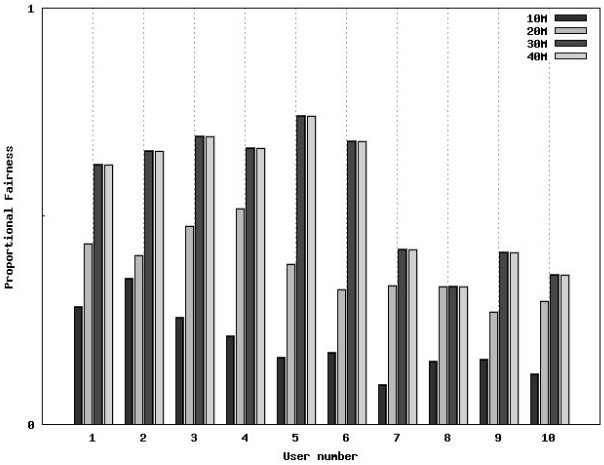


Figure 4. Proportional Fairness, fixed T=2s

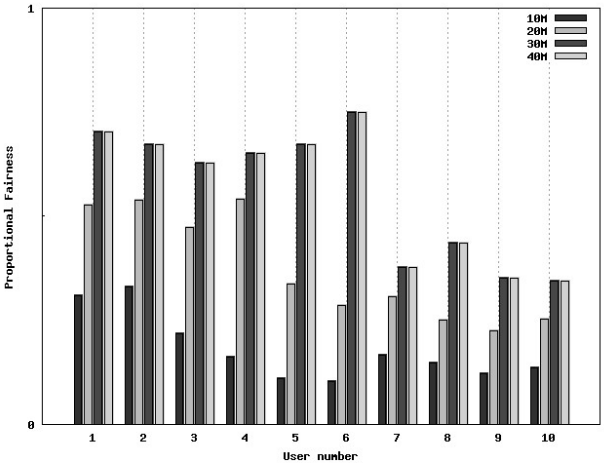


Figure 5. Proportional Fairness, Uniform T [1.6, 2.4]

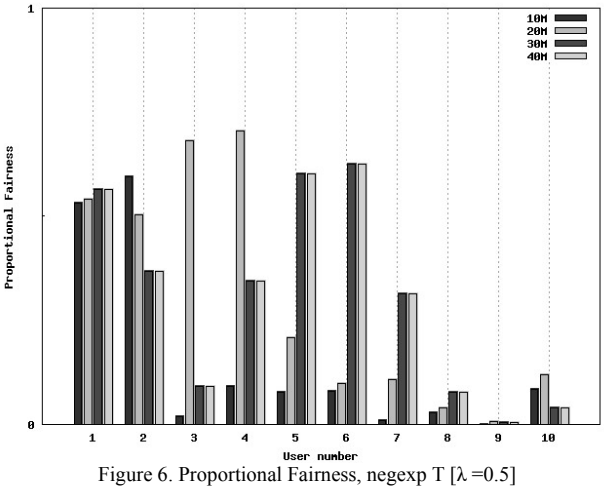


Figure 6. Proportional Fairness, negexp T [λ =0.5]

B. Perceived Fairness

The perceived fairness metric is calculated as the number of quality (rate) level reductions per minute (cf. Figure 7, Figure 8, Figure 9). Here, a low metric value is good – as it would reflect less rate reductions per minute.

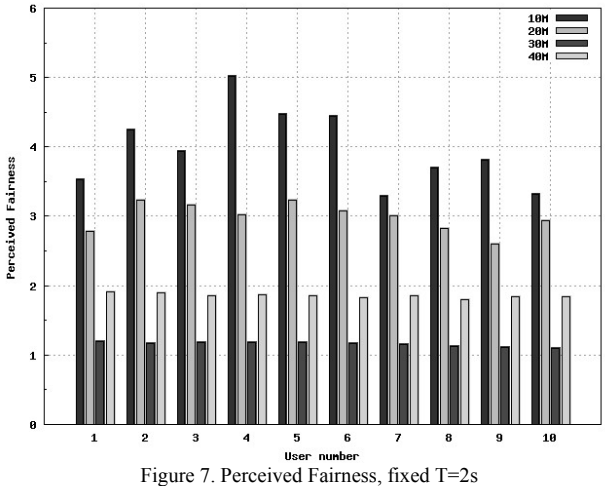


Figure 7. Perceived Fairness, fixed T=2s

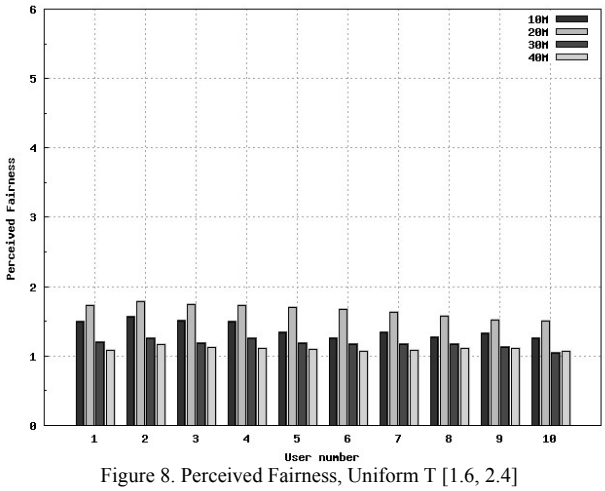


Figure 8. Perceived Fairness, Uniform T [1.6, 2.4]

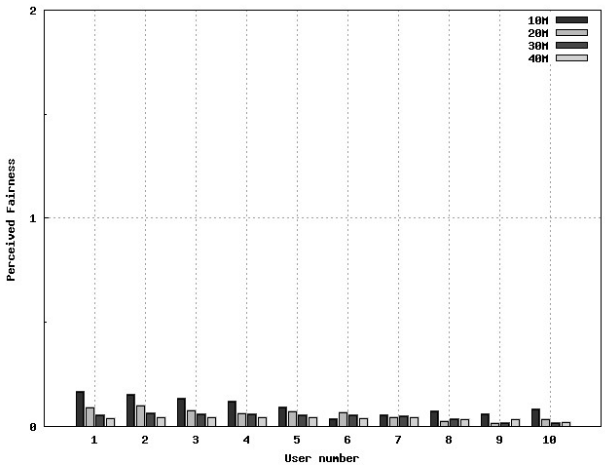


Figure 9. Perceived Fairness, negexp T [λ =0.5]

C. Coefficient of Variation (CV)

The Coefficient of Variation is calculated as Standard Deviation/Mean Value for sessions belonging to a user (cf. Figure 10, Figure 11, Figure 12). Values below 1 indicate low-variance which, is preferred.

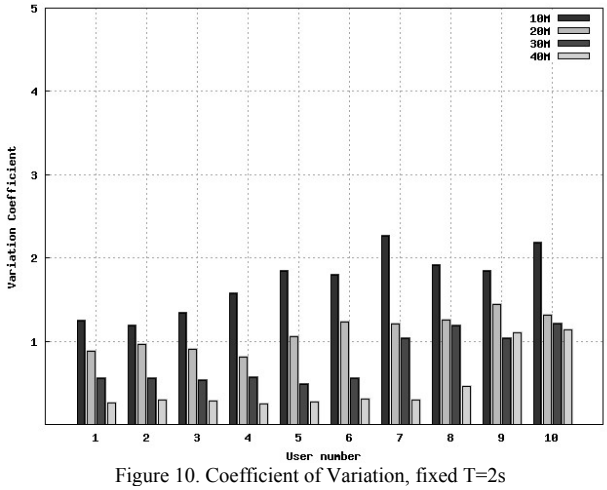


Figure 10. Coefficient of Variation, fixed T=2s

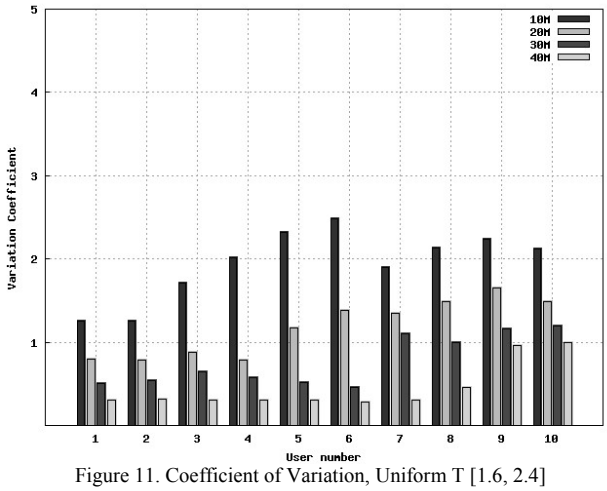


Figure 11. Coefficient of Variation, Uniform T [1.6, 2.4]

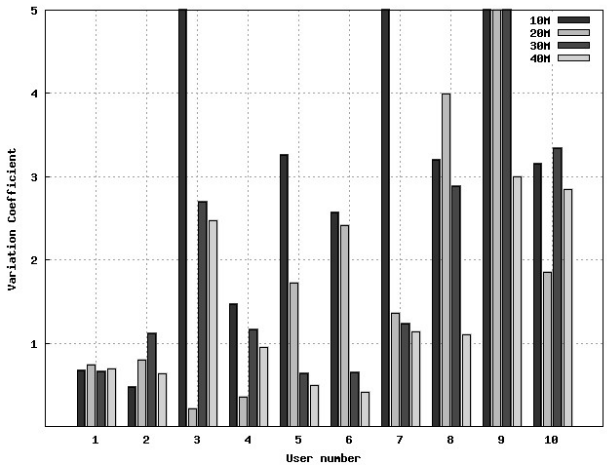


Figure 12. Coefficient of Variation, negexp T [λ =0.5]

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

VI. ANALYSIS

As expected, the randomization of time interval duration does have an effect on the parameters studied. However, the effect is not always positive.

Concerning proportional fairness, the introduction of a uniform *T* variable does not have a significant effect. The result can be viewed as neutral. On the other side, when the negexp *T* variable is used a clear negative effect is observed as the difference between users becomes significant.

For the perceived fairness metric, both the use of a uniform *T* and a negexp *T* have a significant positive effect. The best results are achieved for the negexp case which, gives values well below 1 for all congestion levels and users. It may be considered promising that the effect is especially strong during high times of high congestion (link capacity of 10M and 20M).

Regarding Coefficient of Variation, the results are similar to Proportional Fairness. A uniform *T* give no change, while a negexp *T* gives a negative change.

TABLE I. SUMMARY OF SIMULATION RESULTS

	Proportional Fairness	***Perceived Fairness***	***Coefficient Variation***
uniform T	neutral	positive	neutral
negexp T	negative	positive	negative

The somewhat intuitive explanation to why changes could be expected is that some of the negative effects of a fixed adjustment interval as illustrated in Figure 13 are reduced. In the case of fixed periods, each session would get the same periodic view on the link utilization, always missing or including some other traffic. This gives a certain degree of error in the available bandwidth estimation functions.

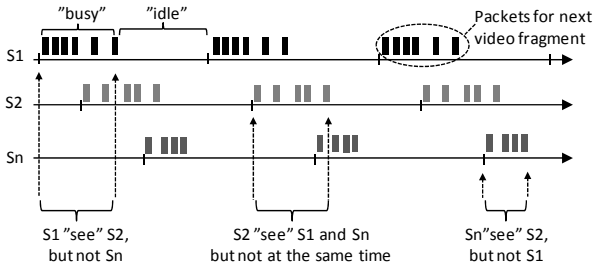


Figure 13. Problem with fixed estimation periods

Each session estimates available bandwidth only during its busy periods (ref Section III Subsection C). This means that in order to get an accurate estimation it is beneficial for it to have overlapping busy periods with as many other sessions as possible.

A. Burst Period Duration

The duration of the busy period for a specific session depends on both its current rate level and the rate adjustment interval. The dependency of the rate level follows from the obvious relation to data volume to be transferred per time unit for a specific rate level, while the dependency of rate adjustment interval follows from the requirement to maintain the same average amount of data received over time.

At the beginning of each interval the client requests the next video fragment for a specific rate level, with duration equal to its rate adjustment interval. This is illustrated in Figure 14 where two sessions running at the same rate level, but with different rate adjustment intervals have different busy period durations.

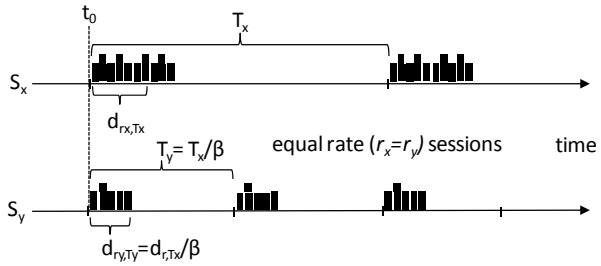


Figure 14. Equal rate sessions with different busy periods

Any two sessions (S_x, S_y) running at the same rate level, will have a relation between their burst period durations expressed by the parameter β. This parameter is given by the following expression (5).

$$\beta = {d_{rx,Tx}}/{d_{ry,Ty}} = {T_x}/{T_y}, \qquad for\ r_x = r_y \tag{5}$$

Using this relationship, we can express (6) the burst period duration $d_{ri,Ti}$ for any session S_i as a function of its rate adjustment interval T_i and a reference burst period duration $d_{ri,T}$.

$$d_{ri,Ti} = d_{ri,T}({T_i}/{T}) \tag{6}$$

The values for $d_{ri,T}$ can presumably be calculated based on information about the codec used for the specific media stream inside each sessions, together with assumption on per session server side capacity. Alternatively one could make measurements on a specific system and establish a $d_{ri,T}$ matrix for all valid values of r_i and the reference T value.

However, if we assume that the server side capacity is not a limitation, and that it will always try to burst with a certain bitrate C_{burst} we can also express the burst period duration $d_{ri,Ti}$ as follows (7).

$$d_{ri,Ti} = \left({r_i\,T}/{C_{burst}}\right)\left({T_i}/{T}\right) = \left({r_i T_i}/{C_{burst}}\right) \tag{7}$$

The maximum value for C_{burst} is natural to think of as the access capacity for the user group / home network, as this is normally the end-to-end bottleneck. However, it is likely

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

that the actual C_{burst} is related to the maximum rate for the specific service.

B. Probability for Burst Period Overlap

For T_i values according to a uniform distribution, the probability $P_{i,r,t}$ for a session i at rate level r to be in its busy period at time t will be according to the following expression (8).

$$P_{i,r,t} = {d_{r,Ti}}/{T_i} = \frac{d_{r,T}\left({T_i}/{T}\right)}{T_i} = {d_{r,T}}/{T} \quad (8)$$

From this, we see that all sessions at a specific rate level has the same probability of being in its busy period at time t. We can then express the probability that all n sessions are in their busy period at time t as follows (9).

$$P_{all\ busy,t} = \left({d_{r_1,T}}/{T}\right)^{c_1} \left({d_{r_m,T}}/{T}\right)^{c_m} \quad (9)$$

The parameter c_m represents the number of sessions at rate level r_m and the sum of all c_m values equals n. From this we see that the probability of any session to see all other sessions during its busy period depends on the session rate level mix, and this probability increases when more sessions are running at high rate levels.

Further on, we recognize that the probability for that a session i has an overlap with each of the other sessions sometimes during its busy period T_i is the integral of $P_{all\ busy,t}$ over the period [0, T_i] which, is easily expressed as the constant $P_{all\ busy,t}$ multiplied by T_i.

We then let a specific session mix be described by the vector $R_{mix}=\{r_1,\ldots,r_n\}$, whereas r_i represents the rate level for session i. Also, for a specific session i let A_i be the group of sessions which, has overlapping busy periods with session i at a specific time t_0, and B_i be the group of sessions for which, it did not have an overlap. In the situation where all sessions have the same rate adjustment interval duration T_i, the probability of that session i has an overlapping busy period with any of the sessions in group B_i at time t_0+T_i is zero. This leads to that while R_{mix} remains unchanged, the view a specific session has of the total traffic will not change. The system state for session i in terms of busy period overlap with other sessions is independent of the state at t_0 and also t in general.

In the case where T_i is not equal for all sessions, but instead are chosen according to some stochastic distribution – the group of sessions which, overlap the busy period of session i at t_0+T_i is not independent of the state at t_0. If we let C_i denote the sub-group of sessions from B_i which, has overlapping busy periods with session i at time t_0+T_i, it can be shown that there is a deterministic relationship between A_i, B_i and C_i.

If we then remember the assumed use of a smoothed average function we see the benefit of this potential additional burst period overlaps in subsequent periods.

C. Dynamics in Burst Period Overlap

When the starting times for each session and their respective rate adjustment intervals (T_i) are considered stochastic processes, the sessions will combine in time in different ways. In order to define the deterministic relationship between overlapping busy periods during subsequent intervals, we need to analyze scenarios where sessions with different rate levels and different rate adjustment interval are combined.

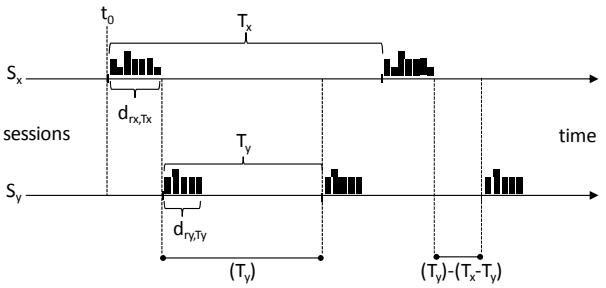


Figure 15. Session S_y staring after S_x (T_y<T_x)

The first scenario (a) to be studied is the one where two sessions (S_x, S_y) with different T_i values (T_x, T_y) are active at the same time. We assume $T_x > T_y$ and that S_y starts immediately after the busy period of S_x finishes as illustrated in Figure 15.

For the two sessions (S_x, S_y) there will be shift in phase between them as a function of time which, makes them have a full or partial busy period overlap at some time. The question is then how many rounds it will take for S_x to see S_y and vice versa. It can be shown that we can express the number of rounds for S_x before it has an overlapping busy period with S_y as follows (10).

$$\begin{aligned} \mathrm{N_{a,x\to y}} &= 1 + \left\lceil \frac{T_y}{T_x - T_y} \right\rceil \\ \text{when } \tfrac{T_x}{2} < T_y < (T_x - d_{rx,Tx} - d_{ry,Ty}) & \\ \mathrm{N_{a,x\to y}} &= 2 \\ \text{when } (T_x - d_{rx,Tx} - d_{ry,Ty}) < T_y < T_x & \end{aligned} \quad (10)$$

In the same way, we can express the number of rounds for S_y before the same overlap of busy period with S_x takes place (11).

$$\begin{aligned} \mathrm{N_{a,y\to x}} &= 1 + \left\lceil \frac{T_x}{T_x - T_y} \right\rceil \\ \text{when } \tfrac{T_x}{2} < T_y < \left(T_x - d_{rx,Tx} - d_{ry,Ty}\right) & \\ \mathrm{N_{a,y\to x}} &= 2 \\ \text{when } \left(T_x - d_{rx,Tx} - d_{ry,Ty}\right) < T_y < T_x & \end{aligned} \quad (11)$$

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

The next scenario (*b*) to be studied is where the sessions (S_x, S_y) are running with different T_i values (T_x, T_y) but now S_y finishes its busy period before S_x (cf. Figure 16).

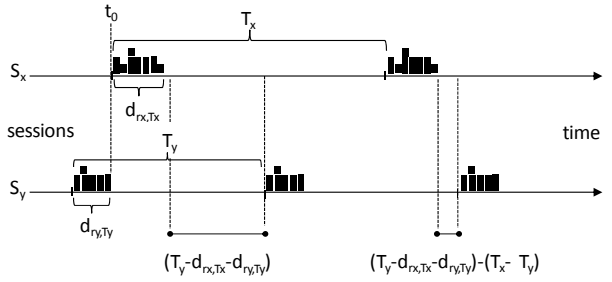


Figure 16. Session S_x staring after S_y ($T_y<T_x$)

The number of rounds it takes for S_x to see S_y is expressed as follows (12).

$$\mathrm{N_{b,x\to y}} = 1 + \left\lceil \frac{T_y - dx - dy}{T_x - T_y} \right\rceil$$
$$\text{when } \tfrac{T_x}{2} < T_y < \left(T_x - d_{r,Ty}\right) \qquad (12)$$
$$\mathrm{N_{b,x\to y}} = 2$$
$$\text{when } (T_x - d_{r,Ty}) < T_y < T_x$$

The number of rounds it takes for S_y to see S_x is expressed as follows (13).

$$\mathrm{N_{b,y\to x}} = 1 + \left\lceil \frac{T_x - dx - dy}{T_x - T_y} \right\rceil$$
$$\text{when } \tfrac{T_x}{2} < T_y < (T_x - d_{r,Ty}) \qquad (13)$$
$$\mathrm{N_{b,y\to x}} = 2$$
$$\text{when } (T_x - d_{r,Ty}) < T_y < T_x$$

It should be noted that for both scenarios there is a special case where $N_{a,y\text{-}x}/N_{b,y\text{-}x}$ and $N_{a,x\text{-}y}/N_{b,x\text{-}y}$ are always 2, i.e., two sessions which, did not have overlapping busy periods at t_0 is guaranteed to have overlapped during the next period for S_x and S_y. For a smoothed average function operating over two periods this is desirable, i.e., whatever it does not see in the first period it is guaranteed to see in the next.

D. Optimization Problem

The expressions for $N_{y\text{-}x}$ and $N_{x\text{-}y}$ contain many variables. These variables are the rate adjustment intervals T_i and the burst period durations $d_{ri,Ti}$ for all sessions. The latter are calculated based on the session rates r_x and r_y and C_{burst} as defined in Section V. These expressions can be used as input to a constrained optimization problem and analyzed as such in order to find maximum and minimum values.

As the starting point for this optimization problem we can focus on the worst case scenario, that would be the number of rounds for S_y before it has an overlap with S_x ($N_{a,y\text{-}x}/N_{b,y\text{-}x}$), which, will always be higher than the number of rounds for S_x before this has an overlap with S_y.

We also see that $N_{a,y\text{-}x}$ will always be greater than $N_{b,y\text{-}x}$ since $T_x>T_y$. This gives us only one expression to analyze for the worst case scenario as follows (14).

