ICNS 2011

The Seventh International Conference on Networking and Services

LMPCNA 2011

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

May 22-27, 2011

Venice/Mestre, Italy

ICNS 2011 Editors

Jaime Lloret Mauri, Polytechnic University of Valencia, Spain

Steffen Fries, Siemens, Germany

Mary Luz Mouronte López, Ericsson, Spain

Ron J. Kovac, Ball State University, USA

ICNS 2011

Foreword

The Seventh International Conference on Networking and Services (ICNS 2011), held between May 22 - 27, 2011 in Venice, Italy, 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 2011 also included:

LMPCNA 2011: The Third 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 2011 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 2011. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

We hope that ICNS 2011 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress in sensor technologies and applications research.

We are certain that the participants found the event useful and communications very open. We also hope the attendees enjoyed the beautiful surroundings of Venice.

ICNS 2011 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
Abdulrahman Yarali, Murray State University, USA
Emmanuel Bertin, France Telecom R&D - Orange Labs, France
Steffen Fries, Siemens, Germany
Sorin Georgescu, Ericsson Research, Canada
Mary Luz Mouronte López, Ericsson, Spain
Nirav Kapadia, Fijitsu America, USA
Patryk Chamuczynski, Technisat Digital R&D, Poland

LMPCNA 2011 Chairs

Kristen DiCerbo, Cisco Systems, Inc., USA
Adam M. Gadomski, ECONA
Ron J. Kovac, Ball State University, USA
Iain Murray, Curtin University of Technology - Perth, Australia
Doru Ursutiu, University "Transilvania"- Brasov, Romania / IAOE
Harry Wang, Cisco Academy Training Centre - Asia Pacific, Australia

ICNS 2011

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
Abdulrahman Yarali, Murray State University, USA

ICNS Industry/Research Chairs

Emmanuel Bertin, France Telecom R&D - Orange Labs, France
Steffen Fries, Siemens, Germany
Sorin Georgescu, Ericsson Research, Canada
Mary Luz Mouronte López, Ericsson, Spain
Nirav Kapadia, Fijitsu America, USA
Patryk Chamuczynski, Technisat Digital R&D, Poland

ICNS 2011 Technical Program Committee

Ryma Abassi, Higher School of Communication of Tunis /Sup'Com, Tunisia
Javier M. Aguiar Pérez, Universidad de Valladolid, Spain
Rui L.A. Aguiar, University of Aveiro, Portugal
Francisca Aparecida Prado Pinto, Federal University of Ceará, Brazil
Ali H. Al-Bayatti, De Montfort University - Leicester, UK
Ali Amer, Saudi Telecom Company - Riyadh, Saudi Arabia
Mario Anzures-García, Benemérita Universidad Autónoma de Puebla, Mexico
Pedro Andrés Aranda Gutiérrez, Telefónica I+D - Madrid, Spain
Michael Bauer, The University of Western Ontario - London, Canada
Micah Beck, University of Tennessee, USA
Carlos Becker Westphall, Federal University of Santa Catarina, Brazil
Luis Bernardo, Universidade Nova de Lisboa, Portugal
Emmanuel Bertin, France Telecom R&D - Orange Labs, France
Jun Bi, Tsinghua University, China
Alex Bikfalvi, IMDEA Networks - Madrid, Spain
Eugen Borcoci, University "Politehnica"of Bucharest (UPB), Romania
Fernando Boronat, Polytechnic University of Valencia, Spain
Jens Buysse, University of Gent, Belgium
Diletta Romana Cacciagrano, Università di Camerino, Italia
Maria Calderon Pastor, Universidad Carlos III, Madrid, Spain
Maria Dolores Cano Baños, Polytechnic University of Cartagena - Campus Muralla del Mar, Spain
Kalinka Castelo Branco, University of São Paulo, Brazil

Patryk Chamuczynski, Technisat Digital R&D, Poland
Bruno Chatras, Orange Labs, France
Hugo Coll, Universidad Politecnica de Valencia, Spain
Todor Cooklev, Indiana University - Purdue University Fort Wayne, USA
Noelia Correia, Universidade do Algarve, Portugal
Félix Cuadrado, Universidad Politécnica de Madrid, Spain
Carlton Davis, École Polytechnique de Montréal, Canada
João Henrique de Souza Pereira, University of São Paulo (USP), Brazil
Mari Carmen Domingo, University of Technology, Spain
Prabu Dorairaj, EMC Corporation - Bangalore, India
Zbigniew Dziong, ETS - Montreal, Canada
Gledson Elias, Federal University of Paraíba, Brazil
Juan Ferreiro, University of Vigo, Spain
Armando Ferro Vázquez, University of the Basque Country/ Euskal Herriko Unibertsitatea, Spain
Juan Flores, University of Michoacan, Mexico
Mostafa Fouda, Tohoku University, Japan
Mário Freire, University of Beira Interior, Portugal
Steffen Fries, Siemens, Germany
Sebastian Fudickar, University of Potsdam, Germany
Adriano Galati, University of Nottingham, UK
Michael Galetzka, Fraunhofer Institute for Integrated Circuits - Dresden, Germany
Alex Galis, University College London, UK
Ivan Ganchev, University of Limerick, Ireland
Miguel Garcia, Universidad Politécnica de Valencia, Spain
Gordana Gardasevic, University of Banja Luka, Bosnia and Herzegovina
Rosario Garroppo, University of Pisa, Italy
Manfred Georg, Washington University in St. Louis / Google, USA
Sorin Georgescu, Ericsson Research, Canada
Marc Gilg, University of Haute Alsace, France
Debasis Giri , Haldia Institute of Technology, India
Ann Gordon-Ross, University of Florida, USA
Jean-Charles Grégoire, INRS - Université du Québec - Montreal, Canada
Dominic Greenwood, Whitestein, Switzerland
Vic Grout, Glyndwr University - Wrexham, UK
Go Hasegawa, Osaka University, Japan
Hermann Hellwagner, Klagenfurt University, Austria
Enrique Hernandez Orallo, Universidad Politécnica de Valencia, Spain
Naohiro Ishii, Aichi Institute of Technology, Japan
Peter Janacik, Heinz Nixdorf Institute / University of Paderborn, Germany
Robert Janowski, PTC, Poland
Imad Jawhar, United Arab Emirates University - Al Ain, UAE
Ying Jian, Google, Inc., USA
Nirav Kapadia, Fujitsu America, USA
Masoumeh Karimi, Technological University of America, USA
Adrain Knoth, Friedrich-Schiller-University Jena, Germany
DongJin Lee, University of Auckland, New Zealand
Juong-Sik Lee, Nokia Research Center - Palo Alto, USA
Leo Lehmann, OFCOM, Switzerland

Xu Li, University of Waterloo, Canada
Fidel Liberal Malaina, University of Basque Country, Spain
Wei-Ming Lin, University of Texas at San Antonio, USA
Jaime Lloret Mauri, Polytechnic University of Valencia, Spain
Chengnian Long, Shanghai Jiao Tong University, P. R. China
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, Meraka Institute - Pretoria / Tshwane University of Technology, South Africa
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
Bernhard Neumair, Karlsruhe Institute of Technology (KIT)/Steinbuch Centre for Computing, Germany
Máirtín O'Droma, University of Limerick, Ireland
Kazuya Odagiri, Advanced Institute of Industrial Technology, Japan
Rainer Oechsle, University of Applied Sciences Trier, Germany
Eugène Pamba Capo-Chichi, University of Franche Comte - Besançon, France
Harry Perros, North Carolina State University, USA
David C. Pheanis, Arizona State University - Tempe, USA
Francesco Quaglia, Sapienza Università di Roma, Italy
Idris A. Rai, Makerere University - Kampala, Uganda
Karim Mohammed Rezaul, Centre for Applied Internet Research (CAIR), NEWI, University of Wales, UK
Miklos Molnar, IRISA, France
Oliviero Riganelli, University of Camerino, Italy
David Rincon Rivera, Technical University of Catalonia (UPC) - Barcelona, Spain
Joel Rodrigues, University of Beira Interior, Portugal
Paolo Romano, Università degli Studi di Roma “La Sapienza”, Italy
Sattar B. Sadkhan, University of Babylon, Iraq
Francisco Javier Sánchez Bolumar, Administrador de Infraestructuras Ferroviarias (ADIF), Spain
Luz A. Sánchez-Gálvez, Benemérita Universidad Autónoma de Puebla, México
Hermann Schloss, University of Trier, Germany
Thomas C. Schmidt, Fachhochschule für Technik und Wirtschaft - Berlin, Germany
René Serral Garcia, Universitat Politècnica de Catalunya, Spain
Fangyang Shen, Northern New Mexico College, USA
Yongning Tang, Illinois State University, USA
Yoshiaki Taniguchi, Osaka University, Japan
Olivier Terzo, Istituto Superiore Mario Boella - Torino, Italy
Carlos Turró Ribalta, Polytechnic University of Valencia, Spain
Dimitrios D. Vergados, University of Piraeus, Greece
Dario Vieira, EFREI, France
Manuel Villén-Altamirano, Universidad Politécnica de Madrid, Spain
Michelle Wetterwald, EURECOM - Sophia Antipolis, - France
Feng Xia, Dalian University of Technology, China
Haiyong Xie, Yale University, USA

Qin Xin, Simula Research Laboratory - Oslo, Norway
Kaiping Xue, USTC - Hefei, China
Ramin Yahyapour, TU Dortmund University, Germany
Homayoun Yousefi'zadeh, University of California - Irvine, USA
Vladimir S. Zaborovsky, Polytechnic University/Robotics Institute - St.Petersburg, Russia
Faramak Zandi, Alzahra University - Tehran, Iran
Sherali Zeadally, University of the District of Columbia, USA

LMPCNA 2011

LMPCNA Advisory Chairs

Kristen DiCerbo, Cisco Systems, Inc., USA
Adam M. Gadomski, ECONA
Ron J. Kovac, Ball State University, USA
Iain Murray, Curtin University of Technology - Perth, Australia
Doru Ursutiu, University "Transilvania"- Brasov, Romania / IAOE
Harry Wang, Cisco Academy Training Centre - Asia Pacific, Australia

LMPCNA 2011 Technical Program Committee

Nalin Abeysekera, Open University of Sri Lanka, Sri Lanka
Joan Arnedo Moreno, Universitat Oberta de Catalunya, Spain
Giancarlo Bo, Technology and Innovation Consultant - Genova, Italy
Doina Bucur, Oxford University, UK
Dumitru Dan Burdescu, University of Craiova, Romania
Maiga Chang, Athabasca University, Canada
Pavel Cicak, Slovak University of Technology, Slovakia
Giuseppe Cinque, Consorzio ELIS - Rome, Italy
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
Frantisek Jakab, AAM Cisco Slovakia (BDA at CEEE) /Technical University of Kosice/DCI, Slovakia
Karol Kniewald, Cisco Systems, USA
Ron J. Kovac, Ball State University, USA
Eugenijus Kurilovas, Vilnius Gediminas Technical University, Lithuania
Jaime Lloret Mauri, Universidad Politécnica de Valencia, Spain
Nicholas G. Moss, The Open University, UK
Iain Murray, Curtin University of Technology - Perth, Australia
Elisabetta Parodi, eXact learning solutions - Sestri Levante, Italy
Josep Prieto-Blázquez, Universitat Oberta de Catalunya, Spain
Jelena Revzina, Transport and Telecommunication Institute, Latvia
Andrew Smith, The Open University, UK
Harry Wang, Cisco Academy Training Centre - Asia Pacific, Australia

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

Comparative Analysis and Tests of Intelligent Streaming Video on Demand for Next Generation Networks: Two Colombian Study Cases

Juan Sebastián Zabala; Harold René Chamorro; Margarita María Díaz; Elvis Eduardo Gaona
Ingeniería Electrónica, Facultad de Ingeniería, Universidad Distrital
Bogotá, Colombia
jszabala@ieee.org; hr.chamo@ieee.org; margarita.diaz@ieee.org ; egaona@udistrital.edu.co

Abstract – **This document proposes to evaluate the transmission of Video on Demand using Intelligent Streaming and Next Generation Networks concepts. Intelligent Streaming is an adaptive methodology to take advantage of bandwidth and the network resources, the streaming system had been tested over the UDNET and RUMBO-RENATA networks during the peak data traffic. Jitter deviation, data packet losses, user datagram protocol efficiency and quality of service are some of the indicators measured. The Intelligent Streaming guarantees the quality of service matching the server with the client needs, without changing the network policies delivery systems.**

Keywords – ***Intelligent Streaming; Video on Demand; Next Generation Networks; TCP and UDP; Network Efficiency.***

I. INTRODUCTION

Communications have generated a change in people's daily habits since Internet is accessible to everyone. Internet has become the main tool for work, networking and leisure activities due to home use as well as the number of smart mobile terminals. Now, the way to share Internet content has changed with the Next Generation Networks (NGN) because the files are sent in data packets [1, 2]. An efficient way to broadcast information is video streaming which has become the main application for video-conferencing, Video on Demand (VOD) and video-aided distance learning [3, 4].

The streaming transmission over the web has been developed for many authors [5, 6], where the reliability of real time playing of video segments is shown. Some of them study the point of traffic in the network making a pre-release of information and broadcast at a defined time. If someone wants to access to specific video, the user has to subscribe and when the information is available the user has to authenticate the subscription [7]. One of the limitations for video streaming is the consumption resources when too many users are connected. Bandwidth efficient algorithms applications show the non - reduction of efficiency system when the level of interaction increases. The interaction available is not affected and the overheads are associated to external sources and not to an algorithm failure [8].

The development exposed in this paper is a transmission VOD using intelligent streaming, increasing the QoS (Quality of Service) and decreasing jitter. In this proposal the QoS is defined according to the features of the client's connection and the streaming server capability, keeping the policy delivery unmodified.

The advantages of intelligent streaming are playing the video while the information is being downloaded and analysing the client's speed connection in order to transmit the data packet in a bit rate appropriately in relation with the channel bandwidth, avoiding delays and getting hung up [10, 11]. The strategy for improving the systems quality and transmission rate is the use of dedicated servers for streaming instead of web servers. These servers are in charge of storing and operating the data packets [12].

The design requirements are a trade-off between latency, memory and the overhead, take advantage of bandwidth, availability of data network and decrease in data packet losses and jitter [13, 14].

This paper is organized as follows; the first part shows the features of a streaming system, the tool chosen for this development, the process for encoding files and the selection of encoding rates for multimedia contents. Next the streaming model proposed for being tested using VOD is described. Also the delivery polices and the operation of the server streaming sub-system. In the last section the test results over the UDNET and RENATA-RUMBO networks are compiled and analysed.

II. STREAMING STUDY FEATURES

A streaming system comprises the user with a specified bandwidth and transmission rate [1], a player that supports video and audio, an Internet Service Provider (ISP), a streaming server with a control, storage and communications subsystem, an encoding and compression unit and video and audio sources. The components are explained briefly next. Fig. 1 depicts a typical structure of streaming.

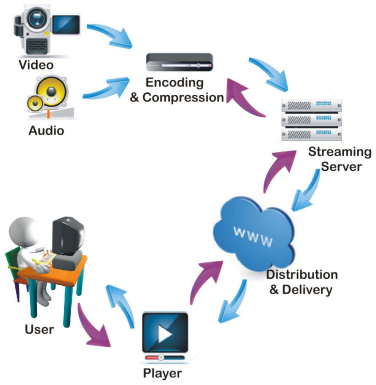


Fig. 1 Streaming system components

User: supports the multimedia contents reception and uncut viewing as well as VCR (Video Cassette Recorder) commands.

Player: is in charge of sending the user requests to the streaming server, the storage received contents in local buffers, encoding contents and synchronize the data for displaying.

Distribution and Delivery: Internet is the platform for exchanging information between *player* and *streaming server.*

Streaming server: the main functions are management of the policies delivery, system storage of the contents and to answer the requests made by the clients.

Encoding and Compression: is in charge of modifying the video and audio sources in a compatible format file with the user player.

Video and audio: are the sources for a streaming system. Their transmission can be on demand or live.

The aim of NGN and IPTV is convergence [15], for this reason the software for developing the application must satisfy this.

The tool Windows Media® has been chosen for its encoding system with a public domain license useful for files format supported by WMP® (Windows Media Player) like *wma* for audio files and *wmv* for video files. WME® (Windows Media Encoder) is a free multimedia contents player; the platform server is Windows Server 2008.

WME® offers a variety of possibilities for video edition and guarantees high quality and high compression file [16, 17]. The most important feature of WME is the capture contents in frames enabling the transmission of encoding sequences for different transmission rates, what is known as Intelligent Streaming. Besides it has other features: hardware acceleration, high definition video, multi-channel sound quality, segment-based encoding, easy to use because it has an intuitive interface, save contents and manage QoS delivery policies with the client. In addition to, it supports protocols for copyright like ISAN (International Standard Audiovisual Number), Ad-ID (Advertising Industry Standard Unique Identifier) and DRM (Digital Rights Management).

For encoding contents using WME it is necessary to get the source in a digital format, select the source, the distribution method and the characteristics of user connection, add the meta-data to the file and apply DRM. The process of encoding contents is:

File Conversion: this option allows for converting an *avi* or *mpeg* format file to compatible formats file with WME like *wmv, wma or wav*.

File Information: in this stage the meta-data are updated in the streaming file.

Direct Encoding: WME® allows it capture multimedia contents directly from the sound card or video card for encoding.

Select Source: the input file must be selected for encoding as well as the file name and the storage directory.

Distribution Method: is selected depending on the encoding formats and the kind of application. The streaming server for this development is WMS® (Windows Media Server).

Connection Characteristics: the transmission rate must be set, depending on the quality of connection. Also is possible set the transmission rate for multiple terminals that will be adjusted by intelligent streaming as is proposed in this document. In Table I. are showed the encoding rates selected.

TABLE I. SELECTION OF ENCODING RATES FOR MULTIMEDIA CONTENTS

TRANSMISSION RATE	FRAME RATE	FRAME SIZE	USE
1128 Kbps	29.97 fps	320 X 240	HIGH VIDEO DEFINITION BROADCAST OVER HIGH BANDWIDTH NETWORKS
548 Kbps	29.97 fps	320 X 240	STANDARD VIDEO DEFINITION BROADCAST OVER MODERATE BANDWIDTH NETWORKS
282 Kbps	29.97 fps	320 X 240	LOW BANDWIDTH NETWORKS
148 Kbps	29.97 fps	320 X 240	NETWORKS WITH DATA TRANSMISSION RATE NEAR TO 500 KBPS

III. MODEL SERVICE CHARACTERIZATION

The streaming systems are classified according to services offered like interactivity and information availability [17]. These characteristics make attractive the service increasing the interest and the desire to interact and learn about specific topics. The most common services offered are “Live” and “On Demand”.

Live: is a transmission that is being generated in real time [16]. On demand: is a transmission that is has its information stored in a buffer, allowing the user to download at any time video selected [4]. The model proposed and tested to VOD and applying Intelligent Streaming is presented in Fig. 2.

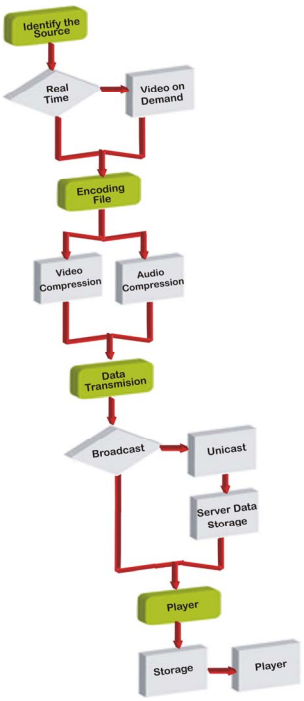


Fig. 2 Intelligent streaming structure

Intelligent streaming is a feature of WMS that scaling the data rate transmission through multiple video tracks [18]. The process for intelligent streaming begins with the request from the user, this request defines if the data source is live or VOD. Then the streaming server start the negotiation of rate transmission defining the QoS policies. In all streaming transmission the audio has priority over the video, because of overlapping is less noticeable in video than audio. Next, media content can be broadcasted in any of the 4 coding rates explained and the requests are analyzed to broadcast the video preserving the audio quality. The next steps are encoding separately audio and video and synchronize them on the client player. Data transmission is the stage that define if it is necessary storage the encoding files in a storage server, it happens when the request is unicast, while a broadcast request send the information to all terminals and each client decide what information wants to watch [16]. The advantage of broadcast is the traffic reduction over the net, the decreasing bandwidth needed and the streaming server processes are steady. The player has a storage system that save a part of files and the playing begins, this stage is a dynamic system that is storing and playing the information [8].

Streaming server delivery policies are the basis to synchronize data transmission ensuring data flow along all network and are related with a control sub-system [19], a communications sub-system and a storage sub-system as is depicted in Fig. 3.

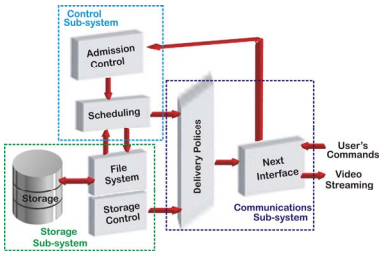


Fig. 3 Server Streaming Sub-systems

The storage sub-system has three sub-modules, storage, file system and storage control. All the encoding files are saved on data storage and the sub-module file system is in charge of enabling the data saving or transmission depending on the control sub-system requests. The sub-module storage control is communicated with the sub-module delivery policies of communications Sub-system, and its function is informing the capabilities of system during the transmission [19].

The control sub-system has the sub-modules of admission control and scheduling. The admission control must ensure the QoS and the resources availability as disk bandwidth, processing bandwidth, bandwidth network and space for storage, for the client. The scheduling sub-module has two sub-modules, disk Scheduling and network scheduling their function is planning how must be the data transference from the Storage sub-system to the memory buffers and from the memory buffers to the network. Communications sub-system has the sub-modules, delivery policies and net interface. The main function is responded to user requests and offers the services according to client capabilities.

IV. Network Analysis Measurement Results

With the known benchmark program GNU Iperf [20], which measures some TCP/IP network features, two networks are studied as study case with the same video data and are evaluated with different performance indices in order to show the effectiveness of the intelligent streaming model proposed.

Fig.4 shows two pictures of the same video from a server to a client – server, the software mentioned evaluate the TCP and UDP performance.

Fig. 4 Real Time Video Streaming Implemented

A. Measurements on the UDNET Network

The first study case is the UDNET Network (Universidad Distrital Network) which is LAN (Local Area Network) and its theoretical transmission bandwidth is of 100 Mbps/s, the evaluation indicates that this network has more efficiency with large amount of information without fragmenting; the first test shows the TCP window rate with 8, 20000, 100000 and 1000000 Kbytes and is observed how the efficiency increase with those windows rates. Fig. 5 shows this network test.

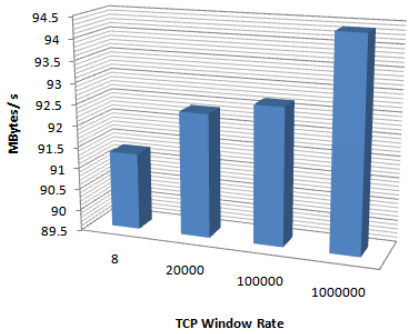


Fig. 5 TCP Effective Bandwidth of UDNET Network.

Data packets have not always the same delay [14], this associated to the jitter effect which relates the time expected when packets arrive, with the time of packets are delivered.

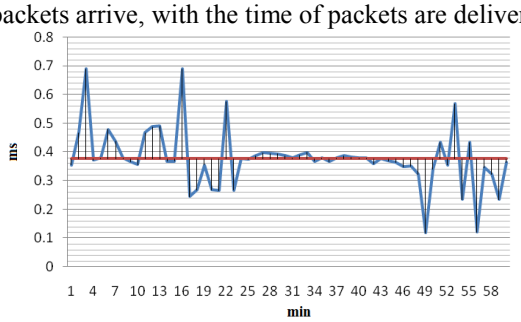


Fig. 6 In blue: UDP Jitter derivation in UDNET network. In red: UDP Jitter average.

Jitter measurement in this application is presented in Fig. 6; the jitter difference between the average is minimal; it shows

that there is no degradation in video transmission. The average level of jitter deviation is 0.38ms, with a maximum value of 0.689ms and a minimum of 0.118ms.

It is well known that packets of data do not have a correct order in delivery and the rate transmission has losses and is also inconstant, however the audio requires this rate to be constant. Jitter buffer at the receiver compensates the effects by its function of trade off between delay and loss [10], these jitter buffers have a variations of 30ms and manage the audio transmission at a constant rate.

If the rate of transmission is slower than supported in buffer, is presented with high losses in packets spoiling the transmission quality, the maximum limit of losses should not exceed 1% [10, 12], due to these data losses which are noticeable in the final user and the service is demoted. The data video on the network studied shows the next behaviour according to the amount of information transferred.

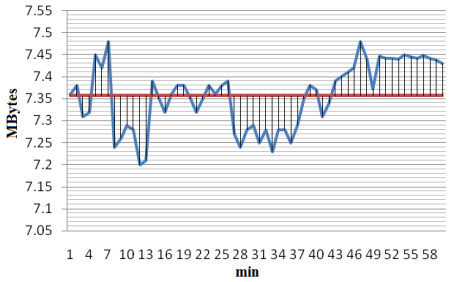


Fig. 7 In blue: Amount of Information Deviation Transferred. In red: Amount of Information Average.

According to the effective average of data transmitted, which is of 7.36 Mbytes, 3600 samples per hour are sent (sample per minute), Fig. 8 shows the variability of losses, obtaining a minimum deviation of average reference.

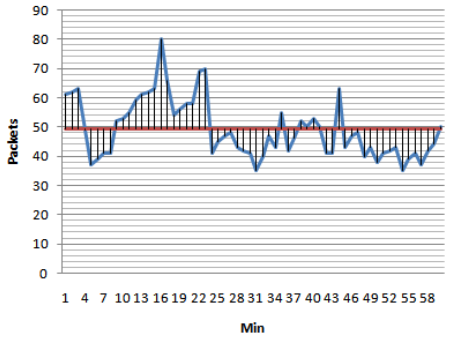


Fig. 8 In blue: Data Packet Sent Losses on the UDNET Network. In red: Information Losses Average.

The packet average lost is of 49 packets per 5351 packets sent, with a maximum lost of 80 and a minimum of 35 packets. In percentage terms the value of data packet losses is about of 0.91%, which is suitable to guarantee a QoS (Quality of Service QoS) in voice and video standard demand of lower value of 1%. Tests on UDNET network shows this fulfillment demanded in this kind of service. Fig. 9 shows the efficiency measurement mentioned.

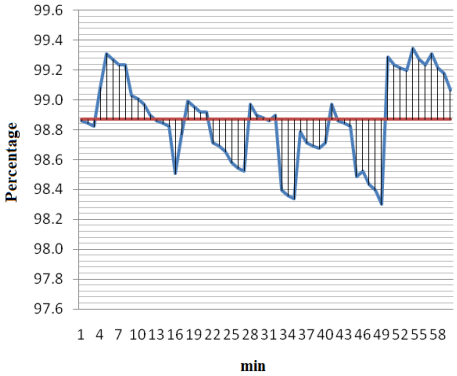


Fig. 9 In blue: UDP Efficency on the UDNET Network. In red: UDP Efficency Average.

The efficiency average in the network faced with observed variations, has a measured value of 99.1% which is in the within margins in a correct service.

Table II shows the main measurements performance indices evaluated in the UDNET Network, where TB, EB and TCP WR mean Theoretical Bandwidth, EB Effective Bandwidth respectively (Mbytes/s) and TCP Window Rate (Kbytes).

TABLE II. UDNET NETWORK MEASUREMENTS SUMMARY

SECONDS	MBYTES	TB	EB	TCP WR	EFFICIENCY
10	109	100	91.3	8	91.3
12	133	100	92.4	20000	92.4
19.3	213	100	92.7	100000	92.7
11.3	127	100	94.4	1000000	94.4
				AVERAGE	92.7

B. *Measurements on the RUMBO-RENATA Network*

Now, the same tests exposed above are applied to the second study case, the RUMBO – RENATA network. Fig. 10 shows the measurement of the network with the same window rates of the first case, this presents a better efficiency transmission packet that is about of 20000 Kbytes and its theoretical transmission rate is of 60Mbytes/s.

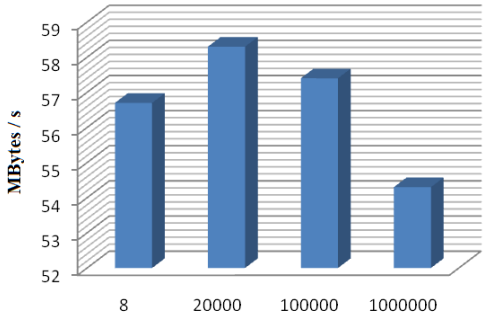


Fig. 10 TCP Effective Bandwidth for RUMBO-RENATA Network.

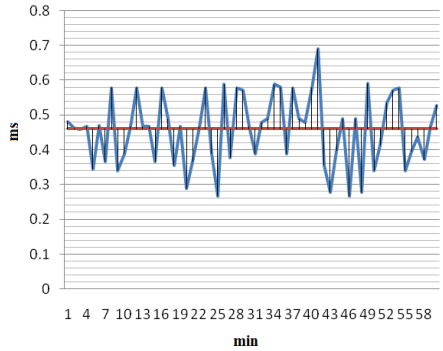


Fig. 11 In blue: UDP Jitter derivation in the RUMBO-RENATA Network. In red: UDP Jitter Average.

Jitter average deaviation measured is 0.46ms, with a maximum of 0.690ms and a minimum of 0.267ms, these values are in the margins to bring a suitable service to VOD, this value is according to the maximum value allowed in buffer jitter mentioned before.

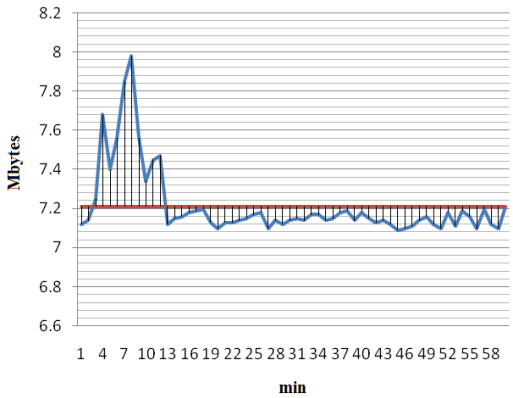


Fig. 12 In blue: Amount of Information Variability Transferred on RUMBO-RENATA Network. In red: Amount of Information Variability Average Transfered.

Fig. 12 shows the effective transference in the network, the average data transferred is 7.21 Mbytes with some variations observed, like in the first case 3600 samples has been sent in one hour.

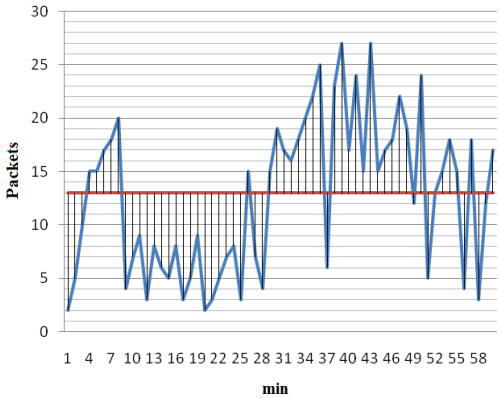


Fig. 13 Network. In blue: Data Packet Sent Losses on the RUMBO-RENATA Network. In red: Information Losses Average.

Data packet loses average is of 13 per 5351 packets sent which represents 0,24%, with a minimum of 2 packets and a maximum of 27, this fits with the parameters established.

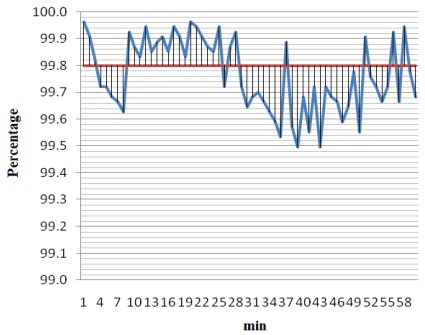


Fig. 14 UDP Efficiency on the RUMBO-RENATA Network. In blue: UDP Efficency on the RUMBO-RENATA Network. In red: UDP Efficency Average.

Average efficiency conexión has a value of 99.8% which implicates a well performance in video transmission.