Maximize: $N_{a,y\to x}$

where

$$N_{a,y\to x} = \begin{cases} 1 + \left\lceil \dfrac{T_x}{T_x - T_y} \right\rceil, if\ T_y < \left(T_x - d_{rx,Tx} - d_{ry,Ty}\right) \\ 2, if\ \left(T_x - d_{rx,Tx} - d_{ry,Ty}\right) < T_y < T_x \end{cases} \qquad (14)$$

subject to:

$$1.6 < T_y, T_x < 2.4 \text{ and } T_x/2 < T_y$$
$$R_x, R_y \in \{250, 750, 1500, 2500, 3500, 5000\}$$
$$d_{r_x,T_x} = {r_x T_x}/{C_{burst}}$$
$$d_{r_y,T_y} = {r_y T_y}/{C_{burst}}$$

The above maximization can then be done for different values of C_{burst}. In the simulations the access speeds used were between 10 and 40Mbps and the maximum session rate was 5Mbps. Based on measurements of real traffic we can see that the C_{burst} is lower than the actual access speed and thereforee values of respectively 5Mbps, 7.5Mbps and 10Mbps were used for C_{burst}.

For the two different alternatives of choosing values for T_i used in the simulations, the uniform approach is easiest to work with in the optimization context since it gives a min and max value for T_i. For the negexp alternative the corresponding range would be $[0,\infty]$ and for this scenario the optimzation problem does not have a useful solution.

The result from solving the optimization problem is shown in Figure 17. The three different burst bitrates (C_{burst}) give surfaces which, are plotted, whereas the highest capacity gives the highest values for $N_{a,y\text{-}x}$.

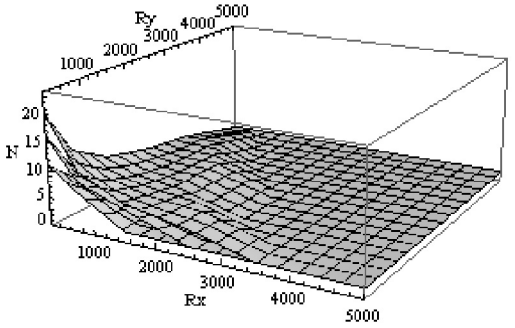


Figure 17. Maximum $N_{a,y\text{-}x}$ for different burst bitrates

We see that in many cases we get an overlap already in the second round, and thereby we improve the basis for the

available bandwidth estimation algorithm. This analysis then strengthens the findings of the simulations.

The more likely explanation to the negexp behavior in terms of good perceived fairness is the somewhat extreme proportional unfairness. By allowing some sessions to be very greedy, one prevents others from increasing at all. This is a stable but proportionally very unfair situation.

In order to improve the available bandwidth estimations further one may consider the well known PASTA principle [22] from queuing theory which, states that a Poisson based Arrival process See Time Averages. This implies that the bandwidth probing should take place not only during the burst periods, but as a process taking samples throughout the whole rate adjustment period.

VII. CONCLUSIONS AND FUTURE WORK

The results show that there is a significant potential of improving perceived fairness as defined and associated QoE for adaptive streams of the category studied. The positive effect of the suggested enhancement to the rate adaption scheme, i.e., using a stochastically determined duration of rate adjustment intervals rather than fixed values is supported by the simulation results and theoretical analysis.

The results also illustrate that when studying the performance of adaptive streaming solutions, it is not enough to only focus on the network centric QoS domain. A change in this domain does not necessary lead to a corresponding change in the QoE domain, and vice versa. The significant improvement in Perceived Fairness, while proportional fairness remained the same for the uniform T case supports this statement.

As future work in this field it is planned to further study objective and no-reference based QoE metrics such as Perceived Fairness which, is possible to correlate over to the QoS and network domain. It is also planned to verify the simulation and analytical results by means of measurements.

VIII. ACKNOWLEDGEMENTS

The reported work is done as part of the Road to media-aware user-Dependant self-aDaptive NETWORKS - R2D2 project. This project is funded by The Research Council of Norway. The work has also been actively supported by TV2, the leading commercial TV broadcaster in Norway. TV2 is among the pioneers in providing a full commercial TV offering over the Internet based on ABR technology.

REFERENCES

[1] A. Zambelli, "IIS smooth streaming technical overview," http://www.microsoft.com/silverlight/whitepapers/, Tech. Rep., March 2009 (last accessed 10.01.2012).

[2] E. Areizaga, L. Perez, C. Verikoukis, N. Zorba, E. Jacob, and P. Odling, "A road to media-aware user-dependent self-adaptive networks," in *Proc. IEEE Int. Symp. BMSB '09*, 2009, pp. 1–6.

[3] S. Akhshabi, A. C. Begen, and C. Dovrolis, "An experimental evaluation of rate-adaptation algorithms in adaptive streaming over http," in *ACM Multimedia Systems (MMSys)*, 2011.

[4] L. De Cicco and S. Mascolo, "An experimental investigation of the akamai adaptive video streaming," in *Proceedings of the 6th international conference on HCI in work and learning, life and leisure: workgroup human-computer interaction and usability engineering*, ser. USAB'10. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 447–464.

[5] R. Kuschnig, I. Kofler, and H. Hellwagner, "An evaluation of tcp-based rate-control algorithms for adaptive internet streaming of h.264/svc," in *MMSys '10*. New York, NY, USA: ACM, 2010, pp. 157–168.

[6] B. J. Villa and P. E. Heegaard, "Monitoring and control of QoE in media streams using the click software router," in *NIK2010, Norway. ISBN 978-82-519-2702-4.*, vol. 1, November 2010, pp. 24–33.

[7] B. J. Villa and P. E. Heegaard,, "Towards knowledge-driven QoE optimization in home gateways," *ICNS 2011*, May 2011.

[8] S. Bhatti, M. Bateman, and D. Miras, "Revisiting inter-flow fairness," in *Broadband Communications, Networks and Systems, 2008. BROADNETS 2008. 5th International Conference on*, sept. 2008, pp. 585 –592.

[9] S. Floyd, "Metrics for the Evaluation of Congestion Control Mechanisms," RFC 5166 (Informational), Internet Engineering Task Force, Mar. 2008.

[10] R. Jain, D. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared systems," DEC, Tech. Rep., 1984.

[11] D. Mitra and J. B. Seery, "Dynamic adaptive windows for high speed data networks with multiple paths and propagation delays," *Computer Networks and ISDN Systems*, vol. 25, no. 6, pp. 663 – 679, 1993, high Speed Networks.

[12] Y. Zhang, S.-R. Kang, and D. Loguinov, "Delayed stability and performance of distributed congestion control," in *Proceedings of the 2004 conference on Applications, technologies, architectures, and protocols for computer communications*, ser. SIGCOMM '04. New York, NY, USA: ACM, 2004, pp. 307–318.

[13] Z. Cao and E. Zegura, "Utility max-min: an application-oriented bandwidth allocation scheme," in *INFOCOM '99*, vol. 2, mar 1999, pp. 793 –801 vol.2.

[14] F. Kelly, "Charging and rate control for elastic traffic," *European Transactions on Telecommunications*, 1997.

[15] S. Kunniyur and R. Srikant, "End-to-end congestion control schemes: utility functions, random losses and ecn marks," *IEEE/ACM Trans. Netw.*, vol. 11, pp. 689–702, October 2003.

[16] A. Jdidi and T. Chahed, "Flow-level performance of proportional fairness with hierarchical modulation in ofdma-based networks," *Comput. Netw.*, vol. 55, pp. 1784–1793, June 2011.

[17] H. Levy, B. Avi-Itzhak, and D. Raz, "Network performance engineering," D. D. Kouvatsos, Ed. Berlin, Heidelberg: Springer-Verlag, 2011, ch. Principles of fairness quantification in queueing systems, pp. 284–300.

[18] W. Sandmann, "Quantitative fairness for assessing perceived service quality in queues," *Journal Operational Research*, pp. 1–34, April 2011.

[19] R. Prasad, C. Dovrolis, M. Murray, and K. Claffy, "Bandwidth estimation: metrics, measurement techniques, and tools," *Network, IEEE*, vol. 17, no. 6, pp. 27 – 35, nov.-dec. 2003.

[20] R. J. Pooley, *An Introduction to Programming in Simula*. Blackwell Scientific Publications. ISBN: 0632014229, 1987.

[21] G. Birtwistle, *Demos - A system for Discrete Event Modelling on Simula*, G. Birtwistle, Ed. School of Computer Science, University of Sheffeld,, July 1997.

[22] R. W. Wolff, "Poisson Arrivals See Time Averages," *Operations Research*, vol. 30, no. 2, pp. 223–231, 1982.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

A Relay-assisted Handover Pre-authentication Protocol in the LTE Advanced Network

Ling Tie
The Department of Information Science and Technique
The Chengdu University
Chengdu, Sichuan, China
tlcd4579@gmail.com

Di He
The Department of Electronic Engineering
The Shanghai Jiaotong University
Shanghai, China
dihe@sjtu.edu.cn

***Abstract*—With the help of relaying, The Long Term Evolution Advanced network is improved on the coverage and capacity. But relaying concept brings many problems on handover management. In this paper, a relay node is introduced in the system structure. We extend the handover procedure to support relaying and transmit security context. We design a relay assisted handover pre-authentication protocol, which happened before the mobile node handovers to the target cell. This article focuses on formal analysis of our proposed security protocol. Finally, an improved strand space and ideal formal analysis method, which includes message authentication code, is introduced. We use it to prove that our proposed protocol can meet the security and authentication proprieties.**

Keywords- Long Term Evolution; relay; handover; security; strand space; ideal

I. INTRODUCTION

The Long Term Evolution Advanced network (LTE-A) is considered as one of the main standards for the 4th generation broadband wireless network. Recently, due to the increasing demand for high transport rate and capacity, the relaying concept is introduced. The implementation of relay can overcome the restriction of coverage, especially at the cellular boundary. With the help of relay node (RN) located at the overlap between two adjacent cells, the user equipment (UE) being served by a source cell can pre-handover to the target cell. The handover interrupt rate and delay will be reduced significantly.

There are many papers have talked about handover schemes happened in the relay LTE-A network. In [1], five relay handover scenarios are categorized in multi-hop cellular network (MCN). Several handover frameworks for relay enhanced LTE Network are introduced in [2]. Some relay handover procedures supporting centralized and decentralized relaying are illustrates. But，these articles do not discuss security and authentication issues.

In the 3GPP draft 36.300 [3], a Relay Node (RN) is connected to an evolution Node B (eNB) wirelessly. The eNB serving the RN is called Donor eNB (DeNB). The Draft 36.300 gives an end-to-end authentication and key agreement (AKA) procedure when the RN first attaches to the LTE. The 3GPP draft 33.816 [4] proposes many solutions that support LTE relay node security. But，these standards do not discuss pre-handover authentication issue.

In this paper, a relay-assisted handover pre-authentication protocol is provided. The UE, RN and Target DeNB authenticate mutually before handover occurs. This protocol will reduce the handover delay significantly. But, this paper focuses on only formal analysis of our proposed security protocol. The performance of our proposed protocol will be discussed in future works.

The rest of the paper is organized as follows. The system architecture is given in Section II. In Section III, the proposed relay assisted handover authentication protocol is presented. The extended strand space is introduced in Section IV. The security is proved using this strand space model. Section V gives a brief introduction of the performance improvement. Section VI concludes this paper.

II. SYSTEM STRUCTURE

The system architecture of the proposed scheme is illustrated in Figure1. During the UE handovers from cell 1 to cell 2, the signal strength between the UE and the sourcing DeNB decreases. When the signal strength falls below certain threshold, handover will happen. However, the UE still remain in the source cellular. The UE will try to find one RN located at the coverage area between the source DeNB and the target DeNB. The RN helps the UE to pre-authentication to the target DeNB.

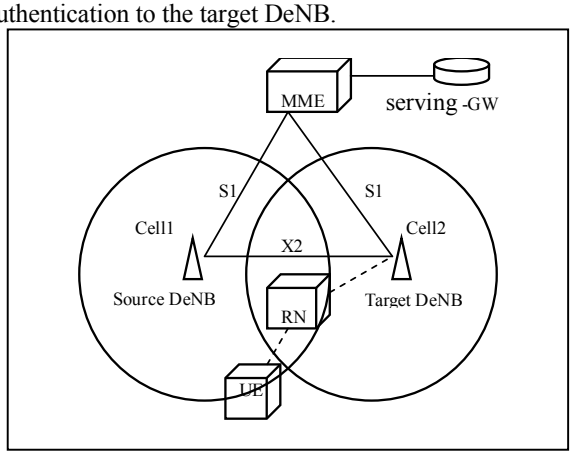


Figure 1. Relay handover structure

The LTE-A consists of the following major components:

- User Equipment (UE): It is mobile host.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

- Relay Network (RN): It provides multi-hop Wireless connectivity from the UE to the Target DeNB.
- Donor eNB (DeNB): It is an eNB serving the RN.

III. RELAY ASSISTED HANDVER PRE-AUTHENTCAITON PROTOCOL

A. Notations

Before we describe the relay assisted pre-authentication protocol, we will specify some of notations used in the protocol, as shown in Table I.

TABLE I. NOTATIONS

symbol	Explanation
k_X	The Key known by the X
$MAC(k_X, h)$	Message authentication code produced by k_X
$\{h\}_{k_X}$	The message is encrypted by the key k_X
N_X	The nonce value produced by X
ID_X	The identification of X
$\|\|$	Concatenation
$KDF(k_X, h)$	Security key produce function by k_X

B. Handover Key Hierarchy

According to the 3GPP draft 33.816 [4], the calculation of k^*_{eNB} and $k_{\mathrm{Re}lay}$ is based on the key hierarchy in Figure 2. The k_{eNB} is achieved from traditional AKA authentication method defined in the draft 33.401 [5].The handover key k^*_{eNB} is produced according to the method published in the draft 33.401 .The relay key is calculated as in (1).

$$k_{\mathrm{Re}lay} = \mathrm{KDF}\,(k_{eNB} \,||\, \mathrm{SSID}_{\mathrm{relay}}) \quad (1)$$

C. Relay Assisted Handover Procedure

Figure 3 shows the relay assisted pre-handover procedure. This procedure is based on [2]. We extended this procedure to include the calculation and transmission of security key. All the handover messages are protected by the key in Figure 2.

The Steps are as follows:

Step 1: The RN sends the measurement control message to the UE.

Step 2: The UE sends the measurement report message to the RN and forward it to the Source DeNB.

Step 3: Based on the measurement report, the Source DeNB makes RN handover decision. If RN handover is allowed, the source DeNB initiates the handover process. The source DeNB will calculate k^*_{eNB} as the handover key for the target DeNB[5]. It can also produce the key $k_{\mathrm{Re}lay}$ for the RN.

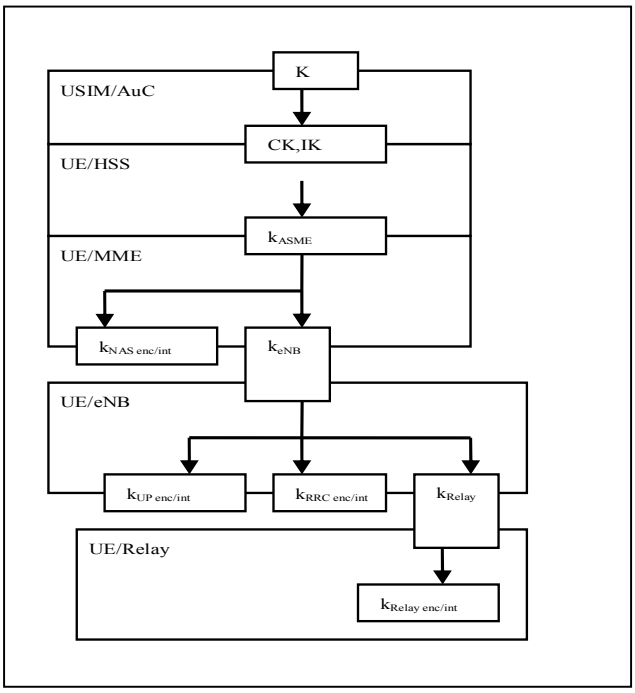


Figure 2. Handover key hierarchy

Step 4: The source DeNB sends the handover request message to the Target DeNB. This message includes k^*_{eNB} and $k_{\mathrm{Re}lay}$, which are protected by the RRC key.

Step 5: The target DeNB accepts the handover request message.

Step 6: After completing admission control, the target DeNB sends the handover request acknowledge message to the source DeNB.

Step 7: The source DeNB sends the handover command message to the RN securely. This message includes $k_{\mathrm{Re}lay}$ for the RN.

Step 8: The RN receives the handover command message and $k_{\mathrm{Re}lay}$ for the RN. It sends the handover command message to the UE, which includes Relay SSID.

Step 9: The UE uses information received from the handover command message to create k^*_{eNB}. It calculates $k_{\mathrm{Re}lay}$ simultaneously. Now，the UE has keys k^*_{eNB} and $k_{\mathrm{Re}lay}$.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

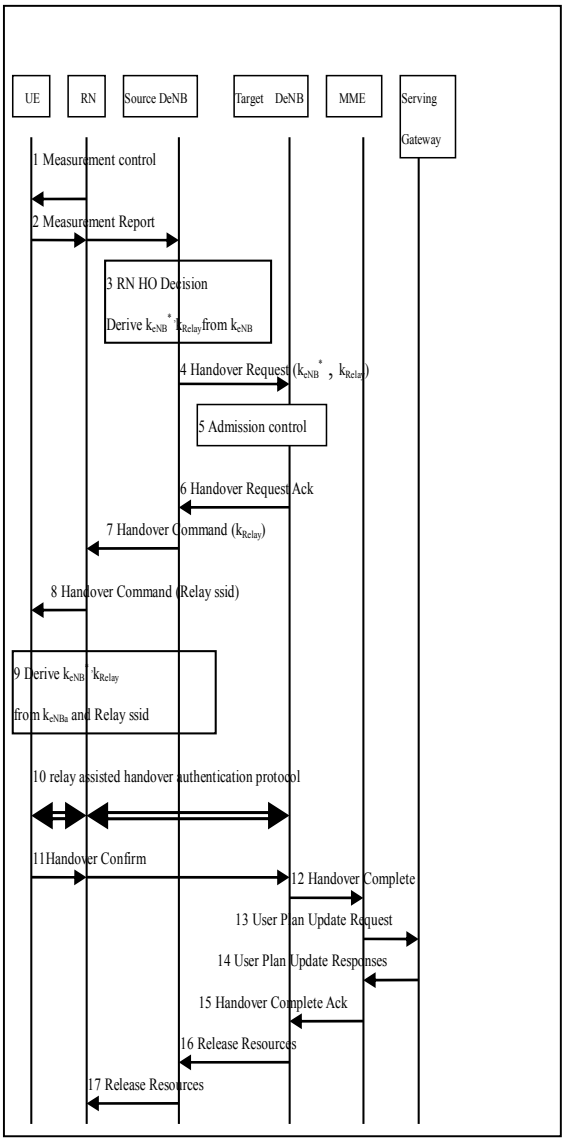


Figure 3. Relay-assisted handover Pre-authentication Sequence

Step 10: The UE achieves relay-assisted handover pre-authentication with the Target DeNB through the RN.

Step 11: The UE sends the handover confirm message to the RN and forward it to the Target DeNB.

Step 12: The Target DeNB sends the handover complete message to the MME.

Step 13: The MME sends the user plan update Request message to the Serving GW.

Step 14: The Serving GW sends the user plan update Response message to The MME.

Step 15: The MME sends the handover complete Ack message to the Target DeNB.

Step 16: The Target DeNB sends the release resource message to the Source DeNB

Step 17: The Source DeNB sends the release resource message to the RN.

D. Relay Assisted Handover Pre-authetication Protocol

In this section, a detailed description of our proposed pre-authentication protocol is given. It is the step 10 in figure 3. The extension of the step 10 sequence in Figure 3 is showed in Figure 4.

Step10. 1: When the UE still remain in the cell 1, it begins the pre-authentication procedure, with the help of RN. The UE sends the relay handover authentication request message to the RN. The UE generates two nonces. $N2_{UE}$ is used for RN, the other $N1_{UE}$ is used for Target DeNB. The nonce is encrypted by the key known only by the RN and the Target DeNB. The UE calculates the message authentication code (MAC) of the RN and the Target DeNB.

Step 10.2: Upon receiving the relay authentication request message, the RN decrypts the message and verifies the MAC using the key k_{Relay}. If the verifications succeed, the RN authenticates the UE.The RN produces a nonce $N1_{RN}$ and generates the MAC using the key k_{Relay} .The nonce and the identification of the relay are encrypted using the key k_{Relay} . Then the RN sends a Target authentication request message to the Target DeNB.

Step 10.3: When this message reaches the Target DeNB, the Target DeNB carries out the same MAC computation. If the verification is correct. The Target DeNB will successfully authenticate the UE and the Relay. The Target DeNB produces two nonces, $N1_T$ and $N2_T$. The Target DeNB constructs Target authentication response messages include the session key k_{U-T} as in (2) share between the UE and Target DeNB and the key k_{R-T} as in (3) share between the RN and Target DeNB. The Target DeNB uses the MAC algorithm to produce two message authentication codes for the RN and the UE.

$$k_{U-T} = \text{KDF}(k_{eNB}^*, N1_T, N1_{UE}+1, ID_T, ID_{UE}) \quad (2)$$

$$k_{R-T} = \text{KDF}(k_{Relay}, N1_{RN}+1, N2_T, ID_{RN}, ID_T) \quad (3)$$

Step 10.4: The RN receives the Target authentication response message and uses the key k_{relay} to decrypt the k_{R-T} as in (3). The RN verifies the MAC and authenticates the Target DeNB. Then, it calculates the key k_{U-R} as in (4) between the UE and RN. Then it adds the MAC of the UE into the

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

relay authentication response message and sends it to the UE.

k_{U-R} =KDF ($k_{\mathrm{Re}lay}$, N2$_{\mathrm{Relay}}$, N2$_{\mathrm{UE}}$+1, ID$_{\mathrm{RN}}$, D$_{\mathrm{UE}}$) (4)

Step 10.5: The UE receives the relay authentication message and decrypts k_{U-T} and k_{U-R} .The UE uses the same MAC algorithm to authenticate the RN and Target DeNB.