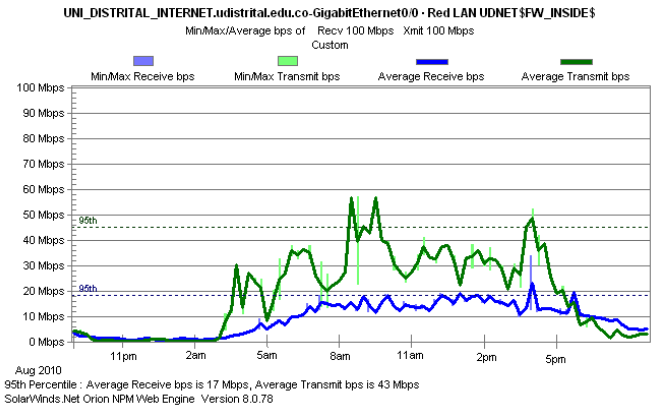


Fig. 15 Transmit and Receive Rate on the UDNET Network. In blue: Average Receive bps on the UDNET Network. In green: Average Transmit bps on the UDNET Network.

TABLE III. RUMBO RENATA NETWORK MEASUREMENTS SUMMARY

SECONDS	MBYTES	TB	EB	TCP WR	EFFICIENCY
10	111	100	56.7	8	94.50
12	130	100	58.3	20000	97.17
19.3	220	100	57.4	100000	95.67
11.3	118	100	54.3	1000000	90.50
				AVERAGE	94.46

From measurement Table is concluded that RENATA-RUMBO Network presents a high level efficiency in video packets transmission using intelligent streaming

V. CONCLUSIONS

Next Generation Networks have become a viable option for firms and has impacted positively in clients which demand them, tests performance in this document demonstrate the simplicity to broadcast multimedia contents with intelligent streaming in networks where it is not apparently supported.

A high efficiency performance of TCP and UDP is obtained in both Intranet networks (92.7%), even though the peak hour traffic (8:00 am – 9:00 am) and it validate the implementation QoS in the two study network cases developed here.

This real time application shows how it is improved the quality of service applying intelligent streaming in some local

study cases networks, without a change in the policies or access network, only requiring network negotiations between client and the server.

It has found that the peak hour is between the 8 and 9 of morning because in this hour the University starts the most Network activities, start to work the administrative area and a big part of University´s computers request access to the servers, as shown in Fig. 15.

Multimedia streaming has been gradually gaining ground from peer to peer networks used to download and make play lists contents, due to streaming technology it does not require high storage capacity of streams, on the other hand, peer to peer requires first, a complete download to reproduce the content.

The TCP Protocol is used only for the service website and the connection control, the pair of protocols RTP/UDP was used for the transmission of audio and video data.

SCTP Protocol was no used because their greatest features as RTO and heartbeats to declare inactive or failed connections would not be exploited, also for this specific application was unlikely that the computer, laptop or smartphone to have 3 or more IP address to take advantage of another great feature of this protocol as the "multihoming".

This paper shows a relatively easy and efficient way to transmit audio a video through a typical network and an academic network of advanced technology using techniques such as intelligent streaming with support for IPv6 using Windows Server 2008, and comparing the results of efficiency in both networks.

ACKNOWLEDGMENT

Special acknowledgment to GITUD (Grupo de Investigación en Telecomunicaciones de la Universidad Distrital) and SITIR (Semillero de Investigación en Tecnologías de Información y Redes de la Universidad Santo Tomas).

VI. REFERENCES

[1] Gubbi, R., "Multimedia streams and quality of service in the next generation wireless home networks," *IEEE International Workshop on* Mobile Multimedia Communications, 1999. (MoMuc ´99) 1999, pp. 232-235, 1999

[2] Mahajan, A.; Soans, S., "Next generation mobile network concepts, technology and sample applications," *International Conference on Wireless Communication and Sensor Computing, 2010. ICWCSC 2010*, pp. 1-6, 2-4 Jan. 2010

[3] Wang, J.R.; Parameswaran, N., "Intelligent Streaming Video Data over the Web," *International Conference on Web Intelligence Proceedings, 2004. WI 2004. IEEE/WIC/ACM* , pp. 744-747, 20-24 Sept. 2004

[4] Hayoung, Yoon; JongWon, Kim; Tan, Feiselia; Hsieh, Robert, "On-demand Video Streaming in Mobile Opportunistic Networks," *Sixth Annual IEEE International Conference on Pervasive Computing and Communications*, *2008. PerCom 2008,* pp. 80-89, 17-21 March 2008

[5] Apostolopoulos, Jhon; Trott, Mitchell; Kalker, Ton; Wai-Tian Tan, "Enterprise Streaming: Different Challenges from Internet Streaming," *IEEE International Conference on Multimedia and Expo, 2005. ICME 2005*, pp. 1386-1391, 6-6 July 2005

[6] Rejaie, R., "On design of adaptive Internet streaming applications: an architectural perspective," *IEEE International Conference on Multimedia and Expo, 2000. ICME 2000*, pp. 327-330, July 30 2000–August 2 2000

[7] Chiu, S.K.H.; Vuong, S.T., "A novel method for flash crowd avoidance in P2P video on demand streaming via pre-release distribution," *International Conference on Advanced Technologies for Communications, 2008. ATC 2008*, pp. 219-222, 6-9 Oct. 2008

[8] J. Moyano., "Difusión de Sesiones Lectivas con Imagen y Video en Red," Escola Politecnica Superior de Casteldefels y Universitat Politécnica de Catalunya, 2006. Catalunya 2006, P-10.

[9] Guang, Tan; Jarvis, S.A., "A Payment-based Incentive and Service Differentiation Mechanism for Peer-to-Peer Streaming Broadcast," *14th IEEE International Workshop on Quality of Service, 2006. IWQoS 2006*, pp. 41-50, 19-21 June 23006

[10] Qadeer, M.A.; Akhtar, N.; Khan, F.; Haque, F., "Monitoring and Analysis of Data Packets Using Data Stream Management System," *International Conference on Computer and Electrical Engineering, 2008. ICCEE 2008*, pp. 214-218, 20-22 Dec. 2008

[11] Wen Zhang; Junwei Cao; Yisheng Zhong; Lianchen Liu; Cheng Wu, "Block-Based Concurrent and Storage-Aware Data Streaming for Grid Applications with Lots of Small Files," *9th IEEE/ACM International Symposium on Cluster Computing and the Grid, 2009. CCGRID 2009*, pp. 538-543, 18-21 May 2009

[12] Sun Dakang; Yan Danfeng; Yang Fangchun, "Research on Packet Tagging Using the Attributes of Data Stream," 2010 *International Conference on Communications and Mobile Computing, 2010. CMC 2010*, pp. 116-120, 12-14 April 2010

[13] Siu-Ping Chan; Kok, C.-W.; Wong, A.K., "Multimedia streaming gateway with jitter detection," *IEEE Transactions on Multimedia*, vol.7, no.3, pp. 585- 592, June 2005

[14] Guanfeng Liang; Ben Liang, "Effect of Delay and Buffering on Jitter-Free Streaming Over Random VBR Channels," *IEEE Transactions on Multimedia,* vol.10, no.6, pp. 1128-1141, Oct. 2008

[15] Hyeokchan Kwon; Sangchoon Kim; Jaehoon Nah; Dongil Seo, "Secure Overlay for Multicast IPTV Streaming Using Trust Rendezvous Point," *International Conference on New Trends in Information and Service Science, 2009. NISS 2009*, pp. 419-422, June 30 2009-July 2 2009

[16] Peng Huang; Chunlei Jiang; Dongcai Qiu, "A method of monitoring transmission quality of streaming media," *11th IEEE International Conference on Communication Technology, 2008. ICCT 2008*, pp. 537-540, 10-12 Nov. 2008

[17] Hashimoto, K.; Shibata, Y., "Design and Implementation of Adaptive Streaming Modules for Multipoint Video Communication," *22nd International Conference on Advanced Information Networking and Applications - Workshops, 2008. AINAW 2008*, pp. 553-560, 25-28 March 2008

[18] Chao Huang; Jintao Li; Hongzhou Shi, "An intelligent streaming media video service system," *IEEE Region 10 Conference on Computers, Communications, Control and Power Engineering, 2002. TENCON 2002,* pp. 5-10, 28-31 Oct. 2002

[19] Hammad, M.A.; Aref, W.G.; Elmagarmid, A.K., "Search-based buffer management policies for streaming in continuous media servers," *IEEE International Conference on Multimedia and Expo, 2002. ICME 2002,* pp. 253- 256, 2002

[20] Rao, N.S.V.; Poole, S.W.; Wing, W.R.; Carter, S.M., "Experimental Analysis of Flow Optimization and Data Compression for TCP Enhancement," *IEEE International Conference on Computer Communications, 2009. INFOCOM Workshops 2009,* pp. 1-6, 19-25 April 2009

 ISBN: 978-1-61208-133-5

A Workflow Platform for Simulation on Grids

Toàn Nguyên, Laurentiu Trifan
Project OPALE
INRIA Grenoble Rhône-Alpes
Grenoble, France
tnguyen@inrialpes.fr, trifan@inrialpes.fr

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

***Abstract*—This paper presents the design, implementation and deployment of a simulation platform based on distributed workflows. It supports the smooth integration of existing software, e.g., Matlab, Scilab, Python, OpenFOAM, ParaView and user-defined programs. Additional features include the support for application-level fault-tolerance and exception-handling, i.e., resilience, and the orchestrated execution of distributed codes on remote high-performance clusters.**

Keywords-workflows; fault-tolerance; resilience; simulation; distributed systems; high-performance computing

I. INTRODUCTION

Large-scale simulation applications are becoming standard in research laboratories and in the industry [1][2]. Because they involve a large variety of existing software and terabytes of data, moving around calculations and data files is not a simple avenue. Further, software and data often reside in proprietary locations and cannot be moved. Distributed computing infrastructures are therefore necessary [6, 8].

This paper details the design, implementation and use of a distributed simulation platform. It is based on a workflow system and a wide-area distributed network. This infrastructure includes heterogeneous hardware and software components. Further, the application codes must interact in a timely, secure and effective manner. Additionally, because the coupling of remote hardware and software components are prone to run-time errors, sophisticated mechanisms are necessary to handle unexpected failures at the infrastructure and system levels. This is also true for the coupled software that contribute to large simulation applications. Consequently, specific management software is required to handle unexpected application and software behavior.

This paper addresses these issues. Section II gives a detailed overview of the implementation using the YAWL workflow management system [4]. Section III is a conclusion.

II. WORKFLOW PLATFORM

A. *The YAWL workflow management system*

Workflows systems are the support of many e-Science applications [1][6][8]. Among the most popular systems are Taverna, Kepler, Pegasus, Bonita and many others [11][15]. They complement scientific software suites like Dakota, Scilab and Matlab in their ability to provide complex application factories that can be shared, reused and evolved. Further, they support the incremental composition of hierarchic composite applications. Providing a control flow approach, they also complement the usual dataflow approach used in programming toolboxes. Another bonus is that they provide seamless user interfaces, masking technicalities of distributed, programming and administrative layers, thus allowing the users and experts to concentrate on their areas of interest.

The OPALE project at INRIA (http://www-opale.inrialpes.fr) is investigating the use of the workflow management system for distributed multidiscipline optimization [3]. The goal is to develop a resilient workflow system for large-scale optimization applications [26]. It is based on extensions to the YAWL system to add resilience and remote computing facilities for deployment on high-performance distributed infrastructures [4]. This includes large-PC clusters connected to broadband networks. It also includes interfaces with the Scilab scientific computing toolbox [16] and the ProActive middleware [17].

Provided as an open-source software, YAWL is implemented in Java. It is based on an Apache server using Tomcat and Apache's Derby relational database system for persistence. YAWL is developed by the University of Eindhoven (NL) and the University of Brisbane (Australia). It runs on Linux, Windows and MacOS platforms [25]. It allows complex workflows to be defined and supports high-level constructs (e.g., XOR- and OR-splits and joins, loops, conditional control flow based on application variables values, composite tasks, parallel execution of multiple instances of tasks, etc) through high-level user interfaces.

 ISBN: 978-1-61208-133-5

Formally, it is based on a sound and proven operational semantics extending the *workflow patterns* of the Workflow Management Coalition [21, 32], implemented and proved by colored Petri nets. In contrast, other workflow management systems which are based on the Business Process Management Notation (BPMN) [27] and the Business Process Execution Language (BPEL) [28] are usually not supported by a proven formal semantics. Further, they usually implement only specific and /or proprietary versions of the BPMN and the BPEL specifications. There are indeed over 73 (supposedly compliant) implementations of the BPMN, as of January 2011, with several others currently being implemented [27], in addition to more than 20 BPEL engine providers. However, BPEL supports the execution of long running processes required by simulation applications, with compensation and undo actions for exception handling and fault-tolerance, as well as concurrent flows and advance synchronization [28].

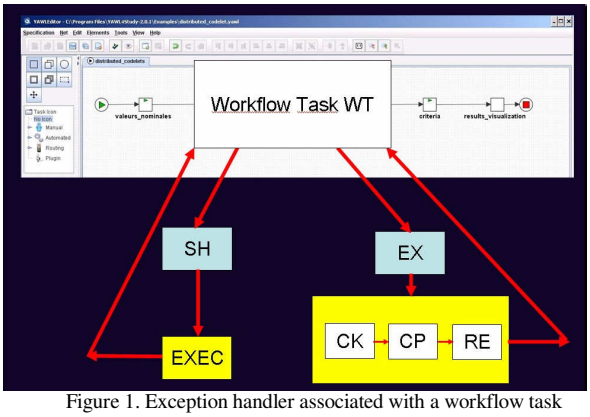


Figure 1. Exception handler associated with a workflow task

Designed as an open platform, YAWL supports natively interactions with external and existing software and application codes written in any programming languages, through shell scripts invocations, as well as distributed computing through Web Services.

It includes a native Web Services interface, custom services invocations through *codelets*, as well as rules, powerful exception handling facilities, and monitoring of workflow executions [13].

Further, it supports dynamic evolution of the applications by extensions to the existing workflows through *worklets*, i.e., on-line inclusion of new workflow components during execution [14].

It supports automatic and step-by-step execution of the workflows, as well as persistence of (possibly partial) executions of the workflows for later resuming, using its internal database system. It also features extensive event logging for later analysis, simulation, configuration and tuning of the application workflows.

Additionally, YAWL supports extensive organizations modeling, allowing complex collaborative projects and teams to be defined with sophisticated privilege management: access rights and granting capabilities to the various projects members (organized as networked teams of roles and capabilities owners) on the project workflows, down to individual components, e.g., edit, launch, pause, restart and abort workitems, as well as processing tools and facilities [25].

Current experiments include industrial testcases for automobile aerodynamics optimization, involving the connection of the Matlab, Scilab, Python, ParaView and OpenFOAM software to the YAWL platform [3]. The YAWL workflow system is used to define the optimization processes, include the testcases and control their execution: this includes reading the input data (StarCCM+ files), the automatic invocation of the external software and automatic control passing between the various application components, e.g., Matlab scripts, OpenFOAM, ParaView.

B. *Exception handling*

The exception handlers are automatically tested by the YAWL workflow engine when the corresponding tasks are invoked. This is standard in YAWL and constraint checking can be activated and deactivated by the users [4].

For example, if a particular workflow task WT invokes an external EXEC code through a shell script SH (Figure 1) using a standard YAWL *codelet*, an exception handler EX can be implemented to prevent from undesirable situations, e.g., infinite loops, unresponsive programs, long network delays, etc. Application variables can be tested, allowing for very close monitoring of the applications behavior, e.g., unexpected values, convergence rates for optimization programs, threshold transgressions, etc.

A set of rules (RDR) is defined in a standard YAWL *exlet* attached to the task WT and defines the exception handler EX. It is composed here of a constraint checker CK, which is automatically tested when executing the task WT. A compensation action CP triggered when a constraint is violated and a notifier RE warning the user of the exception. This is used to implement resilience [26].

The constraint violations are defined by the users and are part of the standard exception handling mechanism provided by YAWL. They can attach sophisticated exception handlers in the form of specific *exlets* that are automatically triggered at runtime when particular user-defined constraints are violated. These constraints are part of the RDR attached to the workflow tasks.

Resilience is the ability for applications to handle unexpected behavior, e.g., erratic computations, abnormal result values, etc. It is inherent to the applications logic and programming. It is therefore different from systems or hardware errors and failures. The usual fault-tolerance mechanisms are therefore inappropriate here. They only cope with late symptoms, at best.

 ISBN: 978-1-61208-133-5

C. *Resilience*

Resilience is the ability for applications to handle unexpected behavior, e.g., erratic computations, abnormal result values, etc. It lies at the level of application logic and programming, not at systems or hardware level. The usual fault-tolerance mechanisms are therefore inappropriate here. They only cope with very late symptoms, at best.

New mechanisms are therefore required to handle logic discrepancies in the applications, most of which are only discovered at run-time [26].

It is therefore important to provide the users with powerful monitoring features and complement them with dynamic tools to evolve the applications according to the erratic behavior observed.

This is supported here using the YAWL workflow system so called "dynamic selection and exception handling mechanism". It supports:

- Application update using dynamically added rules specifying new codes to be executed, based on application data values, constraints and exceptions.
- The persistence of these new rules to allow applications to handle correctly future occurrences of the new case.
- The dynamic extension of these sets of rules.
- The definition of the new codes to be executed using the framework provided by the YAWL application specification tool: the new codes are just new workflows included in the global composite application specification.
- Component workflows invoke external programs written in any programming language through shell scripts, custom service invocations and Web Services.

In order to implement resilience, two particular YAWL features are used:

- Ripple-down-rules (RDR) which are handlers for exception management,
- Worklets, which are actions to be taken when exceptions or specific events occur.

The RDR define the decision process which is run to decide which worklet to use in specific circumstances.

D. *Distributed workflows*

The distributed workflow is based on an interface between the YAWL engine and the ProActive middleware (Figure 2). At the application level, users provide a specification of the simulation applications using the YAWL Editor. It supports a high-level abstract description of the simulation processes. These processes are decomposed into components which can be other workflows or basic workitems. The basic workitems invoke executable tasks, e.g., shell scripts or custom services. These custom services are specific execution units that call user-defined YAWL services. They support interactions with external and remote codes. In this particular platform, the external services are invoked through the middleware interface.

This interface delegates the distributed execution of the remote tasks to the ProActive middleware [17]. The middleware is in charge of the distributed resources allocation to the individual jobs, their scheduling, and the coordinated execution and result gathering of the individual tasks composing the jobs. It also takes in charge the fault-tolerance related to hardware, communications and system failures. The resilience, i.e., the application-level fault-tolerance is handled using the rules described in the previous Sections.

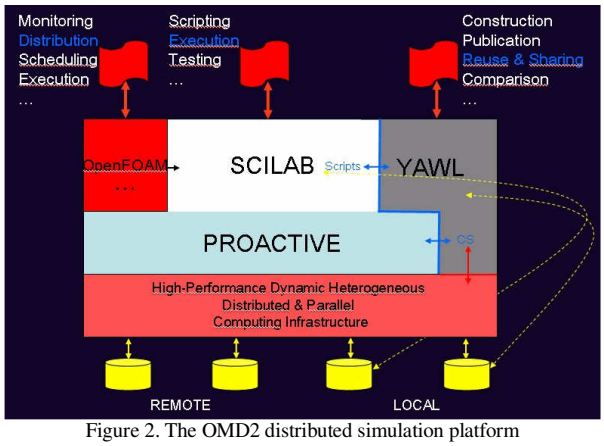


Figure 2. The OMD2 distributed simulation platform

The remote executions invoke the middleware functionalities through a Java API. The various modules invoked are the ProActive Scheduler, the Jobs definition module and the tasks which compose the jobs (Figure 3). The jobs are allocated to the distributed computing resources based upon the scheduler policy. The tasks are dispatched based on the job scheduling and invoke Java executables, possibly wrapping code written in other programming languages, e.g., Matlab, Scilab, Python, or calling other programs, e.g., CATIA, STAR-CCM+, ParaView, etc.

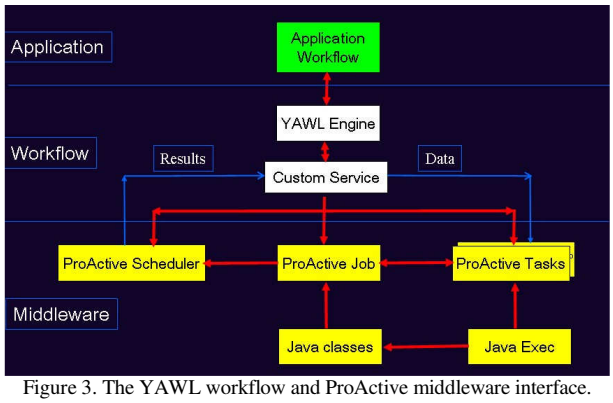


Figure 3. The YAWL workflow and ProActive middleware interface.

Optionally, the workflow can invoke local tasks using shell scripts and remote tasks using Web Services. These options are standard in YAWL.

 ISBN: 978-1-61208-133-5

E. Secured access

In contrast with the use of middleware, there is also a need to preserve and comply with the reservation and scheduling policies on the various HPC resources and clusters that are used. This is the case for national, e.g., IDRIS and CINES in France, and transnational HPC centers, e.g., PRACE in Europe.

Because some of the software run on proprietary resources and are not publicly accessible, some privileged connections must also be implemented through secured X11 tunnels to remote high-performance clusters (Figure 4). This also allows for fast access to software needing almost real-time answers, avoiding the constraints associated with the middleware overhead. It also allows running parallel optimization software on large HPC clusters. In this perspective, a both-ways SSH tunnel infrastructure has been implemented for the invocation of remote optimization software running on high-performance clusters and for fast result gathering.

Using the specific ports used by the communication protocol (5000) and YAWL (8080), a fast communication infrastructure is implemented for remote invocation of testcase optimizers between several different locations on a high-speed (40 GB/s) network at INRIA. This is also accessible through standard Internet connections using the same secured tunnels.

Current tests have been implemented monitoring from Grenoble in France a set of optimizers software running on HPC clusters in Sophia-Antipolis near Nice. The optimizers are invoked as custom YAWL services from the application workflow. The data and results are transparently transferred through secured SSH tunnels.

In addition t the previous interfaces, direct local access to numeric software, e.g., SciLab and OpenFOAM, is available through the standard YAWL custom services using the 8080 communication port and shell script invocations. Therefore, truly heterogeneous and distributed environments can be built here in a unified workflow framework.

F. Interfaces

To summarize, the simulation platform which is based on the YAWL workflow management system for the application specification, execution and monitoring, provides three complementary interfaces that suit all potential performance, security, portability and interoperability requirements of the current sophisticated simulation environments.

These interfaces run concurrently and are used transparently for the parallel execution of the different parts of the workflows. These interfaces are:

- The direct access to numeric software through YAWL custom services that invoke Java executables and shell scripts that trigger numeric software, e.g., OpenFOAM, and visualization tools, e.g., ParaView
- The remote access to high-performance clusters running parallel software, e.g., optimizers, through secured SSH tunnels, using remote invocations of custom services
- The access to wide-area networks through a grid middleware, e.g., ProActive, for distributed resource reservation and job scheduling

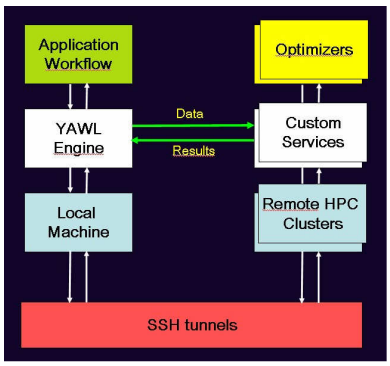


Figure 4. High-speed infrastructure for remote cluster access.

G. Service orchestration

The YAWL system provides a native Web service interface. This is a very powerful standard interface to distributed service execution, although it might impact HPC concerns. This is the reason why a comprehensive set of interfaces are provided by the platform (Section F, above).

Combined altogether and offered to the users, this rich set of functionalities is intended to support most application requirements, in terms of performance, heterogeneity and standardization.

Basically, an application workflow specifies general services orchestration. General services include here not only Web services, but also shell scripts, YAWL custom services implemented by Java class executables and high-level operators, as defined in the workflow control flow patterns of the Workflow Management Coalition [5, 21], e.g., AND-joins, XOR-joins, conditional branching, etc.

The approach implemented here therefore not only fulfills sound and semantically proved operators for task specification, deployment, invocation, execution and. synchronization. It also fulfills the requirements for heterogeneous distributed and HPC codes to be deployed and executed in a unified framework. This provides the users with high-level GUIs and hides the technicalities of distributed, and HPC software combination, synchronization and orchestration.

Further, because resilience mechanisms are implemented at the application level (Section C), on top of the middleware, network and OS fault-tolerance features, a secured and fault resilient HPC environment is provided, based on high-level constructs for complex and large-scale simulations.

The interface between the workflow tasks and the actual simulation codes can therefore be implemented as Web Services, YAWL custom services, and shell scripts

 ISBN: 978-1-61208-133-5

through secured communication channels. This is a unique set of possibilities offered by our approach (Figure 5).

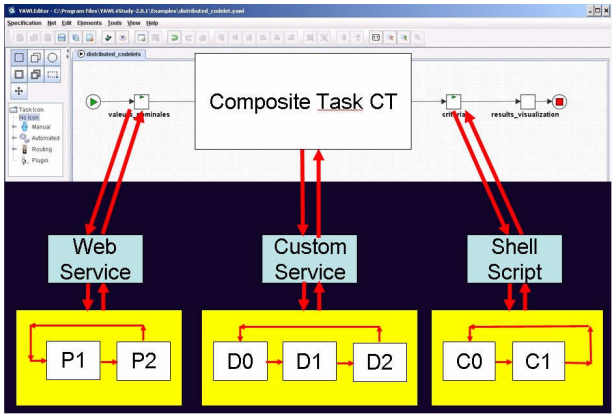


Figure 5. External services interfaces.

H. *Dataflow and control flow*

The dual requirements for the dataflow and control flow properties are preserved. Both aspects are important and address different requirements [6]. The control flow aspect addresses the need for user control over the workflow tasks execution. The dataflow aspect addresses the need for high-performance and parallel algorithms to be implemented effectively.

The control flow aspect is necessary to provide the users with global control over the synchronization and execution of the various heterogeneous and remote software that run in parallel to contribute to the application results. This is natively supported by YAWL.

The dataflow aspect is also preserved here in two complementary ways:

- the workflow data is transparently managed by the YAWL engine to ensure the proper synchronization, triggering and stopping of the tasks and complex operators among the different parallel branches of the workflows, e.g., AND joins, OR and XOR forks, conditional branching. This includes a unique YAWL feature called "cancellation set" that refers to a subset of a workflow that is frozen when another designated task is triggered [3]
- the data synchronization and dataflow scheme implemented by the specific numeric software invoked remain unchanged using a separation of concerns policy, as explained below

The various software with data dependencies that execute based on dataflow control are wrapped in adequate YAWL workflow tasks, so that the workflow engine does not interfere with the dataflow policies they implement.

This allows high-performance concerns to be taken into consideration along with the users concerns and expectations concerning the sophisticated algorithms associated with these programs.

Also, this preserves the global control flow approach over the applications which is necessary for heterogeneous software to cooperate in the workflow.

As a bonus, it allows user interactions during the workflow execution in order to cope with unexpected situations. This would otherwise be very difficult to implement because when unexpected situations occur while using a pure dataflow approach, it requires stopping the running processes or threads in the midst of possibly parallel and remote running calculations, while (possibly remote) running processes are also waiting for incoming data produced by (possibly parallel and remote) erratic predecessors in the workflow. This might cause intractable situations even if the errors are due to rather simple events, e.g., network data transfers or execution time-outs.

Note that so far, because basic tasks cannot be divided into remote components in the workflow, the dataflow control is not supported between remotely located software. This also avoids large uncontrolled data transfers on the underlying network. Thus, only collocated software, i.e., using the same computing resources or running on the same cluster, can use dataflow control on the platform. They are wrapped by workflow tasks which are controlled by the YAWL engine as standard workflow tasks.

For example, the dataflow controlled codes C0 and C1 depicted Figure 5 are wrapped by the composite task which is a genuine YAWL task that invokes a shell script to trigger them.

Specific performance improvements can therefore be expected from dataflow controlled sets of programs running on large HPC clusters. This is fully compatible with the control flow approach implemented at the application (i.e., workflow) specification level. Incidentally, this also avoids the streaming of large data collections of intermediate results through network connections. It therefore alleviates bandwidth congestion.

The platform interfaces are illustrated by Figure 5. Once the orchestration of local and distributed codes is specified at the application (workflow) level, their invocation is transparent to the user, whatever their localization.

I. *Experiments*

The current testcases include vehicle aerodynamics simulation (Figure 6) and air-conditioner pipes optimization (Figure 7). The distributed and heterogeneous platform is also tested with the Gmsh mesh generator (http://geuz.org/gmsh/) and the FAMOSA optimization suite developed at INRIA by project OPALE [34]. It is deployed on HPC clusters and invoked from remote workflows running on Linux workstations.

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

 ISBN: 978-1-61208-133-5

The input is a design vector and the output is a set of simulation results. The various components are invoked by shell scripts. FAMOSA is currently tested by the PSA Automotive Company and ONERA (the French National Aerospace Research Office) for aerodynamics problem solving.

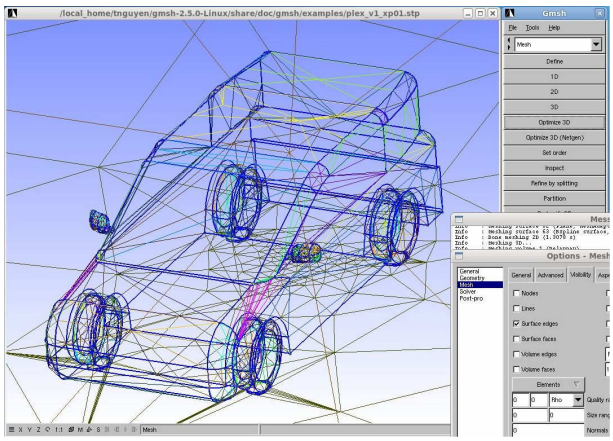


Figure 6. Vehicle mesh for aerodynamics simulation (Gmsh screenshot).

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.

The FAMOSA components are here triggered by remote shell scripts including PBS invocations 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.

Additionally, another experiment uses the distributed simulation platform for testing the heterogeneity of the application codes running on various hardware and software environments. It includes four remote computing resources that are connected by a high-speed network. One site is a HPC cluster. Another site is a standard Linux server. The two other sites are remote virtualized computing resources running Windows and Linux operating systems on different VirtualBox virtual machines that interface the ProActive middleware.

III. CONCLUSION

This paper presents an experiment for deploying a distributed simulation platform on grids. It uses a network of high-performance computers connected by a middleware layer. Users interact dynamically with the applications using a workflow management system. It allows them to define, deploy and control the application execution interactively.

In contrast with choreography of services, where autonomous software interact in a controlled manner, but where resilience and fault-tolerance are difficult to implement, the approach used here is an orchestration of heterogeneous and distributed software components that interact in a dynamic way under the user control [29]. This allows the dynamic interaction in case of errors and erratic application behavior. This approach is also fully compatible with both the dataflow and control flow approaches which are often described as poorly compatible [30, 31, 32] and are extensively used in numeric software platforms.

Because of the heterogeneity of the software and resources, the platform also combines secured access to remote HPC clusters and local software in a unified workflow framework.

This approach is also proved to combine in an elegant way the dataflow control used by many HPC software and the control flow approach required by complex and distributed application execution and monitoring.

A significant bonus of this approach is that the users can define and handle application failures at the workflow specification level. This means that a new abstraction layer is introduced to cope with application-level errors at run-time. Indeed, these errors do not necessarily result from programming and design errors. They may also result from unforeseen situations, data values and boundary conditions that were not envisaged at first. This is often the case for simulations, due to their experimental nature, e.g., discovering the behavior of the system being simulated.

This provides support for resiliency using an asymmetric checkpoint mechanism. This feature allows for efficient handling mechanisms to restart only those parts of the applications that are characterized by the users as necessary for overcoming erratic behavior.

Further, this approach can be evolved dynamically, i.e., when the applications are running. This uses the dynamic selection and exception handling mechanism in the YAWL workflow system. It allows for new rules and new exception handling to be added on-line if unexpected situations occur.

ACKNOWLEDGMENT

This work is supported by the French National Research Agency ANR (*Agence Nationale de la Recherche*), grant ANR-08-COSI-007, OMD2 project (*Optimisation Multi-Discipline Distribuée*).

REFERENCES

[1] Y. Simmhan, R. Barga, C. van Ingen, E. Lazowska and A. Szalay "Building the Trident Scientific Workflow Workbench for Data Management in the Cloud". In proceedings of the *3rd Intl. Conf. on Advanced Engineering Computing and Applications in Science*. ADVCOMP'2009. Sliema (Malta). October 2009. pp 41-50.