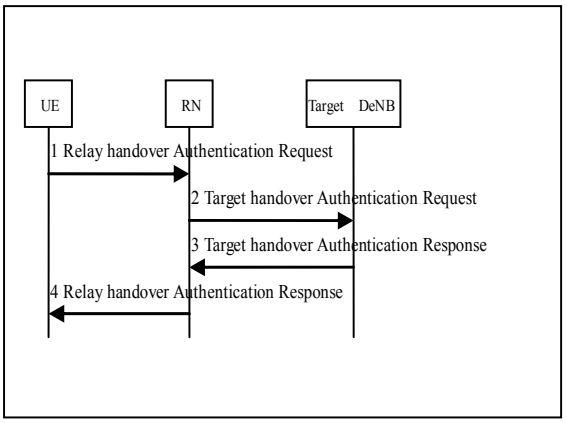


Figure 4. Relay assisted handover pre-authentication protocol

The protocol described in detail below:

UE →RN:

$\{N1_{UE}, ID_{UE}\}_{k^*_{eNB}}, MAC(k^*_{eNB}, N1_{UE} \| ID_{UE})$,

$\{N2_{UE}, ID_{UE}\}_{k_{\mathrm{Re}lay}}, MAC(k_{\mathrm{Re}lay}, N2_{UE} \| ID_{UE})$

RN →Target DeNB:

$\{N1_{UE}, ID_{UE}\}_{k^*_{eNB}}$, $MAC(k^*_{eNB}, N1_{UE} \| ID_{UE})$,

$\{N1_{RN}, N2_{UE}, ID_{UE}, ID_{RN}\}_{k_{\mathrm{Re}lay}}$,

$MAC(k_{\mathrm{Re}lay}, N1_{RN} \| N2_{UE} \| ID_{RN} \| ID_{UE})$

Target DeNB → RN:

$\{k_{U-T}, N1_T, N1_{UE}+1, ID_T, ID_{UE}\}_{k^*_{eNB}}$,

$MAC(k^*_{eNB}, K_{U-T} \| N1_T \| N1_{UE}+1 \| ID_T \| ID_{UE})$

$,\{k_{R-T}, N1_{RN}+1, N2_T, ID_{RN}, ID_T, ID_{UE}\}_{k_{\mathrm{Re}lay}}$,

$MAC(k_{\mathrm{Re}lay}, k_{R-T} \| N1_{RN}+1 \| N2_T$

$\| ID_{RN} \| ID_T \| ID_{UE})$

RN→UE:

$\{k_{U-T}, N1_T, N1_{UE}+1, ID_T, ID_{UE}\}_{k^*_{eNB}}$,

$MAC(k^*_{eNB}, K_{U-T} \| N1_T \| N1_{UE}+1 \| ID_T \| ID_{UE})$,

$\{k_{U-R}, N2_{RN}, N2_{UE}+1, ID_{RN}, ID_{UE}, ID_T\}_{k_{\mathrm{Re}lay}}$,

$MAC(k_{\mathrm{Re}lay}, k_{U-R} \| N2_{RN} \| N2_{UE}+1$

$\| ID_{RN} \| ID_{UE} \| ID_T)$

This protocol happens before the UE handovers to the target cell. Due to space limitations, we do not discuss the performance of the proposed protocol. This paper will pay more attention on security formal analysis.

IV. SECURITY FORMAL ANALYSIS

In this section, we will use strand space formal analysis method to prove that our protocol is secure. Although strand space method [6] is efficient, there are some shortcomings. We extend strand space to include message authentication code (MAC) item.

A. Extend Strand Space

We define the set A of terms. The element in A is the information exchanged among subjects. In particular, we will assume:

- A set T of texts (represent the atomic messages) $T \subset A$.
- A set of cryptographic keys K disjoint from T . $K \subset A$.
- A unary operator $inv : K \to K$.
- Three binary operators

$$encr : K \times A \to A$$
$$join : A \times A \to A$$
$$MAC : K^{'} \times A \to A$$

As usual, we will write $inv(K)$ as K^{-1}, $encr(K,m)$ as $\{m\}_K$, and $join(a,b)$ as ab .

We redefine axioms as follows.

Axiom 1: for $m, m^{'} \in A$ and $k_1, k_2 \in K$, $k_1^{'}, k_2^{'} \in K$

If $\{m\}_{k_1} = \{m^{'}\}_{k_2}$ then $m = m^{'}$ and $k_1 = k_2$;

If $MAC(k_1^{'}, m) = MAC(k_2^{'}, m)$ then $m = m^{'}$ and $k_1^{'} = k_2^{'}$.

Axiom 2: for $m_0, m_0^{'}, m_1, m_1^{'} \in A$, and $k, k^{'} \in K$

(1) $m_0 m_1 = m_0^{'} m_1^{'} \Rightarrow m_0 = m_1^{'} \wedge m_1 = m_1^{'}$

(2) $m_0 m_1 \neq \{m_0^{'}\}_{k^{'}}$

(3) $m_0 m_1 \notin K \cup T$

(4) $\{m_0^{'}\}_k \notin K \cup T$

(5) $m_0 m_1 \neq MAC(k^{'}, m_0^{'})$

(6) $MAC(k^{'}, m_0^{'}) \notin K \cup T$

We extend it to include MAC item according to [7].

Definition 1 A strand space is a set $\sum$ with a trace mapping tr : $\sum \to (\pm A)^*$.

Definition 2 A penetrator trace is one of the following:

M. Text message :< +t>, where $t \in T$

F. Flushing :<-g>

T. Tee : <-g,+g,+g>

G. Concatenation : <-g,-h, +gh>

S. Separating into components:<-gh, +g,+h>

K. Key :< +k>, where $k \in K_P$

E. Encryption: <-k,-h, + $\{h\}_k$ >

D. Deception : <-k,-h. - $\{h\}_k$,+h>

MAC . Message authentication code:

<-k',-h, $MAC(k',h)$ >

Definition 3 Let C be a set of edges, and let N_C be the set of nodes incident with any edge in C . C is bundle if :

(1). C is finite.

(2). If $n_1 \in N_C$ and $term(n_1)$ is negative, then there is a unique n_2 such that $n_2 \to n_1 \in N_C$.

(3). If $n_1 \in N_C$ and $n_2 \Rightarrow n_1$ then $n_2 \Rightarrow n_1 \in N_C$.

(4). C is acyclic.

Definition 4 If C is a bundle and $s \in \sum$, then the C height of s , denoted C- $height(s)$, is the largest $i \le length(tr(s))$, such that $s \in \sum$ and $\langle s, i \in C \rangle$

Definition 5 An infiltrated strand space is a pair $\langle \sum, P \rangle$ with $\sum$ a strand space and $P \in \sum$ such that $tr(p)$ is a penetrator trace for all $p \in P$

Definition 6 The sub-term relation $\sqsubset$ is defined Inductively, so that:

(1) $a \sqsubset t$ for $t \in T$ iff $a = t$

(2) $a \sqsubset k$ for $k \in K$ iff $a = k$

(3) $a \sqsubset \{g\}_k$ iff $a \sqsubset g$ or $a = \{g\}_k$

(4) $a \sqsubset gh$ iff $a \sqsubset h$, $a \sqsubset g$ or $a = gh$

(5) $a \sqsubset MAC(k',g)$ iff $a \sqsubset g$ or $a = MAC(k',g)$

We redefine the ideal and honest idea [8] including the MAC.

Definition 7 If $\kappa \subseteq K$, a κ -ideal of A is a subset I of A such that for all $h \in I$, $g \in A$ and $k \in \kappa$

(1) $gh, hg \in I$

(2) $\{h\}_k \in I$

(3) $MAC(k',h) \in I$

The smallest κ - ideal containing h is denoted $I_\kappa[h]$.

Definition 8 h is a sub-term of g , written $h \sqsubset g$, defined as $g \in I_\kappa[h]$.

Proposition 1 $\sqsubset$ is a transitive ,reflextive relation .

More over ,if $h, g \in A$ and $k, k' \in K$, then

(1) $h \sqsubset h\ g$ and $g \sqsubset g\ h$

(2) $h \sqsubset \{h\}_k$

(3) $h \sqsubset MAC(k',h)$

Definition 9 If $S \subseteq A$ $I_\kappa[S]$ is the smallest κ - ideal containing S .

Definition 10 Support $\kappa \subseteq K$. $s \in A$ is a κ - subterm of $t \in A$, written $s \sqsubset_\kappa t$ iff $t \in I_\kappa[s]$

Proposition 2 if $S \subseteq A$, $I_\kappa[S] = \cup_{x \in S} I_\kappa[x]$

Lemma 1 Let $S_0 = S$,

$$S_{i+1} = \{\{g\}_k, MAC(k',g) : g \in I_\phi[S_i], g \in I_\phi[S_i], k, k' \in \kappa\}$$

then $I_\kappa[S] = \cup_i I_\phi[S_i]$

Proposition 3 Suppose $S \subseteq A$, and every $s \in S$ is simple .If $gh \in I_\kappa[S]$ then either $g \in I_\kappa[S]$ or $h \in I_\kappa[S]$

Proposition 4 Suppose $k, k' \in K$; $S \subseteq A$, and for every $s \in S$, s is simple and is not of the form $\{g\}_k$ or $MAC(k',g)$.

if $\{h\}_k \in I_\kappa[S]$ or $MAC(k',h) \in I_\kappa[S]$,then $h \in I_\kappa[S]$.

Lemma 2 Suppose $k_1' \ne k_2'$, and

$MAC(k_1', h_1) \sqsubset MAC(k_2', h_2)$,

Then $MAC(k_1', h_1) \sqsubset h_2$

Proposition 5 Suppose $k, k' \in K$, $S \subseteq A$,and every $s \in S$ is simple and is not of the form $MAC(k',h)$ or $\{h\}_k$ .If $MAC(k',h) \in I_\kappa[S]$ or $\{h\}_k \in I_\kappa[S]$, then $k \in \kappa$ or $k' \in \kappa$.

Proposition 6 Suppose C is a bundle over A. If m is minimal in $\{m \in C : term(m) \in I\}$, then m is an entry point for I .

Definition 11 A set $I \subseteq A$ is honest relative to a bundle C if and only if whenever a penetrator node p is an entry point for I , p is an M node or a K node.

Theorem 1 Suppose C is a bundle over $A, S \subseteq T \cup K$, $\kappa \subseteq K$, $K \subseteq S \cup \kappa^{-1}$, Then $I_\kappa[S]$ is honest.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

Corollary 1 Suppose C is a bundle, $\mathrm{K} \subseteq S \cup \kappa^{-1}$, and $S \cap \mathrm{K}_{\mathrm{P}} = \phi$. If $term(m) \in I_{\kappa}[S]$ for some $m \in C$, then for some regular node $n \in C$, n is an entry point for $I_{\kappa}[S]$.

Corollary 2 Suppose C is a bundle, $\mathrm{K} \subseteq S \cup \kappa^{-1}$, $S \cap \mathrm{K}_{\mathrm{P}} = \phi$, and no regular node $\in C$ is an entry point for $I_{\kappa}[S]$. Then any term of the form $\{g\}_k$ and $MAC(k', g)$ for $k, k' \in S$ does not originate on a penetrator strand.

B. The Bundle

We will use extended honest and ideal concept to prove the security of our proposed protocol. The goal of this protocol is to mutually authenticate hop-by-hop. The bundle of our proposal is shown in Figure. 5.

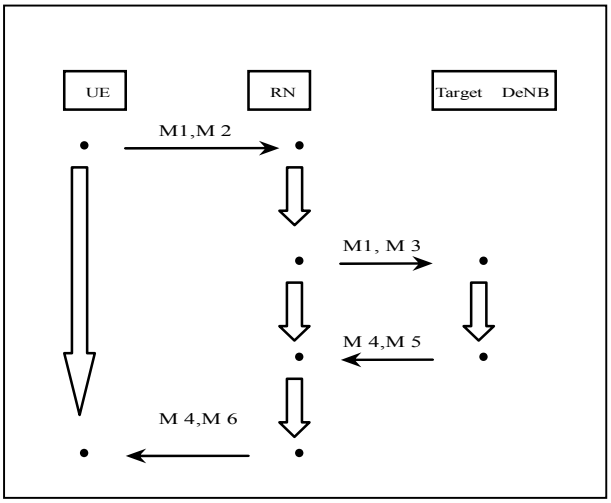


Figure 5. Our propose strand space

M1= $\{N1_{UE}, ID_{UE}\}_{k^*_{eNB}}$ $MAC(k^*_{eNB}, N1_{UE} \| ID_{UE})$

M2= $\{N2_{UE}, ID_{UE}\}_{k_{\mathrm{Re}lay}}$ $MAC(k_{\mathrm{Re}lay}, N2_{UE} \| ID_{UE})$

M3= $\{N1_{RN}, N2_{UE}, ID_{UE}, ID_{RN}\}_{k_{\mathrm{Re}lay}}$,
$MAC(k_{\mathrm{Re}lay}, N1_{RN} \| N2_{UE} \| ID_{RN} \| ID_{UE})$

M4= $\{k_{U-T}, N1_T, N1_{UE}+1, ID_T, ID_{UE}\}_{k^*_{eNB}}$
$MAC(k^*_{eNB}, K_{U-T} \| N1_T \| N1_{UE}+1 \| ID_T \| ID_{UE})$

M5= $\{k_{R-T}, N1_{RN}+1, N2_T, ID_{RN}, ID_T, ID_{UE}\}_{k_{\mathrm{Re}lay}}$
$MAC(k_{\mathrm{Re}lay}, k_{R-T} \| N1_{RN}+1 \| N2_T \| ID_{RN} \| ID_T \| ID_{UE})$

M6= $\{k_{U-R}, N2_{RN}, N2_{UE}+1, ID_{RN}, ID_{UE}, ID_T\}_{k_{\mathrm{Re}lay}}$,
$MAC(k_{\mathrm{Re}lay}, k_{U-R} \| N2_{RN} \| N2_{UE}+1 \| ID_{RN} \| ID_{UE} \| ID_T)$

Definition 12 An infiltrated strand space $(\sum, \mathrm{P})$ is a space if $\sum$ is the union of three kinds of strands:

(1) Penetrator's strand $s \in \mathrm{P}$;

(2) UE 's Strand
$s \in UE[N1_{UE}, N2_{UE}, N1_T, N2_{RN}, ID_{UE}, ID_T, ID_{RN}, k_{U-T}, k_{U-R}]$
with trace, defined to be:
$<+M1, +M2, -M4, -M6>$

(3) Relay's strand
$s \in RN[N2_{UE}, N2_T, N1_{RN}, N2_{RN}, ID_{UE}, ID_{RN}, ID_T, k_{R-T}, k_{U-R}, H, G]$
with trace, defined to be:
$<-H, -M2, +H, +M3, -G, -M5, +G, +M6>$

(4) Target DeNB's Strand
$s \in$ Target DeNB
$[k_{U-T}, k_{R-T}, N1_{UE}, N1_T, N2_T, N1_{RN}, ID_{UE}, ID_{RN}, ID_T, \mathrm{H}]$
with trace defined to be:
$<-H, -M3, +M4, +M5>$

For
$k_{U-T}, K_{R-T}, k_{U-R} \notin \mathrm{K}_{\mathrm{P}}$
$k_{U-T}, K_{R-T}, k_{U-R} \notin \{k_{\mathrm{Re}lay}, k^*_{eNB}\}, k_x = k_x^{-1}$

Proposition 7 The set UE,RN and Target DeNB are disjoint each other.

C. Security Analysis

We first prove that session keys can not be disclosed unless penetrator posses one of the long term keys.

Theorem 2 suppose C is a bundle in strand space $\sum$, $RN \in T_{name}$, session key k_{R-T} are uniquely originating; $k_{R-T} \notin \mathrm{K}_{\mathrm{P}}$; and $s \in \sum_{RN}$ has C - height 3. Let $S = \{k_{\mathrm{Re}lay}, k^*_{eNB}, k_{R-T}\}$ and $\kappa = \mathrm{K} \setminus S$ For every node $m \in C$, $term(m) \notin I_{\kappa}[k_{R-T}]$.

PROOF: By proposition 2, it suffices to prove that stronger statement that for every node m, $term(m) \notin I_{\kappa}[S]$. Since $S \cap \mathrm{K}_{\mathrm{P}} = \phi$, $\kappa = \kappa^{-1}$, and $\mathrm{K} = \kappa \cup S$ by Corollary 1**,** It suffices to show that no regular node m is an entry point for $I_{\kappa}[S]$.

We will argue by contradiction and assume m is a regular node which is an entry point for $I_\kappa[S]$. Since m is an entry point for $I_\kappa[S]$, by the definitions, it follows that $term(m)$ is an element of $I_\kappa[S]$. By proposition 2, this implies that one of the keys $k_{\mathrm{Re}lay}, k^*_{eNB}, k_{R-T}$ is a sub term of $term(m)$. Now, no regular node contains any key $k_{\mathrm{Re}lay}, k^*_{eNB}$ as a sub-term. In fact, the only session keys which occur as sub-terms of $term(m)$ for m regular, are the session keys emanating from the Target DeNB. If m is a positive regular node on a strand s, then $k_{R-T} \sqsubset term(m)$ implies either:

(1) $s \in \sum_{Target\ DeNB}$ and $m = \langle s,2 \rangle$, in which case k_{R-T} is the session key;

(2) $s \in \sum_{RN}$ and $m = \langle s,3 \rangle$ $k_{R-T} \sqsubset H$

In case 2, m is not an entry point for $I_\kappa[S]$, because $H \sqsubset \langle s,1 \rangle$, which is a preceding negative node. So m is not entry point of $I_\kappa[S]$.

So consider case 1. By the unique origination of k_{R-T}, $s = S_{Target\ DeNB}$, so $term(m)$ = M5 or $term(m)$ =M4

By Proposition **3**, either

1) $\{k_{R-T}, N1_{RN}+1, N2_T, ID_{RN}, ID_T, ID_{UE}\}_{k_{\mathrm{Re}lay}} \in I_\kappa[S]$

2) $\{k_{U-T}, N1_T, N1_{UE}+1, ID_T, ID_{UE}\}_{k^*_{eNB}} \in I_\kappa[S]$

3) $MAC(k_{\mathrm{Re}lay}, k_{R-T} \| N1_{RN}+1 \| N2_T \| ID_{RN} \| ID_T \| ID_{UE}) \in I_\kappa[S]$

4) $MAC(k^*_{eNB}, K_{U-T} \| N1_T \| N1_{UE}+1 \| ID_T \| ID_{UE}) \in I_\kappa[S]$

But, the first and the second are impossible. The third and the fourth are impossible by Proposition 5.□

Similarly, we can prove that k_{U-T} and k_{U-R} are secure. So, we can conclude that session keys can not be disposed. Our protocol is secure.

D. Authetication Analysis

In this section we will prove the authentication guarantees to its UE, RN and Target DeNB.

Proposition 8 Consider a bundle C in $\sum$, Suppose $\mathrm{Target\ DeNB} \in T_{name}$ and $k^*_{eNB} \notin \mathrm{K_P}$. Then no term of the form $\{g\}_{k^*_{eNB}}, MAC(k^*_{eNB}, g)$ can originate on a penetrator node in C.

PROOF: Let $S = \{k^*_{eNB}\}$ and $\kappa = \mathrm{K}$. To apply **Corollary 2,** we must check that no regular node is an entry point for $I_\kappa[S]$, or equivalently, the k^*_{eNB} does not originate on any regular node.

A key originates on a regular node only if it is a session key k originating on a Target DeNB strand $s \in \sum_{Target\ DeNB}$. However, by the definition of $\sum_{Target\ DeNB}$, the session key k is never a long term key k^*_{eNB}.

Hence we may apply **Corollary 2** to $I_\kappa[S]$, so any term $\{g\}_{k^*_{eNB}}, MAC(k^*_{eNB}, g)$ can only originate on a regular node. □

Lemma 3 Consider a bundle C in $\sum$, Suppose $\mathrm{R}N \in T_{name}$ and $k_{\mathrm{Re}layB} \notin \mathrm{K_P}$. Then no term of the form $\{g\}_{k_{\mathrm{Re}lay}}, MAC(k_{\mathrm{Re}lay}, g)$ can originate on a penetrator node in C.

Proposition 9

1) If $\{H\}_{k^*_{eNB}}$ originates on a regular strand s, then If $s \in \sum_{Target\ DeNB}$, then $H = k_{U-T} \| N1_T \| N1_{UE}+1 \| ID_T \| ID_{UE}$ and $k_{U-T} \in \kappa$

2) If $\{H\}_{k_{\mathrm{Re}lay}}$ originates on a regular strand s, then

If $s \in \sum_{Target\ DeNB}$ then $H = k_{R-T} \| N2_T \| N1_{\mathrm{Re}lay}+1 \| ID_{\mathrm{Re}lay} \| ID_T \| ID_{UE}$ and $k_{R-T} \in \kappa$

If $s \in \sum_{RN}$ then $H = k_{U-R} \| N2_{RN} \| N2_{UE}+1 \| ID_{RN} \| ID_{UE} \| ID_T$ and $k_{U-R} \in \kappa$

3) If $MAC(k^*_{eNB}, H)$ originates on a regular strand s then If $s \in \sum_{UE}$,then $H = N1_{UE} \| ID_{UE}$

If $s \in \sum_{Target\ DeNB}$, then $H = k_{U-T} \| N1_T \| N1_{UE}+1 \| ID_{UE} \| ID_T$

4) If $MAC(k_{\mathrm{Re}lay}, H)$ originates on a regular strand s, then If $s \in \sum_{RN}$,

$H = N1_{RN} \| N2_{UE} \| ID_{UE} \| ID_{RN}$

Or

$H = k_{U-R} \| N2_{RN} \| N2_{UE} + 1 \| ID_{RN} \| ID_{UE} \| ID_T$

(5) If $MAC(k_{\mathrm{Re}lay}, H)$ originates on a regular strand s, then If $s \in \sum_{T\arg et\ DeNB}$ then

$H = k_{R-T} \| N2_T \| N1_{\mathrm{Re}lay} + 1 \| N2_{UE} + 1 \| ID_{\mathrm{Re}lay} \| ID_T \| ID_{UE}$

PROOF: By the definition of originating, if the term $\{H\}_k$ originates on m, then m is positive.