[2] A. Abbas, High Computing Power: A radical Change in Aircraft Design Process, In proceedings 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, In proceedings of the 6th *Intl. Conf. on Cooperative Design, Visualization and Engineering*. Luxemburg. September 2009. Springer-Verlag. *LNCS 5738*, pp. 341–350 (2009)

[4] A.H.M ter Hofstede, W. Van der Aalst, M. Adams and N. Russell, Modern Business Process Automation: YAWL and its support environment, *Springer* (2010).

 ISBN: 978-1-61208-133-5

[5] N. Russel, A.H.M ter Hofstede and W. Van der Aalst. Workflow Control Flow Patterns. A Revised View. Technical Report. University of Eindhoven (NL). 2006.

[6] E. Deelman and Y. Gil., Managing Large-Scale Scientific Workflows in Distributed Environments: Experiences and Challenges, In proceedings of the 2nd IEEE Intl. Conf. on e-Science and the Grid. Amsterdam (NL). December 2006. pp 131-139.

[7] SUN VirtualBox, User Manual, 2010. http://www.virtualbox.org.

[8] M. Ghanem, N. Azam, M. Boniface and J. Ferris, Grid-enabled workflows for industrial product design, In proceedings of the 2nd Intl. Conf. on e-Science and Grid Computing. Amsterdam (NL). December 2006. pp 88-92.

[9] G. Kandaswamy, A. Mandal and D.A. Reed, Fault-tolerant and recovery of scientific workflows on computational grids, In proceedings of the 8th Intl. Symp. On Cluster Computing and the Grid. 2008. pp 777-782.

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

[11] J. Wang, I. Altintas, C. Berkley, L. Gilbert and M.B. Jones, A high-level distributed execution framework for scientific workflows, In proceedings of the *4th IEEE Intl. Conf. on eScience*. Indianapolis (In). December 2008. pp 156-164.

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

[13] M. Adams, A.H.M ter Hofstede, W. Van der Aalst and N. Russell, Facilitating Flexibility and Dynamic Exception Handling in Workflows through Worklets, Technical report, Faculty of Information Technology, Queensland University of Technology, Brisbane (Aus.), October 2006.

[14] 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.

[15] L. Ramakrishnan et al., VGrADS: Enabling e-Science workflows on grids and clouds with fault tolerance. Proc. ACM SC'09 Conf. Portland (Or.), November 2009. pp 145-152.

[16] M. Baudin, Introduction to Scilab", Consortium Scilab. January 2010. Also: http://wiki.scilab.org/

[17] F. Baude et al., Programming, composing, deploying for the grid. in "GRID COMPUTING: Software Environments and Tools", Jose C. Cunha and Omer F. Rana (Eds), Springer Verlag, January 2006.

[18] http://edition.cnn.com/2009/TRAVEL/01/20/mumbai.overview last accessed: 07/07/2010.

[19] J. Dongarra et al. "The International Exascale Software Project Roadmap". University of Tennessee EECS Technical report UT-CS-10-654. May 2010. Available at: http://www.exascale.org/

[20] R. Gupta, et al. "CIFTS: a Coordinated Infrastructure for Fault-Tolerant Systems". Proc. 38th Intl. Conf. Parallel Processing Systems. Vienna (Au). September 2009.pp 145-154.

[21] The Workflow Management Coalition. http://www.wfmc.org

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

[23] 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. Florence (I). October 2010. pp 133-139.

[24] 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. Sorrento (I). LNCS 3726. September 2005. pp 265-278.

[25] http://www.yawlfoundation.org/software/documentation.The YAWL foundation. 2010.

[26] 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. Florence (I). October 2010. pp 321-329.

[27] Object Management Group / Business Process Management Initiative. BPMN Specifications. http://www.bpmn.org, last accessed: 12/01/2011.

[28] OASIS Web Services Business Process Execution Language. http://www.oasisopen.org/committees/tc_home.php?=wg_abbrev=wsbpel last accessed: 12/01/2011.

[29] Sherp G., Hoing A., Gudenkauf S., Hasselbring W. and Kao O. Using UNICORE and WS-BPEL for Scientific Workfow execution in Grid Environments. Proc. EuroPAR 2009. LNCS 6043. . Springer. 2010. pp 455-461.

[30] Ludäscher B.,Weske M., McPhillips T. and Bowers S. Scientific Workflows: Business as usual ? Proc. BPM 2009. LNCS 5701. Springer. 2009. pp 351-358.

[31] 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. WORKS 2009. Portland (Or). ACM 2009. pp 235-242.

[32] 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. WORKS 2009. Portland (Or). ACM 2009. pp 121-129.

[33] 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

[34] Duvigneau R. and Chandrashekaran P. A three-level parallelization strategy for robust design in aerodynamics. Proc. 20th Intl. Conf. on Parallel Computational Fluid Dynamics. May 2008. Lyon (F). pp 101-108.

 ISBN: 978-1-61208-133-5

Preemptive Channel Allocations for Cellular Networks with Multiple Sectors

Chia-Nan Lin and Tsang-Ling Sheu
Department of Electrical Engineering
National Sun Yat-Sen University
Kaohsiung, Taiwan
cnl@atm.ee.nsysu.edu.tw sheu@ee.nsysu.edu.tw

Abstract—**This paper presents preemptive channel allocations (PCA) for multiple-sector cellular networks, where directional antennas are used to divide the coverage of a cell into a number of same-sized sectors. When traffic in a sector unexpectedly increases, call blocking probability will increase accordingly. To remedy channel insufficient problem in a single sector, two aspects of channel preemptions are utilized. First, to reduce the blocking probability of new calls, the proposed PCA allows a new call to preempt an ongoing call when the ongoing call is located in the overlapping regions of two adjacent sectors or two neighboring cells. Second, the reserved channels not only can be used by the handoff calls, but also by the preempted calls. For the purpose of performance evaluation, we build an analytical model with four-tuple Markov chains. Numerical results show that the proposed PCA scheme improve the system performance in terms of the blocking and preemption probabilities.**

Keywords-Preemptive channel allocations; multiple sectors; cellular networks; blocking probability; Markov chains;

I. INTRODUCTION

Over the past decade, the rapid growth of the cellular technology (2G/3G/3.5G or even the upcoming 4G) has been proven that it can provide high reliability, stability, and ubiquity for personal communications [1]. A basic cellular network is composed of a base station (BS) and numerous mobile terminals (MTs). Channel capacity of a cellular network may become insufficient when MTs are attached to the network or moving between cells or sectors. There have been many previous works focused on the preemption mechanisms for cellular networks. A scheme called Adjusted Multimode Dynamic Guard Bandwidth (AM-DGB) [1] can temporarily block one or more lower-priority calls to guarantee longer connection time for higher-priority calls. A centralized and decentralized preemption algorithm was proposed by Lau *et al*. [3] for a connection-oriented network to minimize the service disruptions of ongoing calls. Recently, there were copious researches on sector-based cellular networks, such as WiMAX and LTE (Long Term Evolution)/LTE-A (Long Term Evolution-Advanced) networks. For example, to improve the throughput and capacity and to alleviate the inter-cell interference, numerous schemes on frequency reuse were proposed. Among them, Lei *et al*. [4] proposed a frequency reuse scheme to divide the available subcarriers into two groups, the super group used for the central region of a cell and the regular group used for the boundary of a cell. Similarly, Ali *et al*. [5] proposed the architecture of a dynamic fractional frequency reused (FFR) cell. Dynamic FFR scheme only partitions subcarriers into two physical groups. In our work, the main objective is to design a novel preemption scheme for ongoing calls residing in the overlapping areas of any two adjacent sectors. The remainder of this paper is organized as follows. Section II introduces the proposed channel preemption algorithms. Performance evaluation model is described in Section III. Section IV shows the analytical results along with discussions. Finally, Section V contains our concluding remarks.

II. PREEMPTIVE CHANNEL ALLOCATIONS

A. Sector-based Cellular Networks

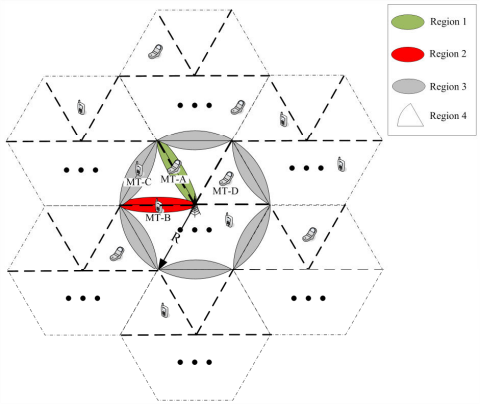


Figure 1. Generalized topology of a sector-based cellular network

A sector-based cellular network consists of multiple sectors divided by directional antennas. Figure 1 shows a generalized topology of a sector-based cellular network, consisting of one central cell and six neighboring cells. In the sector-based cellular network, we assume the cell has radius *R*. Due to the different coverage areas, an MT in a sector may reside in the following four regions. Region 1 (R1): MT resides in the clockwise overlapping region of

 ISBN: 978-1-61208-133-5

two adjacent sectors (e.g., MT-A). Region 2 (R2): MT resides in the counterclockwise overlapping region of two adjacent sectors (e.g., MT-B). Region 3 (R3): MT resides in the overlapping regions of two adjacent cells (e.g., MT-C). Region 4 (R4): MT resides in a sector other than the Regions of 1, 2 and 3 (e.g., MT-D). We design two different types of handoffs; (i) *Inter-sector handoff*: an MT originally residing in the sector is moving out to the neighboring sectors by passing through R1 or R2, and (ii) *Inter-cell handoff*: an MT originally residing in a cell is moving out to one of its six neighboring cells by passing through R3. A cell is divided into N_S sectors, and the *n*-th sector has channel capacity, C_T^n , among which certain amount of channels are purposely reserved for inter-sector/inter-cell handoff and preempted MTs, where *n*=1,2,…… N_S (counted in clockwise direction). Thus, the total channel capacity within a single cell is $C_T = \sum_{n=1}^{N_S} C_T^n$. Let C_{SR}^n represent the channels reserved by the *n*-th sector for inter-sector handoff calls and preempted calls resides in the overlapping region of two adjacent sectors. Let C_{CR}^n and C_{NR}^n represent the channels reserved by the *n*-th sector of the central cell and that of the neighboring cell respectively for inter-cell handoff and preempted calls resides in the overlapping region of two adjacent cells. Accordingly, in a cell, the total channels of $C_{SR} = \sum_{n=1}^{N_S} C_{SR}^n$, the total channels of $C_{CR} = \sum_{n=1}^{N_S} C_{CR}^n$, and the total channels of $C_{NR} = \sum_{n=1}^{N_S} C_{NR}^n$. As a result, in the central cell, the available channels of the *n*-th sector that can be assigned to new calls become $C_A^n = C_T^n - C_{SR}^n - C_{CR}^n$. Then the total available channels of the central cell are $C_A = \sum_{n=1}^{N_S} C_A^n$.

B. Channels for Odd/Even Sectors

To reuse the frequency spectrum, the total carriers in a cell can be divided into two subcarriers: $\{C_T/2, C_T/2\}$ for even number of sectors (e.g., N_S =2, 4, 6……), and $\{C_T/3, C_T/3, C_T/3\}$ for odd number of sectors (e.g., N_S =3, 5, 7……). Thus, for even sectors, $C_T^n = C_T/2$, and for odd sectors, $C_T^n = C_T/3$. Figure 2 illustrates the generalized cases of frequency reuse and channel allocations in a sector-based cellular network.

A preemptive channel allocations (PCA) algorithm is designed for the cellular network with multiple sectors. Under this assumption, three phases of channel preemption, *PCA-cws*, *PCA-ccs*, and *PCA-nbc*, could be invoked by an MT. They are explained one by one as below. *PCA-cws:* When the available channels in the *n*-th sector are used up, a new call generated in the sector can be blocked. However, *PCA-cws* can be invoked by the new call if: (i) one active MT residing in R1 is employing an available channel of the *n*-th sector, and (ii) at least one reserved channel of the clockwise neighboring sector C_{SR}^{n+1} is free, where $n+1 = \begin{cases} 1, \text{if } n = N_S \\ 2, 3, \ldots\ldots N_S, \text{otherwise} \end{cases}$. *PCA-ccs:* When the available channels in the *n*-th sector are used up, and there is no active MT residing in R1 or C_{SR}^{n+1} are also used up, a new call generated in the sector can be blocked because *PCA-cws* is not possible. However, *PCA-ccs* can be invoked by the new call if: (i) one active MT residing in R2 is employing an available channel of the *n*-th sector, and (ii) at least one reserved channel of the counterclockwise neighboring sector C_{SR}^{n-1} is free, where $n-1 = \begin{cases} N_S, \text{if } n = 1 \\ 1, 2, \ldots\ldots N_S - 1, \text{otherwise} \end{cases}$. *PCA-nbc:* When the available channels in the *n*-th sector are used up, and *PCA-cws* and *PCA-ccs* are not possible, then a new call generated in the sector can be blocked. However, *PCA-nbc* can be invoked by the new call if: (i) an active MT residing in R3 is employing an available channel of the *n*-th sector, and (ii) at least one channel of C_{NR}^n is free. We define the following four types of ongoing calls which currently use the available channels in a sector according to the four regions; (i) *i* = the number of ongoing calls which reside in R4 of a sector, (ii) *j* = the number of ongoing calls which reside in R1 of a sector, (iii) *k* = the number of ongoing calls which reside in R2 of a sector, and (iv) *l* = the number of ongoing calls which reside in R3 of a sector. In addition, we use five variables, c_u , c_v , c_{cs} , c_{cw} , and c_{cc} to represent channel increment or decrement in C_{NR}^n , C_{CR}^n , C_{SR}^n , C_{SR}^{n+1} , and C_{SR}^{n-1} , respectively.

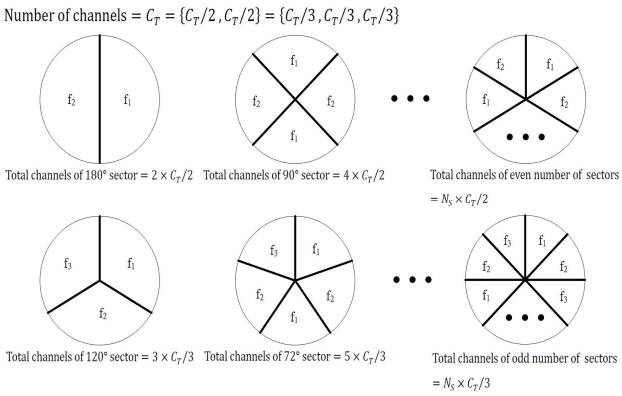


Figure 2. Channel allocations for odd/even sectors

III. PERFORMANCE EVALUATION MODEL OF PCA

In this section, we are interested in evaluating the proposed PCA algorithm on the sector-based cellular networks. Using 4-D (*i, j, k, l*) in a Markovian state, we can capture the characteristics of the proposed PCA.

A. Model Assumptions

The following assumptions are made in the analytical model: (i) it needs one and only one channel for an MT to become active; and (ii) the co-channel interference is ignored when an active MT resides in the overlapping regions of two adjacent sectors or cells. To facilitate our analysis, as shown in Figure 3, we approximate a single hexagon cell with six overlapping areas into an equivalent topology with two concentric circles [6], the outer circle

with radius $a \times R_{eq}$ and the inner circle with radius $b \times R_{eq}$, where $R_{eq} = \sqrt{\frac{3\sqrt{3}}{2\pi}} R \approx 0.91R$, $1 \le a \le 2$, and $0 \le b \le 1$. Hence, R_{eq} is the equivalent radius of the hexagon cell. By adjusting the parameters, *a* and *b*, we can enlarge or shrink the handoff area. As compared to Figure 1, R3 is converted to the area between the outer and the inner circle, and R4 is converted to the area of inner circle by excluding R1 and R2. If we let AR_1, AR_2, AR_3 and AR_4 denote the area ratio of R1, R2, R3 and R4 to the outer circle, respectively, we have $AR_1 = AR_2 = \frac{(\alpha/360°) \times (bR_{eq})^2 \pi}{A_A}$, $AR_3 = \frac{(\theta_T/360°) \times (a^2 - b^2) R_{eq}^2 \pi}{A_A}$, and $AR_4 = \frac{A_A - 2 \times \frac{\alpha}{360°} \times (bR_{eq})^2 \pi - \frac{\theta_T}{360°} \times (a^2 - b^2) R_{eq}^2 \pi}{A_A} = 1 - AR_1 - AR_2 - AR_3$.

Notice that the coverage area of a directional antenna is $A_A = (\theta_T/360°) \times (aR_{eq})^2 \pi$, the angle of a sector is $\theta_S = 360°/N_S$, θ_T is the transmission angle of a directional antenna. If α denotes the angle of two overlapping sectors, then $\alpha = \theta_T - \theta_S$.

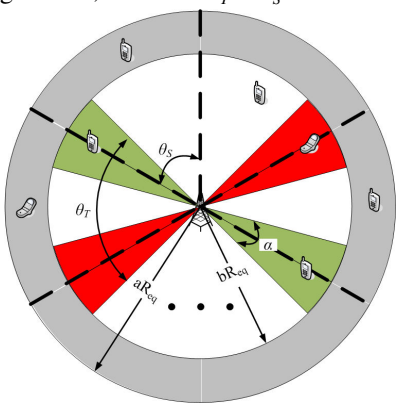


Figure 3. Equivalent topology of a single cell

B. *Markov Chains*

TABLE I. STATE TRANSITION RATES FOR ARRIVAL PROCESS

$A_1 = AR_4 \times \lambda_N$ (I.1)	$A_2 = AR_1 \times \lambda_N$ (I.2)	$A_3 = AR_2 \times \lambda_N$ (I.3)	$A_4 = AR_3 \times \lambda_N$ (I.4)
$A_5 = \begin{cases} \text{Eq.(III.2)} \\ \text{Eq.(III.2)} + \text{Eq.(IV.3)} \\ \text{Eq.(III.2)} + \text{Eq.(IV.6)} \end{cases}$ (I.5)	$A_6 = \begin{cases} \text{Eq.(III.4)} \\ \text{Eq.(III.4)} + \text{Eq.(IV.9)} \\ \text{Eq.(III.4)} + \text{Eq.(IV.11)} \\ \text{Eq.(III.4)} + \text{Eq.(IV.14)} \\ \text{Eq.(III.4)} + \text{Eq.(IV.16)} \end{cases}$ (I.6)	$A_7 = \begin{cases} \text{Eq.(III.7)} \\ \text{Eq.(III.7)} + \text{Eq.(V.9)} \\ \text{Eq.(III.7)} + \text{Eq.(V.10)} \\ \text{Eq.(III.7)} + \text{Eq.(V.11)} \\ \text{Eq.(III.7)} + \text{Eq.(V.12)} \\ \text{Eq.(III.7)} + \text{Eq.(V.13)} \\ \text{Eq.(III.7)} + \text{Eq.(V.14)} \\ \text{Eq.(III.7)} + \text{Eq.(V.15)} \\ \text{Eq.(III.7)} + \text{Eq.(V.16)} \end{cases}$ (I.7)
$A_8 = \begin{cases} \text{Eq.(IV.7)} \\ \text{Eq.(IV.12)} \end{cases}$ (I.8)	$A_9 = \begin{cases} \text{Eq.(III.5)} \\ \text{Eq.(III.5)} + \text{Eq.(V.1)} \\ \text{Eq.(III.5)} + \text{Eq.(V.2)} \\ \text{Eq.(III.5)} + \text{Eq.(V.3)} \\ \text{Eq.(III.5)} + \text{Eq.(V.4)} \end{cases}$ (I.9)	$A_{10} = \begin{cases} \text{Eq.(III.6)} \\ \text{Eq.(III.6)} + \text{Eq.(V.5)} \\ \text{Eq.(III.6)} + \text{Eq.(V.6)} \\ \text{Eq.(III.6)} + \text{Eq.(V.7)} \\ \text{Eq.(III.6)} + \text{Eq.(V.8)} \end{cases}$ (I.10)
$A_{11} = A_1, \forall\, i = i - 1$ (I.11)	$A_{12} = A_2, \forall\, j = j - 1$ (I.12)	$A_{13} = A_3, \forall\, k = k - 1$ (I.13)
$A_{14} = A_4, \forall\, l = l - 1$ (I.14)	$A_{15} = A_5, \forall\, j = j + 1$ (I.15)	$A_{16} = A_6, \forall\, k = k + 1$ (I.16)
$A_{17} = A_7, \forall\, l = l + 1$ (I.17)	$A_{18} = A_8, \forall\, k = k + 1$ (I.18)	
$A_{19} = A_9, \forall\, l = l + 1$ (I.19)	$A_{20} = A_{10}, \forall\, l = l + 1$ (I.20)	

A 4-D Markov chain model with states (*i, j, k, l*) is built to analyze the proposed PCA algorithm on a sector-based cellular network. The transition rates for arrival and departure processes are listed as in Table I and Table II, respectively.

TABLE II. STATE TRANSITION RATES FOR DEPARTURE PROCESS

$D_1 = i \times \mu$ (II.1)	$D_2 = j \times \mu + j \times \mu_{S1}$ (II.2)	$D_3 = k \times \mu + k \times \mu_{S2}$ (II.3)
$D_4 = l \times \mu + l \times \mu_H$ (II.4)	$D_5 =$ Eq. (III.8) (II.5)	$D_6 =$ Eq. (III.9) (II.6)
$D_7 =$ Eq. (III.10) (II.7)	$D_8 = \begin{cases} \text{Eq. (IV.1)} \\ \text{Eq. (IV.4)} \end{cases}$ (II.8)	$D_9 = \begin{cases} \text{Eq. (III.1)} \\ \text{Eq. (III.1)} + \text{Eq. (IV.2)} \\ \text{Eq. (III.1)} + \text{Eq. (IV.5)} \end{cases}$ (II.9)
$D_{10} = \begin{cases} \text{Eq. (III.3)} \\ \text{Eq. (III.3)} + \text{Eq. (IV.8)} \\ \text{Eq. (III.3)} + \text{Eq. (IV.10)} \\ \text{Eq. (III.3)} + \text{Eq. (IV.13)} \\ \text{Eq. (III.3)} + \text{Eq. (IV.15)} \end{cases}$ (II.10)	$D_{11} = D_1, \forall\, i = i + 1$ (II.11)	$D_{12} = D_2, \forall\, j = j + 1$ (II.12)
$D_{13} = D_3, \forall\, k = k + 1$ (II.13)	$D_{14} = D_4, \forall\, l = l + 1$ (II.14)	$D_{15} = D_5, \forall\, i = i + 1$ (II.15)
$D_{16} = D_6, \forall\, i = i + 1$ (II.16)	$D_{17} = D_7, \forall\, i = i + 1$ (II.17)	$D_{18} = D_8, \forall\, j = j + 1$ (II.18)
$D_{19} = D_9, \forall\, j = j + 1$ (II.19)	$D_{20} = D_{10}, \forall\, k = k + 1$ (II.20)	

The call departure rates from one region to another region in a sector and the inter-sector/inter-cell handoff rates are listed in Table III.

TABLE III. DEPARTURES RATES FOR CALLS MOVING BETWEEN REGIONS

Departure rates	Conditions	From/To	Eq.
$j \times \frac{\frac{\alpha}{\theta_T} \times AR_3}{\frac{\alpha}{\theta_T} \times AR_3 + 2 \times AR_4} \times \mu_{d1}$	$j > 0$	R1/R3	(III.1)
$j \times \frac{AR_4}{\frac{\alpha}{\theta_T} \times AR_3 + 2 \times AR_4} \times \mu_{d1}$		R1/R4	(III.2)
$k \times \frac{\frac{\alpha}{\theta_T} \times AR_3}{\frac{\alpha}{\theta_T} \times AR_3 + 2 \times AR_4} \times \mu_{d2}$	$k > 0$	R2/R3	(III.3)
$k \times \frac{AR_4}{\frac{\alpha}{\theta_T} \times AR_3 + 2 \times AR_4} \times \mu_{d2}$		R2/R4	(III.4)
$l \times \frac{AR_1}{2 \times (AR_1 + AR_2 + AR_3)} \times \mu_{d3}$	$l > 0$	R3/R1	(III.5)
$l \times \frac{AR_2}{2 \times (AR_1 + AR_2 + AR_3)} \times \mu_{d3}$		R3/R2	(III.6)
$l \times \frac{AR_4}{2 \times (AR_1 + AR_2 + AR_3)} \times \mu_{d3}$		R3/R4	(III.7)
$i \times \frac{AR_1}{AR_1 + AR_2 + \frac{\theta_S - \alpha}{\theta_S} \times AR_3)} \times \mu_{d4}$	$i > 0$	R4/R1	(III.8)
$i \times \frac{AR_2}{AR_1 + AR_2 + \frac{\theta_S - \alpha}{\theta_S} \times AR_3)} \times \mu_{d4}$		R4/R2	(III.9)
$i \times \frac{\frac{\theta_S - \alpha}{\theta_S} \times AR_3}{AR_1 + AR_2 + \frac{\theta_S - \alpha}{\theta_S} \times AR_3)} \times \mu_{d4}$		R4/R3	(III.10)

The call preemption probabilities under *PCA-cws* and *PCA-ccs* are listed in Table IV, and the call preemption probability under *PCA-nbc* is listed in Table V.

TABLE IV. PREEMPTION RATES FOR NEW CALLS UNDER *PCA-CWS* AND *PCA-CCS* WHEN $i + j + k + l = C_A^n$

Rates	Conditions	Regions	Phase	Eq.
$AR_2 \times \lambda_N$	$j > 0 \&\& c_{cw} > 0$	2	*PCA-cws*	(IV.1)
$AR_3 \times \lambda_N$		3		(IV.2)
$AR_4 \times \lambda_N$		4		(IV.3)
$AR_2 \times \lambda_N \times S^{PCA-cws}$	$j > 0 \&\& c_{cw} = 0$	2		(IV.4)
$AR_3 \times \lambda_N \times S^{PCA-cws}$		3		(IV.5)
$AR_4 \times \lambda_N \times S^{PCA-cws}$		4		(IV.6)
$AR_1 \times \lambda_N$	$k > 0 \&\& c_{cc} > 0$	1	*PCA-ccs*	(IV.7)
$AR_3 \times \lambda_N$	$j = 0 \&\& k > 0 \&\& c_{cc} > 0$	3		(IV.8)
$AR_4 \times \lambda_N$		4		(IV.9)
$AR_3 \times \lambda_N \times (1 - S^{PCA-cws})$	$c_{cw} = 0 \&\& k > 0 \&\& c_{cc} > 0$	3		(IV.10)
$AR_4 \times \lambda_N \times (1 - S^{PCA-cws})$		4		(IV.11)
$AR_1 \times \lambda_N \times S^{PCA-ccs}$	$k > 0 \&\& c_{cc} = 0$	1		(IV.12)
$AR_3 \times \lambda_N \times S^{PCA-ccs}$	$j = 0 \&\& k > 0 \&\& c_{cc} = 0$	3		(IV.13)
$AR_4 \times \lambda_N \times S^{PCA-ccs}$		4		(IV.14)
$AR_3 \times \lambda_N \times (1 - S^{PCA-cws}) \times$	$c_{cw} = 0 \&\& k > 0 \&\& c_{cc} = 0$	3		(IV.15)
$AR_4 \times \lambda_N \times (1 - S^{PCA-cws}) \times$		4		(IV.16)

TABLE V. PREEMPTION RATES FOR NEW CALLS UNDER *PCA-NBC*

WHEN $i+j+k+l=C_A^n$

Preempted rates	Conditions	Regions	Eq.
$AR_1 \times \lambda_N$	$k=0\&\&l>0\&\&c_u>0$	1	(V.1)
$AR_1 \times \lambda_N \times (1-S^{PCA-ccs})$	$c_{cc}=0\&\&l>0\&\&c_u>0$		(V.2)
$AR_1 \times \lambda_N \times S_{NR}^{PCA-nbc}$	$k=0\&\&l>0\&\&c_u=0$		(V.3)
$AR_1 \times \lambda_N \times (1-S^{PCA-ccs}) \times S_{NR}^{PCA-nbc}$	$c_{cc}=0\&\&l>0\&\&c_u=0$		(V.4)
$AR_2 \times \lambda_N$	$j=0\&\&l>0\&\&c_u>0$	2	(V.5)
$AR_2 \times \lambda_N \times (1-S^{PCA-cws})$	$c_{cw}=0\&\&l>0\&\&c_u>0$		(V.6)
$AR_2 \times \lambda_N \times S_{NR}^{PCA-nbc}$	$j=0\&\&l>0\&\&c_u=0$		(V.7)
$AR_2 \times \lambda_N \times (1-S^{PCA-cws}) \times S_{NR}^{PCA-nbc}$	$c_{cw}=0\&\&l>0\&\&c_u=0$		(V.8)
$AR_4 \times \lambda_N$	$j=0\&\&k=0\&\&l>0\&\&c_u>0$	4	(V.9)
$AR_4 \times \lambda_N \times (1-S^{PCA-ccs})$	$j=0\&\&c_{cc}=0\&\&l>0\&\&c_u>0$		(V.10)
$AR_4 \times \lambda_N \times (1-S^{PCA-cws})$	$c_{cw}=0\&\&k=0\&\&l>0\&\&c_u>0$		(V.11)
$AR_4 \times \lambda_N \times (1-S^{PCA-ccs}) \times (1-S^{PCA-cws})$	$c_{cw}=0\&\&c_{cc}=0\&\&l>0\&\&c_u>0$		(V.12)
$AR_4 \times \lambda_N \times S_{NR}^{PCA-nbc}$	$j=0\&\&k=0\&\&l>0\&\&c_u=0$		(V.13)
$AR_4 \times \lambda_N \times (1-S^{PCA-ccs}) \times S_{NR}^{PCA-nbc}$	$j=0\&\&c_{cc}=0\&\&l>0\&\&c_u=0$		(V.14)
$AR_4 \times \lambda_N \times (1-S^{PCA-cws}) \times S_{NR}^{PCA-nbc}$	$c_{cw}=0\&\&k=0\&\&l>0\&\&c_u=0$		(V.15)
$AR_4 \times \lambda_N \times (1-S^{PCA-ccs}) \times (1-S^{PCA-cws}) \times S_{NR}^{PCA-nbc}$	$c_{cw}=0\&\&c_{cc}=0\&\&l>0\&\&c_u=0$		(V.16)

To derive the state transition rates in Table I and Table II, first of all, we need to define the cell service rate and the handoff-area service rate by referring to [7] and [8]. In the model, we assume that the new-call arrival rate is a Poisson process with mean λ_N and the call duration time, T, is exponentially distributed with mean μ^{-1}. Let T_{d1}, T_{d2}, T_{d3}, and T_{d4} represent the dwell time of an ongoing call in R1, R2, R3 and R4, respectively. The service rates of four regions (denoted as μ_{d1}, μ_{d2}, μ_{d3} and μ_{d4}) can be computed as shown in Eq. (1).

$$\mu_{d1}=\mu_{d2}=\frac{2E[V]}{\pi \times bR_{eq}}\times\frac{360^\circ}{\alpha},\mu_{d3}=\frac{2E[V]}{(a-b)\times R_{eq}}\times\frac{360^\circ}{\theta_T},$$
$$\mu_{d4}=\frac{2E[V]}{\pi \times bR_{eq}}\times\frac{360^\circ}{\theta_S-\alpha} \quad (1)$$

When an MT resides in R1 or R2, the probability of moving out the overlapping region of two adjacent sectors is determined by the area ratio of R3 and R4. Thus, the inter-sector handoff rates of an MT residing in R1 and R2, denotes as μ_{S1} and μ_{S2}, can be derived from μ_{d1} and μ_{d2} directly. Similarly, when an MT resides in R3, the inter-cell handoff rate of an MT, μ_H, can be derived from μ_{d3} directly. That is,

$$\mu_{S1}=\frac{AR_4}{\frac{\alpha}{\theta_T}\times AR_3+2\times AR_4}\times\mu_{d1},\mu_{S2}=\frac{AR_4}{\frac{\alpha}{\theta_T}\times AR_3+2\times AR_4}\times\mu_{d2},$$
$$\mu_H=\frac{1}{2}\times\mu_{d3} \quad (2)$$

Let $S^{PCA-cws}$ be the successful-generation probability of a new call when available channels in the n-th sector and the reserved channels of the clockwise neighboring sector becomes zero, but at least one active MT resides in R1. Let $S^{PCA-ccs}$ be the successful-generation probability of a new call when available channels in the n-th sector are used up, *PCA-cws* cannot be invoked, and the reserved channels of the counterclockwise neighboring sector becomes zero, but at least one active MT resides in R2.We have

$$S^{PCA-cws}=\left(\frac{\mu^{-1}-\lambda_N^{-1}}{\mu^{-1}}\right)^{C_{SR}^{n+1}},$$
$$\text{if } (i+j+k+l=C_A^n)\,\&\&\,(j>0)\,\&\&\,(c_{cw}=0) \quad (3)$$
$$S^{PCA-ccs}=\left(\frac{\mu^{-1}-\lambda_N^{-1}}{\mu^{-1}}\right)^{C_{SR}^{n-1}},$$
$$\text{if } (i+j+k+l=C_A^n)\,\&\&\,[(j=0)\,\|\,(c_{cw}=0)]\,\&\&\,(k>0)\,\&\&\,(c_{cc}=0)$$

Likewise, let $S_{NR}^{PCA-nbc}$ be the successful-generation probability of a new call when available channels in the n-th sector are used up, *PCA-cws* and *PCA-ccs* cannot be invoked, and the reserved channels of the n-th sector in the neighboring cell becomes zero, but at least one active MT resides in R3. Let $S_{CR}^{PCA-nbc}$ be the successful-generation probability of a new call when available channels in the n-th sector are used up, *PCA-cws* and *PCA-ccs* cannot be invoked, and the reserved channels of a sector in the central cell becomes zero, but at least one active MT resides in R3. We have

$$S_{NR}^{PCA-nbc}=\left(\frac{\mu^{-1}-\lambda_N^{-1}}{\mu^{-1}}\right)^{C_{NR}^n},$$
$$\text{if } (i+j+k+l=C_A^n)\&\&[(j=0)\|(c_{cw}=0)]\&\&[(k=0)\|(c_{cc}=0)]\&\&(l>0)\&\&(c_u=0) \quad (4)$$
$$S_{CR}^{PCA-nbc}=\left(\frac{\mu^{-1}-\lambda_N^{-1}}{\mu^{-1}}\right)^{C_{CR}^n},$$
$$\text{if } (i+j+k+l=C_A^n)\&\&[(j=0)\|(c_{cw}=0)]\&\&[(k=0)\|(c_{cc}=0)]\&\&(l>0)\&\&(c_v=0)$$

An inter-sector call can use the reserved channels of a sector. Let F_{InterS}^{in} be the failure probability of an inter-sector call which moves from the neighboring sectors to the n-th sector. Similarly, let $F_{InterS}^{out\_cw}$ and $F_{InterS}^{out\_cc}$ be the failure probability of an inter-sector call which moves from the n-th sector to the clockwise sector and to the counterclockwise sector, respectively. We have

$$F_{InterS}^{in}=\left(\frac{\mu^{-1}-(\mu^{-1}-n_1\times\mu_{d1}^{-1}-n_2\times\mu_{d2}^{-1}-n_4\times\mu_{d4}^{-1})}{\mu^{-1}}\right)^{C_{SR}^n}$$
$$=\left(\frac{n_1\times\mu_{d1}^{-1}+n_2\times\mu_{d2}^{-1}+n_4\times\mu_{d4}^{-1}}{\mu^{-1}}\right)^{C_{SR}^n},\ \text{if } [(j>0)\|(k>0)]\&\&(c_{cs}=0)$$
$$F_{InterS}^{out\_cw}=\left(\frac{\mu^{-1}-(\mu^{-1}-n_1\times\mu_{d1}^{-1}-n_2\times\mu_{d2}^{-1}-n_4\times\mu_{d4}^{-1})}{\mu^{-1}}\right)^{C_{SR}^{n+1}} \quad (5)$$
$$=\left(\frac{n_1\times\mu_{d1}^{-1}+n_2\times\mu_{d2}^{-1}+n_4\times\mu_{d4}^{-1}}{\mu^{-1}}\right)^{C_{SR}^{n+1}},\ \text{if } (j>0)\&\&(c_{cw}=0)$$
$$F_{InterS}^{out\_cc}=\left(\frac{\mu^{-1}-(\mu^{-1}-n_1\times\mu_{d1}^{-1}-n_2\times\mu_{d2}^{-1}-n_4\times\mu_{d4}^{-1})}{\mu^{-1}}\right)^{C_{SR}^{n-1}}$$
$$=\left(\frac{n_1\times\mu_{d1}^{-1}+n_2\times\mu_{d2}^{-1}+n_4\times\mu_{d4}^{-1}}{\mu^{-1}}\right)^{C_{SR}^{n-1}},\ \text{if } (k>0)\&\&(c_{cc}=0)$$

For inter-handoff calls, let F_{InterC}^{out} be the failure probability of an inter-handoff call which moves from the central cell to the neighboring cell, and F_{InterC}^{in} be the failure probability of an inter-handoff call which moves from one of the neighboring cell to the central cell. F_{InterC}^{out} and F_{InterC}^{in} can be computed from Eq. (6).

$$F_{InterC}^{out}=\left(\frac{\mu^{-1}-(\mu^{-1}-n_1\times\mu_{d1}^{-1}-n_2\times\mu_{d2}^{-1}-n_3\times\mu_{d3}^{-1}-n_4\times\mu_{d4}^{-1})}{\mu^{-1}}\right)^{C_{NR}^n}$$
$$=\left(\frac{n_1\times\mu_{d1}^{-1}+n_2\times\mu_{d2}^{-1}+n_3\times\mu_{d3}^{-1}+n_4\times\mu_{d4}^{-1}}{\mu^{-1}}\right)^{C_{NR}^n},\ \text{if } (l>0)\&\&(c_u=0) \quad (6)$$
$$F_{InterC}^{in}=\left(\frac{\mu^{-1}-(\mu^{-1}-n_1\times\mu_{d1}^{-1}-n_2\times\mu_{d2}^{-1}-n_3\times\mu_{d3}^{-1}-n_4\times\mu_{d4}^{-1})}{\mu^{-1}}\right)^{C_{CR}^n}$$
$$=\left(\frac{n_1\times\mu_{d1}^{-1}+n_2\times\mu_{d2}^{-1}+n_3\times\mu_{d3}^{-1}+n_4\times\mu_{d4}^{-1}}{\mu^{-1}}\right)^{C_{CR}^n},\ \text{if } (l>0)\&\&(c_v=0)$$

Notice that, in Eq. (5) and (6), n_1, n_2, n_3, and n_4 denote the number of times which an ongoing may pass

through R1, R2, R3, and R4 respectively. In this paper, we let $n_1 = n_2 = n_3 = n_4 = 1$. Finally, let $\pi(i, j, k, l)$ be the steady-state probability in the 4-D Markov chain model. To analytically solve this model, we have to include the initial condition, as shown in Eq. (7), into the state-transition matrix, which can be derived from Tables II and III.

$$\sum_{l=0}^{C_A^n} \sum_{k=0}^{C_A^n - l} \sum_{j=0}^{C_A^n - l - k} \sum_{i=0}^{C_A^n - l - k - j} \pi(i, j, k, l) = 1 \quad (7)$$

C. Performance Metrics

Let P_{PCA}^n be the PCA preemption probability in the n-th sector. P_{PCA}^n can be computed as shown in Eq. (8). $P_{PCA-cws}^n$, $P_{PCA-ccs}^n$, and $P_{PCA-nbc}^n$ respectively represent the preemption probability of new calls under the operation of *PCA-cws*, *PCA-ccs*, and *PCA-nbc* in the n-th sector as shown in Eq. (9), Eq. (10), and Eq. (11), respectively.

$$P_{PCA}^n = P_{PCA-cws}^n + P_{PCA-ccs}^n + P_{PCA-nbc}^n \quad (8)$$

Where

$$P_{PCA-cws}^n = \begin{cases} \sum_{l=0}^{C_A^n} \sum_{k=0}^{C_A^n - l} \sum_{j=1}^{C_A^n - l - k} [\pi(C_A^n - j - k - l, j, k, l)], \text{if } c_{cw} > 0 \\ \sum_{l=0}^{C_A^n} \sum_{k=0}^{C_A^n - l} \sum_{j=1}^{C_A^n - l - k} [\pi(C_A^n - j - k - l, j, k, l) \times S^{PCA-cws}], \text{if } c_{cw} = 0 \end{cases} \quad (9)$$

$$P_{PCA-ccs}^n = \begin{cases} \sum_{l=0}^{C_A^n} \sum_{k=1}^{C_A^n - l} [\pi(C_A^n - k - l, 0, k, l)] \\ + \sum_{l=0}^{C_A^n} \sum_{k=1}^{C_A^n - l} \sum_{j=1}^{C_A^n - l - k} [\pi(C_A^n - j - k - l, j, k, l) \times (1 - S^{PCA-cws})], \text{if } c_{cc} > 0 \\ \sum_{l=0}^{C_A^n} \sum_{k=1}^{C_A^n - l} [\pi(C_A^n - k - l, 0, k, l) \times S^{PCA-ccs}] \\ + \sum_{l=0}^{C_A^n} \sum_{i=1}^{C_A^n - l} \sum_{i=1}^{C_A^n - l - k} [\pi(C_A^n - j - k - l, j, k, l) \times (1 - S^{ACP-cws}) \times S^{PCA-ccs}], \text{if } c_{cc} = 0 \end{cases} \quad (10)$$

$$P_{PCA-nbc}^n = \begin{cases} \sum_{l=1}^{C_A^n} [\pi(C_A^n - l, 0, 0, l)] + \sum_{l=1}^{C_A^n} \sum_{j=1}^{C_A^n - l} [\pi(C_A^n - j - l, j, 0, l) \times (1 - S^{PCA-cws})] \\ + \sum_{l=1}^{C_A^n} \sum_{k=1}^{C_A^n - l} [\pi(C_A^n - k - l, 0, k, l) \times (1 - S^{PCA-ccs})] \\ + \sum_{l=1}^{C_A^n} \sum_{k=1}^{C_A^n - l} \sum_{j=1}^{C_A^n - l - k} [\pi(C_A^n - j - k - l, j, k, l) \times (1 - S^{PCA-cws}) \times (1 - S^{PCA-ccs})], \text{if } c_u > 0 \\ \sum_{l=1}^{C_A^n} [\pi(C_A^n - l, 0, 0, l) \times S^{PCA-nbc}] + \sum_{l=1}^{C_A^n} \sum_{j=1}^{C_A^n - l} [\pi(C_A^n - j - l, j, 0, l) \times (1 - S^{PCA-cws}) \times S^{PCA-nbc}] \\ + \sum_{l=1}^{C_A^n} \sum_{k=1}^{C_A^n - l} [\pi(C_A^n - k - l, 0, k, l) \times (1 - S^{PCA-ccs}) \times S^{PCA-nbc}] \\ + \sum_{l=1}^{C_A^n} \sum_{k=1}^{C_A^n - l} \sum_{j=1}^{C_A^n - l - k} [\pi(C_A^n - j - k - l, j, k, l) \times (1 - S^{PCA-cws}) \times (1 - S^{PCA-ccs}) \times S^{PCA-nbc}], \text{if } c_u = 0 \end{cases} \quad (11)$$

Let P_{nb}^n be the new-call blocking probability in the n-th sector as shown in Eq. (12). Basically, P_{nb}^n consists of two terms which describe *PCA* cannot be invoked. The first term represents the probability that there is no ongoing call residing in R3, and the second term represents the probability that the reserved channels of the neighboring cell become zero.

$$P_{nb}^n = \begin{cases} \sum_{i=0}^{C_A^n} [\pi(i,0,0,0)] + \sum_{j=1}^{C_A^n} [\pi(C_A^n - j, j, 0, 0) \times (1 - S^{ACP-cwt})] \\ + \sum_{k=1}^{C_A^n} [\pi(C_A^n - k, 0, k, 0) \times (1 - S^{ACP-ccr})] \\ + \sum_{k=1}^{C_A^n} \sum_{j=1}^{C_A^n - k} [\pi(C_A^n - j - k, j, k, 0) \times (1 - S^{ACP-cwt}) \times (1 - S^{ACP-ccr})] \end{cases} \quad (12)$$
$$+ \begin{cases} \sum_{l=1}^{C_A^n} [\pi(C_A^n - l, 0, 0, l) \times (1 - S^{ACP-nbc})] + \sum_{l=1}^{C_A^n} \sum_{j=1}^{C_A^n - l} [\pi(C_A^n - j - l, j, 0, l) \times (1 - S^{ACP-cwt}) \times (1 - S^{ACP-nbc})] \\ + \sum_{l=1}^{C_A^n} \sum_{k=1}^{C_A^n - l} [\pi(C_A^n - k - l, 0, k, l) \times (1 - S^{ACP-ccr}) \times (1 - S^{ACP-nbc})] \\ + \sum_{l=1}^{C_A^n} \sum_{k=1}^{C_A^n - l} \sum_{j=1}^{C_A^n - l - k} [\pi(C_A^n - j - k - l, j, k, l) \times (1 - S^{ACP-cwt}) \times (1 - S^{ACP-ccr}) \times (1 - S^{ACP-nbc})] \end{cases}$$

IV. Numerical Results

TABLE VI. PARAMETERS USED IN THE ANALYTICAL MODEL

Parameters	Values
Total channel capacity in a cell (C_T)	36
C_{SR}^n, C_{CR}^n, C_{NR}^n	1, 3
Call duration time ($1/\mu$)	500 sec
Distance from the hexagon center to any vertex (R)	1000 m

The parameters and values listed in Table VI were used when running the MATLAB tool. To investigate the impact of the traffic in the networks, we define traffic load as $\rho = \lambda_N / C_A^n \mu$. Figure 4 shows the new-call blocking probability as ρ increases from 0.3 to 1.2. It is interesting to notice that new-call blocking probabilities can be significantly reduced under the cell with four or five sectors due to the expanded available channels. There is another phenomenon worthy to observe that new-call blocking probability will be slightly increased when the speed of MT decreases from 10 to 50 km/h. The reason is because that by referring to Eq. (1), low-speed MT will increase the dwell time in the sector. In other words, the channel occupancy time of low-speed MT (e.g., V = 10 km/h) is much longer than that of high-speed MT (e.g., V = 50 km/h).

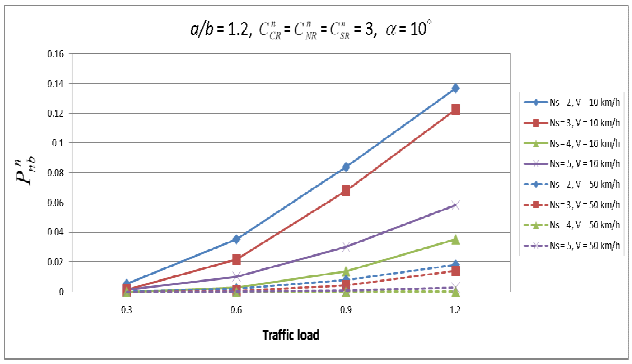


Figure 4. New-call blocking probability versus traffic load

Figure 5 shows the preemption probability of PCA in the n-th sector (P_{PCA}^n) as the angle of two overlapping sectors increases from 10° to 25°. It is observed that P_{PCA}^n is increased more rapidly as α increases when a cell is divided into five sectors. This is because increasing α in a cell with more sectors (e.g., $N_S = 5$) has higher possibility to let the MT residing in R1 be preempted than

with less sectors (e.g., $N_S = 2$). Another interesting phenomenon is that P_{PCA}^n is continuous increased when the reserved channels becomes more (e.g., $C_{SR}^n = 3$) in a sector, because the preempted calls have higher possibility to sustain their connections when PCA mechanism is invoked.

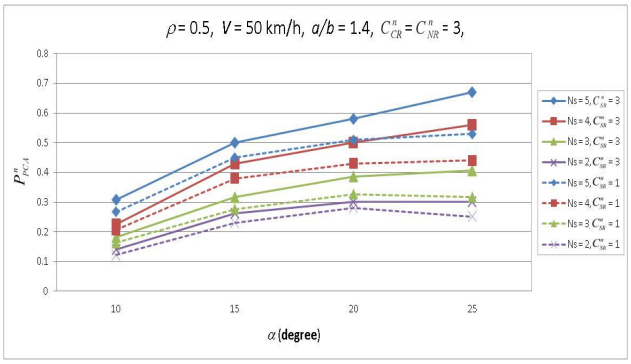


Figure 5. Preemption probability of PCA versus the angle of two overlapping sectors

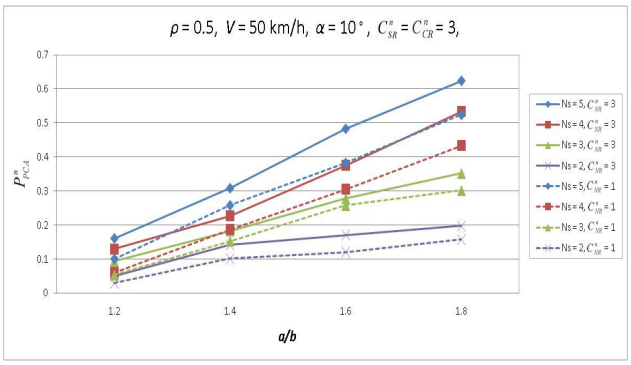


Figure 6. Preemption probability of PCA versus the ratio *a/b*

Figure 6 shows the preemption probability of PCA in the *n*-th sector is increased as the inter-cell overlapping region ratio increases from 1.2 to 1.8. In the figure, it is observed that P_{PCA}^n at $N_S = 5$ is higher than that at $N_S = 2$. This result reveals that more number of sectors have higher possibility to invoke the preemption scheme because of the more overlapping regions. We can also observe that by simply increasing C_{NR}^n from 1 to 3 can significantly increase the preemption probability. It should be noticed that although incrementing C_{NR}^n can increase the increasing preemption probability, it may adversely increase the new-call blocking probability, since available channels in the neighboring cell could be reduced.

Finally, let us investigate the average speed of MT versus P_{PCA}^n when the overlapping regions are changed. Figure 7 shows the preemption probability of PCA is decreased as the speed of MT increases from 20 km/h to 80 km/h due to the channel occupancy time (by referring to Eq. (1)). By fixing $N_S = 4$ and $C_{SR}^n = C_{CR}^n = C_{NR}^n = 3$, we can observe that the reserved channels are quite enough for the preempted calls to execute PCA mechanism. Thus, when the overlapping regions of two sectors or cells are increased, P_{PCA}^n is still increased.

V. CONCLUSIONS

This paper has presented an analytical model of adaptive channel preemption (PCA) for sector-based cellular networks. Three different preemption phases, *PCA-cws*, *PCA-ccs*, and *PCA-nbc* were proposed to fully utilize the capacity of the cellular networks with multiple sectors. One of the novelties presented in this paper is right in that the proposed PCA allows a new call to preempt an ongoing call when the latter is located in the inter-sector or inter-cell overlapping region. Analytical results have revealed two annotations: (i) the reserved channels can not only used by the inter-sector/inter-cell handoff calls but also used by the preempted calls, and (ii) the low-speed MT makes more impact on the new-call blocking probability than the high-speed MT due to the longer channel occupancy time.

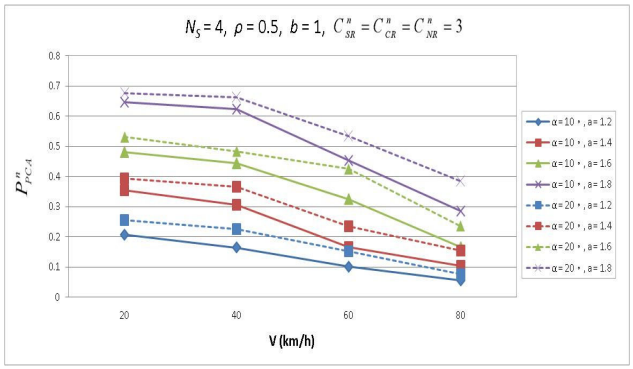


Figure 7. Preemption probability of PCA versus speed of MT

REFERENCES

[1] H. Hu, J. Zhang, X. Zheng, Y. Yang, and P. Wu, "Self-configuration and self-optimization for LTE networks," IEEE Communications Magazine, Vol. 48, Issue 2, Page(s): 94–100, Feb. 2010.

[2] O. Yu, E. Saric, and A. Li, "Adaptive Prioritized Admission over CDMA," IEEE Wireless Communications and Networking Conference, Vol. 2, Page(s): 1260–1265, Mar. 13–17, 2005.

[3] C. H. Lau, B.-H. Soong, and S. K. Bose, "Preemption With Rerouting to Minimize Service Disruption in Connection-Oriented Networks," IEEE Transactions on Systems, Man and Cybernetics, Vol. 38, Issue 5, Page(s): 1093–1104, Sep. 2008.

[4] H. Lei, X. Zhang, and D. Yang, "A Novel Frequency Reuse Scheme for Multi-Cell OFDMA Systems," IEEE 66th Vehicular Technology Conference, VTC-2007 Fall, Sep. 30–Oct. 3, 2007.

[5] S. H. Ali, V. C. M. Leung, "Dynamic frequency allocation in fractional frequency reused OFDMA networks," IEEE Transactions on Wireless Communications, Vol. 8, Issue 8, Page(s): 4286–4295, Aug. 2009.

[6] T.-L. Sheu and J.-H. Hou, "On the Influences of Enlarging and Shrinking the Soft Handoff Coverage for a cellular CDMA System," Journal of Information Science and Engineering (JISE), Vol. 23, No. 5, Page(s): 1453–1467, Sep. 2007.

[7] J. Wang, Q.-A. Zeng and D. P. Agrawal, "Performance analysis of a preemptive and priority reservation handoff scheme for integrated service-based wireless mobile networks," IEEE Transactions on Mobile Computing, Vol. 2, Page(s): 65–75, Jan.–Mar. 2003.

[8] W. Li, H. Chen, and D. P. Agrawal, "Performance analysis of handoff schemes with preemptive and nonpreemptive channel borrowing in integrated wireless cellular networks," IEEE Transactions on Wireless Communications, Vol. 4, Issue 3, Page(s): 1222–1233, May 2005.

 ISBN: 978-1-61208-133-5

Ubiquitous Home-Based Services

Jean-Charles Grégoire
INRS-EMT
Montreal, CANADA
gregoire@emt.inrs.ca

***Abstract*—Remotely accessing services or content based at home is increasingly required as high speed wireless networks become more widespread and mobile terminals more capable. Still, providing such access in reliable and secure fashion presents challenges, especially since media is involved. We explore here how this can be done in a SIP-based framework, taking into account more recent developments in media architectures such as IMS, from extension to home-based (DSL, cable) access to new means of exchanging information between end users through messaging. We demonstrate how MSRP is used to that effect.**

Keywords - SIP; SDP; MSRP; Home monitoring; Home Services.

I. Introduction

Cheap, ubiquitous Internet access, from hotspots to greater affordability of wireless (3G and beyond) services, means that users can be connected to the Internet in almost continuous fashion. This however does not translate into universal access to services as specific, remote access terminals (e.g. RIM's blackberry) remain the norm. We thus tend to see the creation of mobile-device specific variants of common services, or services created specifically for mobility (again, RIM's service)[2], [6]. Even for newer devices (e.g. the iPhone), there tends to be a distinction between a hotspot-based use and a cellular-based use.

We can argue that there are really two different markets at play here, one based on the mobile terminal, "always" connected, the other that of the mobile computer, served by hotspots in a context such that, for the user, connectivity is indistinguishable from the home network, at least as long as massive data transfers are not involved. In the case of the mobile terminals, the restrictions in the nature of the service which can be accessed are manifest: while some are infrastructure-based, most applications are essentially terminal-based, with simple client-server behaviour, and activated on demand or periodically, typically the "app" market for new devices. In either case, user to user (IP–based) communications are elusive.

The emergence of middleware for mobile services, such as the IP Multimedia Subsystem (IMS) puts another twist on this issue, as they allow the creation of new services with proper mechanisms to overcome restrictions that mobility and/or restricted bandwidth access can impose. These operator-based services can come in competition with Internet-based services and this is actually a topic of some controversy, although this is not our focus here.

Both models, Internet-based service specific or operator-based middleware multi–service, present restrictions in the delivery of services. In the first case, we depend on a silo model, where only services deployed on Internet servers are available, with little—or proprietary—means for extension. While this model serves some applications such as social networks or personal communications rather well, it has clear limits in terms of integration (e.g., [3], [9]. The middleware-based model is more flexible in that respect, but users themselves usually have no possibility to provide personal extensions. In both cases, access to personal information is quite limited, restricted to repositories, or confined within applications (e.g. pictures).

Our focus here is on providing access to home-based services or information from remote terminals, as well as allowing home-based applications to communicate remotely with owners, in a secure way, where both parties can mutually authenticate and protect their communications.

In this paper, we show how current SIP–related features actually provide most of the required support for such services, with minimal extra effort. Such an approach has advantages over network-based services as it can more easily enable direct (user to user) communications. It also avoids holding information in the network for the user, which may have security and legal ramifications. Finally, it also allows us to take advantage of established mechanisms to bypass devices which restrict communications.

In section II, we start with an overview of the different elements upon which our argument is built. Section III presents our view of home-based services. Section IV discusses all the issues which need to be resolved. Section V illustrates how messaging mechanisms can be used to transport various forms of date. Our solution is discussed in Section VI and we draw our conclusions in Section VII.

II. Background

In this section we present the key elements required to understand the foundations of our work. We assume that the reader will be familiar with most of the technological underpinnings and we keep this presentation succinct.

A. Home Monitoring Services

Remote access to home services from a wireless terminal is hardly a new concept. For example, we find in [6] a description of the use of off–the–shelf protocols and programming tools to implement alarm monitoring. More recently, wireless operators

 ISBN: 978-1-61208-133-5

have started to offer such services, again centred around monitoring and alarms. In AT&T's case[2], for example, the application was proprietary and required users to deploy specific hardware, which included a remotely operable video camera and various sensors. Motion, door and window activity, water leakage, and temperature changes are cited as common examples.

Building a home sensor network is certainly no longer a challenge, and it is also straightforward to program alarms based on monitored values. The issue is rather the interconnection of this network — or a home-based driving application — with the remote user. In AT&T's case, the application had a web interface and the user had to connect to the server remotely via IP to access the services, essentially enabling access to a web server from any terminal, including cellular phones with such a capacity.

However, while conceptually straightforward, remote access to home–based servers is blocked by many operators, and IP addresses may change through time. Furthermore such a form of remote–access is open to various forms of attacks, as typically befalls web servers.

B. Other services

While sensors/actuators and video surveillance are the most often cited examples of home applications, there are many other possibilities we can imagine, such as access to various forms of content, including audio and video, or pictures. Such access can take different forms, as we shall see later. Accessing content directly from the home is important to alleviate such issues as protecting copyrighted, personal or sensitive information.

C. SIP & SDP

The Session Initiation Protocol (SIP)[10] is the foundation of media services. SIP is a signalling protocol which supports negotiation of parameters for the establishment of an end-to-end session for multimedia communications. The Session Description Protocol (SDP)[7] is used to present parameters.

While SIP was originally proposed for multimedia services, we must take notice that it resolves many issues that arise in home connectivity and enriched, interactive end-to-end communications. The challenge is to identify whether it offers all the flexibility we need for home services and, in the next section, we clarify our expectations in that respect.

III. Home Services

There is no single definition of what home services can be, so we must define what we mean in this context. We have seen earlier examples of monitoring, alarms and surveillance. We broaden this definition with entertainment. We must insist here that we focus on remote services, namely services which must be accessible (but not exclusively) remotely.

Figure 1 presents a schematics view of home services and their connection to the outside world. We consider a network for home devices, with possibly separate dedicated networks for sensors based on proximity technology such as variants of 802.15 (Bluetooth, Zygbee). Communications with the Internet go through a gateway device, which acts as an SIP User Equipment (UE); this device would integrate other functions described below.

Note also that we can have internal home communications as well as communications between the home and an external user. Home communications can be device to device, device to person or person to device. These communications need not be SIP-based, and can be supported through proprietary means. We shall come back to this issue later.

A. Remote services

Home Monitoring: Monitoring is a classical example of remote home automation. This includes remotely receiving alarms notification, reading sensor, setting actuators but also possibly reading documents, such as a shopping list of a family memo and receiving a video stream.

From an Internet-based service perspective, such services do not present many challenges. Access and security are the key issues, but the functionality required to manipulate sensors and actuators and the network resources required are readily available.

Home Entertainment: We mean here access to media sources, such as music and video, from a home server, not unlike what is achievable through Apple's iTunes software in a LAN.

Such an offering is more challenging. We need to be able to browse directories and activate transmission of a specific content. It may be necessary to choose a suitable codec—or suitable parameters/profile—for the medium. Depending on the quality desired, as well as the degree of interactivity required, bounded bandwidth and delay constraints may exist.

B. Some support

We require to make some assumptions about support functions for these services.

Connectivity: We assume that all services are supported by a home IP network, wired and/or wireless. Monitoring devices on a wireless sensor network could be accessible indirectly, i.e. through a control centre which itself would be part of a home network.

Access: For uniformity, we suppose that internal/external access to services is organized through a home-based portal. It receives requests and redirects them to the appropriate device and answers back to the query device. It must also keep track of whether requests are internal—within the home, or external. In the latter case, it would also have to act as a relay for media communications.

Presence: Because alarms are to be sent unrequested, it is important to know whether the user is inside or outside the home to notify her with the suitable means. The portal must therefore also register presence information for the user and forward requests accordingly. Our assumption is that, unless the user is registered internally, the portal will attempt to reach her externally. In any case, all events will always be logged and the logs available for consulting.

 ISBN: 978-1-61208-133-5

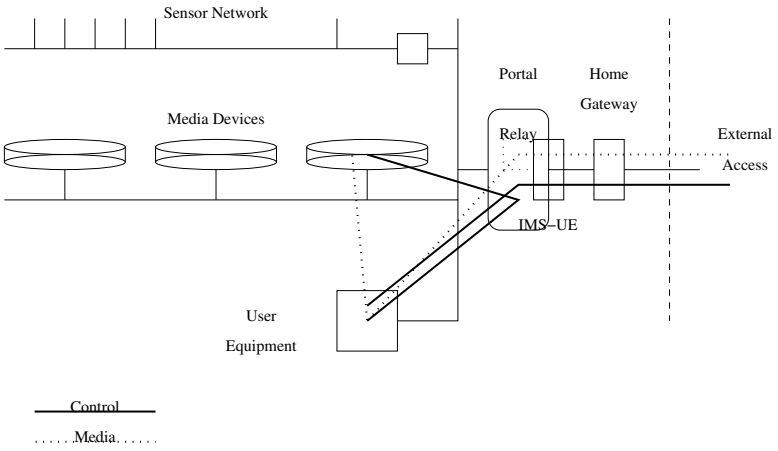


Figure 1. Home Services

Inter-Networking: To allow external access, the portal must be reachable from the outside network. As we have seen in examples above, this can sound like a simple statement, but there are practical limits: network connectivity is one, as is security and possibly quality of service.

Networking restrictions can be imposed by the operator or also the networking equipment used to connect the home network to the Internet: use of private addresses or firewall blocking impose typical limits. Networking devices will typically allow traffic to originate from the home network towards the Internet, but block incoming requests for connection.

When private addresses are used translation devices (i.e. NAT boxes) may impose restrictions on application traffic. It is necessary to translate private to public addresses, including communication ports. If we have SIP traffic, for example, this requires that its SDP content be modified.

IV. Issues

The simple challenge we are confronted with is to provide access to home services remotely. Some would argue that it could easily be done directly, in a typical Internet end–to–end (e2e) model, but there are many practical restrictions to such a model and we propose that there are benefits to take advantage of the access to the IMS infrastructure and its features. Most important, network-based support is required to circumvent restrictions imposed by the presence of middleboxes, which we have mentioned before but revisit below in closer relationship with SIP/SDP. Beyond transport-level connectivity, we must also consider user to user connectivity, i.e. that either home and user can initiate communications at any time.

A. Middleboxes

Middleboxes are network devices which impair communications in some way, either for security reasons, such as firewalls, or for address reuse, such as NATs. Each create specific problems. In the first case, TCP connection establishment can be blocked in one direction and authorized in the other. Still, once established, traffic can freely flow on both directions although this may require modification of signalling content—in our case SDP bodies.

Extensions to SDP provide support to help alleviate the problem; they are specific annotations in SDP bodies which are read and possibly manipulated by middleboxes. For TCP transport, it is possible to set an attribute (*a=setup:*) with the values of *active*, *passive*, *actpass* or *holdconn*. These values announce whether or not the end point can set up the connection or not, does not care one way or the other (actpass), or whether the establishment should be suspended for the time being.