If $s \in \sum_{T\arg et\ DeNB}$ then $m = \langle s,2 \rangle$. Thus the encrypted subterm of $term(m)$ $\{k_{U-T}, N1_T, \| N1_{UE} + 1 \| ID_{UE} \| ID_T\}_{k^*_{eNB}}$ is of from (1)

If the term $MAC(k,H)$ originates on m, then m is positive. If $s \in \sum_{UE}$ then $m = \langle s,1 \rangle$. The subterm of this term is of the form (3).

If $s \in \sum_{RN}$, If the term $MAC(k,H)$ originates on m, then m is positive. Then the positive nodes of the s is $m = \langle s,3 \rangle$ and the sub-term of this term is of the form (4). □

Corollary 3 Suppose s is a regular strand of $\sum$

(1) IF $\{k_{U-T}, N1_T, N1_{UE} + 1, ID_{UE}, ID_T\}_{k^*_{eNB}}$ originates on s, then $s \in \sum_{T\arg et\ DeNB}$. The term originates on the node $\langle s,2 \rangle$ and k_{U-T} originates on s.

(2) If $\{k_{R-T}, N1_{RN} + 1, N2_T, ID_{RN}, ID_T, ID_{UE}\}_{k_{\mathrm{Re}lay}}$ originates on s, then $s \in \sum_{T\arg et\ DeNB}$. The term originates on the node $\langle s,2 \rangle$, and k_{R-T} originates on s.

(3) If $\{k_{U-R}, N2_{RN}, N2_{UE} + 1, ID_{RN}, ID_{UE}, ID_T\}_{k_{\mathrm{Re}lay}}$ originates on s, then $s \in \sum_{RN}$, the term originates on the node $\langle s,3 \rangle$ and k_{U-R} originates on s.

(4) If $\{N1_{UE}, ID_{UE}\}_{k^*_{eNB}}$ originates on s, then $s \in \sum_{UE}$, then the term $\{N1_{UE}, ID_{UE}\}_{k^*_{eNB}}$ originates on node $\langle s,1 \rangle$

(5) If $\{N2_{UE}, ID_{UE}\}_{k_{\mathrm{Re}lay}}$ originates on s, then $s \in \sum_{UE}$, $\{N2_{UE}, ID_{UE}\}_{k_{\mathrm{Re}lay}}$ originates on node $\langle s,1 \rangle$

(6) If $MAC(k^*_{eNB}, N1_{UE} \| ID_{UE})$ originates on s, then $s \in \sum_{UE}$, $MAC(k^*_{eNB}, N1_{UE} \| ID_{UE})$ originates on node $\langle s,1 \rangle$

(7) If $MAC(k_{\mathrm{Re}lay}, N2_{UE} \| ID_{UE}\}$ originates on s, then $s \in \sum_{UE}$ $MAC(k_{\mathrm{Re}lay}, N2_{UE} \| ID_{UE}\}$ originates on node $\langle s,1 \rangle$

(8) If $MAC(k_{\mathrm{Re}lay}, N1_{RN} \| N2_{UE} \| ID_{RN} \| ID_{UE})$ originates on s, then $s \in \sum_{RN}$ $MAC(k_{\mathrm{Re}lay}, N1_{RN} \| N2_{UE} \| ID_{RN} \| ID_{UE})$ origin ates on node $\langle s,2 \rangle$

(9) If $MAC(k_{\mathrm{Re}lay}, k_{R-T} \| N1_{RN} + 1 \| N2_T \| ID_{RN} \| ID_T \| ID_{UE})$ originates on s, then $s \in \sum_{T\arg et\ DeNB}$ $MAC(k_{\mathrm{Re}lay}, k_{R-T} \| N1_{RN} + 1 \| N2_T \| ID_{RN} \| ID_T \| ID_{UE})$ originates on $\langle s,2 \rangle$

If $MAC(k_{\mathrm{Re}lay}, k_{U-R} \| N2_{RN} \| N2_{UE} + 1 \| ID_{RN} \| ID_{UE} \| ID_T)$ originates on s, then $s \in \sum_{RN}$, the $MAC(k_{\mathrm{Re}lay}, k_{U-R} \| N2_{RN} \| N2_{UE} + 1 \| ID_{RN} \| ID_{UE} \| ID_T)$ originates on $\langle s,3 \rangle$.

PROOF: Since s is regular, $s \in \sum_{UE} \cup \sum_{RN} \cup \sum_{T\arg et\ DeNB}$ Apply proposition 9. □

The following theorem asserts that if a bundle contains a strand $s \in \sum_{UE}$ then under reasonable assumptions, there are regular strand $s \in \sum_{RN}$, $s \in \sum_{T\arg et\ DeNB}$, Which agrees on the UE, RN, Target DeNB

Theorem 3 Support C is a bundle in $\sum$; $UE \neq T\arg et\ DeNB \neq RN$; $N1_{UE}, N2_{UE}$ is uniquely originating in C; and $k^*_{eNB}, k_{\mathrm{Re}lay} \notin \mathrm{K_P}$ If $s \in \sum_{UE}$ has C - height 2, then there are regular strands :

(1) $s \in \sum_{RN}$ of height 3 at least

(2) $s \in \sum_{Target\ DeNB}$ of height 2

PROOF: According to the trace of $s \in \sum_{UE}$

Since $k^*_{eNB}, k_{Relay} \notin \mathrm{K_P}$, by Lemma 3 $-M6$ originates on a regular node in C. By Corollary 3**,** this node belongs to a strand $s \in \sum_{RN}$ which has C - height 3 at least.

Since $k^*_{eNB}, k_{Relay} \notin \mathrm{K_P}$, by Lemma 3, $-M4$ originates on a regular node in C. By Corollary 3, this node belongs to a strand $s \in \sum_{Target\ DeNB}$ which C - height 2. □

Theorem 4 Support C is a bundle in $\sum$; $UE \neq Target\ DeNB \neq RN$, $N1_{RN}, N2_{UE}, N1_{UE}$ are uniquely originating in C; and $k^*_{eNB}, k_{Relay} \notin \mathrm{K_P}$ If $s \in \sum_{Target\ DeNB}$ has C- height 2 , then there are regular strands :

(1) $s \in \sum_{RN}$ of height 2 at least

(2) $s \in \sum_{UE}$ of height 1 at least

PROOF: According to the trace of

$s \in \sum_{Target\ DeNB}$

Since $k^*_{eNB}, k_{Relay} \notin \mathrm{K_P}$, by lemma 3, $-M3$ originates on a regular node in C. By Corollary 3 **,** this node belongs to a strand $s \in \sum_{RN}$ which C - height 2 at least.

Since $k^*_{eNB}, k_{Relay} \notin \mathrm{K_P}$, by lemma 3 $-M1$ originates on a regular node in C By Corollary 3, this node belongs to a strand $s \in \sum_{UE}$ which C - height 1 at least. □

Theorem 5 Support C is a bundle in $\sum$; $UE \neq Target\ DeNB \neq RN$; $N2_{UE}, N1_{RN}, N1_T$ are uniquely originating in C; and $k_{Relay} \notin \mathrm{K_P}$ If $s \in \sum_{RN}$ has C -height 3 , then there are regular strands :

(1) $s \in \sum_{Target\ DeNB}$ of height 2

(2) $s \in \sum_{UE}$ of height 1 at least

PROOF: According to the trace of $s \in \sum_{RN}$

Since $k_{Relay} \notin \mathrm{K_P}$, by lemma 3, $-M5$ originates on a regular node in C. By Corollary 3 , this node belongs to a strand $s \in \sum_{Target\ DeNB}$ with C - height 2. Since $k_{Relay} \notin \mathrm{K_P}$, by lemma 3, $-M1$, originates on a regular node in C. By Corollary 3, this node belongsto a strand $s \in \sum_{UE}$ have height 1 at least. □

So we can conclude that UE, RN and Target deNB can authenticate each other.

V. PERFORMANCE EVALUATION

In this section, we will discuss the performance of our proposal scheme.

In the traditional handover scheme, the UE will tear down the connection with the Source eNB first. When the UE moves into the target cellular, it will establish the connection with the Target eNB. The handover messages are transmitted from the UE to the Source eNB, then to the MME and finally to the serving GW, as in figure 1. UE and Target eNB will finish end-to-end authentication using AKA protocol.

But , in the relay-assisted handover procedure, the handover messages are transmitted among the source DeNB, RN and Target DeNB. With the help of the RN, the handover information does not need to be transmitted on the S1 interface. The handover delay will be reduced significantly.

In our extended relay-assisted handover procedure, security keys are carried on the handover messages. They are protected by the RRC key. Some security keys are transmitted to the target DeNB and RN before the handover happens.

With the help of the RN, wireless connection is established among the UE, RN and the Target DeNB. Before the UE handovers to the target DeNB, Our proposed pre-authentication protocol can be executed among the UE, RN and Traget DeNB hop-by-hop. Because this protocol is happened before handover, the overhead is not calculated on handover delay. When the UE handovers to the target DeNB, it does not need to run the AKA protocol from scratch. The UE only needs to finish local authentication process with the target DeNB. The handover authentication delay will be reduced.

VI. CONCLUSION

Relaying is key technique in future LTE-A network. Relay node is introduced to extend coverage and capacity. In order to enable relaying, handover procedure, architecture and protocol have to be modified. This paper introduced a new relay assisted handover mechanism. Handover messages are exchanged among UE, RN and DeNB. We consider the security issue about the relay-assisted handover procedure. Before the UE moves into the target cellular, security contexts are transferred on the handover messages to the target cellular. With the aid of the relay nodes, the UE performs pre-authentication protocol when the UE still remain in the source cellular. The UE, RN and Target DeNB mutual authenticate using hop-by-hop communications. When the UE handovers to the target cellular, it does not need to perform end-to-end authentication from scratch. Handover authentication delay is reduced significantly. The security formal analysis is our main task. We also extend traditional strand space including message authentication

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

code. We use the extended ideal and honest idea to prove the security of our pre-authentication protocol. But, in this paper, we do not discuss the handover delay and loss rate. In future work, we will evaluate the overhead of our scheme using simulation and analytical methods.

ACKNOWLEDGMENT

This research work is supported by the National Natural Science Foundation of China (NSFC) under Grant No. 60802058, and the SMC young scholar sponsorship of Shanghai Jiao Tong University.

REFERENCES

[1] C. Sunghyun, E. W. Jang, and J. M. Cloffi, "Handover in multihop Cellular network," IEEE Communication Magzine, Vol. 47, pp. 529–551, July 2009.

[2] O. Teyeb, V. V. Phan, B. Raaf, and D. Redana, "Handover framework for relay enhanced LTE networks," IEEE International Conference on communications workshops, pp. 1–5, June 2009.

[3] The 3GPP draft 36.300, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2".

[4] The 3GPP draft 33.816, "Feasibility Study on LTE Relay Node Security."

[5] The 3GPP draft 33.401 , "3GPP System Architecture Evolution (SAE); Security architecture."

[6] F. J. Thsyer Fabreqa, J. C. Herzog, and J. D. Guttman, "Strand spaces: Why is a Security Protocol Correct," proceedings of IEEE Symposium on security and Privacy, pp. 160-171, May 1998.

[7] Y. Li Li, P. Dai Yuan, and G. Yue Xiang, "Analysis and Improvement of Sensor Networks Security Protocol," Journal on Communications ,Vol . 32, No. 5, pp. 139–145, May 2011.

[8] F. J. Thayer Fabreqa, J. C. Herzog, and J. D. Guttman, "Strand Spaces: Honest Ideals on Strand spaces," Proceedings of the 11th IEEE Computer Security Foundations Workshop, pp. 66–77, June 1998.

,

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

A Flow Label Based QoS Scheme for End-to-End Mobile Services

Tao Zheng, Lan Wang, Daqing Gu
Orange Labs Beijing
France Telecom Group
Beijing, China
e-mail: {tao.zheng; lan.wang; daqing.gu}@orange.com

***Abstract* - As a network evolution goal, IPv6 is deployed in mobile network, including access network, core network and mobile carrier IP network. IPv6 introduction in mobile network will impact on Quality of Service (QoS) of mobile services. In this paper, a flow label based QoS scheme is proposed to improve QoS in mobile network. This scheme utilizes the flow label in IPv6 packet header to indentify the services and flows and make the mobile network and carrier IP network entities to perceive the existence of flows and control them. Particularly, the experiment results based this scheme indicate that it has the finest granularity of QoS and minimizes the affected flows with the help of flow label.**

Keywords-flow label; QoS; TFT; carrier IP network

I. INTRODUCTION

With the development of IP-based carrier network, Long Term Evolved (LTE) evolution and the rise of mobile Internet and multimedia services, QoS, especially Internet Protocol (IP) QoS is becoming more and more important in mobile networks.

The 3rd Generation Partnership Project (3GPP) TS 23.107 specification [1] defines the QoS architecture in Universal Mobile Telecommunications System (UMTS) as shown in Figure 1.

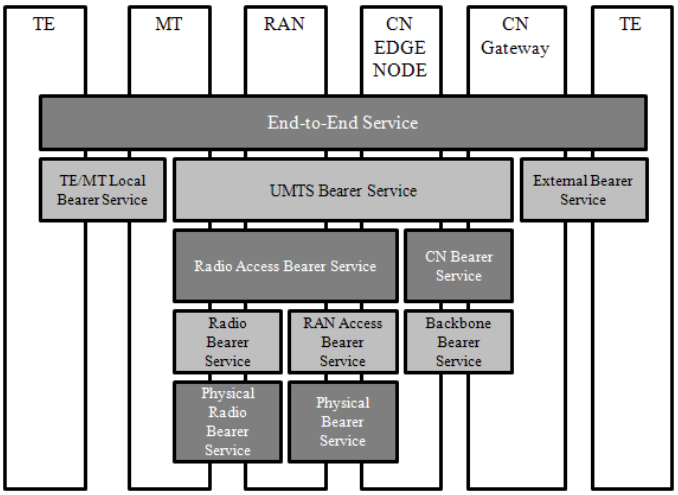


Figure 1. QoS architecture defined by 3GPP

The service on the higher layer consists of bearer services on the lower layer. Accordingly, the higher service's QoS is guaranteed by lower services' QoS. The End-to-End Service is finally mapped to Terminal Equipment/Mobile Terminal (TE/MT) Local Bearer Service, Physical Radio Bearer Service, Physical Bearer Service of Radio Access Network (RAN), Backbone Bearer Service, and External Bearer Service.

There are four different QoS classes defined by 3GPP: conversational class, streaming class, interactive class and background class. The main distinguishing factor between these QoS classes is how delay sensitive the traffic is: Conversational class is meant for traffic, which is very delay sensitive, while Background class is the most delay insensitive traffic class.

LTE utilizes a class-based QoS concept, which reduces complexity while still allowing enough differentiation of traffic handling and charging by operators. Bearers can be classified into two categories based on the nature of the QoS they provide: Minimum Guaranteed Bit Rate (GBR) bearers and Non-GBR bearers. The Figure 2 shows the QoS architecture in LTE [2].

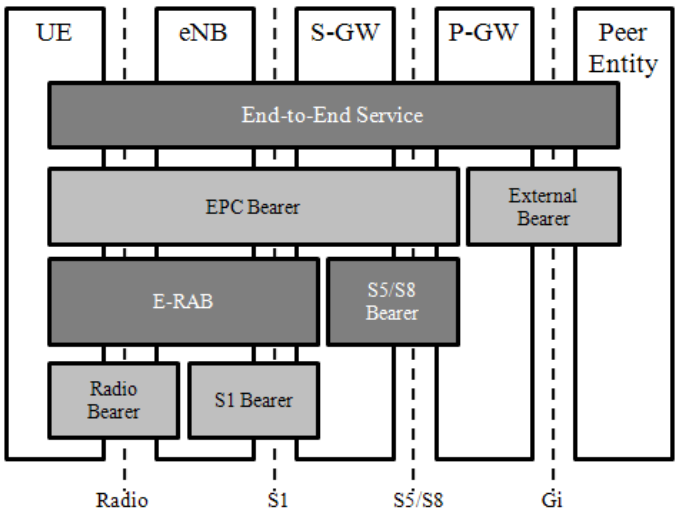


Figure 2. QoS architecture in LTE

The QoS architecture in LTE is similar to that in UMTS, as shown in Figure 1. The End-to-End Service is finally mapped to several bearers between mobile network entities.

In PS domain of UMTS and LTE, Traffic Flow Template (TFT) [3] mechanism is used to provide QoS guarantee. The TFTs contain packet filter information that allows the User

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

Equipment (UE) and Gateway GPRS Support Node/Public Data Network Gateway (GGSN/P-GW) to identify the packets belonging to a certain IP packet flow aggregate. This packet filter information is typically a 5-tuple that contains the source and destination IP addresses, source and destination ports as well as a protocol identifier (e.g., User Datagram Protocol (UDP) or Transmission Control Protocol (TCP)). Figure 3 describes the TFT architecture in LTE. The UE and the P-GW (for GPRS Tunneling Protocol (GTP)-based S5/S8) or Serving Gateway (S-GW) (for Proxy Mobile IP (PMIP)-based S5/S8) use packet filters to map IP traffic onto the different bearers. The TFTs are typically created when a new Evolved Packet System (EPS) bearer is established, and they are then modified during the lifetime of the EPS bearer.

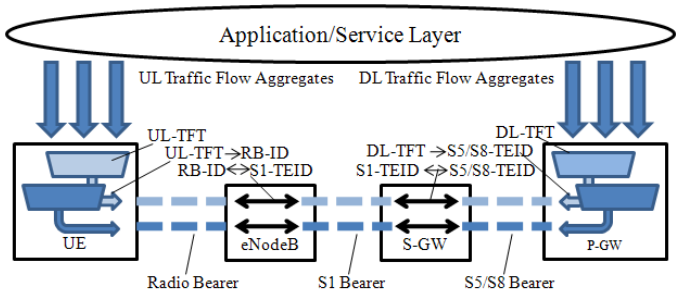


Figure 3. TFT reference network architecture

In the mobile IP carrier network, Differentiated Services (Diffsev) mechanisms or the Resource Reservation Protocol (RSVP) protocol can be used for QoS policy enforcement and resource reservation purposes at present. Multiple Protocol Label Switching Virtual Private Network/Traffic Engineering (MPLS VPN/TE) also can be used to provide QoS in IP carrier network. Such QoS policies and mechanisms design are usually engineered in advance.

From the above introduction, we can conclude that the current QoS mechanism used in mobile network is based on bearers (in access and core mobile networks) and Diffserv or RSVP techniques (in IP carrier network). With the deployment of LTE and the explosion of mobile Internet traffic, it's necessary to provide fine-grained control of traffic. Actually, each bearer includes some flows, such as http sessions from different websites, different applications in same Packet Data Protocol (PDP) context/Evolved Packet Core (EPC) bearer; there are more flows in same service class, e.g., active Voice over IP (VoIP) calls have the same QoS level in IP carrier network. So it's hard to perceive the existence of flows, and the controlled granularity can not be flows.

This paper is organized as follows. In Section 2, we introduce the current situation of IPv6 flow label. In Section 3, a flow label based QoS scheme is presented. In Section 4, an experiment based-on this scheme is implemented and experiment results are analyzed. Finally, Section 5 summarizes the conclusions.

II. IPv6 Flow Label and Applications

A sequence of packets sent from a particular source to a particular unicast, anycast, or multicast destination constitute a flow. In IPv4 network, the 5-tuple of the source and destination addresses, ports, and the transport protocol type is able to identify a flow.

IPv6 has introduced a field named flow label. The 20-bit flow label in the IPv6 header is used by a node to label packets of a flow. General rules for the flow label field have been documented in Request for Comments (RFC) 3697 [4]. But how to apply this field in real-world network is still an open issue since the research work on flow label is far from enough.

Some Internet Engineering Task Force (IETF) RFCs and drafts have been proposed to update flow label works.

In fact, some published proposals for use of the IPv6 flow label are incompatible with the original RFC 3697. Furthermore, very little practical use is made of the flow label, partly due to some uncertainties about the correct interpretation of the specification. In [5], the authors present some changes to the specification in order to clarify it, and to introduce some additional flexibility are discussed.

The draft "IPv6 flow label Specification" [6] is trying to update the RFC 3697. It specifies the IPv6 flow label field and the minimum requirements for IPv6 nodes labeling flows, IPv6 nodes forwarding labeled packets, and flow state establishment methods.

In [7], the author surveys various published proposals for using the flow label and shows that most of them are inconsistent with the standard. Methods to address this problem are briefly reviewed.

Various use cases have been proposed that infringe flow label rules. In [8], the authors describe how those restrictions apply when using the flow label for load balancing by equal cost multipath routing, and for link aggregation, particularly for IP-in-IPv6 tunneled traffic. In [9], the authors give an application that how the IPv6 flow label can be used in support of layer 3/4 load balancing for large server farms.

Flow identification is of vital importance for end-to-end QoS provision in mobile network. Many QoS management techniques can be achieved, e.g., Connection Admission Control (CAC), scheduling, etc, and some technologies have come out, e.g., MPLS, Scalable Core (SCORE), etc. with the ability to deal with per-flow and aggregated flow. At present, none of them is applied in mobile networks. In this paper, we proposed a QoS scheme utilized IPv6 flow label to provide a flow granularity QoS mechanism for end-to-end mobile services. Compared with other flow-based QoS techniques, our scheme is simpler and has minimal impacts on mobile networks.

III. Proposed Flow Label Based QoS Scheme

Previous QoS mechanisms in mobile network are based-on bearers, the packets contains in one same bearer share same QoS profile and be treated in the same way. According to the definition of a flow, there should be some flows in a same

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

bearer. It's hard to handle each flow in such current QoS mechanisms. It's the same for the mobile IP carrier network.

We propose to employ IPv6 flow label to identify flows instead of bearers in mobile access, core and IP carrier networks.

A. Overview

In mobile network, the IP packets are transported in GTP Tunnel. Figure 4 shows the IP packet structure through GTP Tunnel.

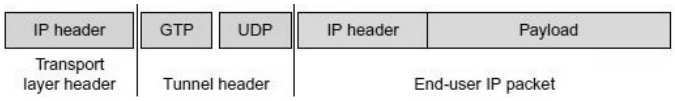


Figure 4. IP packet structure in GTP Tunnel

There are two IP headers, which correspond to the transport layer and the end-user IP packet respectively. That is, two IPv6 flow label fields can be handled by mobile network entities and IP carrier network entities respectively.

In our QoS scheme, the 20-bit IPv6 flow label is divided three parts from left to right: Access Point Name (APN) part, service part, and flow part. There parts are generated and stored by different entities. The flow label prefix definition can be adjusted by mobile operators according to their current situation.