Another attribute, *a=connection:*, allows to specify whether a new connection must be established or an existing one can be reused. It supports modifying the parameters of an established connection without having to tear down and re-establish a new TCP session.

We must note that the protocol does not support the establishment of several TCP connections for the same medium. On the other hand, the secured form of TCP, TLS, is also available for transport.

B. Presence & Reachability

Alarms are sent from the home to the user and the user can contact her home to access sensor status and media. This requires that:

- User and home must have names well-known to each other,
- The home knows whether the user is “present” in the network and,
- Both user and home can initiate connections, which implies that,
- Both user and home can access each other’s address.

Names are important because home and user need to be able to reach each other, i.e. initiate data transfer. This is done trivially if both are customers of the same service network, but generalized with URIs. Presence should also indicate whether the user is reachable at all.

 ISBN: 978-1-61208-133-5

C. Relays

Middlebox traversal can be sufficient to achieve user to network communications, but may not be sufficient to achieve end to end connectivity, e.g. if both end users are behind firewalls. In this case, application relays in the network have to be used. Such architecture is commonly used by services such as Skype[12] and are also fundamental to the architecture of the Asterisk[11] soft PBX.

The use of such relays raise several issues of security. They require proper authentication, protection against hijacking or DoS. Note that there is also a chicken and egg issue at work here: To enable a relay, there must be a way to discover it to force its presence on exchanges. This can be done through a separate discovery process, or through registration mechanisms *à la* SIP: either communications are permanently enabled between both ends, or an enabling signalling channel is established which allows to negotiate and setup proper connections.

The relay may provide added value to the communication. Minimally, it can be buffering and flow control, in case of mismatched performance in the links. Media conversion (transcoding) can also be performed.

D. Information transmission

The remaining issue is the transmission of information from end to end. This includes:

- Commands and values for sensors and alarms;
- Menu, menu selection;
- A/V streaming and streaming control, e.g. play/pause.

A protocol is therefore required to carry this information.

V. MSRP

The Message Session Relay Protocol (MSRP[4]) is a protocol to support session-oriented instant messaging. It is text-based, connection-oriented and supports exchange of arbitrary (binary) MIME-encoded content. Unlike SIP's page–mode messages, MSRP allows messages of any length and structure.

Unlike other messaging protocols, MSRP is integrated with SIP and its offer-answer mechanism, and thus blends naturally into IMS. Note here that we have three protocols present in MSRP exchanges:

- SIP carries the information required to negotiate the exchange between endpoints, possibly through relays;
- SDP is used to capture this information, including data format, ports, transport used, etc.;
- MSRP formats the IM messages, supports chunks, fragmentation, success reports, etc.

The specific use of SDP and MSRP is illustrated below.

A. Basic MSRP Operations

The following example, borrowed from [4], illustrates key elements of the use of MSRP; it is a typical first step in a SIP transaction between Alice and Bob.

```
INVITE sip:bob@biloxi.example.com SIP/2.0
To: <sip:bob@biloxi.example.com>
From: <sip:alice@atlanta.example.com>;tag=786
Call-ID: 3413an89KU
Content-Type: application/sdp

c=IN IP4 atlanta.example.com
m=message 7654 TCP/MSRP *
a=accept-types:text/plain
a=path:
  msrp://atlanta.example.com:7654/jshA7weztas;tcp
```

The c field sets the address (Internet, IPv4) of the source point. The m field specifies an IM protocol, based on MSRP, and the port used for communications. The a fields contain MSRP-specific information, including encoding supported. The presence of "path" information in mandatory.

The field values "TCP/MSRP" and "TCP/TLS/MSRP" have been added to the SDP protocol for explicit support of MSRP. They support two forms of transport for MSRP content, one plain TCP the other one encrypted.

Note that, with MRSP and unlike other use of SDP, the attributes—and more specifically the a=path attributes—rather than the information contained on the c and m lines are to be used to determine where to connect. Also note that a TCP connection can be used for several different transfers.

Bob's answer could be the following:

```
SIP/2.0 200 OK
To: <sip:bob@biloxi.example.com>;tag=087js
 From: <sip:alice@atlanta.example.com>;tag=786
Call-ID: 3413an89KU
Content-Type: application/sdp

c=IN IP4 biloxi.example.com
m=message 12763 TCP/MSRP *
a=accept-types:text/plain
a=path:msrp://biloxi.example.com:12763/
   kjhd37s2s20w2a;tcp
```

The answer contains Bob's contact information which matches Alice's, i.e. IP address (or name) and port, together with protocol.

And Alice's final answer:

```
ACK sip:bob@biloxi SIP/2.0
To: <sip:bob@biloxi.example.com>;tag=087js
From: <sip:alice@atlanta.example.com>;tag=786
Call-ID: 3413an89KU
```

We see that this exchange follows the normal SIP 3–way handshake of INVITE, OK and ACK. After this, both Alice and Bob can open a TCP connection and exchange MSRP messages over it. MSRP has SEND methods and acknowledgement. The SEND method supports sending fragments of large messages. It is also possible to specify the nature of the content of the message.

```
MSRP a786hjs2 SEND
To-Path: msrp://biloxi.example.com:12763/
   kjhd37s2s20w2a;tcp
From-Path: msrp://atlanta.example.com:7654/
   jshA7weztas;tcp
Message-ID: 87652491
Byte-Range: 1-25/25
Content-Type: text/plain

Hey Bob, are you there?
```

 ISBN: 978-1-61208-133-5

```
-------a786hjs2$
```

All messages sent are acknowledged with a copy of the transaction identifiers present in the message header, as well as a copy of information present in the SDP body: to-path and from-path.

MSRP has several provisions for reporting on message sent. It is possible to request in the header whether or not a report should be sent in situations of success or failure. Reports use the message ID to differentiate between different transmissions.

B. URIs, Paths and Relays

MSRP endpoints are identified by URIs, with am msrp (or msrps, when carried by TLS) prefix, as seen in the example above. From-Path, To-Path fields in MSRP contain sequences of URIs, which are relays to the final destination. Beyond the protocol used, URIs have features we are accustomed to from other uses of SIP. Rather than a fully qualified name, it is also possible to use IP addresses.

An endpoint that uses one or more relays will indicate that by putting a URI for each device in the relay chain into the SDP path attribute. The final entry will point to the endpoint itself. The other entries will indicate each proposed relay, in order.

Since both ends of communications can be isolated behind security devices, it may be necessary to communicate through relays, not unlike what is done for SIP. In our specific case, we would consider the use of a single relay. In the following section, we see how it can be inserted in the communication, and its practical benefits.

VI. Discussion

We propose that both a home user agent and the remote user are both customers of the same IMS infrastructure. End to end communication establishment is done by the basic mechanisms of IMS. Both parties know each other's name and correct authentication is guaranteed by IMS. e2e signalling is thus quite trivially established between parties. The issues remaining are the transparency of the home services (for the IMS infrastructure) and the support for information exchange.

Services: Home services and their nature are essentially transparent for the IMS operator: media exchanges can be no different from typical usage, while notifications, menus and operations are embedded in MSRP messages and encoded in, say, XML, in a simple command–parameter format. The following example shows a sequence of sensor information.

```
<?xml version="1.0" encoding="ISO-8859-1"?>
  <status date=31/01/2009>
    <sensor>
      <name code="1">Kitchen</name>
      <value>empty</value>
    </sensor>
    <sensor>
      <name code="2">Living Room</name>
      <value>empty</value>
    </sensor>
      ...
  </status>
```

A generic application can associate operations and (GUI) presentation based on XML documents exchanged through a SIP UA. This application has a home and a remote flavour. At home, it interacts with devices and with the user either through the UA or through a local menu. On the remote terminal, it is only interacting with the user.

We are not investigating the application any further here as it presents no specific challenge.

Communications: The main hurdle we face is the possible presence of devices restricting the establishment of communications in one direction. While IMS-related standards[5] are designed to circumvent such restrictions, we may still require that a relay be present in the network; this relay acts as a back to back user agent (B2BUA), typical in SIP architectures. Note that this relay has two dimensions: signalling, and media. IMS is structured in such a way that a signalling relay is not necessary, beyond what is supplied by the CSCF. Yet in some circumstances, the use of a B2BUA has been mandated (e.g. [1]).

Media is a more critical issue, especially when TCP is used for transport, which is also why MSRP relays [8] were created. Typical UDP-based SIP communications are initiated from the user to the network, with the first REGISTER message, which would be allowed to traverse NATs and firewalls and set the path for future SIP exchanges. TCP connections must be initiated from one side only. Our alternative is either to use a B2BUA, or simply an MSRP relay.

The relay issue is important for another reason: the provider must not hold any personal information for the user, unlike typical Internet "service in the cloud" models, beyond subscription information. We must therefore exclude architectures where a storage server would act as a temporary repository. Note however that communications between home and relay, and user and relay can be encrypted, but other solutions must be found if strict end-to-end confidentiality is required.

We propose that a B2BUA would be required for all communications, i.e., media and data. While some forms of communication could be authorized by middleboxes and not others, it is simpler to use a single connectivity model for all.

Relay Discovery: An issue with the B2BUA is to 1) decide whether or not it is necessary and 2) discover its location.

For the first problem, it is simpler to impose its systematic use, as we have just discussed. For the second one, S-CSCF filters must be used to route the call through the B2BUA. This can simply be done by assigning homes a special class of URIs, and recognizing a communication between the user and the home.

Configuration: Home User Agent and Remote User must share some information for proper inter-operation, such as list of known devices/sensors and other supported media services, e.g. audio, video or pictures. We would typically create a remote configuration based on that of the home application and transfer it to the remote terminal.

 ISBN: 978-1-61208-133-5

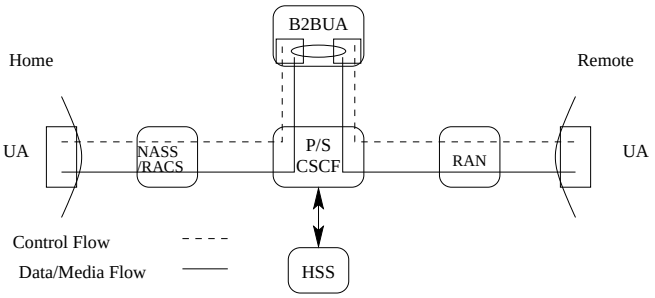


Figure 2. e2e View

They would also obviously know each other's URI and could also share keys for mutual authentication and message encryption.

All these issues are beyond the IMS infrastructure's influence, however.

Integration: Figure 2 illustrates an example of end-to-end connectivity. The home is connected to the IMS infrastructure via a Network Attachment Subsystem (NASS), associated with a Resource and Administration Control Subsystem (RACS) which can perform network security operations. At the other extremity, the user would have a mobile terminal exploiting a radio access network (RAN).

Home-User SIP sessions are switched by the CSCF function towards the B2BUA which bridges requests and connects data/media flows. As we have explained above, the use of the B2BUA can be transparent to the users and inserted in the signalling path through the S-CSCF filters.

The functionality required of the B2BUA for the data/media path is minimal, and content dependent. Audio/Video codecs are negotiated end-to-end and media frames, carried over UDP, need only be relayed towards their destination.

MSRP data is carried over TCP and presents a different problem. While it would be possible to collect TCP segments and relay them directly, it is more appropriate to collect well-formed messages and forward them, as would an MSRP relay. Again, it is possible to use encryption to keep message content private if necessary.

Overall, we see that the infrastructure we need for our communications is well within the IMS model. Since filtering is involved, the participation of the IMS operator is required, although we should put a caveat there: all IMS services (A/V communications, Messaging) are straightforward, except that operator support is required to overcome networking restrictions imposed in some domains. While we can imagine that, in some circumstances, offered IMS services could be integrated into a suitable application, it may also well be the case that a B2BUA would have to be deployed in the operator's network, with a matching service offering. Considerations for a suitable business model are beyond the scope of this paper, however.

VII. Conclusions

We have shown a SIP-based model to support home-based services and how it is possible to use an IMS infrastructure to deploy such basic tools. Beyond established A/V services, the use of MSRP, for data exchange, combined with a B2BUA in the operator's network are sufficient to allow the user to safely exchange information between home and remote locations. The application itself can be designed independently, for example on an XML basis, while benefiting from IMS' services. The scheme proposed is overall rather straightforward and would support applications of various degree of complexity.

Further work is required to study how to support streaming more efficiently, or closer to an Internet model, since we have here IMS' interactive model. We believe this would require special support in a network B2BUA.

Finally, we should be able to bridge the gap between home-internal and home-external communications, if only to be able to transparently reuse the same devices. This is also the focus of further investigations.

References

[1] Open Mobile Alliance. Instant message using simple. [OMA-TS-SIMPLE_IM-V1_0-20080312-D, Sep. 2008.
[2] AT&T. AT&T launches remote home monitoring video service nationwide. last visited March 15th, 2011, http://www.att.com/gen/press-room?pid=4800&cdvn=news&newsarticleid=23003.
[3] Paula Bernier. Verizon tests home monitoring and control service. Last visited March 15th, 2011, http://www.techzone360.com/topics/techzone/articles/134094-verizon-tests-home-monitoring-control-service.htm, January 2011.
[4] B. Campbell, R. Mahy, and C. Jennings. The Message Session Relay Protocol (MSRP). Request for Comments (RFC) 4975, September 2007.
[5] ETSI. Telecommunications and internet converged services and protocols for advanced networking (tispan); resource and admission control sub-system (racs): Functional architecture. ETSI Standard 282 003 B3.2.0, Nov. 2008.
[6] David Fox. Home monitor on a cell phone, o'reilly, last visited march 15th, 2011.
[7] M. Handley, V. Jacobson, and C. Perkins. SDP: Session Description Protocol. Request For Comment (RFC) 4566, July 2006.
[8] C. Jennings, R. Mahy, and A.B. Roach. Relay extensions for the message session relay protocol (msrp). Request for Comments (RFC) 4976, September 2007.
[9] Meye. A ground-breaking home-monitoring service. Last visited March 15th, 2011, http://www.meye.com.my/.
[10] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks M. Handley, and E. Schooler. SIP: Session Initiation Protocol. Request For Comment (RFC) 3261, June 2002.
[11] www.digium.com. Asterisk is a registered trademark of digium inc.
[12] www.skype.com. Skype is a registered trademark of skype ltd,.

 ISBN: 978-1-61208-133-5

Improvement of Job Scheduling for Automatic Chain Processing in Radio Occultation Context

Lorenzo Mossucca, Olivier Terzo
Istituto Superiore Mario Boella
Via P. C. Boggio 61,
Torino, Italy
Email: {mossucca, terzo}@ismb.it

Manuela Cucca, Riccardo Notarpietro
Politecnico di Torino
Corso Duca degli Abruzzi 24,
Torino, Italy
Email: {manuela.cucca, riccardo.notarpietro}@polito.it

***Abstract*—The new Italian GPS receiver for Radio Occultation has been launched from Satish Dhawan Space Center (Sriharikota, India) on board of the Indian Remote Sensing OCEANSAT-2 satellite. The Italian Space Agency has established a set of Italian universities and research centers to develop the Web Science Grid, an infrastructure based on grid computing, that is implemented for the overall processing Radio Occultation chain. In consideration of the complexity of our scenario due to the modules involved and difficulties of geographically dispersed nodes, after a brief description of the algorithms adopted, that can be used to characterize the temperature, pressure and humidity, the paper presents an improvement of job scheduling in order to further decrease the elaboration time. Two applications to manage automatically the Radio Occultation data are described: Local and Global scheduler, one for worker nodes and one for the master node. Also the estimated processing time and actual processing are shown.**

***Keywords*-radio occultation; grid computing; local scheduler; global scheduler; job scheduling.**

I. INTRODUCTION

The GPS Radio Occultation (RO) is an emerging remote sensing technique for the profiling of atmospheric parameters (first of all refractivity, but also pressure, temperature, humidity and electron density, see [1] and [2]). It is based on the inversion of L_1 and L_2 GPS signals collected by an ad hoc receiver placed on-board a Low Earth Orbit (LEO) platform, when the transmitter rises or sets beyond the Earth's limb. The relative movement of both satellites allows a "quasi" vertical atmospheric scan of the signal trajectory and the profiles extracted are characterized by high vertical resolution and high accuracy. The RO technique is applied for meteorological purposes (data collected by one LEO receiver placed at 700 km altitude produce 300÷400 profiles per day, worldwide distributed) since such observations can easily be assimilated into Numerical Weather Prediction models. Anyway, it is also very useful for climatological purposes, for gravity wave observations and for Space Weather applications. Starting from the first operational RO mission on board the German CHAMP satellite [6], there are presently several other satellite missions carrying on-board a RO payload. The most important are RO experiments on-board the European METOP-1 mission [3] and on-board the USA/Taiwan COSMIC constellation mission. Several other missions are planned for the next future. In particular, during the 2009 autumn season, the Indian OCEANSAT-2 mission carrying on-board the Italian ROSA (Radio Occultation Sounder of the Atmosphere) GPS receiver was launched. In the framework of this opportunity, the Italian Space Agency [4] funded a pool of Italian Universities and Research Centers for the implementation of the overall RO processing chain, which is called ROSA-ROSSA (ROSA-Research and Operational Satellite and Software Activities). The ROSA-ROSSA was integrated in the operational ROSA Ground Segment by an Italian Software enterprise (INNOVA, located in Matera, Italy), and the ROSA ground segment is operating in Italy (at the ASI Space Geodesy Center, near Matera) and in India (at the Indian National Remote Sensing Agency [5], near Hyderabad) starting from the 2009 autumn season. This version implements RO state-of-the-art algorithms and, for the first time, it was developed and it runs on a distributed hardware and software infrastructure exploiting a grid computing strategy, which is called Web Science Grid (WSG). The paper is structured as follows: Section 2 is devoted to a more detailed description of the ROSA-ROSSA software. This section is given in order to better set up the scientific application which exploits grid processing strategies. Section 3 describes motivations. Section 4 presents the structure of our system and scheduling description. Section 5 contains considerations about the time execution obtained by the system based on grid computing. Section 6 draws the conclusions.

II. THE PROCESSING CHAIN OF RO OBSERVATIONS

The ROSA-ROSSA software implements state-of-the-art RO algorithms which were already available from the scientific group and are during the validation phase before their final transfer inside the official Ground Segment of the ROSA Radio Occultation receiver. The processing chain, which is subdivided into seven different software modules (namely Data Generators-DG), is executed in a sequential

 ISBN: 978-1-61208-133-5

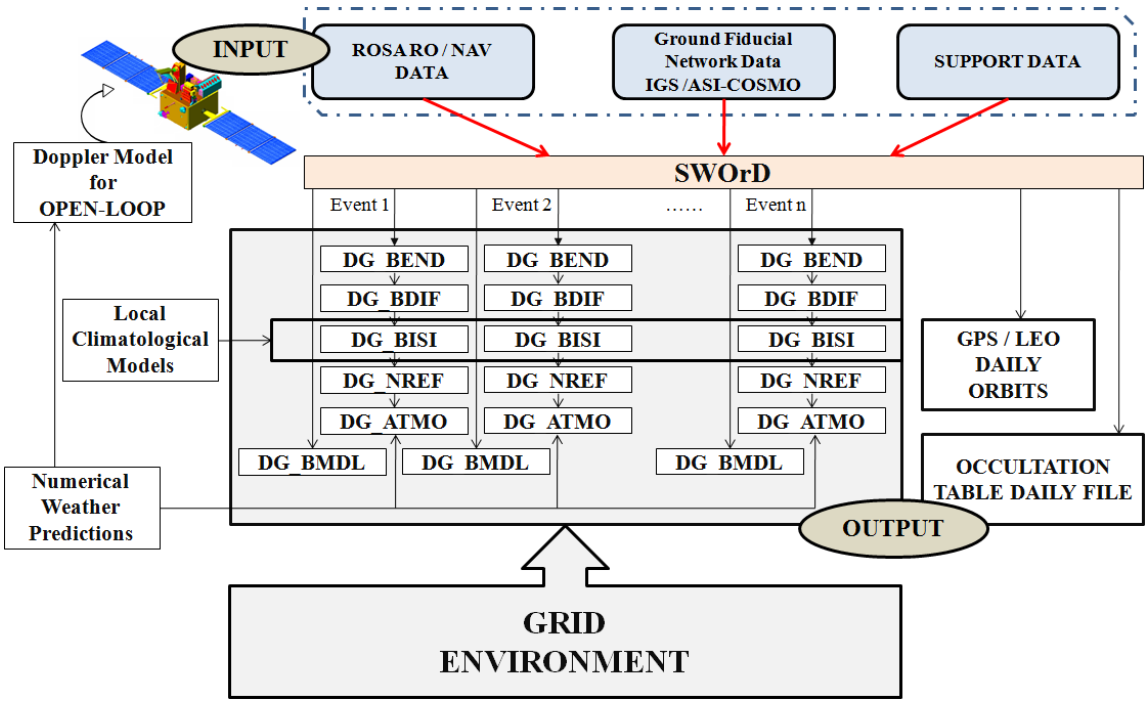


Figure 1. ROSA-ROSSA Overall Chain.

mode. Figure 1 shows a simple diagram of the processing chain and of the corresponding data-flow. Before delving into how the various parts work, a DGs explanation is given, in order to focus the data types to process. Starting from ROSA Level 1.a engineered data coming from the ROSA on-board OCEANSAT-2 platform observations, from the ground GPS network and from other support data, the ROSA-ROSSA is able to produce data at higher levels, using a data processing chain defined by the following Data Generators. SWOrD is a software module that fully supports orbit determination, orbit prediction, and implements Level 2 data generation connected with the ROSA sensor on-board OCEANSAT-2. Input data for SWOrD are ROSA GPS navigation and Radio Occultation observations, ground GPS network data and other support data. It generates the following output data:

- Estimated rapid orbits and predicted orbits for the GPS constellation;
- Estimated rapid orbits and predicted orbits for the OCEANSAT-2 platform;
- 50 Hz closed-loop and 100 Hz open-loop excess phases and signal amplitude data for each single occultation event;
- Tables showing estimated and predicted (up to 6 hours in advance) occultation (Data Level 2.c).

The BMDL Data Generator predicts a bending angle and impact parameter profile (Level 2.d data) usable as input in the ROSA on-board software excess doppler prediction module for open-loop tracking. For each ”predicted” occultation event, latitude and longitude of the geometrical tangent points (the nearest point of each trajectory to the Earth’s surface, evaluated through predicted orbits) is used to compute bending angle and impact parameter profile from interpolated numerical weather prediction models (bending angle and impact parameter are geometrical parameter univocally identifying each trajectory followed by the RO signal. See Figure 2 for details). Predicted bending angle and impact parameter profiles α (a) (2.d Data Level) are stored in ASCII data files containing bending angles and impact parameters together with the UTC time stamp, one file for each event. Input data for DG_BMDL are 1b.a, 1b.b (predicted GPS and LEO orbits, respectively) and 2.c (Predicted Occultation Tables), together with ECMWF world forecasts for the synoptic times valid for the future observed occultation event. The BEND Data Generator provides ”raw” bending angle and impact parameter profiles α(a) computed on GPS occulted signals on both GPS frequencies L_1 and L_2, by using a Wave Optics approach below a certain altitude (generally in troposphere). Above that altitude threshold, standard Geometrical Optics algorithms are applied. Raw bending angle and impact parameter profiles α(a) (Data Level 3.a) are stored for each event in ASCII data files. Inputs for DG_BEND are 2.a data (L_1 and L_2 excess-phases and related orbit data) and 2.b data (L_1 and L_2 signal amplitudes).

 ISBN: 978-1-61208-133-5

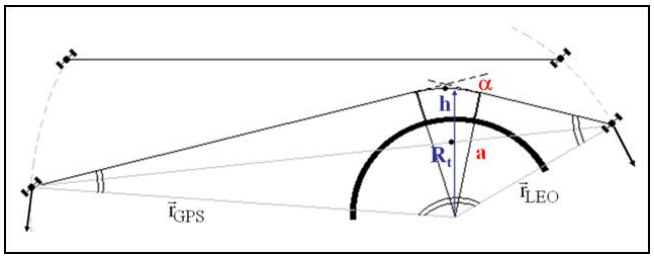


Figure 2. Radio Occultation geometry. The quasi instantaneous trajectory can be defined by the following geometrical parameters: the bending angle α, the impact parameter a. R_t is the local Earth's radius and h the tangent point height.

The BDIF Data Generator provides (for each event) a bending angle and impact parameter profile, on which the ionospheric effects have been compensated for. This DG processes both L_1 and L_2 bending angle and impact parameters profiles (Data Level 3.a) given as input, in order to minimize the first order ionospheric dispersive effects. Outputs for DG_BDIF are bending angle and impact parameter iono-free profiles (Data Level 3.b). The BISI Data Generator provides profiles of bending angle versus impact parameter optimized in the stratosphere above 40 km. In the ROSA-ROSSA, data coming from a Numerical Weather Prediction Model (ECMWF analysis) are used in place of climatological data for implementing the statistical optimization procedure necessary to reduce the high noise level left to the signal after ionospheric first order compensation applied by the previous DG_BDIF. DG_BISI processes bending angle and impact parameter profiles obtained from Data Level 3.b. Output for DG_BISI are bending angle and impact parameter profiles optimized in the stratosphere (Data Level 3.c). The NREF Data Generator provides (for each event) the refractivity profile and dry air temperature and pressure profiles. This DG is able to process iono-free and properly initialized bending angle and impact parameter profiles (Data Level 3.c) in order to compute the corresponding dry air "quasi" vertical atmospheric profiles (Data Level 3.d). The ATMO Data Generator allows to evaluate the temperature and the water vapour profiles using forecasts or analysis obtained by numerical weather prediction. This DG receives on input Level 3.d data files and produces on output Level 3.e data files, which contain the total temperature and total pressure profiles in terms of wet and dry components.

III. Architecture Motivations

The main purpose is to create a flexible architecture in order to manage the radio occultation data and to reduce their processing time. The system guarantees the entire processing chain automatically that consists of seven DGs executed sequentially as explained before. In a learning phase, we evaluated that for each day, the events number to process are about 250, on a single machine the elaboration

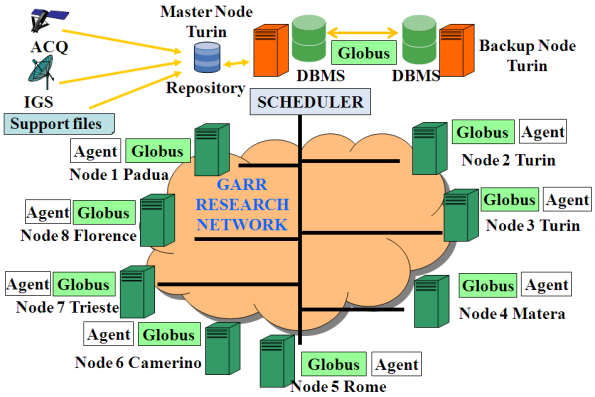


Figure 3. Web Science Grid.

time for the entire chain processing, is approximatively 40 hours. The idea of using a distributed environment arose from the need to reduce this processing time because these makes it difficult to use the results. The WSG has been developed with the goal of simplifying this task, by providing implementations of various core services provided by Globus Toolkit and deemed essential for high-performance distributed computing. Furthermore, it allows engineers and physicists of the project to have a tool for processing and sharing data, independently from the university in which they are.

IV. Architecture Design

The Web Science Grid (WSG) is an integrated system devoted to handle and process RO data of the OCEANSAT-2 ROSA on board sensor.

A. Web Science Grid architecture

The WSG is composed by the subsystems(see Figure 3): middleware, central repository, relational database, scheduler, agents and applications. The general purpose of our project is: sharing the computational resources, transferring a great amount of files and submitting jobs from several different organizations of the scientific community located in different places in Italy. All these operations are processed in an automatic way without any user interaction. The pool of nodes consists of 10 nodes with 2 processors each, 2 GB RAM, 64 bit machines , and on all these machines run Linux (Ubuntu). The nodes are located geographically in Italy, for accuracy to:

- Istituto Superiore Mario Boella (Turin);
- Polytechnic University of Turin (Turin);
- University of Padua (Padua);
- Sapienza University (Rome);
- University of Camerino (Macerata);
- International Center of Theoretical Physics (Trieste);
- Italian Space Agency (Matera);
- Institute for Complex System (Florence).

 ISBN: 978-1-61208-133-5

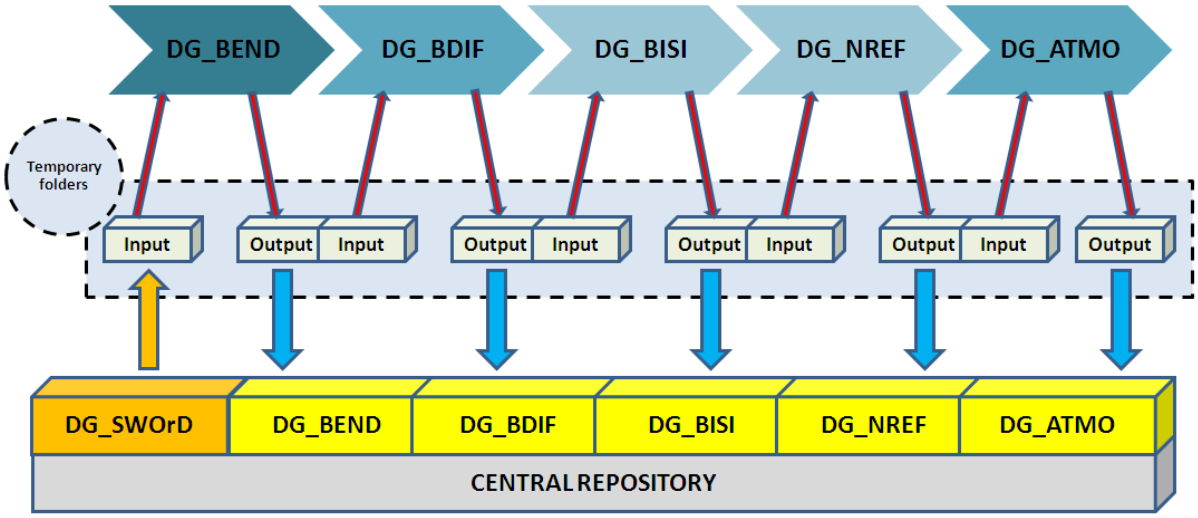


Figure 4. Automatic processing chain

The Globus Toolkit has been used as middleware [7] and [8], 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 [11].

B. Automatic chain

Our software allows to run the chain automatically; it is composed from two schedulers: one who listens to the master node, called global scheduler, and checks for files ready for execution and sends them to worker nodes, according to the scheduling rules, the other, called local scheduler, listens on the worker nodes, and when it receives a file executes and returns the result file obtained on the master node [9] and [10]. In Figure 4, data flow is depicted, the first transaction takes place on master node; it receives the files directly from the satellite and performs the first step in the chain, i.e., SWOrD, generating about 256 files that are placed in the folder the next step, DG_BEND. When there are files in the folder DG_BEND, the global scheduler checks nodes available by querying the database, and sends files to them. Global scheduler provides for automated scheduling of any input files. It uses all machines belonging to the grid to distribute work load and to provide a backup system for all critical tasks within the system. The choice of how to share the file to run is based on 2 sets of scheduling rules, one concerning the available nodes and one derived from an analysis of the file to run. An agent is installed on each node, is used to monitor the availability of each service on the node and periodically, it sends its general status to the database on master node, if all services are active the node is in condition to receive a job. For the selection of nodes available and ready to run, the global scheduler checks on the database directly instead of querying each machine. When the worker node sees a file in its folder, starts the processing procedure that will generate an output file that will be sent to master node in the folder next step, i.e., DG_BDIF. This procedure is performed for every steps of the chain, the operation is as follows: from SWOrD, the DG n-1 generates the output file that will be the input files of DG n, and so on. On worker nodes, each execution is performed in a temporary folder, so that, in case of error, identify the type of error made and then to reprocess the file. Two types of errors can occur: the first for lack of data in the file due to the satellite reception, the second for network failures or node crash. Only in the last case it is worth recover the process, and it is enough reprocess il file. Anyway, each process has a timeout, if within a fixed time processing has not been completed, the process is killed. An important component of this architecture is the database, which allows us to monitor any action of the grid. Regarding the automatic chain, each transaction is stored on the database when it starts running, when it ends, input files, output files, the node that has run and type of error, if it has generated them. The database also contains information on the status of each node and are available to receive the file to run, this allows us to understand whether there are network problems, so if the node is reachable.