The APN part (first several bits) of IPv6 flow label is defined and stored in Home Location Register/Home Subscriber Server (HLR/HSS) and each APN has one IPv6 flow label prefix. The lower APN part means higher QoS level.

The service part (middle several bits) of IPv6 flow label is provided by Application Function (AF), located in service platform, or GGSN/P-GW for the services without service platform, for example, Internet services. The lower service part means higher QoS level.

The flow part (rest bits) of IPv6 flow label is generated by terminal for different flows.

When the mobile terminal or mobile network launches a bearer, the APN part of flow label will be generated and sent from HLR/HSS to the terminal after the requested APN passes the authentication. When the terminal sends a service request to AF or GGSN/P-GW receives a service request to Internet, the service part of IPv6 flow label will be generated and sent to the terminal. At last, the terminal generated the flow part and forms the full flow label for each flow.

In mobile access and core network, Radio Network Controller (RNC), Service GPRS Supporting Node (SGSN)/S-GW and GGSN/P-GW entities process the packets according to the IPv6 flow label in the end-user packet header. The packets with lower IPv6 flow label prefix will have priority in processing. When the congestion happens, the packets with higher IPv6 flow label will be discarded at first and then the packets with same IPv6 flow label of discarded packets will be discarded. This mechanism guarantees that the service with high QoS level has priority and the impacted flows are minimal when congestion happens.

When the packets are sent the IP carrier network, the IPv6 flow label in end-user IP packet header will be copied to the transport layer header. The mobile IP carrier network will process the packets according to IPv6 flow label by the same way used in mobile network.

B. Entities Functions Related to Flow Label

- HLR/HSS: store the APN part of IPv6 flow label; and provide the APN part to the terminal after the requested APN passes the authentication.
- UE: generate the IPv6 flow label suffix for each flow; form the full IPv6 flow label with corresponding prefix provided by HLR; and be responsible for verifying the APN and service part of IPv6 flow label by TFT filters if necessary.
- AF: store the service part of IPv6 flow label; and provide the service part to the terminal after the service request is accepted by the service platform.
- RNC, SGSN/S-GW: process IPv6 flow label in end-user IP packet header; and copy flow label in end-user IP packet header to transport layer header and generate transport layer header.
- GGSN/P-GW: process IPv6 flow label in end-user IP packet header; copy flow label in end-user IP packet header to transport layer header and generate transport layer header; store the service part of IPv6 flow label for the services without service platform; provide the service part to the terminal when receiving a service request from terminal; and be responsible for verifying the APN and service part of IPv6 flow label by TFT filters if necessary.
- IP carrier router: process IPv6 flow label in transport layer header.

C. Process Flow of Flow Label Generation

The whole process flow related to IPv6 flow label generation is described in Figure 5, where only shows the UE-initiated services. First, UE initiates a bear establishment request to network entities. Then, according to the functions described in Section III B and the QoS profile of the service requested by UE, the network entities generate the corresponding parts of IPv6 flow label separately and provide them to the UE. Finally, the UE generates the flow part of flow label for each service flow and forms the full IPv6 flow label for communications.

D. Security Consideration

Because IPv6 flow label implies the QoS level, the security of flow label is vital important. The security problems come form two aspects: the flow label should be assigned according to the corresponding QoS level (e.g., QoS level of APN and service); and the flow label should be guaranteed not to be modified illegally during the transport.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

In our proposed scheme, the APN part and service part of flow label are stored in HLR/HSS, GGSN/P-GW and AF. These entities belong to the mobile operators and can be trusted. Mobile terminals can get the APN and service parts of flow label and have the chance to modify them by itself, for example, give a small flow label to a service with lower QoS level. For avoiding the IPv6 flow label prefix to be modified illegally by mobile terminals, special TFT filters is added to UE and GGSN/P-GW. These filters are responsible for verifying whether the IPv6 flow label prefix is matched to the corresponding APN and service part, which had been provided to terminal. If the mismatch thing happens, the corresponding packet will be discarded and a wrong message will be sent to the terminal.

The consistency and integrity of flow label during the transport can be guaranteed by security mechanisms used in current mobile network and IP carrier network, such as VPN, firewalls, Access Control List (ACL), etc.

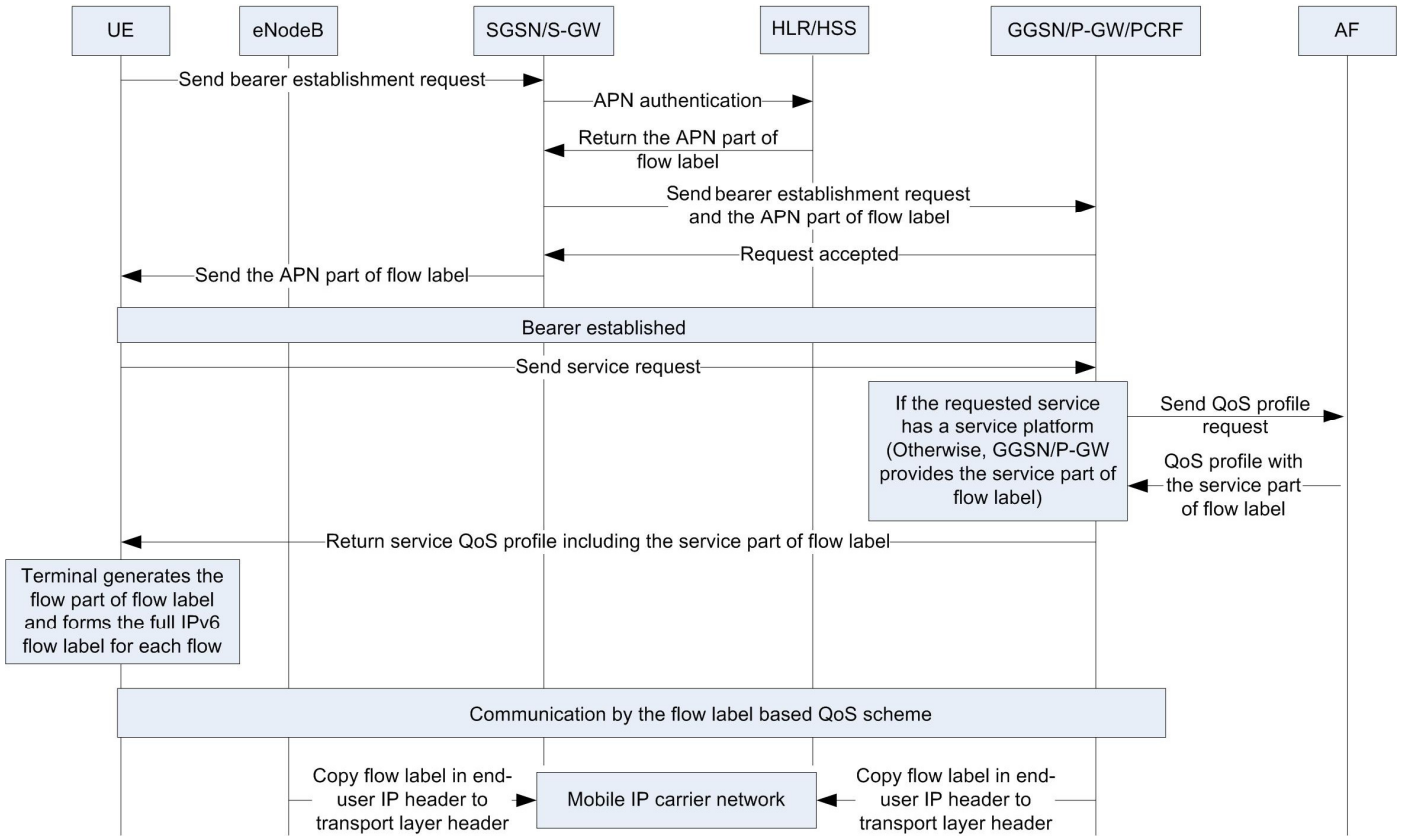


Figure 5. The process flow generating IPv6 flow label

IV. EXPERIMENT RESULT

In our experiment, the proposed flow label based scheme is evaluated by the simulation platform based-on OPNET software. According to the generation and QoS control methods described in section III, we programmed network entities functions on the simulation platform. The experiment we launched has two scenarios. One scenario tested different flows in a same PDP Context/EPC bearer transported in the mobile network, and another tested different flows with same QoS profile transported in the mobile IP carrier network.

The experiment platform includes mobile terminals, mobile network, mobile IP carrier network and external network. There are three mobile terminals and a simply mobile access and core network composed by NodeB, RNC, SGSN and GGSN. Four routers constitute the mobile IP carrier network. And the external network includes a switch, two File Transfer Protocol (FTP) servers and three VoIP clients (as the peer VoIP terminals). The topology of the experiment is described in Figure 6.

The services tested in this experiment platform are FTP downloading and VoIP, which have their own APN and QoS profiles. The QoS level of VoIP service is higher than that of FTP downloading. And three mobile terminals communicate with three peer VoIP clients respectively with same QoS level. For the FTP service, the QoS level of FTP server2 is higher than that of FTP server1.

The IPv6 flow label used in the experiment is defined as following. The first 2bits of IPv6 flow label belongs to APN part, and two of four combinations are assigned to APNs used in VoIP and FTP services. The next 2bits is allocated to service part, and VoIP service only occupies one combination and two

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

levels of FTP have two of combinations. The rest bits of flow label belong to the flows assigned by mobile terminals.

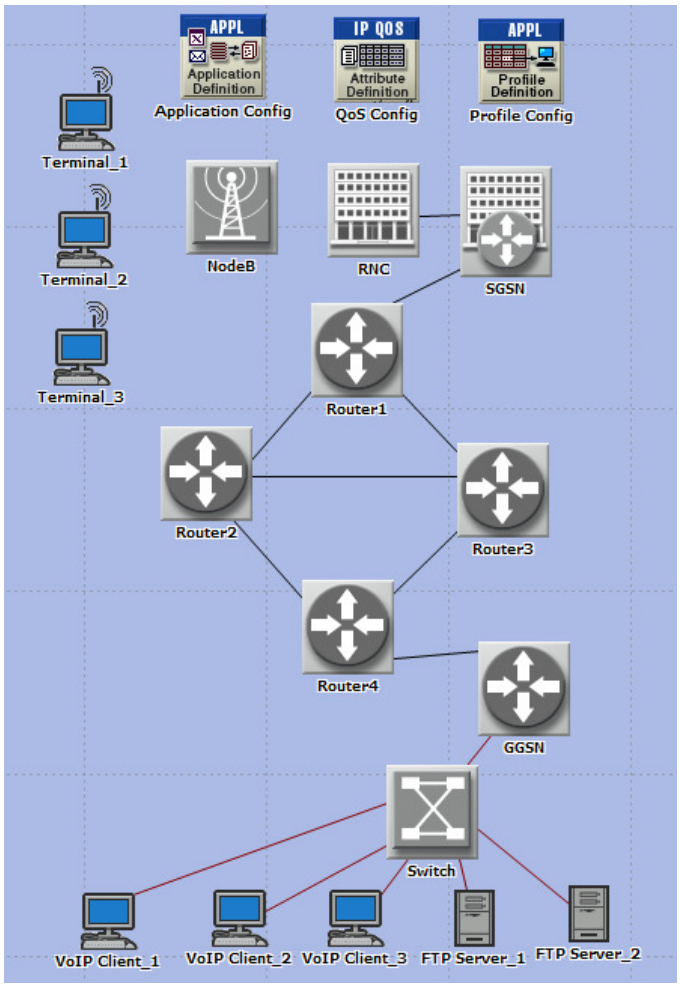


Figure 6. The network topology of the experiment

A. Experiment Results in Mobile Network Scenario

In this scenario, we tested different flows in same bearer. Mobile terminal1 established a bearer used for FTP downloading service. In this bearer, it communicated with two FTP servers respectively and the bandwidths were all 1.5Mbps. The packets of two flows arrived to all the interfaces randomly.

The bandwidth of all links was 10Mbps except for the link between SGSN and Router1, which was 4Mbps firstly and then reduced to 2Mbps for simulating network congestion.

We tested two cases: without and with the proposed flow label based QoS scheme. The test results are shown as following.

1) Case1 without the proposed QoS scheme: The Figure 7 shows the changes of two FTP flows after network congestion happened.

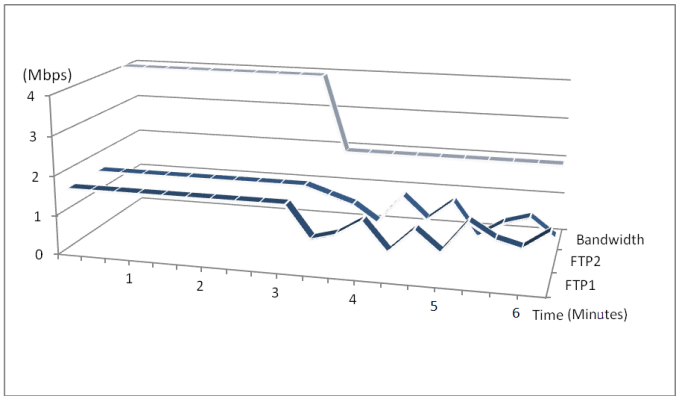


Figure 7. Test result in case1 of scenario1

We can conclude that the two FTP flows were all impacted due to the network congestion because that they were transported in a same bearer and the core network could not distinguish them and treated them with a same QoS policy.

2) Case2 with the proposed QoS scheme: The Figure 8 shows the change of two FTP downlodings after network congestion happened.

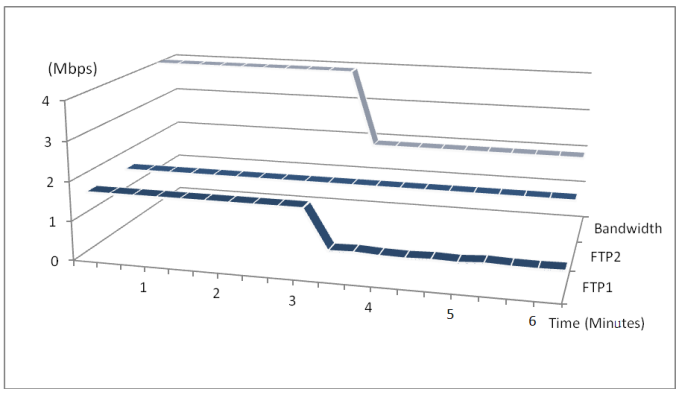


Figure 8. Test result in case2 of scenario1

In this case, we applied the flow label based QoS scheme. The core network entities could distinguish different flows by IPv6 flow label even they existed in same bearer and guaranteed the flow with high QoS priority. So when network congestion happened, the bandwidth of FTP2 was still in 1.5Mbps and that of FTP1 reduced to almost 0.5Mbps.

B. Experiment Results in Mobile IP Carrier Network Scenario

In this scenario, we tested different flows with same QoS profile transported in mobile IP carrier network. Three mobile terminals communicated with three peer VoIP clients respectively with a same QoS level. It supposed that the bandwidth of VoIP was all 1Mbps. The packets of three flows arrived to all the interfaces randomly.

The bandwidth of all links was 10Mbps except for the two links between Router1 and Router2, and between Router1 and Router3, which were all 2Mbps firstly and then the link

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

between Router1 and Router3 was down for simulating network congestion.

We tested two cases: without or with the proposed flow label based QoS scheme. The test results are shown as following.

1) Case1 without the proposed QoS scheme: The Figure 9 shows the changes of three VoIP flows after the link was down.

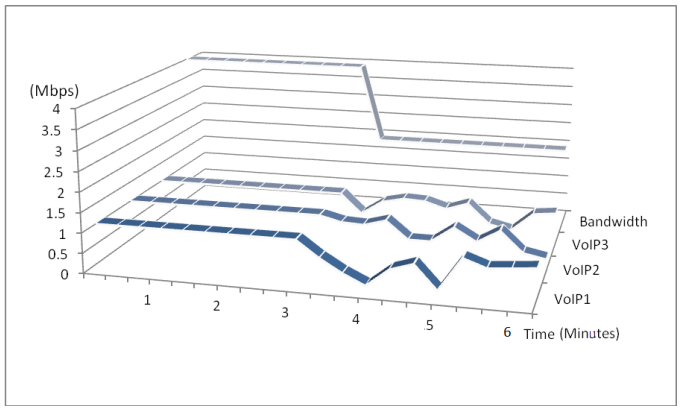


Figure 9. Test result in case1 of scenario2

For the test results, we can get that the three VoIP flows were all be impacted due to the network congestion because that they had same QoS level and the mobile IP carrier network could not distinguish them and treated them with a same QoS policy.

2) Case2 with the proposed QoS scheme: The Figure 10 shows the changes of three VoIP flows after the link was down.

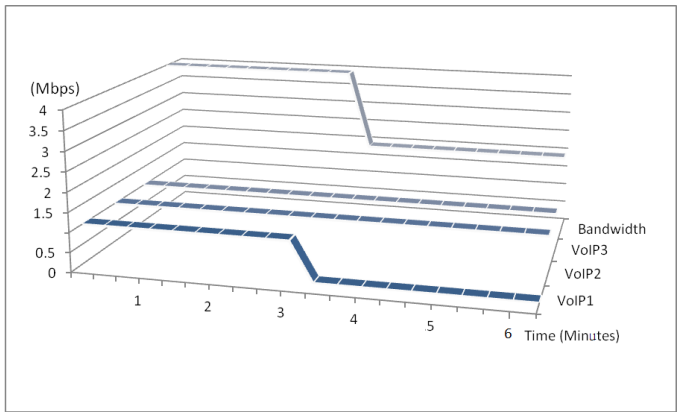


Figure 10. Test result in case2 of scenario2

In this case, we applied the flow label based QoS scheme in mobile IP carrier network. The carrier network entities could distinguish different flows by IPv6 flow label in transport layer header. When the link was down and the network congestion happened, routers could distinguish different VoIP flows and discarded the packets belonged to the same flow at first to minimize impacts to other flows.

V. CONCLUSION AND FUTURE WORK

In this paper, we investigated QoS architecture of mobile network and the current situation of IPv6 flow label. We proposed a flow label based QoS scheme, where flow label can be used to distinguish of packet flow and mobile network and IP carrier network can control the traffic more accurately and provide the finest granularity QoS. Experiment results show that our proposed QoS scheme is able to achieve better fine-grained control than existing ones.

Two services with several traffic flows are simulated on our experiment platform. In the future work we consider increasing the volume of traffic and new services. In addition we will research the performance of this scheme under mobility environment.

REFERENCES

[1] 3GPP TS 23.107, Quality of Service (QoS) concept and architecture, 2009.

[2] Dino Flore, LTE RAN architecture aspects, 3GPP IMT-Advanced Evaluation, 2009.

[3] 3GPP TS 23.401, General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access, 2009.

[4] J. Rajahalme, A. Conta, B. Carpenter, and S. Deering, IPv6 flow label specification, IETF RFC 3697, March 2004.

[5] S. Amante, B. Carpenter, and S. Jiang, Rationale for update to the IPv6 flow label specification, IETF Internet draft, July 2011.

[6] S. Amante, B. Carpenter, S. Jiang, and J. Rajahalme, IPv6 flow label specification, IETF Internet draft, July 2011.

[7] Q.Hu and B. Carpenter, Survey of proposed use cases for the IPv6 flow label, IETF RFC 6294, June 2011.

[8] B. Carpenter and S. Amante, Using the IPv6 flow label for equal cost multipath routing and link aggregation in tunnels, IETF Internet draft, July 2011.

[9] B. Carpenter and S. Jiang, Using the IPv6 flow label for server load balancing, IETF Internet draft, October 2011.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

Impacts of IPv6 on Robust Header Compression in LTE Mobile Networks

Daniel Philip VENMANI *†
Marion DUPREZ
*Orange Labs-
France Telecom R&D,
Issy les Moulineaux, France
danielphilip.venmani@orange.com
marion.duprez@orange.com

Houmed IBRAHIM
Yvon GOURHANT
Orange Labs-
France Telecom R&D
Lannion, France
houmed.ibrahim@orange.com
yvon.gourhant@orange.com

Marie-Laure BOUCHERET
†INP - ENSEEIHT Toulouse
Toulouse Cedex 7, FRANCE
marie-Laure.boucheret@enseeiht.fr

***Abstract*— LTE is an all-IP based 3GPP architecture, meaning that the transport in the network is based on IP, as post-Release 5 UMTS, unlike the former 3GPP architectures like GSM, Release 5 UMTS whose transport is based on ATM. Hence, fore-runners in this field have deployed with IPv4 as the basic protocol for addressing and transport, although the deployment of LTE is still in its initial phase and trial runs are executed by various operators. But findings and results prove that the exhaustion of IPv4 will not make it possible anymore IPv4 addresses for this new technology to take its full fledge. Hence, this led to the necessity of considering IPv6 as the protocol for addressing and transport. The primary reason to perceive IPv6 is its scalability feature meaning that it supports large address spacing. Now, with this in mind, when IPv6 is considered in the LTE architecture, the possible impacts on the network are investigated in depth in this paper. This is began by considering IPv6 in transport and application level in the different network entities in the LTE architecture like e-Node B, Serving-GW, PDN-GW and the transition impacts from IPv4 to IPv6 is analyzed. Based on preliminary empirical evaluation, our conclusion is that despite the fact that IPv6 offers large address spacing, the fact that the size of the IPv6 header is 20 bytes more than the header of IPv4, leads to complications as well.**

Keywords- IPv4, IPv6, LTE, ROHC.

I. INTRODUCTION

Next generation mobile communication systems are driven by demands that are expected to provide higher data rates and better link quality with the ability to support real-time and non-real time applications compared to the existing systems. The User Equipments (UE) nowadays are able to provide various internet applications and services that raise the demand for high speed data transfer and Quality of Service (QoS) and the dependency on Internet Protocol (IP) addresses [1] becomes a vital ingredient to rolling out services. Orthogonal Frequency Division Multiple Access (OFDMA) [2] and Single Carrier Frequency Division Multiple Access (SC-FDMA) [2] are strong multiple access candidates for the uplink of the International Mobile Telecommunications-Advanced (IMT-Advanced). These multiple access techniques in combination with the rise of the Mobile Internet will be utilized to reach the targeted IMT-Advanced system performance. Thus, IP capabilities are coming front and center for many operators. 3GPP [3] has responded to this trend with an all-IP core network called the Evolved Packet Core (EPC) and new packet-optimized access technologies like Evolved UMTS Terrestrial Radio Access Network (E-UTRAN), otherwise known as Long Term Evolution (LTE) [4]. LTE, whose radio access is called E-UTRAN, is expected to substantially bring improved user experience with full mobility. With the emergence of IP as the protocol of choice for carrying all types of traffic, LTE is scheduled to provide support for IP-based traffic with end-to-end QoS. With IP being the basic protocol for transport, the issue of Internet Protocol version 4 (IPv4) [5] exhaustion is considered to be pit-stop towards the implementation of LTE networks widely. Hence, the choice of adapting to another version of IP, which is Internet Protocol version 6 (IPv6) [6], [7], is deliberated to be the need of the hour.