V. IMPROVING PERFORMANCES

All DGs of the processing chain have been tested during a learning phase; for a single event it obtained the percentage values in Figure 5, and for a daily events in Figure 6. In the two graphs, the difference is due from that, for each hour SWOrD generates only one event for DG_BMDL

 ISBN: 978-1-61208-133-5

and instead from DG_BEND to DG_ATMO it generates about nine events. Number of input, output and time for elaboration have been considered. It has been assumed that SWOrD has already been executed, then it is outside the calculation processing. The Eq. 1 and Eq. 2 represent an estimation time of elaboration and cover both a non-distributed (N = 1) and distributed architecture ($N > 1$).

$$T_p = T_e * \eta \quad (1)$$

$$T_p = \frac{1}{N} \sum_{i=1}^{\eta} (T_{ei} + \beta) \quad (2)$$

Where:

$$T_p = TotalProcessTime$$
$$T_e = EventProcessTime$$
$$T_{ei} = EventiProcessTime$$
$$\eta = NumberofROEvents$$
$$N = NumberofGridNodes$$
$$\beta = FileTransferTime$$

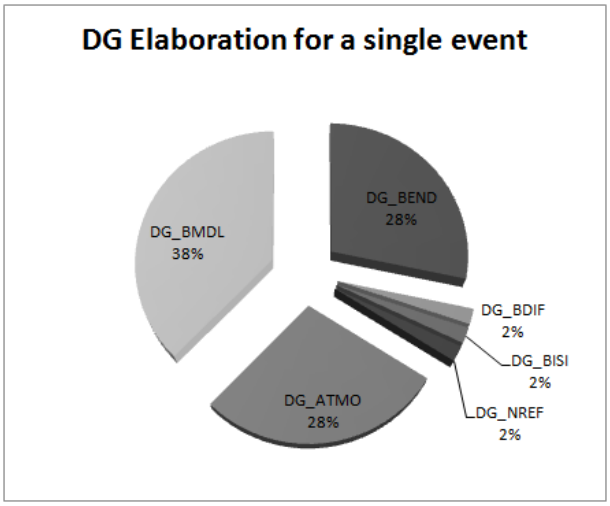


Figure 5. DG Elaboration for a single event.

In Figure 7, the execution time trend is estimated, when the number of nodes and events is increased. When only one node is available, the total execution time for a daily files is 1752 minutes (about 29 hours), instead increasing the number of nodes, the execution decrease further, just note that with 2 nodes is 912 (about 15 hours). An important point when a single event is processed is that there is no gain time in grid environment; rather time is higher because we must consider the transfer time; it has a sizeable gain time only when a set of files are processed.

In Table I, processing time detected for daily data elaboration is considered; it depicts how to change the processing time when worker nodes increase. Certainly, the benefits of the grid is ensure the elaboration the overall chain

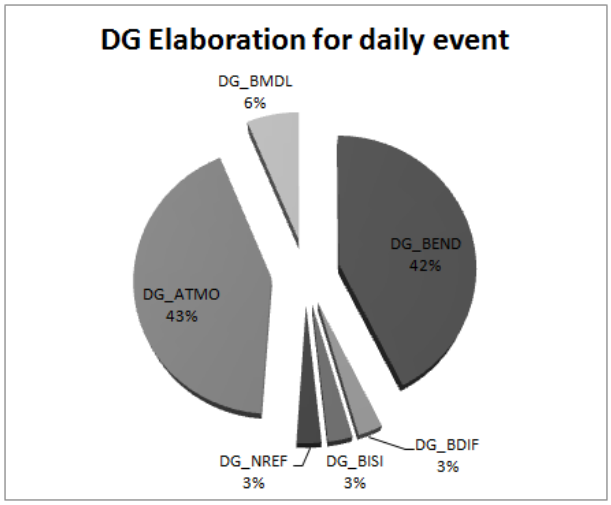


Figure 6. DG elaboration for daily events.

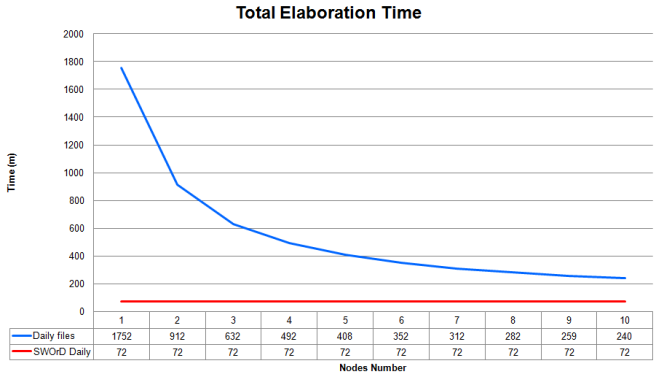


Figure 7. Estimated processing time for daily data (about 250 events)

in less time, instead, in distributed system where worker nodes are geographically located, it can have disadvantage in the network layer, in case of network failures or slow connections, to overcome this problem only internal nodes are available for elaboration.

NODES NUMBER	PROCESSING TIME
10	2h20m
8	3h06m
6	4h00m
4	6h30m

Table I
DETECTED PROCESSING TIME FOR DAILY DATA (ABOUT 250 EVENTS)

VI. CONCLUSIONS

The ROSA-ROSSA software implements Radio Occultation technique, which run for the first time on a grid computing infrastructure, called Web Science Grid and elaborations time are described. This paper want to be an

 ISBN: 978-1-61208-133-5

example of application where you can use grid computing. In frameworks such as Radio Occultation, where the amount of data to be processed is significant, the use of a distributed architecture as the grid can be the best choice. We have focused on a way to manage the assignment nodes for execution in automatic way without any human interaction through a local and a global scheduler. As future works we plan the extension of the proposed architecture to clusters available across the European Grid Infrastructure (EGI) and we are studying a solution for EC2 environment by Amazon to allow to further increase available computing power.

VII. ACKNOWLEDGMENTS

The authors are grateful the Italian Space Agency (ASI) for supporting this project within contract I/006/07/0 and to all the ROSA-ROSSA partners for their contributions.

REFERENCES

[1] Melbourne, W., *The application of spaceborn gps to atmospheric limb sounding and global change monitoring*, JPL Publ, pp. 18-94, 1994

[2] Kursinski, E.R., *Observing Earth's atmosphere with radio occultation measurements*, Journal Geophys. Res., pp. 429-465, 1997

[3] Luntanama, J.P., *Prospects of the EPS GRAS Mission for operational atmospheric applications*, Bull. Am. Met. Soc, pp. 1863-1875, 2008

[4] Italian Space Agency (ASI), *http://www.asi.it/*, 2010

[5] Indian National Remote Sensing Agency (ISRO), *http://www.isro.org/*, 2010

[6] Wickert, J., *The radio occultation experiment aboard CHAMP: Operational data processing and validation of atmospheric parameters*, Journal Meteorol. Soc. Jpn, pp. 381-395, 2004

[7] Foster, I. and Kesselman, C., *The Grid2: Blueprint for a New Computing Infrastructure*, Morgan Kaufmann, pp. 38-63, 2003

[8] Berman, F., Fox, G. and Hey, G., *Grid Computing Making the Global Infrastructure a Reality*, Wiley, pp. 117-170, 2005

[9] Dimitriadou, S. and Karatza, H., *Job Scheduling in a Distributed system Using Backfilling with Inaccurate Runtime Computation*, The international conference on complex, intelligent and software intensive system, pp. 329-336, 2010

[10] Xhafa, F., Pllan, S. and Barolli, L., *Grid and P2P Middleware for Scientific Computing Systems*, The international conference on complex, intelligent and software intensive system, pp. 409-414, 2010

[11] Brunett, S., Czajkowski, K., Fitzgerald, S., Foster, I., Johnson, A., Kesselman, C., Leigh, J., Tuecke, S., *Application Experiences with the Globus Toolkit*, The Seventh International Symposium on High Performance Distributed Computing, pp.81-88, 1998

 ISBN: 978-1-61208-133-5

IEEE 802.11n MAC Mechanisms for High Throughput: a Performance Evaluation

Miguel A. García, M. Ángeles Santos and Jose Villalón,
Albacete Research Institute of Informatics. Universidad de Castilla – La Mancha.
Campus Universitario, 02071. Albacete, SPAIN.
{Mangel.Garcia, MariaAngeles.Santos, JoseMiguel.Villalon}@uclm.es

Abstract— **Nowadays IEEE 802.11 standard is the most widely used one in wireless LAN (WLAN) technology. One of the key reasons is the continuous amendments presented by the IEEE 802.11 working group. One of these amendments (IEEE 802.11n) was approved to enhance 802.11 for higher throughput operation. IEEE 802.11n is an ongoing next-generation wireless LAN standard that supports a very high speed connection with more than 100 Mb/s data throughput measured at the medium access control layer. In this paper we examine the major improvements introduced by IEEE 802.11n MAC: aggregation, block acknowledgement, and reverse direction. We show the impact of each parameter in the network performance.**

Keywords-component; IEEE 802.11n; aggregation; block ACK; reverse direction

I. INTRODUCTION

These days, the wireless LAN (WLAN) technology is usually deployed using the IEEE 802.11 standard. One of the main factors for the popularity of the IEEE 802.11 is the continuous amendments. The IEEE 802.11 working group has always strived to improve this wireless technology through creating new amendments to the base 802.11 standard. The amendments try to solve the low efficiency of its medium access control (MAC) and physical (PHY) layer protocols, which restrict its applications to support high data rate multimedia services. Current WLAN systems endure difficulties due to the increasing expectations of end users and volatile bandwidth Delay-boundary demands from new higher data rate services, such as high-definition television (HDTV), video teleconferencing, multimedia streaming, voice over IP (VoIP), file transfer, and online gaming.

In 2002, the IEEE 802.11 standard working group established the high-throughput study group (HTSG) with the aim to achieve higher data rate solutions by means of existing PHY and MAC mechanisms [2, 3]. Its first interest was to achieve a MAC data throughput over 100 Mb/s using the 802.11a standard. However, the objective proved to be infeasible. So, in September 2003, the HTSG set off the IEEE 802.11n ("n" represents next-generation) resolution to compose a high-throughput (HT) extension of the current WLAN standard would increase the transmission rate and would reduce the unavoidable overhead. The main goal of the IEEE 802.11n task group (TGn) was to define an amendment that had a maximum data throughput of at least 100 Mb/s (measured at the MAC layer) and at the same time, to allow coexistence with legacy devices. To achieve high throughput in 802.11 wireless networks, the most commonly used method is to increase the raw data rate in the PHY layer. For this propose IEEE 802.11n [4] include multiple input multiple output (MIMO) antennas with orthogonal frequency division multiplexing (OFDM) and various channel binding schemes. Moreover, IEEE 802.11n expands the channel bandwidth to 40MHz to increase the channel capacity.

However, higher PHY rates do not necessarily translate into corresponding increases in MAC layer throughput. Indeed, it is well known that the MAC efficiency of 802.11 typically decreases with increasing PHY rate [5], [6]. To solve this limitation, IEEE 802.11n defines new mechanisms to increase the network performance.

The main contribution of this work is to provide an understanding of the three IEEE 802.11n MAC layer major enhanced mechanisms: aggregation, block acknowledgement, and reverse direction. Most previous works on IEEE 802.11n performance evaluation only explored aggregation mechanism [7, 8]. In [9] the authors evaluate both physical and MAC enhanced.

The rest of this paper is structured as follows: in Section 2 we describe a brief outline of the current IEEE 802.11 standard, followed by a discussion of its maximal throughput limitations. We describe the IEEE 802.11n in Section 3. We carry out a performance evaluation in Section 4 by means of extensive simulation. Section 5 concludes this paper.

II. IEEE 802.11

A. Overview of IEEE 802.11 PHY

The IEEE 802.11 PHY layer specification concentrates mainly on wireless transmission. The original specification was first approved in 1997 [1] and includes a primitive MAC architecture and three basic over-the-air communication techniques with maximal raw data rates of 1 and 2 Mb/s. Because of their fairly low data bandwidths, further amendments have been proposed throughout the years: IEEE 802.11a [10], 802.11b [11], and 802.11g [12]. Both 802.11a and 802.11b were finalized in 1999 and support raw data rates up to 11 Mb/s and 54 Mb/s, respectively. In June 2003, a third PHY specification (802.11g) was introduced, with similar maximum raw data rate as 802.11a but operating in separate frequency bands. For this period, there were many

This work was supported by the Spanish MEC and MICINN, as well as European Comission FEDER funds, under Grants CSD2006-00046 and TIN2009-14475-C04. It was also partly supported by the Council of Science and Technology of Castilla-La Mancha under Grants PEII09-0037-2328 and PII2I09-0045-9916.

 ISBN: 978-1-61208-133-5

amendments and countless research works for improved PHY specifications that mostly aim to provide reliable connections and higher data rates. This is mainly because there is a continuous rapid increase in user demand for faster connections. In spite of establishing novel techniques that theoretically can be used for higher data transmission rates, the throughput outcomes at the MAC data are surprisingly low and in most cases, half of what the underlying PHY rates can offer.

B. IEEE 802.11 MAC

The MAC architecture is based on the logical coordination functions, which determine who accesses to the wireless medium at each time. In the legacy IEEE 802.11 standard, there are two types of access schemes: the mandatory distributed coordination function (DCF), which is based on the carrier sense multiple access with collision avoidance (CSMA/CA) mechanism; and the optional point coordination function (PCF), which is based on a poll-and-response mechanism. These MAC schemes are inadequate to resolve differentiation and prioritization between frames and multimedia applications such as VoIP and audio/video conferencing with strict performance constraints. Due to these applications have become widely popular, a new extension was vital. In late 2005, IEEE 802.11 TG approved the IEEE 802.11e amendment [13] to provide an acceptable level of quality of service (QoS) for multimedia applications. The 802.11e proposes the hybrid coordination function (HCF), which uses a contention-based channel access method, known as enhanced DCF channel access (EDCA). EDCA has the ability to operate simultaneously with a polling-based HCF controlled channel access (HCCA). In addition to the differentiation and prioritization that IEEE 802.11e offers, the transmission opportunity (TXOP) was introduced in order to improve MAC efficiency. A TXOP is an interval of time in which multiple data frames can be transferred from one station to another (also known as bursting). During a TXOP period the station can transmit multiple data frames without entering the backoff procedure, reducing the overhead due to contention and backoff period. Along with frame bursting, another type of acknowledgment (ACK), known as block ACK, was established. Receivers can acknowledge multiple received data frames efficiently by using just a single extended ACK frame.

C. Throughput Limitations

To understand the inefficiency of IEEE 802.11 over higher data rates, we must briefly describe the legacy DCF. A successful packet transmission in DCF is illustrated in Fig. 1. When a station has a data frame (MAC service data unit, MSDU) to transmit, MAC headers are added to form MPDUs. A station may start to transmit after having determined that the channel is idle during an interval of time longer than the distributed interframe space (DIFS). Otherwise, once the transmission in course finishes and in order to avoid a potential collision with other active stations, if the channel is busy, the station will wait a random interval of time (the backoff_time) before start to transmit. The station will be able to begin transmission as soon as the backoff counter reaches zero. In order to know if a transmission has been successful, the destination station should respond to the source station with an ACK in an interval of time equal to the short interframe space (SIFS).

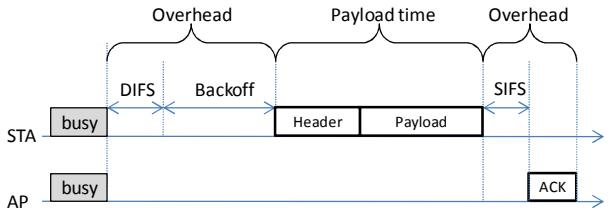


Figure 1. Legacy IEEE 802.11 operation.

By looking into the procedure of a packet transmission, we note that the channel is inefficiently used by the DCF. During the transmission procedure, transmission time is divided into a DIFS, a Contention Window backoff time, the PPDU transmission time, a SIFS, and the ACK frame transmission time. The PPDU transmission time can be further divided into two parts: 802.11 header and data payload transmission time. Other than the payload transmission portion is the overhead. The overhead of the DCF mechanism results in the inefficiency of the channel utilization, and thus limits the data throughput. When the payload is small, the overhead is relatively large and is less efficient. The percentage of the overhead among all usable airtime increases as the physical transmission rate increases. This fact causes that he overhead limits the achievable data throughput. In a higher data rate scenario, although the frame transmission time is reduced, the part of the overhead is unchanged due to the backward compatibility issue. As a result, to achieve higher throughput in 802.11 reducing the percentage of overhead is critical.

III. IEEE 802.11N

Although 802.11e adds the support of QoS, TXOP and block ACK, the inefficiency of channel utilization in legacy 802.11 MAC is not fully solved. To satisfy the current need of the high-speed wireless network access, the major target of IEEE 802.11n, is to provide a high throughput mechanism based on state of art design while allowing the coexistence of legacy 802.11 devices. To meet the requirements of “high throughput”, two possible methods can be applied. The first one is to increase the data rate in the PHY layer, and the second one is to increase the efficiency in the MAC layer. Based on the foundation of 802.11a/b/g/e, many new features in PHY and MAC layers are introduced to enhance the throughput of IEEE 802.11 WLAN.

A. MIMO-OFDM physical layer

To achieve high throughput in 802.11 wireless networks, the most commonly used method is to increase the raw data rate in the PHY layer. IEEE 802.11 uses two mechanisms to increase this data rate: MIMO technology and a channel bandwidth that is twice as size (from 20 MHz to 40 MHz). IEEE 802.11n expands the channel bandwidth to 40MHz in

 ISBN: 978-1-61208-133-5

order to increase the channel capacity. However, IEEE 802.11n operates in OFDM scheme with MIMO technique [6]. MIMO can effectively enhance spectral efficiency with simultaneously multiple data stream transmissions. In theory, channel capacity gain could be up to the number of transmitting antennas without additional bandwidth or power. The power of the MIMO system relies on using space-time coding and the channel information for intelligent transmission. Multiple antennas could help to transmit and receive from multiple spatial channels simultaneously. Multipath wireless fading channel results in poor performance in legacy 802.11 PHY scheme. Hence, 802.11n PHY applies MIMO technique to improve performance over multipath environment. With this enhancement in the PHY layer, the peak PHY rate can be boosted up to 600 Mbps to meet the IEEE 802.11n high throughput requirement.

B. *Aggregation*

Increasing the data rate of PHY layer alone is not enough to achieve the desired MAC layer throughput of more than 100 Mbps due to rate independent overheads. We have described the overhead in legacy IEEE 802.11 MAC, which has been partly solved by the TXOP technique introduced by the 802.11e amendment. Aggregation may further enhance efficiency and channel utilization. The aggregation mechanism combines multiple data packets from the upper layer into one larger aggregated data frame for transmission. Overhead in multiple frame transmissions is reduced since the header overhead and interframe time is saved.

In IEEE 802.11n MAC, the aggregation mechanism is designed as two-level aggregation scheme, and hence two types of aggregation frames are defined: aggregate MAC protocol service unit (A-MSDU) and aggregate MAC protocol data unit (A-MPDU). The aggregation mechanism is able to function with A-MPDU, A-MSDU, or using both of them to form two-level aggregation. A-MSDU is composed of multiple MSDUs and is created when MSDUs are received by the MAC layer. To ease the de-aggregation process, the size of a MSDU, including its own subframe header and padding, must be multiple of 4 bytes. Two parameters are used to form an A-MSDUs: the maximum length of an A-MSDU (3839 or 7935 bytes by default), and the maximum waiting time before creating an A-MSDU. Aggregated MSDUs must belong to the same traffic flow (same TID) and have the same destination and source. Broadcasting and multicasting packets are excluded.

In the second level, multiple MPDUs are aggregated into an A-MPDU, which is created before sending the MSDU (or A-MSDU) to the PHY layer for its transmission. Unlike the A-MSDU, the MAC layer does not wait for additional time before the A-MPDU aggregation. It only uses the available MPDUs in the queue to create A-MPDUs. The TID of each MPDU in the same A-MPDU might be different. The maximum size limit of A-MPDU is 65535 bytes. In an A-MPDU, each MPDU has an MPDU delimiter at the beginning and padding bytes at the end. These bytes ensure that the size of each MPDU is multiple of 4 bytes. Delimiter is used to separate the MPDUs in an A-MPDU. The de-aggregation process first checks the CRC integrity. If the CRC check is passed, the MPDU will be de aggregated and sent to upper layer.

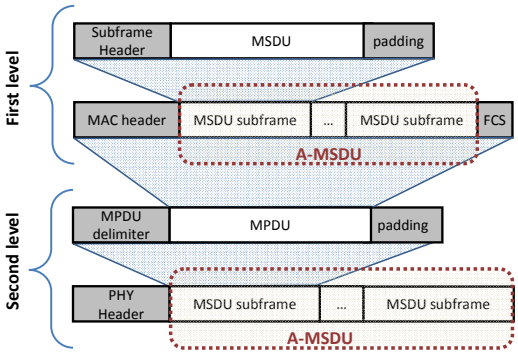


Figure 2. Aggregation in IEEE 802.11n.

The two-level aggregation mechanism is shown in Fig. 2. In the first level, MSDUs received by the MAC layer from the upper layer are buffered for a short time until A-MSDUs are formed according to their TID, destination, source, and the maximum size of A-MSDU. Then, the complete A-MSDUs and other non-aggregate MSDUs are sent to the second level to form an A-MPDU. Due to compatibility reasons, every MPDU in A-MPDU should not exceed 4095 bytes. It must be taken into account that 802.11n aggregation does not support frame fragmentation. Only complete A-MSDUs or MSDUs, not the fragments of A-MSDUs or MSDUs, could be contained in an A-MPDU. The whole aggregation mechanism completes when A-MPDU is created.

C. *Block ACK*

Originally, the block ACK operation incorporates the TXOP mechanism, as previously described in the 802.11e MAC design. The block ACK mechanism is further enhanced in 802.11n to be applied with the aggregation feature. Although a larger aggregation frame may significantly reduce the overhead in a transmission, the frame error rate is higher as the size of the frame increases. Large frames in a high bit-error-rate (BER) wireless environment have a higher error probability and may need more retransmissions. The network performance might be degraded. To overcome this drawback of the aggregation, the block ACK mechanism is modified in 802.11n to support multiple MPDUs in an A-MPDU. Fig. 3 shows the block ACK mechanism. When an A-MPDU from one station is received and errors are found in some of the aggregated MPDUs, the receiving node sends a block ACK which only acknowledges the correct MPDUs. The sender only must retransmit those non-acknowledged MPDUs. Block ACK mechanism resolves the drawback of large aggregation in the error-prone wireless environment and further enhances the performance of 802.11n MAC (Fig. 3).

Block ACK mechanism only applies to AMPDU, but not A-MSDU. That is, when an MSDU is found to be incorrect, the whole A-MSDU needs to be transmitted for error recovery. The maximum number of MPDUs in an A-MPDU is limited to 64 as one block ACK bitmap can only

 ISBN: 978-1-61208-133-5

acknowledge at most 64. The original block ACK message in IEEE 802.11e contains a Block ACK bitmap field with 64 × 2 bytes. These two bytes record the fragment number of the MSDUs to be acknowledged. However, fragmentation of MSDU is not allowed in 802.11n A-MPDU. Thus, those 2 bytes can be reduced to 1 byte, and the block ACK bitmap is compressed to 64 bytes. This is known as compressed block ACK. Compared with 802.11e, the overhead of block ACK bitmap in 802.11n is reduced. Moreover, IEEE 802.11n introduces the use of implicit block ACK. With this mechanism is not necessary to request the sending of ACK block, reducing overhead.

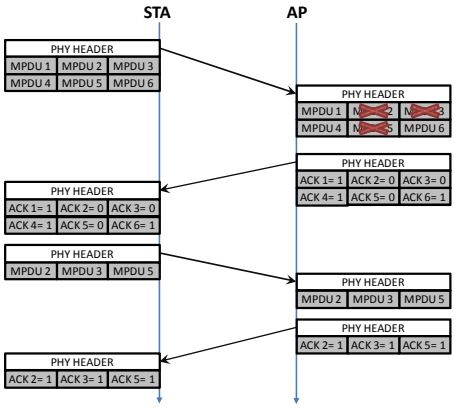


Figure 3. Block ACK with aggregation.

D. *Reverse Direction*

Reverse direction mechanism is a novel breakthrough to enhance the efficiency of TXOP. In conventional TXOP operation, the transmission is uni-directional from the station holding the TXOP, which is not applicable in some network services with bi-directional traffic like VoIP, videoconference and on-line gaming. The conventional TXOP operation only helps the forward direction transmission but not the reverse direction transmission. For application with bi-directional traffic, their performance is degraded by the random backoff and contention of the TXOP. Reverse direction mechanism allows the holder of a TXOP to allocate the unused TXOP time to its receivers and hence, enhancing the channel utilization and performance of reverse direction traffic flows.

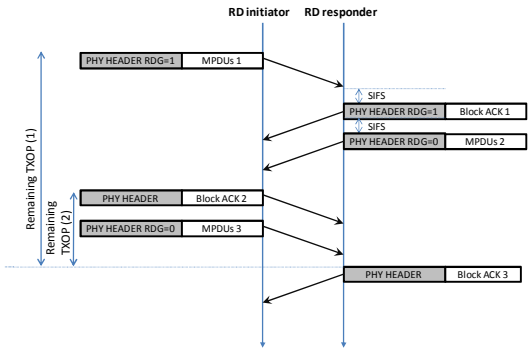


Figure 4. Reverse direction.

The reverse direction operation is illustrated in Fig. 4. In reverse direction operation, two types of stations are defined: RD initiator and RD responder. RD initiator is the station which holds the TXOP and has the right to send reverse direction grant (RDG) to the RD responder. RDG is marked in the 802.11n header and is sent with the data frame to the RD responder. When the RD responder receives the data frame with RDG, it responds with RDG acknowledgement if it has data to be sent, or without RDG if there is no data to be sent to the RD initiator. If the acknowledgement is marked with RDG, the RD initiator will wait for the transmission from the RD responder, which will start after a SIFS or a reduced inter-frame space (RIFS) once the RDG ACK is sent. RIFS can be used in the scheme when no packet is expected to be received after the transmission, which is the case here. If there is still data to be sent from the RD responder, it can mark RDG (which represents MORE DATA) in the data frame header to notify the initiator. The RD initiator still has the right to accept the request. To allocate the remaining TXOP, the initiator will mark the RDG in the acknowledge message or the next data frame. To reject the new RDG request, the initiator just ignores it.

The major enhancement of the reverse direction mechanism is the delay reduction in reverse link traffic. These reverse direction data packets do not need to wait in queue until the station holds a TXOP but can be transmitted immediately when the RD responder is allocated for the remaining TXOP. This feature can benefit a delay-sensitive service like VoIP. We will show a performance enhancement in the simulation section.

IV. PERFORMANCE EVALUATIONS

In this section, we carry out a performance analysis on the effectiveness of the IEEE 802.11n standard. We examine the major improvements introduced by IEEE 802.11n MAC: aggregation, block acknowledgement, and reverse direction. We show the impact of each parameter in the network performance.

A. *Scenario*

In our simulations, we model an IEEE 802.11n wireless LAN using OPNET Modeler tool 10.0 [14]. We use a wireless LAN consisting of several wireless stations and an access point connected to a wired node, which serves as sink for the flows from the wireless domain. All the stations are located within a basic service Set (BSS), i.e., every station is able to detect a transmission from any other station. The parameters of the wired link have been chosen to ensure that the bandwidth bottleneck of the system is within the wireless LAN. Each wireless station operates at 300 Mbit/s IEEE 802.11n mode and we assume the use of an ideal channel. All the stations use a MIMO configuration with 2x2 antennas.

For all the scenarios, we have assumed a bi-directional and constant bit-rate application. This application has an average rate of 8 Mbps and a packet size equal to 1000 bytes. We start by simulating a WLAN consisting of two wireless stations. We then gradually increase the network load by adding the number of stations each time. We increase the

 ISBN: 978-1-61208-133-5

number of stations 2 by 2 starting from 2 and up to 20. In this way, the offered load is increased from 32 Mbps (16 Mbps in the AP and 16 Mbps in the stations) up to 320Mbps. The traffic sources are randomly activated within of the interval [1,1.5] seconds from the start of the simulation. Throughout our study, we have simulated two minutes of operation of each particular scenario. Our measurements start after a warm-up period allowing us to collect the statistics under steady-state conditions. Each point in our plots is an average over thirty simulation runs, and the error bars indicate the 95% confidence interval.

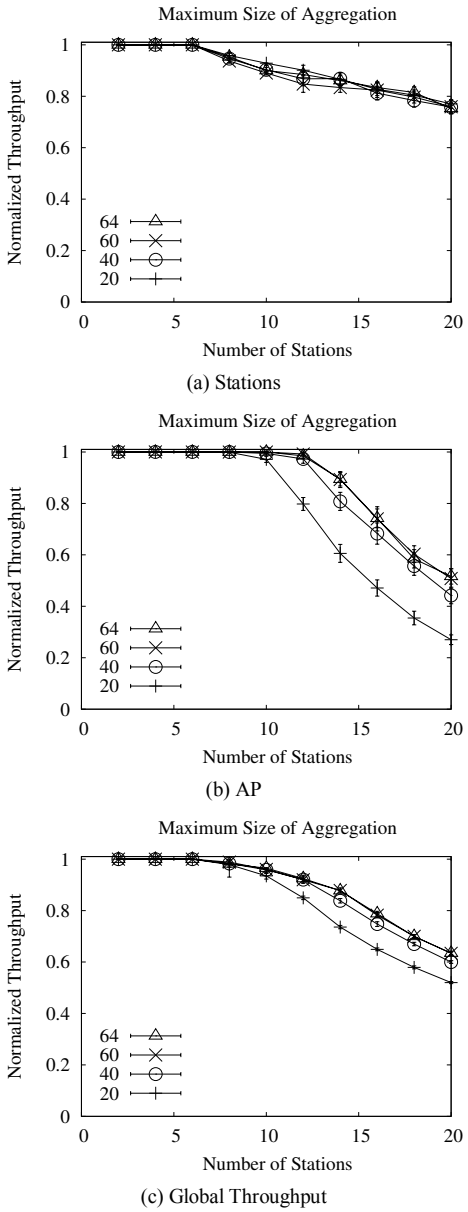


Figure 5. Performance evaluation for different sizes of aggregation.

For the purpose of our performance study we are selected the normalized throughput. The normalized throughput is calculated as the percentage of the offered load actually delivered to destination.

B. Results

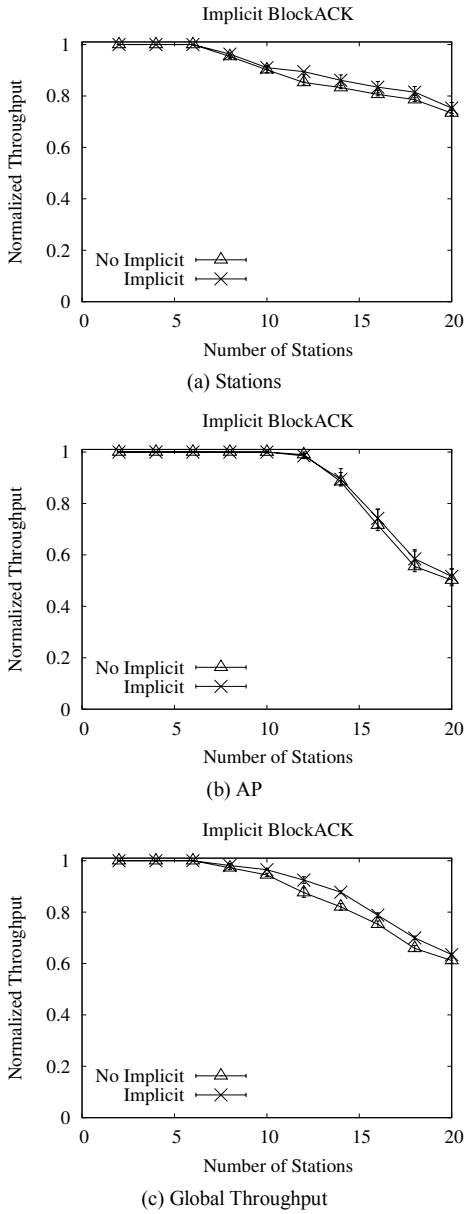


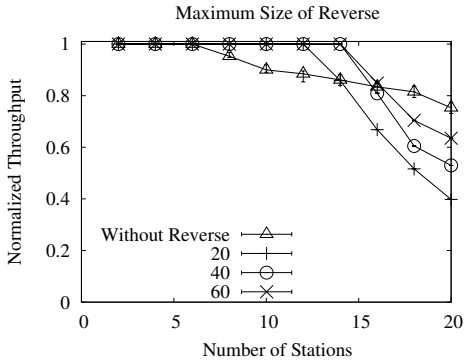
Figure 6. Performance evaluations with implicit block ACK..