With IPv6 in mind, in this paper, we analyze into various impacts on the LTE networks from a mobile network operators' point of view. The particularity of this paper is to analyze the impacts of IPv6 on the Radio Access Networks, i.e., to look into the impacts on the e-Node B in the LTE architecture. It gets deeper into the topic discussion where the analysis takes its root into the concepts of analyzing the impacts of IPv6 focusing on Robust Header Compression (ROHC) [8]-[12] comparing it with the existing IPv4 addressing scheme and looking into the possible changes that could occur when IPv6 comes into existence. The main conclusion of this paper is that, IPv6 besides its various numerous advantages also brings side effects that have to be taken care of. The issue of incompatibility between the existing IPv4 protocol and the IPv6 protocol could be one of the major problems that should be dealt with great care, besides considering the ways to handle the 20 extra header bytes of IPv6. Therefore, it must be considered in all phases of the development and deployment process by network operators and equipment vendors.

The rest of the paper is structured as follows. Section II elaborates our motivation to carry out research within this area stating briefly the impacts of IPv4 address exhaustion. It then describes the role of IP within the context of LTE/EPC networks, extending further identifying the role of IPv6 within LTE/EPC networks. Continuing further is section III which presents a brief state of the art on ROHC mechanisms, then the impact of IPv6 address on ROHC and finally the performance of ROHC with IPv4 and IPv6 packets are evaluated empirically, concluding the paper.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

II. NECESSITY FOR IPv6 IN LTE

An address goes through a number of stages on the path to deployment. Originally the address block is a parameter set of the underlying protocol, and the intended purpose of segments of the address space is described in an address architecture. The number of IPv4 addresses, while vast, is finite. The '/8' IANA blocks for Regional Internet Registries (RIRs) is 0% as of February 03, 2011 [13]. Subsequently, the addressing pool available to RIRs for assignment to Internet Service Providers is anticipated to run-out in the following 2-3 years. Most of the IPv4 address exhaustion mitigation strategies rely on network service providers to act as gatekeepers to selectively issue temporary IPv4 addresses to users. Allocating temporary addresses has technical problems, such as limiting users to existing applications. The impact of IPv4 address exhaustion includes policy issues, where it can be used in a predatory manner to keep competitive services out of the reach of a service provider's customer base. The third generation (3G) mobile network on its own easily consumed a majority of the available addresses. Running out of addresses does not mean the IPv4-based Internet will suddenly stop working. Nevertheless, it does mean it will be difficult, if not impossible, to distribute new IP addresses to new or expanding enterprises. Such a limitation will have clear impacts on commerce and innovation.

A. *Impacts of IPv4 Address Exhaustion*

There are two simultaneous approaches to addressing the run-out problem: delaying the IPv4 address exhaustion, and introducing IPv6 in operational networks. Delaying the public IPv4 address exhaustion involves assigning private IPv4 addressing for end-users, as well as extending an IPv4 address (with the use of extended port ranges). Mechanisms such as a Network Address Translator (NAT) and "A+P" [14] are used at the provider premises (as opposed to customer premises in the existing deployments) to manage IP address assignment and access to the Internet. In a mobile network, the IPv4 address assignment for a Mobile Node (MN) is performed by the Mobile Network Gateway [15]. In the 3GPP network architecture, this assignment is performed in conjunction with the Packet Data Network (PDN) connectivity establishment. A PDN can be understood to be the end-to-end link from the MN to the MNG. There can be one or more PDN connections active at any given time for each MN. A PDN connection may support both IPv4 and IPv6 traffic (as in a dual-stack PDN in 4G LTE networks) or it may support either one only (as in the existing 3G UMTS networks). The IPv4 address is assigned at the time of PDN connectivity establishment, or is assigned using the Dynamic Host Configuration Protocol (DHCP) after the PDN connectivity is established. This IP address needs to be a private IPv4 address which is translated into a shared public IPv4 address in order to delay the exhaustion of public IPv4 addresses as IPv6 is being deployed. Hence, there is a need for private - public IPv4 translation mechanism in the mobile network. In the Long-Term Evolution (LTE) 4G network, there is a requirement for an always-on PDN connection in order to reliably reach a mobile user in the All-IP network. If this PDN connection were to use IPv4 addressing, a private IPv4 address is needed for every MN that attaches to the network. This could significantly affect the availability and usage of private IPv4 addresses. Alternatively, the always-on PDN connection may be assigned with an IPv6 prefix (typically a /64) at the time of connection establishment, and an IPv4 address is assigned only on-demand (e.g., when an application binds to an IPv4 socket interface). This is feasible on the same (dual-stack) PDN in LTE networks (with short DHCP lease times), or with on-demand IPv4 PDNs. On-demand IPv4 PDN and address management can be effective in conserving IPv4 addresses; however, such a management could have some implications to how the PDN and addresses are managed at the MN. On the other hand, in the existing 3G UMTS networks, there is no requirement for an always-on connection (a 'link' from the MN to the MNG in 3G UMTS is referred to as a Packet Data Protocol (PDP) context/connection) even though many Smart Phones seldom relinquish an established PDP context. And, the existing (so-called pre-Release-8) deployments do not support the dual-stack PDP connection. Hence two separate PDP connections are necessary to support IPv4 and IPv6 traffic. Even though some MNs (especially the Smart Phones) in use today may have IPv6 stack, such a capability is not tested extensively and deployed in operational networks. Given this, it is reasonable to expect that IPv6 can only be introduced in the newer MNs, and that such newer MNs still need to be able to access the (predominantly IPv4) Internet.

B. *IP in LTE/EPC Mobile Networks*

The concept of Fixed-Mobile convergence is already on its verge of deployment. The primary reason for this is the use of the IP transport layer for both wired and wireless networks. These converged networks will be the building blocks for "All-IP Networks". As stated previously, LTE evolution calls for a transition to a "flat", all-IP core network with open interfaces, called the Evolved Packet Core or EPC. While the EPC has been defined in conjunction with LTE, it is an open next generation packet core for all networks, including 2.5G, 3G, 4G, non-3GPP, and even fixed networks. LTE network, slightly differing from the traditional architectures, with the base station controller (BSC) or radio network controller (RNC) integrated into the access or core layers in a dual network structure. Base stations which are e-Node B are connected to the EPC through IP, and services are accessed through gateways. The traditional circuit switched domain is removed and service access, bearing, switching, coordination, charging and control are packet domain and IP-based. Therefore, this leads to the so-called mobile network IP transformation to enable the traditional technologies to co-exist with the emerging IP-based LTE technologies. This IP transformation can be realized through three steps as follows:

- First comes the IP transformation of interfaces. IP transmission can be used between 3G base stations and BSCs. In this case, lease and construction costs are reduced in traditional time division multiplexer

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

(TDM) transmission, and sufficient bandwidth is provided for high-speed data services. In the GSM system, the IP transformation of A interfaces can reduce TransCoder (TC) and network costs, enabling TransCoder Free Operation (TrFO) and enhanced voice quality. Interface IP transformation has less impact on the entire network architecture and is easy to achieve.

- The second stage involves the IP transformation of the kernel. As the keys for mobile network IP transformation, prerequisites to avoid failure are strong network capabilities and a thorough knowledge of transmission and data communications. Data sent from a base station to the BSC through IP is not switched or decoded, but is transmitted to the core network directly through an IP switch. Highly-integrated digital signal processing (DSP) and multi-kernels can be applied to enhance equipment performance, reduce power consumption and save resources.
- The final stage describes the IP transformation of services. When network entities and the entire network are transformed to IP, service access can be simplified to a connection between servers and gateways. With the help of an OSS/BSS system, mobile network operators can deploy and manage telecom services just as Internet service providers run their Web services. The IP transformation of the mobile network is an important step for LTE All-IP and flat network architecture, and also a preparation for LTE network architecture.

C. IPv6 in LTE/EPC mobile networks

The considerations from the preceding paragraphs thus led to the following observations. First, there is a need to support private IPv4 addressing in mobile networks in order to address the public IPv4 run-out problem. This means there is a need for private - public IPv4 translation in the mobile network. Second, there is support for IPv6 in both 3G and 4G LTE networks already in the form of PDP context and PDN connections. Also, mobile Internet access from smart phones and other mobile devices is accelerating the exhaustion of IPv4 addresses. It goes without saying that to realize LTE, it needs IPv6.

III. IMPACTS OF IPV6 ON ROHC

As the networks evolve to provide more bandwidth, the applications, services and the consumers of those applications and services, all compete for that bandwidth. For network operators it is important to offer a high QoS in order to attract more customers and encourage them to use their network as much as possible. Hence among many, one of the advantages could be achieving higher average revenue per user (ARPU).

A. Introduction to IP Header Compression

In many services and applications like Voice over IP (VoIP), interactive games, multimedia messaging etc, the payload of the IP packets is almost of the same size or even smaller than the header. Over the end-to-end connection, comprised of multiple hops, these protocol headers are extremely important but over just one link (hop-to-hop) these headers can be compressed and must be uncompressed at the other end of the link. It is possible to compress those headers, providing in many cases more than 90% savings (described in Section IV), and thus save the bandwidth and use the expensive resource efficiently. Thus, IP header compression is the process of compressing excess protocol headers before transmitting them on a link and uncompressing them to their original state on reception at the other end of the link [16]. It is possible to compress the protocol headers due to the redundancy in header fields of the same packet as well as consecutive packets of the same packet stream. IP header compression thus provides a reduction in packet loss and improved interactive response time by compressing the IP headers. On low bandwidth networks, using header compression results in better response times due to smaller packet sizes, i.e., improved RTT values can be observed. A small packet also reduces the probability of packet loss due to bit errors on wireless links resulting in better utilization of the radio spectrum. It has been observed that in applications such as video transmission on wireless links, when using header compression the quality does not change in spite of lower bandwidth usage. For voice transmission, the quality increases while utilizing lower bandwidth.

B. ROHC Scheme

The compression mechanism for IP headers described in the previous section for IP are not considered robust because they do not perform well on links with high error rates and long round trip times like the wireless links and do not take into account that some applications may actually benefit that delivering packets with errors. Therefore, Robust Header Compression emerged from the need to standardize a single, solid and extendable header compression protocol that performed well over links with high error rates and long link round trip times, taking into account the problems shown by its predecessors. ROHC scheme uses window based least significant bits encoding for the compression of dynamic fields in the protocol headers. Due to its feedback mechanism, ROHC is robust on wireless links with high BER and long RTT. It can achieve compression up to 1 byte and thus it is more efficient than other compression schemes. Even though it is complex compared to earlier schemes, it is suitable for wireless networks, which use the very expensive radio spectrum resource.

C. ROHC mechanism

The fundamental challenge in header compression for transmission over wireless links is to maintain the correct context at the decompressor in the face of quite frequent bit errors in the received packets. ROHC supports three different modes for maintaining the context in different wireless systems [9]. The unidirectional mode is designed for systems without a feedback channel from the decompressor to the compressor, i.e., where the decompressor can not acknowledge the correct receipt of context information. To

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

overcome this limitation, the compressor periodically retransmits the context information. The bidirectional optimistic mode and the bidirectional reliable mode are designed for systems with a feedback channel from the decompressor to the compressor, i.e., where the decompressor can acknowledge the correct receipt of context information and/or send negative acknowledgements to request the retransmission of context information. With the bidirectional optimistic mode, bit errors in the compressed header are detected with a 3-bit cyclic redundancy check (CRC) code. When the CRC check fails the decompressor generally discards the affected packet and attempts to repair its context either locally or by requesting a context update from the compressor. The reliable mode extends the optimistic mode by a more complex error detection and correction which uses a larger number of coding bits. Fig. 1 illustrates ROHC mechanism.

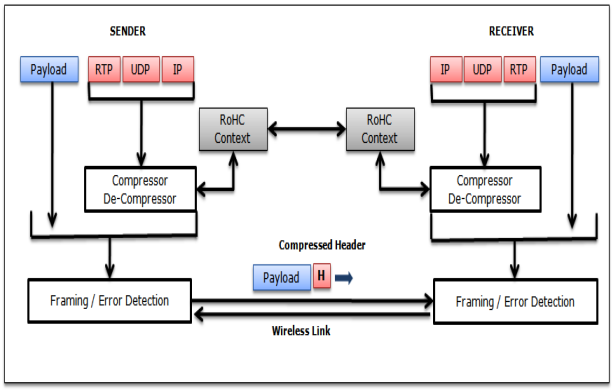


Figure 1. ROHC mechanism

The ROHC compressor replaces the RTP/UDP/IP headers by its own, much smaller header. On the receiver side the decompressor transforms the ROHC header into the original protocol layer headers. A step by step illustration of header compression using ROHC is shown in Fig. 2. A multimedia stream packet composed for an IP network transmission typically consists of a 20 byte IP header (considering it to be an IPv4 address), an 8 byte UDP header, and a 12 byte RTP header. The IPv6 version requires a 40 byte IP header, so the total RTP/UDP/IP header size can sum up to 60 bytes. When voice frame is an audio AMR codec of 12.2 Kbps, it travels on RTP protocol over UDP, besides themselves being very small. Payload is 20 to 60 bytes with a RTP/UDP/IP header of 40 bytes (IPv4=20 bytes; UDP=12 bytes; RTP=8 bytes). Then due to the high relation between header size and payload size, the transmission of VoIP packets is not an efficient process. Being VoIP a protocol to service a playback application (voice playback) its maximum end to end delay should be less that 150-200ms; where 150ms is considered to be the best optimal value. This is to guarantee the good quality of the sound to be transmitted. The efficiency of transmission is low. For transmitting 20-60 bytes, a header of 40 bytes is needed, that results in a relation of 200% to 26.67%. It must be noticed that, because of this significant header's size, the necessary throughput for a VoIP call would be 28.8 kbps (with IPv4) or 36.8kbps (with IPv6) whereas the current service of voice call in circuit switch domain needs a throughput of 12.2 kbps. Thus, the support of this type of packet corresponds to a waste of radio resources and implies the need of performing a compression of RTP/UDP/IP header to reduce the ratio between header and payload's sizes and consequently the necessary throughput, which is carried out by ROHC.

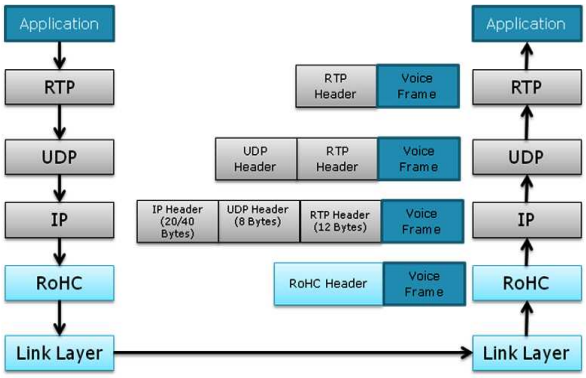


Figure 2. ROHC in a protocol stack

D. *Impact of IPv6 on ROHC in LTE*

While considering the different protocol layers and their functions in the LTE architecture, in the user-plane, the Packet Data Convergence Protocol (PDCP) layer is responsible for compressing and decompressing the headers of user plane IP packets using ROHC to enable efficient use of air interface bandwidth. PDCP specification applies header compression between the e-Node B and the UE in Release 8 onwards. This is depicted in Fig. 3.

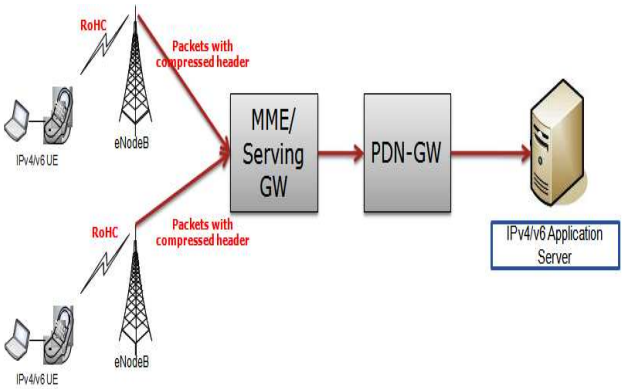


Figure 3. ROHC applied in LTE architecture

Incorporation of header compression between two nodes places restrictions on underlying link layer. IP payload length must be inferred from the underlying link layer. Decompressor must receive the packets in the same order that the compressor sends them. Packets are not duplicated by the link layer between the compressor and decompressor. Header Compression requires extra resources on nodes that

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

instantiate the compression algorithms. This can include additional memory required on nodes for storage of context information and also additional processing required on nodes for compression and decompression of packets. As ROHC always resides above the link layer, the other Internet components do not notice the usage of a compression scheme, but the wireless service provider can take advantage of a significant reduction of the required bandwidth. ROHC requires from the link layer that the packets are sent in a strictly sequential order. Also the packets are not allowed to contain routing information (single hop restriction). As the current LTE architecture implements ROHC with IPv4, the implementation of IPv6 introduces concerns related to expanded packet headers as the size of packet header doubled from 20 Bytes (IPv4) to at least 40 Bytes (IPv6). In addition to the above mentioned case, the incorporation of network-layer encryption mechanism which includes Internet Protocol Security (IPSec) nearly doubles IP operational overhead. Hence, the methods that reduce this expanded overhead will increase user throughput and/or the number of users a network can support.

IV. PERFORMANCE EVALUATION

An empirical evaluation considering three different types of packets to analyze the effect of overhead and to illustrate clearly the use of Robust Header Compression is performed in this section. For the sake of illustrating the same, let us consider an uncompressed IPv4 packet, uncompressed IPv6 packet and another packet that is compressed using ROHC mechanism.

1) Considering a Payload that could be the size of the maximum allowable MTU size of the network.

a) Case 1: Uncompressed IPv4 packet with an IPv6 payload.

Here, the case deals with the scenario when the UE tries to send a packet with an IPv4 header and with an IPv6 payload of size of 1442 bytes (here 1442 bytes is considered as the MTU for the payload since the total payload of the whole packet should sum upto 1500 bytes). Also, the packet is encapsulated with a RTP and UDP header. Fig.4 below illustrates this pictorially.

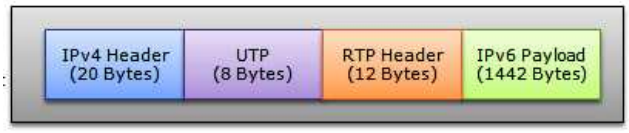


Figure 4. Uncompressed IPv4 packet with an IPv6 payload

Overhead = Total Header bytes / Total bytes transmitted = (40 / 1442) * 100 = 2.77%

b) Case 2: Uncompressed IPv6 packet with an IPv6 payload.

Here, the UE tries to transfer a packet with an IPv6 header with an IPv6 payload. Hence, the packet format and structure looks like in the Fig. 5.

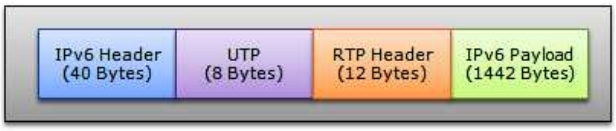


Figure 5. Uncompressed IPv6 packet with an IPv6 payload

Overhead = Total Header bytes / Total bytes transmitted = (60 / 1442) * 100 = 4.16%

c) Case 3: Packet compressed using ROHC mechanism with an IPv6 payload.

Assuming robust header compression algorithm is applied to the RTP/UDP/IP header, overhead calculations can be made:

Overhead = Total Header bytes / Total bytes transmitted = (2 / 1442) * 100 = 0.14%

2) Considering a VoIP Payload

Here, inorder to illustrate effect of ROHC in the voice packets, we consider VoIP payload with audio codecs. The chosen codec is AMR with the mode corresponding to a throughput of 12.2kbps. Each 20ms, a 32 bytes long packet is generated by the codec. After encapsulation by the protocols described before, the packet to transmit on radio interface is much bigger than the initial AMR payload packet. Hereafter is illustrated the overhead introduced with all these encapsulations.

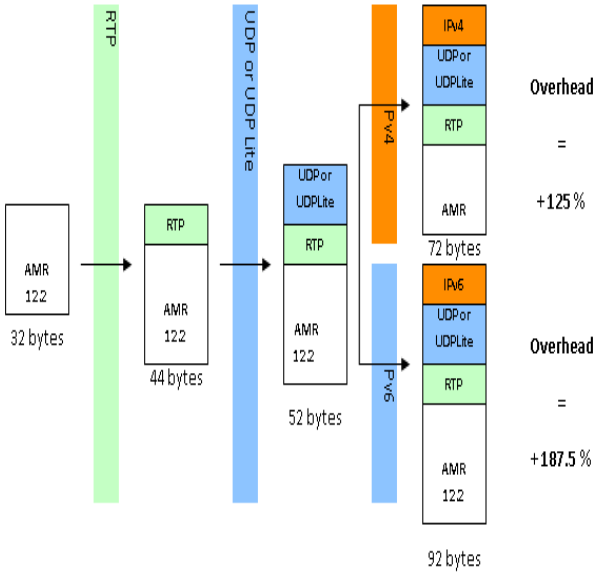


Figure 6. Overhead introduced with VoIP with AMR 12.2 Kbps

a) Case 1: Uncompressed IPv4 Header with an VoIP payload.

Overhead = Total Header bytes / Total bytes transmitted = (40 /32) * 100 = 125%

b) Case2: Uncompressed IPv6 Header with an VoIP payload.

Overhead = Total Header bytes / Total bytes transmitted = (60 /32) * 100 = 187.5%

c) Case3: VoIP Packet Compression with ROHC.