Figure 5 shows the normalized throughput with different maximum size of aggregation. In this simulations we have fixed the first level of aggregation (A-MSDU) changing the second level of aggregation (A-MPDU). We evaluate several maximum sizes for the aggregated frames (20, 40, 60 and 64 packets). The figure shows that the larger the aggregation, higher performance is achieved. This improved performance is higher in the AP (see Figure 5.b). This is because the AP has more packets to transmit that the stations so the AP will

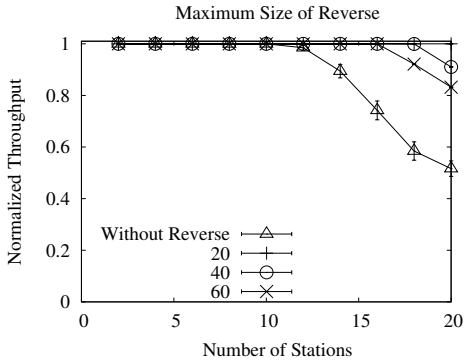
 ISBN: 978-1-61208-133-5

use the higher size of aggregation. This result is expected since a higher aggregation size leads to a lower overhead.

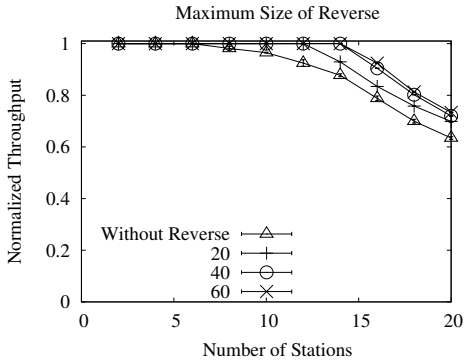
Figure 6 shows the effect of using the implicit block ACK. In these simulations, the maximum size of aggregation is fixed to 60. The figure shows that the normalized throughput increases when an implicit ACK block is used. This is due to the reduction in overhead. When the implicit block ACK is used, the station does not request confirmations.



(a) Stations



(b) AP



(c) Global Throughput

Figure 7. Performance evaluations for different sizes of reverse direction..

Finally, the impact of use the reverse direction is shown in Figure 7. This figure shows that the use of reverse direction improves network performance. This result also was expected because if bidirectional applications are transmitted, the destination station takes advantage of the TXOP that belongs to the sending station. However, the size of the reverse should not be too large, because it gives too much traffic to the destination station

V. CONCLUSIONS

We have investigated the performance of IEEE 802.11n MAC protocol. The three enhanced 802.11n MAC mechanisms: aggregation, block acknowledgement and reverse direction have been discussed. We have implemented an 802.11n module in Opnet Modeler. We designed several simulation scenarios in order to evaluate the influence of the different enhanced mechanisms. The simulation results have shown that the aggregations, implicit block ACK and reverse direction mechanisms reduce the overhead allowing an increase of the network performance.

REFERENCES

[1] IEEE Std. 802.11 WG, “Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications,”. 1999.

[2] V. Jones, R. DeVegt, and T. Jerry, “Interest for Higher Data Rates (HDR) Extension to 802.11a,” IEEE 802.11n working doc. 802.11-02-081r0, Jan. 20023.

[3] J. Rosdahl, “Draft Project Authorization Request (PAR) for High Throughput Study Group,” IEEE 802.11n working doc. 802.11-02/798r2, Mar. 2003.

[4] IEEE P802.11n, “Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Enhancements for Higher Throughput,” 2009.

[5] Magis Networks White Paper, ”IEEE 802.11 e/a Throughput Analysis”, 2004, www.magisnetworks.com.

[6] Y. Xiao and J. Rosdahl, “Performance analysis and enhancement for the current and future IEEE 802.11 MAC protocols”, ACM SIGMOBILE Mobile Computing and Communications Review (MC2R), special issue on Wireless Home Networks, Vol. 7, No. 2, Apr. 2003, pp. 6-19.

[7] Lin Y, Wong VWS (2006) WSN01-1: frame aggregation and optimal frame size adaptation for IEEE 802.11n WLANs. IEEE Global Telecommunications Conference, San Francisco, December.

[8] Xiao Y (2005) IEEE 802.11n: enhancements for higher throughput in wireless LANs. IEEE Wirel Commun 12(6):82–91:

[9] Chih-Yu Wang and Hung-YuWei, “IEEE 802.11n MAC Enhancement and Performance Evaluation”. Mobile Netw Appl, Vol. 14, No. 6, pp. 760-771.

[10] IEEE Std. 802.11 WG Std 802.11a, “Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: High-speed Physical Layer in the 5 GHz Band”, 1999

[11] IEEE Std. 802.11 WG Std 802.11b, “Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Further Higher-Speed Physical Layer Extension in the 2.4 GHz Band”, 1999.

[12] IEEE Std. 802.11 WG Std 802.11g, “Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Further Higher-Speed Physical Layer Extension in the 2.4 GHz Band”, 2003.

[13] IEEE Std. 802.11 WG Std 802.11e: “Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Medium Access Control (MAC) Quality of Service Enhancements", 2005

[14] OPNET Technologies Inc., http://www.opnet.

 ISBN: 978-1-61208-133-5

GeoWiFi: A Geopositioning System Based on WiFi Networks

Jaime Lloret, Jesus Tomas, Alejandro Canovas, Irene Bellver
Instituto de Investigación para la Gestión Integrada de Zonas Costeras
Universidad Politécnica de Valencia, Spain
jlloret@dcom.upv.es, jtomas@dcom.upv.es, alcasol@posgrado.upv.es, irbelser@upvnet.upv.es

***Abstract*—GPS can only be used in outdoors where the satellite signals can be received, so it cannot be used to know the position of a device inside buildings. Cellular networks can be used to locate devices indoors, but the mobile phone network is owned by an enterprise, so a regular user is not allowed to know this information. So, new positioning systems are needed for urban zones and indoor locations. In this paper, we present a geopositioning system that uses public and private WiFi networks placed in cities in order to estimate the location of the user. Then we present the architecture of the system and the prototype implementation. Finally, we will show the developed protocol operation.**

Keywords-geopositioning; WiFi location; Wifi placement; WiFi tracking.

I. INTRODUCTION

Nowadays, the society demands new location systems to adapt the new services provided by the information technology, which demonstrates the acceptation and projection of the society in the Information Technologies. Some years ago, the Global Positioning System (GPS) [1] involved a revolution in the positioning systems (navigators, navy control, anti-thief systems, etc.). It is actually the most used positioning system in the world because it can be used worldwide. Its main features are:

- Global position thanks to a constellation of 30 satellites of the Army of USA.
- It is possible to include a random mistake for security reasons.
- There is a precision up to 5 meters.
- It is indispensable to have direct view from the devices to the satellites, so it cannot operate inside the buildings, tunnel, narrow streets, and in adverse atmospheric conditions.

However, although it is a mature system, this technology has an important limitation: it is not operative in indoors and places where there is a roof, and even in urban environments because a direct view to the satellite is required. GPS have several well-know limitations [2]. These limitations are found in:

- GPS signal reception
- GPS signal integrity
- GPS signal accuracy

GPS measurements are influenced by several types of errors: satellite errors, signal propagation errors, receiver errors and those provided by the GPS geometry. There have been several researchers with the purpose of solving GPS limitations.

Because the cellular telecommunications infrastructure is very large in urban zones, it can be used to locate mobile devices inside the cities and towns. Moreover, mobile phone networks offer mobile coverage inside the buildings. But cellular networks are owned by an enterprise, so a regular user is not allowed to know this information. Its main features are:

- The mobile device searches all cells that is able to connect with. This information is used to estimate its position.
- The system has high coverage and can be use in all places where there are mobile networks.
- It is cheap because it uses a network that is used for other purposes.
- It is quite imprecise, from 200 to 20.000 meters (depending on the number of cells). The precision is enhanced as a function of the density of cells.
- There are more precise systems that involve signal levels and delay times. Nevertheless, this information is only available by the mobile service operator.

Recently, a new location technology based on WiFi networks has been developed as an alternative. This technology uses the radio signals provided by all WiFi networks found in the urban zone in order to estimate the position of the device. In addition, the number of private and public access points is increasing in the cities and the WiFi coverage is getting higher, thus higher precision can be provided. Because WiFi networks can cross several walls, a continuous service can be offered in outdoors as much as building indoors [3]. The position based on WiFi has the following features:

- It works inside the buildings.
- The precision is variable and depends on the number of WiFi networks. It achieves up to 2 meters of accuracy.
- It requires a training phase in order to measure the signal level from a large number of geographic positions. The system could not run properly where the training process has not been performed.
- The set of signal levels measured in a moment is known as a WiFi pattern.

This new technology will favor a wide range of new applications oriented to people positioning and tracking. Recent studies show how a campus that has deployed a Wireless Local Area Network (WLAN), allow the network administrators to place the services in the most appropriate

 ISBN: 978-1-61208-133-5

sites, and even the behavior of the people could be studied and their mobility tracked [4]. Otherwise, GPS technology is commonly used for cars with high mobility or pedestrian (when they are in rural zones or hill walking).

Positioning systems based on WiFi networks is a promising technology, but its deployment presents many Issues:

- The radio coverage of the access points depend on the walls crossed and on the obstacles in the in the line of sight between the emitter and the receiver. Moreover, in indoors, the power received in each point depends on many factors (even differs between different IEEE 802.11 variants) [5].
- Generally, where the WiFi access points are physically placed is unknown. Moreover, their position can vary dynamically, new access points could appear and others could disappear. Therefore, a training system or a baseline data is indispensable to characterize the network [6].
- There is not any WiFi position standard and normalized position processes in existence.
- It requires complex algorithms to locate a device starting only from radio signals. Nowadays, the research community is discussing on which is the most adequate algorithm [7].

In order to solve these issues and difficulties, some algorithms and specific and innovative processes should be developed. In this paper we will show a new geopositioning system based on public and private WiFi networks for urban zones. We have developed several algorithms in order to have an accurate system.

The remainder of this paper is structured as follows. In Section 2 we will discuss some existing related work. Our architecture is explained in section 3. Section 4 details the deployed prototype. The protocol operation is shown in section 5. Finally, Section 6 draws our conclusion and future work.

II. RELATED WORK

As we have stated before, the global positioning systems (GPS) are not efficient in outdoor places where there are buildings, because there are shadow areas where there is no satellite signal. This related work section has been split in three main parts. The first one shows WiFi positioning systems for outdoors. The second one shows some alternatives or proposals that combine GPS with WiFi and other technologies. The third one shows some indoor WiFi-based positioning systems.

Y. Cheng et al. evaluate the feasibility of building a wide-area 802.11 Wi-Fi-based positioning system in [8]. Then, they explore how a user's device can accurately estimate its location using existing hardware and infrastructure and with minimal calibration. They evaluate the estimation accuracy of a number of different algorithms (many of which were originally proposed in the context of precise indoor location) in a variety of scenarios. Although the accuracy of their system is lower than existing positioning systems, it requires substantially lower calibration overhead than existing indoor positioning systems and provides easy deployment and coverage across large metropolitan areas. They conclude the paper letting us know that in dense urban areas Place Lab's positioning accuracy is between 13–20 meters.

In [9], B. Li et al. analyze two WiFi positioning technologies for outdoors: trilateration and fingerprinting. Then, they carry out a fingerprinting study case in the Sydney central business district (CBD) area where WiFi APs are densely deployed. The fingerprint of a specific place can be used to identify the location. The key idea of the fingerprinting approach is to map location-sensitive parameters of measured radio signals in areas of interest. Their results show that the fingerprinting-based positioning system works well for outdoor localization, especially when directional information is utilized, with errors in the tens of meters. The same authors propose a location fingerprinting method in wireless LAN (WLAN) positioning in [10].

In [11], Amalina Abdul Halim developed an outdoor WiFi positioning system that can estimate the location of mobile devices based on the signal strength broadcasted by the access points. One of the main innovations of this work is the device location algorithm, which uses the K-NN algorithm (based on the nearest neighbor). A prototype was applied to the bus services of the University of Technology of MARA (UiTM), which allows estimating the location of the bus and the arrival time at the next stop.

Hereinafter we will see some positioning proposals that combine GPS with WiFi and other systems.

In [12], M. Weyn and F. Schrooyen combine WiFi localization with satellite-based navigation to form a WiFi-Assisted-GPS ubiquitous solution to make GNSS useful in urban and indoor environments. This combination is analyzed from a mathematical perspective. The WiFi location is based on fingerprints. To minimize the initialization and the training of the WiFi-positioning a self mapping system is needed.

The authors of the paper in [13] used the Time Difference of Arrival (TDOA) measurement, generated by the pseudorange observations of two visible satellites GPS with the WiFi fingerprint technology. The authors show that the integration of both technologies can improve the positioning accuracy by more than 50% when the method is applied.

In [14], Pornpen Ratsameethammawong and M. L. Kulthon Kasemsan study cell phone and WiFi positioning systems in environments where standard GPS fails. They focus their main research on a location system based on the Wi-Fi signals received from the existing hardware infrastructure of their University. They conclude that is better to use combined systems.

In [15], B. Li propose a client-based mobile phone location tracking by the combination of GPS, Wi-Fi and Cell location technology. Their proposal use vector calculations to track and locate mobile phones whereabouts are introduced. The combined methods make the tracking and locating of moving mobile phone more accurate and more effective despite the fact that GPS signal is not available.

 ISBN: 978-1-61208-133-5

In [16], A. Kealy et al. study the potential of positioning WiFi to provide solutions in indoor environments. They present the practical results generated from a case study comparing commercial WiFi positioning systems.

Because indoor WiFi positioning systems have been studied in depth, we will show some methods, some of them published by the same authors of this paper.

The location systems in indoor environments have difficulties as the walls, interferences, multipath effect, humidity, temperature variations, etc. In [17], the authors describe two approaches where wireless sensors could find their position using WLAN technology inside a floor of a building. Both approaches are based on the Received Signal Strength Indicator (RSSI). The first approach uses a training session and the position is based on a heuristic system using the training measurements. The second approach uses triangulation model with some fixed access points, but taking into account wall losses and signal variations.

This idea of combining several positioning systems to locate a device is used in [18]. The authors proposed a new stochastic approach which is based on a combination of deductive and inductive methods whereby wireless sensors could determine their positions using WLAN technology inside a floor of a building. Their goal is to reduce the training phase in an indoor environment, but, without a loss of precision. They conclude their paper showing that their method is better than the existing ones.

None of the papers aforementioned use exclusively unknown WiFi access points to estimate the position of the devices in outdoors.

III. SYSTEM ARCHITECTURE

This section presents the architecture of the proposed geopositioning system. There are three main entities: the devices to be located (mobile devices), the one used to estimate the position (positioning manager), and the application infrastructure. It is shown in figure 1.

A. Mobile device

This is the device that is desired to be located. It is typically a mobile phone with a wireless interface, although it could be a PDA or a laptop. A software application will be installed in the mobile device. It will be running hidden and transparent to the user, so the device can be used regularly as a phone, PDA or laptop. It can be a device exclusively developed for location purposes. It can be included in a bracelet or a wristband, or could be embedded in any type of device that is wanted to be located.

The operation of the mobile device is as follows. It will connect to the positioning manager periodically and the positioning manager will reply with the information used for to estimate its location. This information consists on the radio signal strength received from several access points of the WiFi network (hereinafter called WiFi pattern). Optionally, the GPS position could also be sent if the device allows it and if the device is under several GPS satellites coverage. If the device is a mobile phone, it could also use a third location system: the location provided by the cells of the mobile phone providers.

B. Positioning manager

The main function of this entity is to geoposition the mobile devices by using the information received from them. It can also act as a database to allow the vertical software applications ask the position of any mobile device. The positioning system can be divided in three modules:

- Positioning module

The positioning server receives the positioning information from the mobile devices periodically. Using this information, the positioning module estimates the position of the device and stores it in a database.

Typically, this information consists of the signal strength level received from the access points that are closed to the mobile device. It could also receive the GPS position or the closest mobile cells.

In order to find the position from the WiFi signals, we use the algorithm described in [18], which has been developed by the authors of this paper. Its main drawback is that this method needs a previous training process, so it is only applicable when the mobile device is placed in a previously training area.

In the case of GPS, a positioning module is not needed because it directly provides longitude and latitude coordinates.

The mobile cells positioning case can only used if the device is a mobile phone. We have configured our system to use this option only if the others are not available. The information used to estimate the position is using the closest cells signals. We are aware that this method could have errors of about 100 meters [19]. Last research includes artificial neural networks in the location methods.

- Enquiry module

It provides service to the vertical software applications. It allows to ask the current position of a mobile device, or to ask for a historic list of its position. The information provided is the code of the device, longitude, latitude, high, accuracy estimation, time and date. This module will be implemented as a web service. XML documents will be exchanged using the HTTP protocol. The web service will be used by the vertical applications or by third parties.

- Training module

A training phase is indispensable for the positioning module, so it is performed in the training module. The WiFi patterns, jointly the position coordinates, are provided during this phase. Its purpose is to create a big database where a coordinate is associated with the received signal strengths.

The training process will be carried out in differently in outdoor than in indoor. In outdoor, a device with WiFi and GPS is needed. This device will be transmitting both GPS and WiFi data to the training module in the area to train. The indoor training process is more complex. It has to be carried out by a person that has to be indicating its position inside the building while it is shifting. A comparison between trilateration and training methods is shown in [6]. However, there are methods that let us estimate the position inside the buildings by using the training carried out outside the buildings. These methods are quite less accurate.

 ISBN: 978-1-61208-133-5

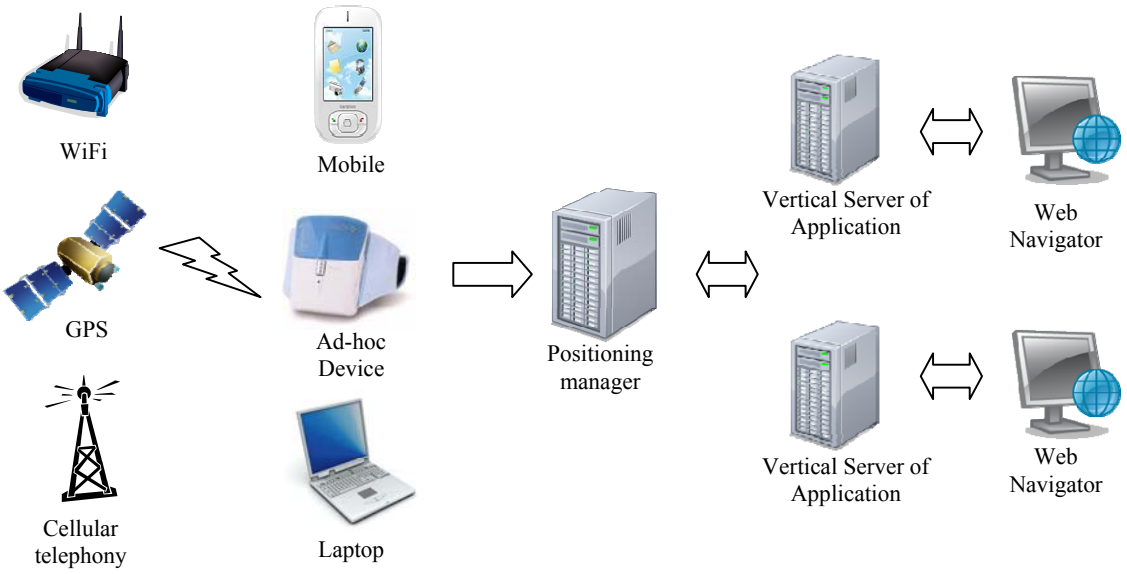


Figure 1. Positioning System Architecture

C. *Applications infastructure*

Each vertical application will have specific needs that will be solved in this entity. All software applications will acquire the positioning information using the enquiry module. The applications infrastructure is formed by an Apache Web Server (using PHP programming) and a MySQL database. We have used AJAX (Asynchronous JavaScript And XML) to create interactive applications in our system.

Figure 2 shows the architecture operation.

IV. PROTOTYPE IMPLEMENTATION

In this section we describe the elements needed to implement the prototype. As a basis the architecture described in the previous section will be used.

We have deployed two positioning devices with different features in order to cover higher number of applications.

A. *Cellular Mobile Device*

Nowadays, the cellular mobile device has become an indispensable tool in the personal communications. Thus, a location software application installed in the phone may open a wide range of new applications.

In order to use the device, it should have a WiFi interface and the software application should be compatible with the mobile phone operative system. In order to reach to a high number of mobile devices we have developed our software application to different operative systems. These Operative systems are the following:

- Microsoft .NET (Windows Mobile)
- Google Android
- Apple iPhone

Many mobile devices are able to obtain their location by using additional information sources than WiFi signals. The software application detects if the device has GPS and identifies the closest mobile phone cells (in case of a cellular phone). Then the device will send this jointly with the WiFi patters. This information is sent periodically to the positioning module in regular intervals which are configurable by the application and depends on the mobility of the user (larger time intervals will be used when there is less mobility).

B. *Specific Device*

There are many cases were a mobile phone cannot be used so another type of device is needed. Some examples are a positioning wristband or bracelet, or an anti-theft device. Because of it, a specific device for geopositioning is also needed in the market.

In this case we have chosen a SoC (Solution on Chip), from G2 enterprise [20], to develop the device. We used G2C543 Wi-Fi SoC model. It embeds a WiFi interface, a CPU and an I/O ports in a unique Chip. It has the following features:

- 32-bits CPU
- Internal memory: 128 KB RAM and 512 KB ROM
- A flexible interface for external sensors.
- An operative System with TCP/IP protocol stack, security and IEEE 802.11 b/g.
- A power management subsystem to save energy.

G2 enterprise has also a module called Epsilon Module Family PB which can be used with G2C543 to enhance its features [21].

- 8 Mb Memory Flash
- Low consumption WiFi Antenna.
- Real time clock.
- Power regulator for alkaline batteries.
- Reduced dimensions.

Table I shows a comparison of the main features of the two proposed devices for our geopositioning system.

 ISBN: 978-1-61208-133-5

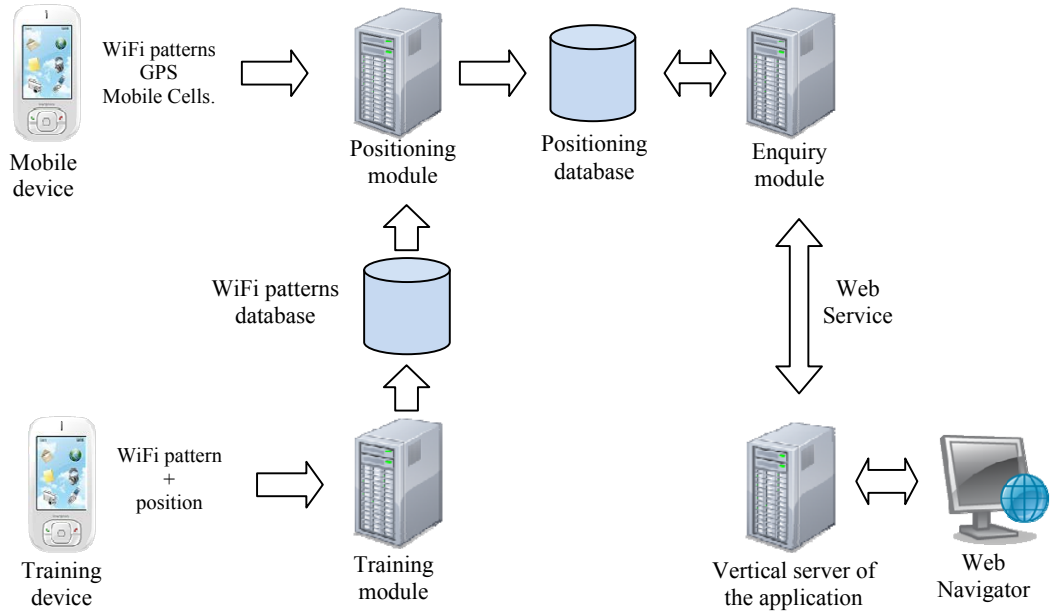


Figure 2. Iteraction between positioning manager modules.

TABLE I. DEVICE FEATURES COMPARITON

	Cellular Mobile Device	**Specific Device**
Type of device	Phone + positioning	Focused only on positioning
Coverage area	Medium city	Building or campus
Type of connection	Internet (GPRS/UMTS) + WiFi	Only WiFi
Application Examples	Teenager control / mobile workers …	Movement Control / equipment management

V. PROTOCOL OPERATION

In this section we describe the operation of the developed protocol. First, a user connects to the vertical server of application using a web navigator. This web page allows him/her to authenticate in the database or register as a new user. The web page is shown in Figure 3. When there is a new user, he/she must click on the "New User" button and Figure 4 appears. It shows the information needed for registration. This procedure allows registering himself. Then, the user must choose the group to join. So after being authenticated, it receives a popup asking for the joining group (see figure 5). This popup even allows creating a new group if the user will. Finally, if the device is not included in the system, it must be added by clicking in the "New device" button. Figure 6 shows the information that must be filled up to add a new device in the system. Only the users with enough privileges can add new devices to the system. If the device is yet added, this step is not needed because the system will know the device (each device has a unique WiFi interface MAC address which is also associated to a user or users).

Finally, after having filled up the user and the group and having the device recognized in the system, the GeoWiFi main window is opened (see figure 7). It shows the device list and a map with the placement of all devices in a group. We have programmed the application to reply some enquires. Thus, a user can know the last position for each device in its group. Moreover, the trace of a device can be also estimated. The map screen shows the placement of each device in real time. The "OK" button leaves the application.

A registered user, that is already registered in a group and is using a device that is included in the system, will follow the message flow shown in figure 8. Only four messages are exchanged between the Vertical server of the application and the user.

VI. CONCLUSION

In this paper, we have shown the implementation of a Geopositioning System Based on WiFi Networks. It uses the radio signal strength of the WiFi networks in order to estimate the position of the users.

The system uses a WiFi pattern method that allows having the most accurate system. Thus, a previous training phase is needed before setting up the system.

The developed system allows knowing the position or tracking teenager, mobile workers, and equipment movement.

We have shown the web pages used for authentication, registering and for displaying the information estimated by the system.

Our next step is to provide higher security in the user authentication procedure and to provide a intrusion detection system in order to grant access only original MAC addresses (not copied or cloned).

 ISBN: 978-1-61208-133-5

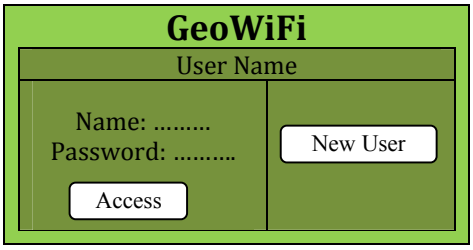


Figure 3. Firs access to the vertical server of application

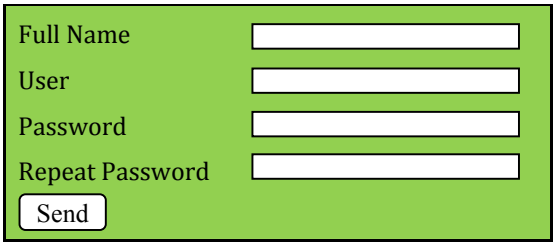


Figure 4. New user registration

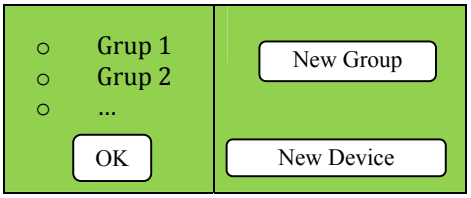


Figure 5. Group selection

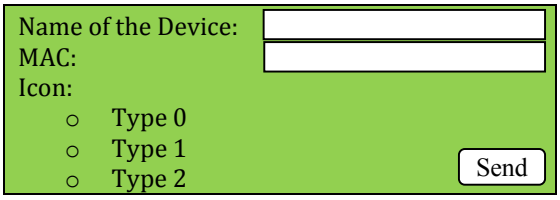


Figure 6. New device window

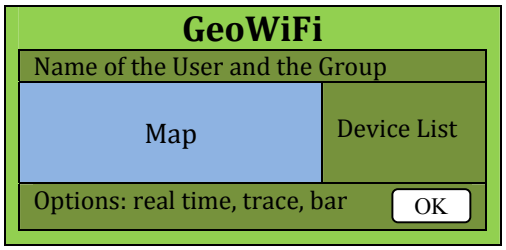


Figure 7. GeoWiFi main window

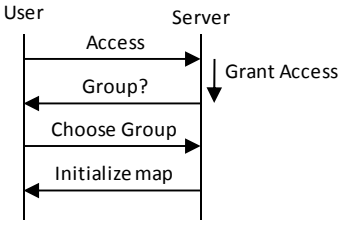


Figure 8. Messages Flow

REFERENCES

[1] B. Hofmann-Wellenhof, H. Lichtenegger, and J. Collins, Global Positioning System. Theory and practice, Springer, Wien (Austria), 347 pages.1993.

[2] Alfred Kleusberg and Richard B. Langley, The Limitations of GPS. GPS World, March/April 1990. pp 50-52.

[3] Diana Bri, Hugo Coll, Sandra Sendra, and Jaime Lloret. Providing Outdoor and Indoor Ubiquity with WLANs, chapter of the Handbook of Research on Mobility and Computing: Evolving Technologies and Ubiquitous Impacts, IGI Global, Pp. 1155-1168, 2011.

[4] S. Sendra, M. Garcia, C. Turro, and J. Lloret, User's Macro and Micro-mobility Study using WLANs in a University Campus, International Journal On Advances in Internet Technology, Vol. 4, Nº 1&2, 2011. In press

[5] S. Sendra, M. Garcia, C. Turro, and J. Lloret, WLAN IEEE 802.11 a/b/g/n Indoor Coverage and Interference Performance Study, Int. J. on Advances in Networks and Services, Vol. 4, N 1&2,2011. In press.

[6] Miguel Garcia, Carlos Martinez, Jesus Tomas, and Jaime Lloret, Wireless Sensors self-location in an Indoor WLAN environment, 1st Int. Conf. on Sensor Technologies and Applications (Sensorcomm 2007), Valencia (Spain), October 14-20, 2007. Pp.146-151.

[7] M. Garcia, F. Boronat, J. Tomás, and J. Lloret, The Development of Two Systems for Indoor Wireless Sensors Self-location, Ad Hoc & Sensor Wireless Networks, Vol. 8, Issue 3-4, Pp. 235-258.

[8] Y. Cheng, Y. Chawathe, A. LaMarca, and J. Krumm, Accuracy characterization for metropolitan-scale Wi-Fi localization, MobiSys 2005, Seattle, Washington, USA, 6-8 June 2005. Pp. 233–245.

[9] B. Li, I. Quader, and A.G. Dempster, "On Outdoor Positioning with Wi-Fi", Journal of GPS, vol. 7, issue. 1, pp. 18-26. 2008.

[10] B. Li, Y. Wang, H.K. Lee, A.G. Dempster, and C. Rizos. Method for yielding a database of location fingerprints in WLAN, IEE Proceedings Communications, vol. 152, no. 5, pp. 580-586. 2005.

[11] Amalina Abdul Halim, Wifi positioning system, Faculty of Information Technology and Quantitative Science, Mara University of Teknology Shah Alam. Thesis. May 2006.

[12] M. Weyn and F. Schrooyen, A Wifi Assisted GPS Positioning Concept, 3rd European Conf. on the Use of Modern Information and Communication Technologies, Gent, 13-14 March 2008. Pp 479-486.

[13] Binghao Li, Yong Khing Tan, and Andrew G. Dempster. Using two GPS satellites to improve WiFi positioning accuracy in urban canyons. IEET Communications. 2011. In press.

[14] Pornpen Ratsameethammawong and M.L. Kulthon Kasemsan. Mobile Phone Location Tracking by the Combination of GPS, Wi-Fi and Cell Location Technology. Communications of the IBIMA. Vol. 2010, Article ID 566928, 7 pages. 2010

[15] B. Li, A. G. Dempster, and C. Rizos. Positioning in environments where GPS fails. FIG Congress 2010, Sydney (Australia), 11-16 April 2010. Pp 1-18.

[16] A. Kealy, B. Li, T. Gallagher, and A. Dempster. Evaluation of WiFi Technologies for Indoor Positioning Applications. Surveying & Spatial Sciences Institute Biennial International Conference, Adelaide, South Australia, 28 Sep. - 2 Oct., 2009. Pp. 411-421. 2009.