Overhead = Total Header bytes / Total bytes transmitted = (3 /32) * 100 = 9.375%

Therefore, to meet the requirements in terms of delay, jitter and latency for interactive communication, like VoIP, where delay is also caused due to sampling and packetization, the transmission must be minimized. When using the Real-time Transport Protocol (RTP) and IP/UDP/RTP headers for encapsulating voice samples, the ratio between the IP/UDP/RTP header portion and payload size is typically 2:1 and 3:1 for IPv4 and IPv6 respectively. Using ROHC in such scenarios will increase the wireless capacity by the factor of 3 and 5 for IPv4 and IPv6 respectively. The table below summarizes the different possible compressed gain values that can be obtained from various combinations of the headers calculated empirically as above.

TABLE I. HEADER COMPRESSION GAINS

Protocol Headers	Total Header Size (Bytes)	Minimum Compressed Header Size (Bytes)	Compressed Gain %
IPv4/TCP	40	4	90
IPv4/UDP	20	1	96.4
IPv4/UDP/RTP	40	1	97.5
IPv6/TCP	60	4	93.3
IPv6/UDP	48	3	93.75
IPv4/UDP/RTP	60	3	95

Although, from the results above we can conclude that the percentage compressed gain for IPv4 header is greater when compared to the compressed gain of IPv6 header, ROHC yields benefits in both IPv6 and IPv4. Infact, there are greater benefits with IPv6 due to a fixed-size header and static fields leading to even better compression efficiency gains. For e.g., a typical RTP/UDP/IPv4 has static fields of 25 Octets and dynamic fields of 15 Octets and a typical RTP/UDP/IPv6 has a static fields of 49 Octets and dynamic fields of 11 Octets. Therefore, IPv6 compressed headers are smaller than IPv4 compressed headers, as only fewer octets are dynamic. Additionally, there is no fragmentation in the network with IPv6 in the presence of path MTU discovery in IPv6, making every datagram compressible in IPv6. It should be noted, however, that smaller packets offer a smaller target for bit errors. So the packet loss rate for any compression method should be lower than for uncompressed packets. So if ROHC has only a very small probability of loss of sync between compressor and/or decompressor state machines, there should be a small reduction in overall packet loss rate between applications. This is a minor effect. The main purpose of ROHC is just to not increase the packet loss rate between applications.

V. CONCLUSIONS

The evolution from existing networks to LTE should be a smooth and gradual process through mobile network IP transformation. The most interesting question to be addressed here is, is this shortage of IPv4 addresses, a problem or an opportunity? Hoarding IPv4 addresses and postponing IPv6 deployment means that the countries laying away IPv4 risks becoming an island in the global next-generation Internet. Our preliminary results here show that, inorder to facilitate quick prototyping and rapid implementation, it is very important to consider even small details such the impact of ROHC within LTE. Our evaluation identifies the global addressing problem in the context of efficient network utilization that helps network designers and vendors to carefully design their system. With IPv6 widely seen as crucial for the continued operation and growth of the Internet, it is critical in mobile networks in particular. It goes without saying that to realize this vision, LTE needs IPv6. IPv6 is a minor aspect in the big LTE scheme of things but is essential for its success as a truly global and pervasive means of communications.

REFERENCES

[1] An Architecture for IP Address Allocation with CIDR, RFC 1518.

[2] H.Holma and A.Toskala: LTE for UMTS: OFDMA and SC-FDMA based radio access, John Wiley and Sons, Jun 2, 2009 - Technology & Engineering.

[3] http://www.3gpp.org/ , available online as on January 2012.

[4] Universal Mobile Telecommunications System (UMTS): LTE: Architecture and functional description (3GPP TS 36.000 series, Release 9)

[5] Internet Protocol DARPA Internet Program Protocol Specification, RFC 791.

[6] Internet Protocol, Version 6 (IPv6) Specification, RFC 2460.

[7] Internet Protocol Version 6 (IPv6) Addressing Architecture RFC 3513.

[8] E. T. David , H. Andreas, D. Andreas, and D. Gero, “Robust header compression (ROHC) in next generation network processors,” IEEE/ACM Transactions on Networking (TON), vol. 13 , August 2005..

[9] The RObust Header Compression (ROHC) Framework, RFC 4995.

[10] RObust Header Compression (ROHC): A Compression Profile for IP , RFC 3843.

[11] G. Boggia, P. Camarda, and V.G. Squeo, “ROHC+: A new header compression scheme for TCP streams in 3G wireless systems,” IEEE ICC'02, pp.3271-3278, April 2002.

[12] http://datatracker.ietf.org/wg/ROHC/charter/, available online as on January 2012.

[13] http://inetcore.com/project/ipv4ec/index_en.html, available online as on January 2012.

[14] IP Network Address Translator (NAT) Terminology and Considerations, RFC 2663.

[15] www.ietf.org/html/draft-ietf-v6ops-v6-in-mobile-networks, available online as on January 2012.

[16] W. Yichuan, L. Sun, and M. Jian, "Adaptive robust TCP/IP header compression algorithm for 3G wireless networks," IEEE Wireless Communications and Networking Conference, WCNC 2004, vol. 2, pp. 1046 - 1050, March 2004.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

Summaries of Lecture Recordings Used As Learning Material in Blended Learning

Sari Mettiäinen, Anna-Liisa Karjalainen
Tampere University of Applied Sciences
Tampere, Finland
sari.mettiainen@tamk.fi,
anna-liisa.karjalainen@tamk.fi

***Abstract*-In order to enhance learning, it is necessary to develop flexible learning methods in conjunction with traditional classroom teaching. Information and communications technology has developed enormously during the past years and but its potential as a teaching aid has not yet, however, been fully exploited in education. This paper describes the opinions of Finnish nursing students' on the usefulness of lecture recordings in nursing education at Tampere University of Applied Sciences. Faculty from the nursing education degree program of Tampere University of Applied Sciences produced thirty-one 10-62 minute lecture summaries or course content syntheses in five subjects. They were produced by four teachers and listened by 4 student groups in 2010-2011. In the study, 93 students completed a questionnaire, with a response rate of 52%. According to the results of the questionnaire, slightly more than half of the students listened to the recordings via computer and slightly less than half listened as podcasts on mobile device. Students stated that the recordings helped them to understand the key content of lectures and that they were of additional value to learning process when compared to the traditional PowerPoint teaching slides used in contact lectures. Opinions varied, however, as to whether students learned as much by listening to the recordings as they did by attending contact lectures. The lecture recordings enabled better time management, but the recordings did not, however, significantly affect the time used for preparing for examinations or for participating in contact lectures. The students found listening to the recordings an interesting experience and most of them would also like to listen to them also in the future. In a conclusion, the results of this project show that offline recorded lecture summaries are a good support for education either as additional course material or as a partial substitute for contact teaching. The lecture recordings support students' learning and are an agreeable learning method for students.**

Keywords-m-learning; podcast; lecture recording; lecture summary.

I. INTRODUCTION

In this paper, we describe the results of our pilot project which started with the support of EU-funding and where we tested the possibility of using summarized lecture recordings in order to enhance blended learning.

The purpose of the empirical study was to find out how nursing students used lecture recordings, as well as their experiences of mobile learning (m-learning) and usefulness of recordings in education.

The recording of contact lectures in a summarized form and then providing students with the recordings may promote learning, The teacher can use the recordings with several groups and, thus, save resources for the supervision of learning. More and more students are working and studying, and therefore, it is necessary that accessibility to higher education is expanded. One way to achieve this is to offer students video and audio recordings to help them in time management.

In the following sections, we will describe how lecture recording technology is currently used and how the use of lecture recordings in education has been described in literature. We will then present a case study on student views and its findings. The conclusions section presents a short summary of the pilot project and the most significant findings compared with the findings of earlier literature.

II. EDUCATION TECHNOLOGY CURRENTLY USED

It is necessary to develop flexible learning methods together in conjunction with traditional classroom education. The possibilities offered by the information and communications technology that has been developed have not yet been fully exploited in education. Tampere University of Applied Sciences has piloted the use of lecture recordings as support for education. The aim was to develop a mobile and practicable model.

The cost of IT support will rise to a high level if free programs are used for lecture recording and IT staff is needed at diverse phases of the process. As part of the project we acquired the Echo360 system [1]. A disadvantage of the system was its relatively high price but an advantage was the easy process. The Echo360 system includes two lecture-recording tools. One is the classroom model, where the system records the contact lectures using either the AdHoc or timing method. The other is the Personal Capture program, in which case the teacher records the lecture independently on the computer with the client and headsets. The teacher does not need to worry about the technology in either model. The system transfers the recordings to a server

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

in three formats: as podcast, vodcast, and flash video. The system makes the links and, thus, IT support staff is not needed [1].

Podcasting means the distribution of audio and video files at the digital format via Internet. By means of podcasting feeds, sound and video recordings can be ordered to a media player program such as iTunes. It is possible to synchronise podcasts from the computer to a mobile device such as an MP3 player or mobile phone. The term podcasting emerged from the use of Apple's portable audio player the *iPod* [2].

Currently, Tampere University of Applied Sciences provides podcast feeds and video links only for own students in the course management system. The student can select the link or feed according to the terminal used. The student can watch lectures on the computer or transfer podcasts to a mobile device as video or sound recordings (vodcast and podcast).

III. LECTURE RECORDINGS IN BLENDED LEARNING

Lecture recordings can have different roles in education. The most common use is to provide recordings of complete lectures for the purpose of review and revision (substitution use) when they can be used to enhance student understanding of the course content. The second most common use is to provide additional learning material, summaries of lectures or syntheses of course content, to broaden and deepen student understanding (supplementary use) [2].

Lectures can be created in many different ways. Poor lectures can leave students bored and frustrated, but good lectures can inspire [3]. Isaacs [4] observed that lecturing is often characterised by the transfer of the lecturer's notes to the students' notepads without any thinking about or processing of the information. Anyway, the lecture should be more than just the transfer of information. The lecture should inspire and enthuse students, assist them in understanding the complex material and guide them through the topic [5]. Effective lectures can provide the excitement of intellectual discovery through the presentation of challenging and provocative ideas. The lecturer can relate the lecture content to the students' prior knowledge and relate it to real life examples thus, making the knowledge more meaningful [6].

Recordings can trigger memories of the lecture that aid understanding better than just looking at the notes. Listening provides variety for traditional reading. Knowing that there is a possibility to “go back” at any time was a great idea according to students even if the students did not use the files as much they had planned [7].

Student access to audio recordings during assessment times might also help to reduce failure rates. The recordings could help students to understand at topic when it is difficult to take in new concepts and methods the first time they are presented or if students miss the lecture for one reason or another. They can also free the student from having to make notes during the lecture [7].

The mobile device offers the possibility to study based on individual needs and to download the materials needed in the student’s own learning process. The benefit of mobile devices is that they can be used in on-the-go situations, such as driving a car, travelling by bus or train, walking, and exercising. Studying can be practised outside the classroom along with other activities [8]. When the student can concentrate on learning at a suitable time, it may increase the motivation to study.

Learning demands a lot of work and concentration from the student. Because of this, learning in on-the-go situations also includes dangers. The learner may pay attention to other things rather than learning at times and the learning process can becomes susceptible to diverse distractions [9]. However, the student’s thoughts may also become distracted in the classroom.

Even if lecture recordings make it possible to study in different places, the student can also choose the traditional way of learning. According to Lee and Chan’s study [10], students treated podcast listening as a formal learning activity that needed undivided attention and concentration within a designed study location, e.g. at home and not as m-learning while doing other tasks.

In most cases, lecture recordings include multimedia content such as audio, video, and visual aids [7]. In Brittain et al.’s [11] study 66% of students (N=70) preferred the audio-only format instead of the video file or audio synchronized with PowerPoint slide images when all the recordings were exported and available in those three formats. Of those students who used media files, 75% listened to them at home. In spite of this, the results clearly indicated that students preferred the mobility of audio recordings rather than video. Most of the students tended to download the files close to the relevant examination date (44.4 %), one fourth as soon as they were available, and the rest of the students infrequently [11].

Huntsberger and Stavitsky [12] found that 40% of students used podcasts as a replacement for textbooks. This replacement of textbooks with podcasts challenges teachers to create recorded material that does not provide too much detail in order not to make books or other core texts seem unnecessary.

On the whole, it is challenging to make meaningful lecture recordings. Attention has to be paid to the education content because lecture recordings must not be too extensive and the content must be short and independently related to the wider entity to be learned [8]. The appropriate length of lecture recordings is 20-30 minutes. According to feedback given by students, it would be better to record content some other time that during the lectures or they would need editing because it is confusing when there is a discussion in class on some topic that lasts for several minutes [7]. Do summarized lecture recordings better meet the needs of the students?

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

IV. THE EFFECTS OF LECTURE RECORDINGS ON LEARNING RESULTS

Does the availability of lecture recordings have an impact on lecture attendance? It is a question that needs long-term research before it can be answered. According to the findings of von Konsky et al. [13], students showed a tendency to listen to the recording of a missed lecture but the students who achieved a high grade tended to supplement lecture attendance with recordings more than students who achieved a low grade. According to Balfour's [7] study, there was no consistent relationship between audio file use and the end-of-module examination. However, students liked audio recordings and gave them a valuable role in the learning process. The use of lecture recordings supports the learning of some students, but there is also a large variation in other successful learning patterns [13].

Foley [14] used lecture recordings before the contact lessons in advanced computer science courses. He found that the group of students who listened to the lectures via a laptop or mobile device before attending the lecture performed about 10 percent better on the test than the group that only had the traditional classroom lecture. The lecture content, homework, and examinations were the same for both groups. Foley thought that there was more time for meaningful discussion in the classroom once the lecture was out of the way [14].

In Brittain et al.'s [11] study 85% of the dentistry students who used lecture recordings thought that the recordings had a positive effect on their grade. In total, 91% of the students used lecture recordings to review the lectures they had already attended.

According to a student survey (N=687) in the USA, the most popular responses to the question: "How did the recorded lectures benefit you?" were that they helped students to review the classroom material, helped students to prepare for examinations, clarified confusing topics, and allowed the students to learn independently. Other responses were that they improved the overall learning experience, helped to use time more efficiently, and were an alternative to attending class. Of the respondents, 81% watched the recordings once a week and they agreed that watching the recorded lecture increased their understanding of the topic. A total of 88% of the respondents said that they would like to have more lecture recordings in their courses [15].

V. CASE STUDY

Finnish nursing teachers at Tampere University of Applied Sciences have used the Echo 360 system since the fall of 2010. During the first year, four lecturers recorded their lectures with Echo360 Personal Capture software. The topics of the lectures were anatomy and physiology, the basics of cardiology and vascular nursing, social policy, nursing administration, and thesis methodology. The number of the recordings was a total of 31. The length of the recordings varied from 10 minutes to 62 minutes. Most of the lectures included Power Point slides and the lecturer's voice explaining the topic. The teachers planned their lectures to form short entities and recorded them without an audience in the classroom.

The lecture recordings were integrated with formal learning courses for full-time students. The students had some contact lectures in each course and some of the topics were taught totally using virtual methods with the lecture recordings. Students had time for listening in their schedule and the links to the recordings were available in the e-learning platform Moodle. The students had the possibility to listen to the recordings on a computer or on a mobile device as podcasts. They were given guidelines for downloading the links to iTunes and the mobile device.

After the examination all the students were sent a link to the survey questionnaire by email. They were told that the aim of the survey was to evaluate a new way of learning by using ICT technology. The students were told that answering the survey was voluntary and they were able to answer anonymously. The survey consisted of 14 structural quizzes and three open ended questions.

The quantitative data of the structural quizzes were analysed using frequency distributions, and the results were described using percentage distributions.

VI. FINDINGS

A. *Information on the respondents*

The recordings were available for four nursing student groups. In all, 179 students listened to the recordings. The survey questionnaire was answered by 93 students, and the response rate was 52%. Most of the students (86%) were first-year nursing students and the rest were third-year students. Most of them were under 25-year-old females.

B. *Students' recording listening habits*

Most of the students listened to the recordings only by computer (55%) and 42 % of them used podcasts. A total of 3% of the students did not listen to the recordings at all. The majority of the students felt that it was easy to download the lectures to podcasts with the instructions given, but for a fifth of them downloading sounded so complex that they did not even try. Some 17% of the students could download the recordings without prior instruction.

Of the respondents, 42% listened to the recordings at home, a third while jogging or walking, and 13% while driving. The recordings were also listened to while doing housework, during free periods at school, in the gym, and while sunbathing. While listening to the recordings, 8% of

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

the students made notes, 70% followed the slides during listening, and 19% just listened to the podcasts.

Of the respondents, 53% listened to the recordings during the examination week and 37% as soon as they were published. Some 62% said that they liked that they could listen to the lecture recordings when they wanted. A further 14% experienced difficulties in independently finding time to listen to the lectures, and 9% said that they had wanted the listening to be more scheduled.

C. Student opinions on the usefulness of lecture recordings

Of the students, 69% agreed fully that the recorded lecture brought additional value compared to viewing only PowerPoint slides. A total of 84% of the respondents agreed fully or partly that the lecture recordings helped them to better understand the key contents compared with reading a book. Student opinions varied on whether they learned as much by listening to the lecture recordings as they did by attending the contact lectures (Figure 1).

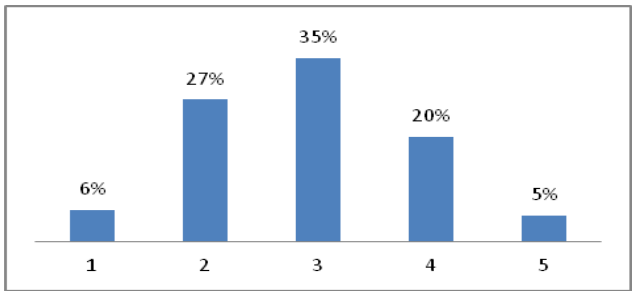


Figure 1. "I learn as much by listening to lecture recordings as by attending contact teaching lessons. "1. Fully disagree – 5. Fully agree".

Two-thirds of the respondents stated that they had used the recordings to review difficult matters. A third of the students said that they had read the textbooks less when the lecture recordings were available. The respondents felt that the use of the lecture recordings had not affected the time used for preparing for the examinations. A fifth of the respondents said that they attended fewer lectures when the lecture recordings were available. Half of the respondents thought that the use of the lecture recordings enabled working and studying at the same time, and two thirds stated that the recordings enabled a combination of studying and family life. A third of the respondents considered that the use of the lecture recordings enabled the completion of several courses at the same time (Figure 2).

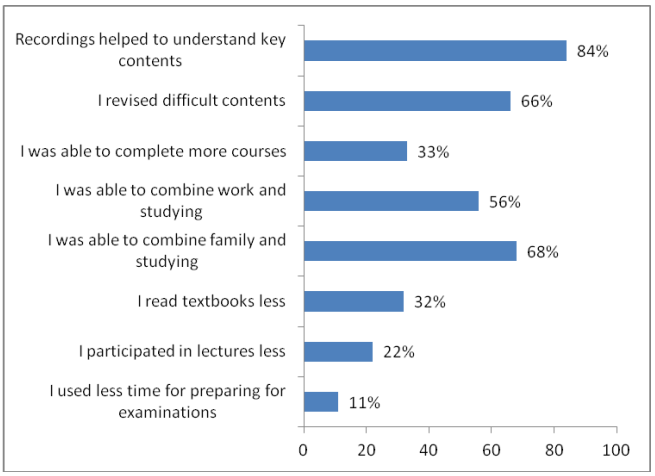


Figure 2. Students' perspectives on the usefulness of summaries of lecture recordings.

The majority of the students considered the lecture recordings as interesting, of suitable length, and a nice variety for the studying (Table 1). Of the respondents, 73% wanted to listen to lecture recordings also in future and the rest would possibly like to listen to them. No one answered that they would not want to have lecture recordings anymore.

TABLE 1. STUDENTS' OPINION OF LECTURE RECORDINGS

Opinion	*%*
Interesting	55
Suitable length	76
Nice variety for the studying	80

VII. CONCLUSION

This paper studied the use of teachers' lecture summaries as education material, not the recording of authentic classroom lectures. The lecture summaries partly compensated for contact teaching lessons and partly were support material for education. When students listen to a lecture as homework it provides a possibility to create a lecture in a new more active way that was piloted by at least one lecturer with good feedback. When offline recorded lectures are combined with other homework, it is possible to create a distance learning method. However, it is important to consider the total amount of the students work and be realistic how much the students are able to study independently. Lecture recordings are an effective teaching and learning method from the lecturer's point of view. The learning material can be easily shared with several groups once it has been produced.

The students stated that the lecture summaries were of an appropriate length (10-62 min) and interesting. According to the results, the recordings helped students to understand the

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

key contents and that the students used them for revising the difficult contents. These results were also found in Balfour's study (2006) [7]. The majority of the students listened to the recordings during the examination week, as the students did in the study of Brittain et al. (2006) [11]. Earlier studies have studied the connection between listening to lecture recordings, the learning results, and the course grades (Brittain 2006 [11], von Konsky 2009) [13], but this connection was not studied in this paper, as it is affected by many concurrent factors, such as the teacher, time, other studies, and the development of educational contents.

Half of the students listened to the lecture recordings on the computer and at home concentrating on studying only, and not as m-learning on-the-go situations even if it was possible, as shown by Lee & Chan's results (2007) [10]. Further study is required to see if m-learners felt studying was different than those who listened via computer.

Mobile learning is new for students and it takes time to learn new studying methods. This problem probably concerns attitudes more than skills because students felt that the used technologies were easy to use even for those who had not experience of podcasts beforehand. Students also have different learning styles, as was shown by the varied opinions on whether they learned as well by listening to the lecture recordings as they did by attending the contact teaching lessons.

Many students work and have a family and, thus, it is important to develop new blended learning methods. Lecture summary recordings as learning materials are one way to develop the education contents. Students have experienced them as being meaningful and useful and they support at least some of the students' learning.

In our case some lecture recordings were used to replace the contact lessons. This method can save the time of the lectures and the students when they do not need to travel to the campus. In Finland we have quite long distances to some campuses of the University of Applied Sciences. It is need to create some activities how to motivate students to concentrate studying, as according to findings some of the student said they had wanted the listening to be more scheduled.

A further challenge is to study what makes a good podcast, what elements do students like and what helps them learn? Another challenge is to study teachers' experiences of constructing lecture recording summaries and using them in education. The copyrights of the lecturers are a challenge question to solve in Finland. We do not have yet a culture as some universities for example in America have, where professors are willing to share their lecturers in Open University. On the other hand, if recordings are produced in Finnish language they are relevant only for the students in Finland. However, we believe in future we are more open-minded and able to share our material more.

REFERENCES

[1] Echo 360. http://echo360.com/blended-learning/instructors/. <retrieved: January, 2012>.

[2] O. McGarr. A review of podcasting in higher education: Its influence on the traditional lecture. Australasian Journal of Educational Technology 25(3), 309-321, 2009. http://www.ascilite.org.au/ajet/ajet25/mcgarr.html. <retrieved: January, 2012>.

[3] H. Edwards and B. Smith, G. Introduction. In H. Edwards, B. Smith & G. Webb (Eds), Lecturing: Case studies, experience and practice (pp. 1-10). London and Philadelphia: Kogan Page. 2001.

[4] G. Isaacs. Lecture note-taking, learning and recall. Medical Teacher, 11(3), 295-302, 1989.

[5] S. Moore, C. Armstrong, and J. Pearson. Lecture absenteeism among students in higher education: A valuable route to understanding student motivation. Journal of Higher Education Policy and Management, 30(1), 15-24, 2008.

[6] S. Dolnicar. Should we still lecture or just post examination questions on the web?: The nature of the shift towards pragmatism in undergraduate lecture attendance. Quality in Higher Education, 11(2), 103-115, 2005. [verified 30 May 2009] http://ro.uow.edu.au/commpapers/299/. <retrieved: January, 2012>.

[7] J. A. D. Balfour. Audio Recordings of Lectures an an E-Learning Resource. Higher Education Academy. 2006. http://www.heacademy.ac.uk/cebe/resources/detail/events/becon_06_p24_audio_recordings. <retrieved: January, 2012>.

[8] M. Leino, H. Turunen, M. Ahonen, and L. Levonen. Mobiililaitteet oppimisen ja opetuksen tukena. 2002. http://tievie.oulu.fi/koulutusresurssit/artikkelit/leino_ym_2002.pdf. <retrieved: January, 2012>.

[9] M. Reagan. In Search on the Learning Bubble. 2000. http://studio.tellme.com/newsletter/dialed20001102.html. <retrieved: January, 2012>.

[10] M. Lee, and A. Chan. Pervasive, lifestyle-integrated mobile learning for distance learners: An analysis and unexpected results from a podcasting study. Open Learning: The Journal of Open and Distance Learning, 22(3), 201-218, 2007.

[11] S. Brittain, P. Glowacki, J. Van Ittersum, and L. Johnson. Podcasting Lectures. Formative evaluation strategies helped identify a solution to a learning dilemma. EDUCAUSE Quarterly Articles. 2006. http://www.educause.edu/EDUCAUSE+Quarterly/EDUCAUSEQuarterlyMagazineVolum/PodcastingLectures/157413. <retrieved: January, 2012>.

[12] M. Huntsberger, and A. Stavitsky. The new "podagogy": Incorporating podcasting into journalism education. Journalism and Mass Communication Educator, 61(4), 397-410, 2006.

[13] BR.von Konsky, J. Ivins, and SJ. Gribble. Lecture attendance and web based lecture technologies: A comparison of student perceptions and usage patterns.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

Australasian Journal of Educational Technology 25(4), 581-595, 2009.

[14] P. McCloskey. Students Learn Better Via iPod Versus Lecture. 2007. http://campustechnology.com/articles/2007/03/gt-prof-students-learn-better-via-ipod-versus-lecture.aspx. <retrieved: January, 2012>.

[15] Capturing Student Perspectives About Lecture Recordings. 2010. http://www.echo360.com/pdf/Echo360_CSPALR_whitepaper.pdf. <retrieved: January, 2012>.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

A Worldwide Descriptive Analysis of Educational Technology Use

Ron Kovac
Ball State University
Muncie, USA
e-mail: rkovac@bsu.edu.

Kristen E. DiCerbo
Independent Researcher
Phoenix, USA
e-mail: kdicerbo@cisco.com

Nikki Kisor
Ball State University
Muncie, USA
e-mail: Jbeck@theascp.org

Jason Beck
Ball State University
Muncie, USA
e-mail: nkisor@theascp.org

Abstract— **The use of technology in the education vertical has been with us for as long as educational curricula have. The technology has changed, but the application of the tools to teaching and learning has been constant. From blackboards, to overhead projectors to now, information and communication technologies have permeated the classroom to various extents. Although there are many measures and studies of the use and effectiveness of educational technology in our educational systems, most of these have been limited geographically, politically and economically. These limitations have hindered a broad reaching viewpoint of the use of the technologies and have stopped comparisons and contrasts between different theatres of our world. This study attempted to measure the use of educational technology worldwide within a specific worldwide program. The gathered data allows comparisons and contrasts between use worldwide and within the various technological sectors present in today's marketplace. The study was conducted in Fall of 2010 with users of the Cisco Networking Academy Program. These users all follow the same curricula, roughly; so, the variability of programs is held relatively constant. Preliminary findings were that use of technology with each theatre of the world was relatively constant with some indications that lower GDP countries had more extensive use of "social networking" software tools and more consideration of flexibility and agility in the classroom.**

Keywords – Educational Technology; Technology Use; Global assessment.

I. INTRODUCTION

The field of education has been with mankind since our first discovery of fire. Passing down the information to later generations was as crucial then as it is now to perpetuate the species, expand the political/cultural environment and maintain the government and language of the then existing region or country. Starting with story telling, cave pictures, and show and tell, the educational systems has always done it best to improve the methods of transfer of information from those who know to those who want and need to know.

In more recent terms, the application of tools to the teaching/learning process has absorbed the name educational technology. Currently, meaning some digitally based electronic tool, educational technology does incorporate everything from the blackboard to the movie, to the overhead. Depending on the decade we live in defines the technology that is currently in vogue with the educational technology space. The 1960s had the overhead projector, the 1970's the movie projector, and the 1980's the start of the personal computer revolution, and so on.

Attention has also always been paid to measuring the use and effectiveness of these tools. How much is enough? How effective is it in transferring information to students? How does it adapt to current pedagogical styles of teachers? The questions are numerous and therefore the attention researchers pay this is high (especially in the doctoral program research produced worldwide). Unfortunately, due to economic, time and logistic reasons most of these studies have been limited. The limitations are mostly geographic (studying a state, or possibly country), but also include curricular and other limitations.

Using the Cisco Networking Academy Program (CNAP) as a base, this study surveyed the global population (broken-down by theatres) to assess the use of current educational technology. The Cisco Networking Academy Program (CNAP) is a program first promulgated by Cisco Systems Inc, in the late 1990's to enhance the awareness of, and training in the Internetworking field. Internetworking being the core of the current Internet with all its comingled and associated parts and pieces. Starting with two beta sites in the United States it has grown to over 10,000 academies in over 160 countries of the world. Available in many languages, the CNAP program is brought forward in an on-line fashion with a heavy emphasis on hands-on learning, formative assessment and instructor involvement. The program is also heavily supported by more traditional book publications, Interactive course guides and various other tools and techniques (Cisco.netacad.net).

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

The results of this survey, in its descriptive form, hold implications for education in general, but more specifically within the educational technology space. Implications can also be found for vendors and other service providers who support the initiatives of the educational movement. The results of this study also hold merit for further study of equity in access for educational programs worldwide.

The organization of this paper follows with the main body description of the review of related literature. Following describes our methodology, sample population and results. A conclusion, nomenclature and references conclude the paper.

II. MAIN BODY

A. Review of Literature and Background

Classroom access to Technology

Studies conducted to look at classroom technology have focused on two areas: technology access and technology implementation. From these studies, cultural and environmental factors determined the viability of classroom technology integration [2]. Other studies concluded that attitudes and limited skills of legislators, administrators, teachers, and students, detracted from classroom technology integration [4].

The integration of technology in the classroom is not limited to just legislatures, teachers, and students but external factors such as family environment. It is also not limited nationally as a United States problem, but a world-wide problem. In a study titled *Factors Influencing Technology in Teaching: A Taiwanese Perspective*, concerns about students' school activities came from families. One teacher in the study reported that concerns from families arose from not having an Internet connection in the home or a computer. Therefore, time for accessing computers was limited to school time. The teacher concluded that since time for accessing computers at school was scarce, most students needed to have a computer at home to work on assignments [5].

Another report completed suggested that, "Technology use in classrooms is often employed for all the wrong reasons—such as convenience, pressure from school administrators, the belief that students need to be entertained, and so on" [3]. The authors believe that the technology must be facilitated in a way that students are using the technology and not the instructors [3]. When technology is used as a creative tool rather than for the distribution of content and information, it allows students to participate more willingly with tasks and to create their own work, rather than regurgitating the information back to the teacher [6, 10].

It is not the Technology it is the People

In a recent study conducted by the Harvard Graduate School of Education they examined the ratio of one computer to one student. According to the study "…the presence of 1:1 laptops did not automatically add value and their high financial costs underscore the need to provide teachers with high-quality professional development to ensure effective teaching" [1].

In a study conducted by Jing Lei, looking at the quality of technology versus the quantity of technology used, Lei came to the conclusion that it is not the quantity of the technology used, but the quality of using the technology effectively [4]. The study also examined different types of technologies had different outcomes. In their analysis general technology use was confidently related with student technology aptitude, while subject specific technology use was adversely related with student technology aptitude [4].

The main component in this review of literature is the 2008 analysis and review titled *A Framework for Addressing Challenges to Classroom Technology Use.* In this article, six factors were recognized that influenced technology application and the instructor's ability to positively incorporate it into the classroom. The six factors are, "(a) legislative factors, (b) district/school-level factors, (c) factors associated with the teacher, (d) factors associated with the technology-enhanced project, (e) factors associated with the students, and (f) factors inherent to technology itself" [2]. The frequent changes in policy and lack of research by legislatures often lead to poorly designed policies that discourage technology use in the classroom [2]. Some of these same issues with legislature are true for the district level factors. Teachers do not receive enough support resources to favor and use technology-based learning in the classroom [2]. Students present their own challenges to technology-based learning, such as lack of skills, limited prior experience and attitudes towards technology [2, 12].

B. Methodology

1. Participants

The context of this research is the Cisco Networking Academies (CNA; see [13]) public/private partnership between Cisco and over 10,000 educational institutions in over 160 countries. Cisco, previously called Cisco Systems, is the world's largest maker of computer and data networking hardware and related equipment. Cisco provides partnering schools with free on-line curriculum and on-line assessments to support local school instructors in teaching ICT skills in areas related to PC repair and maintenance as well as computer and data network design, configuration, and maintenance in alignment with entry-level industry certifications. The value of the program from the perspective of corporate social responsibility was discussed by [7], while the logical origins of the e-learning approach have been described by [8]. Behrens, Collison, and DeMark [9, 11] provide a conceptual framework for the many and varied aspects of the assessment ecosystem in the program. This research focuses on four courses in a sequence that prepares students with the skills required for

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

the Cisco Certified Network Associate (CCNA) certification.

More specifically the survey was made available to all English speaking programs in the full geographic scope of the CNAP program. Participation was limited to those going through the CCNA Curricula (Both Discovery and Exploration) and not to the various other curricula offered. Participation in the survey was voluntary and therefore not random. The Course Management Systems of the program was used for notification and dissemination of the survey tool This allowed easy access to the survey for all participants and demographic data was collected via the log on account information from the users.

The sample size and demographics show in the following table.

TABLE 1 SAMPLE DEMOGRAPHICS

Theatre	Sample Size	Pop %
Africa	8.1	5
Asia Pac	13.1	17
Greater China	2.8	7
CEE	11.5	7
European Markets	11.1	19
Japan	0	1
IA and C	23.1	17
ME	11.0	7
Russia and CIS	4.7	2
US/Can	14.5	17

Education Level	Sample Size	Pop %
Sec/High School	14	15
Com/Tech College	21	35
College/University	62	44
Other	3	5

3. Results

A survey was administered to students and instructors (who self selected) querying them about their use and presence of educational technology. The survey was offered world-wide to all participants and was submitted back by 1,064 instructors and 1,136 students. Below is the break-down of some of the key findings.

When comparing the use of technology by theatres several interesting facts came to light. The United States is not the only dominate user of technology. The survey first examined how regions perceived their Internet access speed. Close to 80 percent believed their Internet speed was moderate to fast or better. Europe, Asia, Australia, and the Americas all reported above or close to this number. The only regions that perceived their Internet speed to be less than adequate was the Middle East and Africa. In the table below, the percentages for each region are listed.

TABLE 2: PERCEPTION OF INTERNET SPEED (MODERATE TO FAST OR BETTER SPEED)

Region	Percentage
United States	87%
Asia Pacific / Australia	72%
America's minus US	73%
Africa	57%
Europe	86%
Middle East	51%

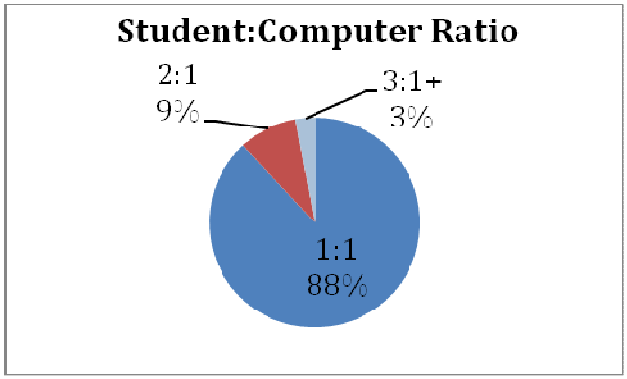
Even though the majority of the theatres report to have moderately fast and higher Internet speed, most report using Windows XP as the operating system of choice for classroom computers. Table 3 shows worldwide data regarding OS used.

TABLE 3: OPERATING SYSETMS USE IN THE WORLD

Operating Systems	Percentage
Windows XP	68%
Windows 7	19%
Linux	4%
Windows Vista	5%
Other	2%
Windows prior to XP	1%
Mac	<1%

The results also produced interesting observations in computer to student ratios. Figure 1 shows that 88 percent of the world has a 1:1 computer to student ratio.

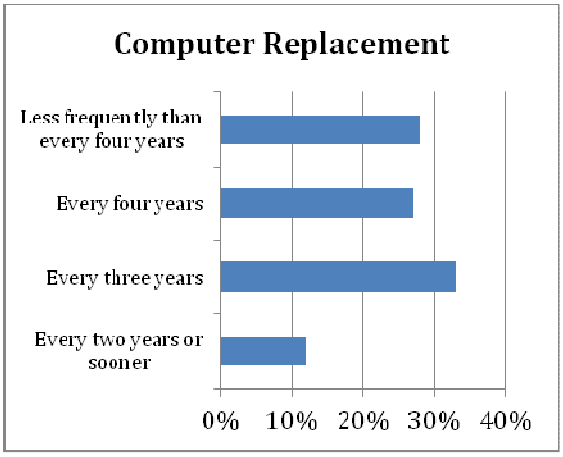
FIGURE 1: COMPUTER STUDENT RATIO



Further analysis showed the refresh ratio for computer technology globally compared. Based on the results listed below in Figure 2, very few schools have the ability to replace computers sooner than three years.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

FIGURE 2: FREQUENCY OF COMPUTER REPLACEMENT



Based on the survey results, it appears that Internet speed internationally is not an issue. However, the issue seems to lie with acquiring newer computers and software for students to use.

The survey asked participants about how often and which web 2.0 technologies they used in the classroom. The instructor survey results are in Table 4. The student survey results are located in Table 5.

Based on the results, students and instructors both share many similarities when it comes to use of web 2.0 technologies. The students and instructors both tend to use social networking the most on a daily basis. Based on the responses received, reading blogs, visiting social networking sites, watching videos and listening to music ranked high on student and instructor web 2.0 uses.

TABLE 4: INSTRUCTOR WEB 2.0 USE

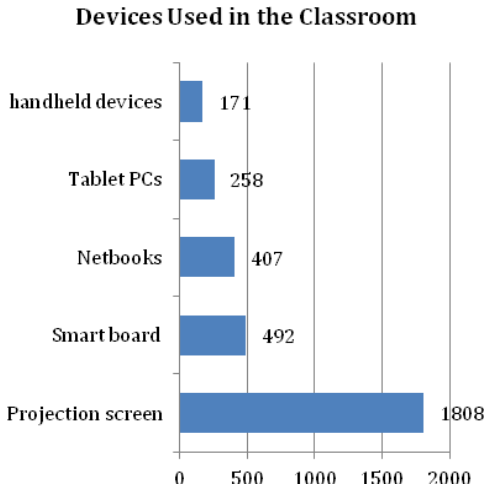
Web 2.0 Technologies	Daily	Weekly	Monthly
read online forums	27	62	85
visits social networking sites	23	49	62
read blogs	18	42	62
watch videos (YouTube)	12	50	73
listen to or download music	11	34	54
comment on social network	9	31	48
listen to podcasts	7	26	51
publish or update own website	7	25	44
update status social network	7	23	43
update blog	5	19	32
upload photos	3	12	30
upload video	3	9	20

TABLE 5: STUDENT WEB 2.0 USE

Web 2.0 Technologies	Daily	Weekly	Monthly
visits social networking sites	46	69	80
read online forums	31	65	84
watch videos (YouTube)	28	64	83
comment on social network	24	53	70
listen to or download music	24	52	73
read blogs	21	49	69
update status social network	18	47	67
update blog	11	29	45
listen to podcasts	10	31	54
publish or update own website	7	22	41
upload photos	7	26	49
upload video	5	18	36

Instructors and students were asked what technology devices were implemented in the classroom. With an overwhelming response projection screens are predominately used to teach students during the learning process. It was quite a surprise how little the other technology devices were applied.

FIGURE 3: DEVICES USED IN THE CLASSROOM



III. CONCLUSION

A. *Discussion*

The purpose of this study was to assess the use of educational technology worldwide within the purview of the Cisco Network Academy Program. In addition, it provided the opportunity to compare and contrast use and type of

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1

educational technology within various theatres of the world. As the results are examined, it is important to keep in mind that survey participants were members of the CNAP program, a specific curriculum within the broader educational realm.

It is interesting to note, that Laptop PC and Cell phone rated the most frequently obtained technology for students worldwide. Desktop was slightly less held with notebooks and tablets being much lower on the scale. Worldwide comparison shows little difference. Difference between student and instructor technology use showed little difference.

The survey also queried participants on the access to and type of LMS/CMS used in their educational environment (Learning/Course Management System). Results show that 50% of users do not have access to an L/CMS. The remainder use mostly open source tools.

The survey was quite comprehensive and the limited real estate of the article prohibits more extensive discussion of the descriptive results or the analytical results. Further analysis will be forthcoming.

B. Limitations

It should be noted that participants who filled out the survey were self-selected. Therefore, because of the lack of randomness, a truly valid nature of the results cannot be established. In addition, the instructor and students who filled out the survey were only English speaking, so only native speakers were excluded. Results therefore cannot absolutely be attributed to the data obtained. Finally, we rely on instructors and students to accurately report their environment, their perceptions and their actions. It is unknown the extent to which instructors and students may mislabel their classes. Additionally, this study only held participants within the CNAP program. This may hold a biases view as this is a technology rich environment.

C. *Conclusions*

Our reporting enclosed is a first blush at a descriptive analysis of the use of educational technology worldwide. There are continued questions about the comparisons of theatre use and impact of educational technology. Further study and more use of inferential statistics will allow a closer look at the data and may provide further information on theatre performance. This study examined this question with a global sample of instructors and students. There was some suggestion that use of educational technology is affected by world theatre, and further study will attempt to group countries by GDP to see if that has a major effect on use of effectiveness of Educational technology.

IV. NOMENCLATURE

The context of this research is the Cisco Networking Academies (CNA; see [13]), a public/private partnership between Cisco and over 10,000 educational institutions in over 160 countries. Cisco, previously called Cisco Systems, is the world's largest maker of computer and data networking hardware and related equipment. Cisco provides partnering schools with free on-line curriculum and on-line assessments to support local school instructors in teaching ICT skills in areas related to PC repair and maintenance, as well as computer and data network design, configuration, and maintenance in alignment with entry-level industry certifications. This research focuses on four courses in a sequence that prepares students with the skills required for the Cisco Certified Network Associate (CCNA) certification.

V. REFERENCES

[1] Dunleavy, M., Dexter, S., Heinecke, W.F. *Journal of Computer Assisted Learning*. Oct 2007, Vol. 23 Issue 5, pp. 440-452.

[2] Groff, J. and Mouza, C. (2008). *A framework for addressing challenges to classroom technology use*. AACE Journal, 16(1), 21-46.

[3] Herrington, Jan, Kervin, Lisa. *Educational Media International*, Sep 2007, Vol. 44, Issue 3, pp. 219-236.

[4] Lei, Jing. *British Journal of Educational Technology*, May2010, Vol. 41 Issue 3, pp. 455-472.

[5] Lih-Juan, Chan Lin, Jon-Chao Hong, Jeou-Shyan Horng, Shih-Hui Chang, and Hui-Chuan Chu. *Innovations in Education & Teaching International*, Feb2006, Vol. 43, Issue 1, pp. 57-68.

[6] Gerbic, P., Stacey, E., Anderson, B., Simpson, M., Mackey, J., Gunn, C. and Samarawickerma, G. (2009). Blended learning: Is there evidence for its effectiveness? In *Same places, different spaces. Proceedings ascilite Auckland 2009*.

[7] Parkinson, D., Greene, W., Kim, Y., and Marioni, J. Emerging Themes of Student Satisfaction in a Traditional Course and a Blended Distance Course. *TechTrends*, *47*(4), 22-28.

[8] Spooner, F., Jordan, L., Algozine, B., and Spooner, M. (1999). Evaluating instruction in distance learning classes. Journal of Educational Research, 92, pp. 132-140.

[9] Voos, R. (2003). 'Blended learning: What is it and where might it take us?' Sloan-C View, 2(1), 2–5.

[10] Privateer, P. M. (1999). Academic technology and the future of higher education: strategic paths taken and not taken', The Journal of Higher Education, 70(1), 60–79.

[11] Smith, A. and Moss, N. (2010). Large scale delivery of Cisco Networking Academy Program by blended distance learning. In: IARIA, 2010 Sixth International Conference on Networking and Services, 6-11 March 2010, Cancun, Mexico.

[12] Cohen, J. (1988). Statistical power for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

[13] http; cisco.com/go/netacad/

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-186-1