[17] M. Garcia, F. Boronat, J. Tomas, and J. Lloret, The Development of Two Systems for Indoor Wireless Sensors Self-location, Ad Hoc & Sensor Wireless Networks. Vol. 8, Issue 3-4, June 2009. Pp. 235-258,

[18] J. Lloret, J. Tomas, M. Garcia, and A. Canovas, Hybrid Stochastic Approach for Wireless Sensors Self-Location in Indoor Environments, Sensors, Vol. 9, Issue 5, Pp. 3695-3712, May 2009.

[19] C. Drane, M. Macnaughtan, and C. Scott, Positioning GSM Telephones. IEEE Communications Magazine, April 1998. Pp. 46-59.

[20] G2 Microsystems Website. At http://www.g2microsystems.com/ [Last access January 20, 2011]

[21] G2 Epsilon Module Family PB Website. At http://www.g2microsystems.com/products/modules.html [Last access January 20, 2011]

 ISBN: 978-1-61208-133-5

Estimation of Packet Loss Probability from Traffic Parameters for Multimedia over IP

Ahmad Vakili
Energy, Materials, and Telecommunications (EMT)
Institut national de la recherche scientifique (INRS)
Montreal, Quebec, Canada
Email: vakili@emt.inrs.ca

Jean-Charles Grégoire
Energy, Materials, and Telecommunications (EMT)
Institut national de la recherche scientifique (INRS)
Montreal, Quebec, Canada
Email: jean-charles.gregoire@emt.inrs.ca

***Abstract*—For network service providers, assessing and monitoring network parameters according to a Service Level Agreement (SLA) as well as optimal usage of resources are important. Packet loss is one of the main factors to be monitored especially when IP networks carry multimedia applications. Measuring network parameters will be more valuable when it is accurate and online. In this paper, we investigate a method to estimate packet loss probability (*plp*) under several conditions and improve the quality of the estimation over established techniques by introducing a new formula. In this method, the estimation of the *plp* in the intermediate nodes is based on the input stochastic traffic process. Different traffic situations and node buffer sizes are simulated by NS-2 and the accuracy of the method is investigated. The simulation results show that our new formula significantly improves the quality of the *plp* estimate.**

***Keywords*—Packet loss probability; estimation; stochastic traffic process.**

I. INTRODUCTION

In telecommunications, performance is assessed in terms of quality of service (QoS). QoS is measured either in terms of technology (e.g., for ATM, cell loss, variation, etc.) [1] or at some protocol level (e.g., packet loss, delay, jitter, etc.) [2].

Today, increased access to Internet networks as well as broadband networks have made possible and affordable the deployment of multimedia applications such as Internet telephony (VoIP), video conferencing, and IP television (IPTV) by academia, industry, and residential communities. Therefore the quality assessment of media communication systems and the parameters which affect this quality have been an important field of study for both academia and industry for decades. Due to the interactive or online nature of media communication and the existence of applicable solutions to deduce the effect of delay and jitter (e.g., deployment of a jitter buffer at the end user node [3], [4]), data loss is a key issue which should be considered. If there is a possibility for online accurate measuring of the amount of packet loss, then the network service providers can take the appropriate action to satisfy the contractual Service Level Agreement (SLA) or to improve and troubleshoot their service without receiving end user feedback.

Packet loss often happens because of congestion. In other words, buffer overflow at the outgoing interface in intermediate network nodes causes the packet loss. Since measuring the packet loss ratio at the intermediate nodes in high speed networks does not seem applicable in real time, some recent research has focused on estimation of packet loss probability (plp) [5]-[8].

According to central limit theory, the aggregated input traffic at intermediate nodes in the network core can be described with a gaussian model [9], [10]. Based on the Large Deviation Theory (LDT) and the large buffer asymptote approach, the plp can be estimated by a stochastic process. Since the input traffic is described by a gaussian process, the latter can be identified by online measure of the mean and variance of the input traffic. In this paper, the plp is estimated by the input traffic and the information which was measured in the past. In other words, we use some online and offline measuring data for accurate online estimation and improve on earlier results. The overall architecture of measurement, estimation, and control loop to keep the quality of service/experience within the SLA bounds is shown in Fig. 1.

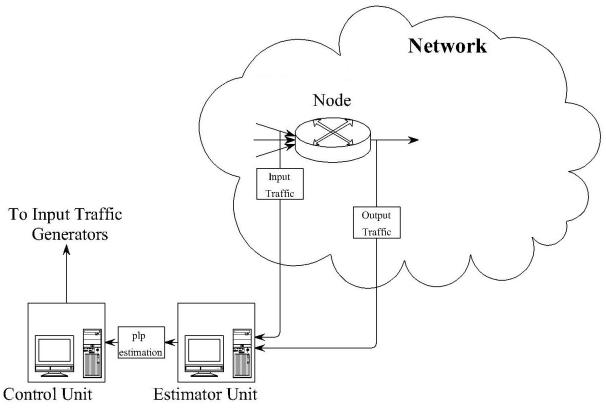


Fig. 1. Measurement, estimation, and control loop schematic.

The rest of the paper is organized as follows: Section II describes recent studies on plp estimators and introduces our improved estimator. Section III presents the testbed and our simulations. Numerical results and comparison that demonstrate the effectiveness of our new estimator are presented in Section IV. Section V concludes the paper and points to our future work.

 ISBN: 978-1-61208-133-5

II. Packet Loss Probability Estimator

There are several approaches to estimate packet loss probability. Sending probe packets periodically through the path and processing the returned signals for predicting the performance of path (e.g., packet loss ratio, delay, etc.) is one of the recent methods for estimating the plp [5], [11]. The disadvantage of this method is to increase the burden of probe packets' bit rate to the available bandwidth when greater accuracy is requested.

Estimation of plp based on stochastic input traffic process is another approach in this field [7], [8], [14]. In this method some important assumptions are taken as follows: 1) Measurement and estimation take place at intermediate nodes in high-speed core links of networks and therefore the input traffic is a mix of a large number of individual traffics and thus the gaussian process model is considered to represent the stochastic input traffic process [9], [10]; and 2) the size of the buffer should not be large, otherwise the queue process is not exponential and the behaviour of the traffic in large buffers cannot be approximated by a logarithmically linear behaviour [12], [13], so the input traffic process cannot estimate plp.

Following the gaussian model assumption for the input traffic, the effective bandwidth in this model [15] is given by:

$$eb(\theta, t) = \mu + \frac{\theta}{2t} VarZ(t) \tag{1}$$

and

$$VarZ(t) = \sigma^2 t^{2H} \tag{2}$$

where θ is the *space* parameter, t is the *time* parameter which corresponds to the most probable duration of the buffer congestion period prior to overflow, μ is defined as the *traffic mean*, *Z(t)* is the stochastic process gaussian distributed with a mean of zero, *Var* represents the second moment of *Z(t)*, σ^2 is the *variance* of the random variable, and *H* is the *Hurst* parameter.

Based on the classical assumption for input traffic, the H parameter is set to 0.5 [7]. So the effective bandwidth can be simplified into:

$$eb(\theta, t) = \mu + \frac{\theta}{2}\sigma^2 \tag{3}$$

In [16], Chang has proven that plp can be calculated by the following equation based on LDT:

$$ln(P_{loss}) = -\theta^* b - ln(\mu\theta^*) \tag{4}$$

where θ^* is the solution for

$$\lim_{t\to\infty} eb(\theta, t) = c \tag{5}$$

and c is the finite value (i.e., the bandwidth). By solving (3) and (5) and replacing θ^* in (4), P_{loss} can be obtained from:

$$ln(P_{loss}) = -\frac{2(c-\mu)}{\sigma^2} - ln\left(\frac{2\mu(c-\mu)}{\sigma^2}\right) \tag{6}$$

In line with other similar studies [7], [8], we change the base of the logarithm function from e to 10. Thus, (6) can be replaced by:

$$log(P_{loss}) = -\frac{2(c-\mu)}{\sigma^2} log(e) - log\left(\frac{2\mu(c-\mu)}{\sigma^2}\right) \tag{7}$$

Replacing μ and σ with their measurement value $\bar{\mu}(k)$ and $\bar{\sigma}(k)$ changes (7) to the following equation:

$$log(P_{loss}) = -\frac{2(c-\bar{\mu}(k))}{\bar{\sigma}^2(k)} log(e) - log\left(\frac{2\bar{\mu}(k)(c-\bar{\mu}(k))}{\bar{\sigma}^2(k)}\right) \tag{8}$$

where $\bar{\mu}(k)$ and $\bar{\sigma}(k)$ are defined as:

$$\bar{\mu}(k) = \frac{1}{N}\sum_{i=0}^{N-1} \bar{\alpha}(k-i) \tag{9}$$

and

$$\bar{\sigma}^2(k) = \frac{1}{N-1}\sum_{i=0}^{N-1} [\bar{\alpha}(k-i) - \bar{\mu}(k)]^2 \tag{10}$$

where $\bar{\alpha}(k)$ is the measured input packet rate in the kth time interval and N is the number of time intervals for calculating the average of the mean and variance of the packet rate.

In the rest of the paper let $epl(k)$ denote the $log(P_{loss})$, which is estimated by the formulas above, and $plp(k)$ denotes the logarithm of real packet loss probability during the time slot $[k, k+1)$ which can be expressed by:

$$plp(k) = log\left(\frac{l(k)}{\alpha(k)}\right) \tag{11}$$

where $l(k)$ is the number of lost packets during the time slot $[k, k+1)$ and $\alpha(k)$ is the number of packets that arrive during the time slot $[k, k+1)$.

Some estimation errors are expected due to the assumption made for the stochastic traffic process and the simplifications and approximations employed in (8). Zhang et al. introduce in [8] a Reactive Estimator (re) which is constructed as:

$$re(k) = epl(k) + \frac{1}{n}\sum_{l=1}^{n} [plp(k-l) - re(k-l)] \tag{12}$$

where $epl(k)$ and $plp(k)$ are calculated via (8) and (11), respectively. This estimator uses the measured $\{plp(k-l), l = 1, 2, ..., n\}$ data for reducing the error between re and plp.

A careful examination of (12) reveals that the error will be decreased to the amount of difference between re and plp, whereas the error is really the difference between epl and plp. We therefore present a new, improved estimator, cre, defined as:

$$cre(k) = epl(k) + \frac{1}{n}\sum_{l=1}^{n} [plp(k-l-m) - epl(k-l-m)] \tag{13}$$

where m is the number of interval periods after which the data of plp is available from (11).

With this new estimator, the required time for measuring and calculating the plp is represented by m in (13), where the mean of errors between epl and plp during a moving window (i.e., n time intervals) in the past (i.e., m time intervals ago) is added to epl to estimate the new plp. Therefore, the duration of time interval is independent from measurement and calculation speed of plp, whereas in former estimator (re) the minimum duration of time interval was equal or greater than the required time for measuring and calculating plp (it is assumed in (12) that the measured plp is available after one interval time). In other words, m in (13) makes the new estimator flexible about duration of measuring time interval.

To investigate the accuracy and applicability of the re and epl estimators and to compare their performance with that of our proposed estimator, cre, we propose to conduct simulations. In these simulations, the effects of different configurations of network traffic and packet loss ratio on estimators' performance are examined, which will be discussed in detail in Sections III and IV.

III. Simulation Testbed

The NS-2 software [17] is used to simulate the network. The network topology which is simulated is shown in Fig. 2.

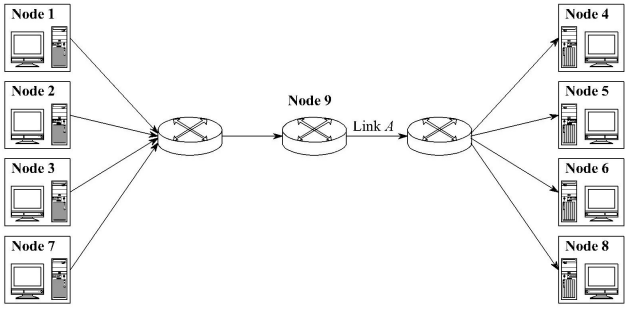


Fig. 2. Testbed topology.

An MPEG2 traffic flow is generated by node 1 and the RTP protocol is deployed for transferring video data to node 4. Node 2 generates the voice traffic flow which is coded by G.729. This data is transferred to node 5. Node 3 and node 6 are designed to generate the common Internet traffic flow for background traffic and make the aggregated traffic situation closer to the gaussian distributed traffic for stochastic input traffic process. The Tmix module in NS-2 is utilized in node 3 and 6 in order to generate realistic Internet network traffic [18]. The protocol deployed for communications between nodes 3 and 6 is TCP. Since the background traffic is TCP-based traffic, congestion (i.e., buffer overflow and loss) affects traffic flows, which leads to a situation similar to that of a real Internet network traffic. Nodes 7 and 8 generate the on-off traffic to randomly increase the packet loss probability. Measurement of the input and output traffics is performed at node 9. Since the focus is on node 9, the bandwidth of all links except link A is set to 100 Mbps and the buffer size of all nodes except node 9 is set to 500 packets. We vary the size of the buffer of node 9 from 5 packets to 100 packets to examine different router configurations. The bandwidth of link A, to generate different amounts of packet loss, varies between 8 Mbps to 10 Mbps. With these settings loss takes place only in node 9. When the bandwidth of link A is set to 10 Mbps and nodes 7 and 8 do not generate any traffic, the packet loss probability will be about 0.1 percent and when the bandwidth is decreased to 8 Mbps, the packet loss probability in node 9 increases to about 1 percent which is closer to the amount where effect of loss on media communication quality becomes considerable. By turning on the traffic of nodes 7 and 8 at some short periods of time, the packet loss probability reaches 7 percent which is an unacceptable amount of loss packet ratio for media communications. In the next section the numerical values of the different estimators in these situations will be examined.

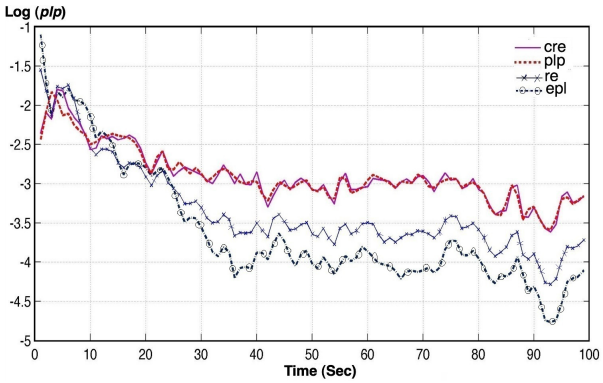


Fig. 3. Measurement and estimation of packet loss probability when plp is about -3.

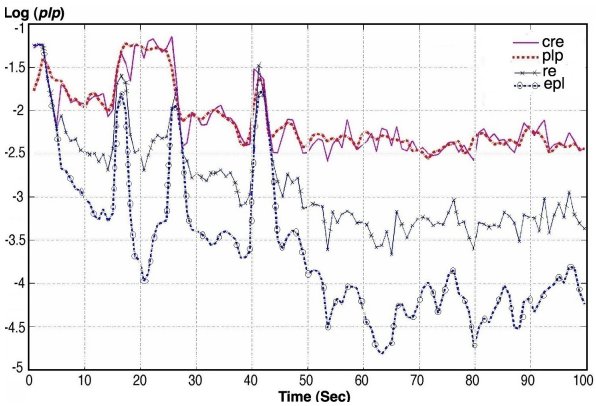
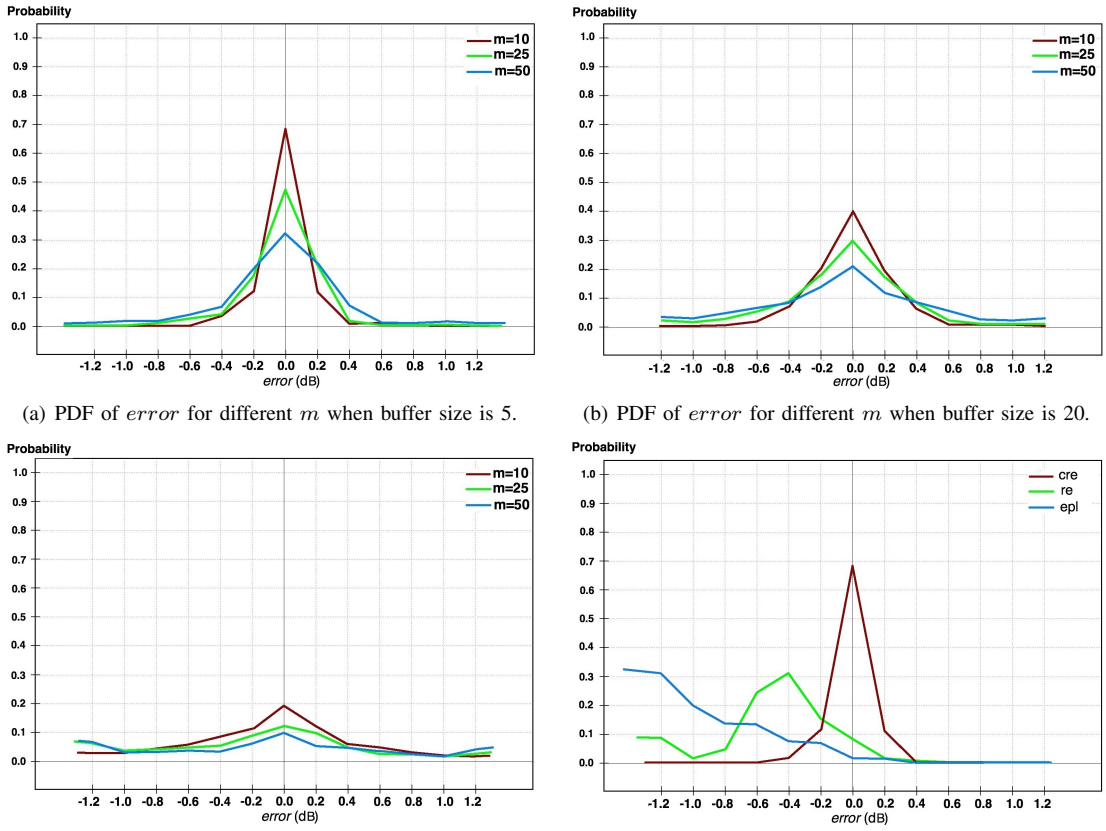


Fig. 4. Measurement and estimation of packet loss probability when plp is about -2.

IV. Numerical Results Analysis

First, all the mentioned estimators (i.e., epl, re, and cre) are evaluated in a situation where the bandwidth of link A is 10 Mbps and there is no traffic coming from nodes 7 and 8.

 ISBN: 978-1-61208-133-5



(a) PDF of $error$ for different m when buffer size is 5.

(b) PDF of $error$ for different m when buffer size is 20.

(c) PDF of $error$ for different m when buffer size is 100.

(d) PDF of $error$ for different estimators when buffer size is 5 and m=10.

Fig. 5. The comparison of PDF of $error$ for different conditions.

As shown in Fig. 3, the measured plp is around -3 (packet loss probability $\simeq 10^{-3}$). The accuracy of proposed estimator (cre) to estimate the plp compared to other estimators is demonstrated in this figure. With this amount of loss, although there is an offset between the plp, re, and epl, but re and epl follow the variations of plp and it can be seen as some soundness of the use of re and epl as the packet loss probability estimator but with a considerable error. In all experiences the time interval is 20 ms. In Fig. 3 cre is calculated according to (13) where m is 50. It means cre uses plp data measured one second before.

Since 10^{-3} can be negligible for loss packet ratio in media communication, we change the network conditions to increase the loss ratio and then re–evaluate the accuracy of estimators. To achieve this situation, the bandwidth of the link A is decreased to 8 Mbps. Fig. 4 shows the results of this experience: during the time periods of $[15, 25]$ and $[40, 41]$, nodes 7 and 8 add network traffic and bring the loss ratio close to 7 percent ($log(plp) = -1.5$). As Fig. 4 shows, the effect of simplification and approximation in (7) and (12) on the operation of epl and re methods for the bigger loss ratio is more apparent.

As mentioned before, buffer size affects the plp and the accuracy of estimators [12], [13]. The bigger the buffer size, the lesser plp and the accuracy of estimation. The effect of buffer size on estimation methods, re and epl, has been examined in [7] and [19] respectively. Beside the size of buffer, m, in (13), also affects the accuracy of cre estimation and it is determined by the speed of measuring and processing packet loss. We define the $error$ as the difference between estimated and measured plp, and study this parameter in different configurations to shows the effect of the network situation on estimation accuracy. Fig. 5 shows the probability density function of $error$ when buffer size is 5, 20, and 100 packet and m is 10, 25, and 50 ($m = 50$ means using a plp measured 1 s before), and the effect of the buffer size on estimation. Considering the effect of buffer size on estimation derived from (8), it appears that the accuracy of estimation (cre) will improve if the role of the measured plp is increased. Therefore, (13) is changed to:

$$cre(k) = p \times epl(k) + \frac{1}{n}\sum_{l=1}^{n}[plp(k-l-m) - p \times epl(k-l-m)] \quad (14)$$

where p is the proportional coefficient and is less than 1. To increase the importance of the second term in (13), n is increased from 3, which is recommended in [8], to 10 and to decrease the effect of first part, p is set to $\frac{2}{3}$. For an smaller

 ISBN: 978-1-61208-133-5

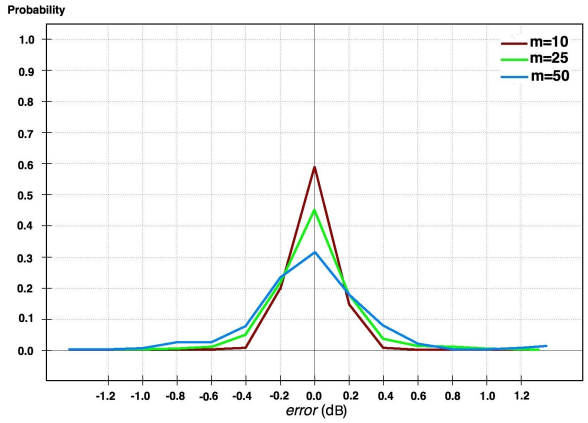


Fig. 6. PDF of $error$ for estimator which uses (14) when buffer size is 100.

p, when a considerable variation happens to plp (e.g., at 40 s in Fig. 4), the estimator (cre) cannot follow the plp properly and the $error$ will be significant.

Fig. 6 shows the $error$ when buffer size is 100 and (14) is used for estimation. Comparing Fig. 6 and Fig. 5c, the effectiveness of the changes in estimation is clear.

To conclude, the advantages of proposed estimator compared to other estimators are: 1) increasing the accuracy of estimation by using the measured parameters properly, 2) being flexible about duration of measuring time interval, and 3) estimating the plp reasonably accurately in case of large buffer.

V. Conclusion

One of the most important issues in multimedia quality of experience is packet loss, which has an especially critical role in interactive communications. Accurate online network-based measurements of loss are necessary to give Service Providers the means to estimate the quality received by a user and to give them an opportunity to take remedial action to satisfy the contractual SLA. Increased use of multimedia communications in the Internet has led to a renewed interest in the measure and estimation of loss, in the form of the plp, in modern communication networks. More specifically, recent studies have focused on estimation of the plp by measurement of input traffic based on LDT and the large buffer asymptote. In this paper, we have reviewed the theory behind plp estimation. By changing the way we use the measurement of output traffic of the node in which loss happens, we have introduced a new formula which significantly improves the quality of the estimate. To study the accuracy of the estimates, we have used the NS-2 simulator and real input traffic at the measurement node. Overall, the simulation results demonstrate the effect of different configurations, such as buffer size, on the estimates. The analysis of the results shows the improvement of accuracy in plp estimation achieved by our new calculation method.

For future research, we plan to investigate how it can be possible to estimate the end user's perception, aka the Quality of Experience (QoE). Along this line of research, we plan to study the methods of estimation of other network parameters (e.g., delay and jitter) to utilize them as the input of QoE measurement.

References

[1] D. McDysan, *QoS & Traffic Management in IP & ATM Networks*, McGraw-Hill, 2000.

[2] W. C. Hardy, *VoIP Service Quality: Measuring and Evaluation Packet-Switched Voice*, McGraw-Hill, 2003.

[3] B. Oklander and M. Sidi, "Jitter buffer analysis", Proc. of 17th IEEE International Conference on Computer Communications and Networks, pp. 1–6, August 2008.

[4] H. Hata, "Playout buffering algorithm using of random walk in VoIP", Proc. of IEEE International Symposium on Communications and Information Technology, pp. 457–460, October 2004.

[5] S. Tao and R. Guerin, "On-line estimation of internet path performance: an application perspective", Proc. of 23rd IEEE Conference on Computer Communications, Vol. 3, pp. 1774–1785, March 2004.

[6] R. Serral-Gracia, A. Cabellos-Aparicio, and J. Domingo-Pascual, "Packet loss estimation using distributed adaptive sampling", Proc. of IEEE Workshop on End-to-End Monitoring Techniques and Services, pp. 124–131, April 2008.

[7] D. Zhang and D. Ionescu, "A new method for measuring packet loss probability using a Kalman filter", IEEE Transaction on Instrumentation and Measurement, Vol. 58, No. 2, pp. 488–499, February 2009.

[8] D. Zhang and D. Ionescu, "Reactive estimation of packet loss probability for IP-based video services", IEEE Transaction on Broadcasting, Vol. 55, No. 2, pp. 375–385, June 2009.

[9] R. van de Meent and M. Mandjes, "Evaluation of user-oriented and black-box traffic models for link provisioning", Proc. of the 1st EuroNGI Conference on Next Generation Internet Networks Traffic, pp. 380–387, April 2005.

[10] J. Kilpi and I. Norros, "Testing the Gaussian approximation of aggregate traffic", Proc. of the 2nd ACM SIGCOMM Workshop on Internet Measurement, pp. 49–61, November 2002.

[11] S. Tao, K. Xu, A. Estepa, T.F.L. Gao, R. Guerin, J. Kurose, D. Towsley, and Z.L. Zhang, "Improving VoIP quality through path switching", Proc. of 24th IEEE Conference on Computer Communications, Vol. 4, pp. 2268–2278, March 2005.

[12] D. P. Heyman and T. V. Lakshman, "What are the implications of long-range dependence for VBR-video traffic engineering?", IEEE/ACM Transactions on Networking, Vol. 4, No. 3, pp. 301–317, June 1996.

[13] B. K. Ryu and A. Elwalid, "The importance of long-range dependence of VBR video traffic in ATM traffic engineering: myths and realities", Proc. of ACM SIGCOMM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, pp. 3–14, August 1996.

[14] C. Lambiri, D. Ionescu, and V. Groza, "A new method for the estimation and measurement of traffic packet loss for virtual private networks services", Proc. of the 21st IEEE Conference on Instrumentation and Measurement Technology, Vol. 1, pp. 401–406, May 2004.

[15] F. Kelly, *Notes on effective bandwidths*, in Stochastic Networks: Theory and Applications, Oxford University Press, pp. 141–168, 1996.

[16] C. Chang, *Performance Guarantees In Communication Networks*, New York: Springer-Verlag, 2000.

[17] UC Berkeley, LBL, USC/ISI, and Xerox PARC, *Network Simulator NS-2*, http://www.isi.edu/nsnam/ns, [Accessed: 6 May 2011].

[18] M.C. Weigle, P. Adurthi, F. Hernandez-Campos, K. Jeffay, and F.D. Smith, "Tmix: A tool for generating realistic application workloads in NS-2", ACM SIGCOMM Computer Communication Review, Vol 36, No 3, pp. 67–76, July 2006.

[19] D. Zhang and D. Ionescu, "On packet loss estimation for virtual private networks services", Proc. of 13th IEEE Conference on Computer Communications and Networks, pp. 175–180, October 2004.

 ISBN: 978-1-61208-133-5

Recent Trends in TCP Packet-Level Characteristics

Per Hurtig
Department of Computer Science
Karlstad University, Karlstad
Email: per.hurtig@kau.se

Wolfgang John
Department of Computer Science and Engineering
Chalmers University of Technology, Göteborg
Email: wolfgang.john@gmail.com

Anna Brunstrom
Department of Computer Science
Karlstad University, Karlstad
Email: anna.brunstrom@kau.se

***Abstract*—Up-to-date TCP traffic characteristics are essential for research and development of protocols and applications. This paper presents recent trends observed in 70 measurements on backbone links from 2006 and 2009. First, we provide general characteristics such as packet size distributions and TCP option usage. We confirm previous observations such as the dominance of TCP as transport and higher utilization of TCP options. Next, we look at out-of-sequence (OOS) TCP segments. OOS segments often have negative effects on TCP performance, and therefore require special consideration. While the total fraction of OOS segments is stable in our measurements, we observe a significant decrease in OOS due to packet reordering (from 22.5% to 5.2% of all OOS segments). We verify that this development is a general trend in our measurements and not caused by single hosts/networks or special temporal events. Our findings are surprising as many researchers previously have speculated in an increased amount of reordering.**

***Keywords*-traffic measurement; TCP; reordering;**

I. INTRODUCTION

The flexibility and versatility of the Internet architecture allows protocols and applications to be developed and deployed quickly. These properties have thus enabled a rapid evolution of the Internet. To support further development of the Internet it is important to investigate and highlight trends in its evolution.

In this paper, we follow up on our previous observations on backbone traffic [1] by comparing general packet characteristics between 35 traces from 2006 and 35 novel traces from 2009. We have complemented the basic packet-level characteristics with an analysis of out-of-sequence (OOS) TCP segments. A TCP segment is said to be OOS when it arrives at a receiver that is expecting another segment. Segments can arrive OOS for different reasons, including retransmissions, network duplication and packet reordering. Although TCP is designed to deal with OOS segments, the performance might suffer. For instance, packet reordering is a problem as it causes receivers to emit duplicate acknowledgments (dupACKs) back to the sender. A high degree of reordering can therefore cause TCP to falsely assume packet loss, leading to unnecessary retransmissions and invocation of the congestion control, which substantially lowers the throughput. Interestingly, a number of novel networking technologies that now are being deployed use mechanisms that create reordering as a side effect. For instance, mobile ad hoc networks (MANETs) often use routing mechanisms that create reordering in their mode of operation. It is therefore important to track the development of reordering over the last years.

Studies on packet-level characteristics have been published before, especially during the early 2000's. Figures about packet size distribution and transport protocol decomposition of Internet traces from different wide-area measurement locations (OC3-OC192) have been presented repeatedly since 1997 [2]–[5]. Also usage of TCP options has been presented on data collected until 2000 [6] and 2004 [5]. However, since our summary about packet characteristics of backbone data from 2006 [1] there have been no publications with complete packet-level details of wide-area Internet traffic. Nevertheless, there have been studies specialized towards certain aspects of TCP. For instance, Maier et al. [7] present transport protocol features such as TCP option deployment and configuration on Internet traffic data from DSL connections of a large European ISP in 2008 and 2009. Qian et al. [8] compared TCP flow sizes and also tried to infer the evolution of a number of TCP implementation details in traces from 2001 and 2008. Also measurements of TCP out-of-sequence (OOS) segments have been conducted in several studies between 1999 and 2008 [9]–[13]. To our knowledge, however, there have not been any studies that have collected data from the same location, at different points in time, to detect possible trends. We believe that it is important to keep the research community updated with this type of basic information, which enables accurate and realistic simulation models to support refinement and development of network protocols and devices.

In Section II the data collection and processing is described. Section III presents and compares general IP packet characteristics, such as packet size distributions and TCP option deployments. Section IV compares different TCP OOS deliveries, and details the trends of OOS due to packet reordering. The frequency of TCP OOS segments appears to be quite stable between the measurements, affecting about 17% of all TCP connections. However, we were surprised to see that OOS segments due to packet reordering have decreased significantly in 2009. Finally, Section V provides

 ISBN: 978-1-61208-133-5

a number of concluding remarks.

II. Data Collection and Processing

The two data sets compared in this paper were collected in the time from April to November 2006 and January to June 2009 (see Table I) on backbone links in the Swedish University Network (SUNET). Altogether, 70 traces were collected, each trace consisting of ten minutes of bi-directional traffic. The measurement times of the 35 traces collected in 2006 where spread out over a period of eight months, and the 2009 collection times over a period of six months. We collected the traffic on OC192 (10Gbit/s) links on two different generations of SUNET. The 2006 data was collected in GigaSUNET, a ring architecture with a central Internet exchange point in Stockholm. The measurement location was on a OC192 ring which was the primary link from the region of Gothenburg to the main Internet outside Sweden. The link mainly carried traffic from major universities and large student residential networks, but also from a regional access point exchanging SUNET traffic with local ISPs.

The 2009 data was collected in the upgraded SUNET (OptoSUNET), a star structure over leased fiber. All customers are redundantly connected to a central Internet access point. Besides some local exchange traffic, the traffic routed to the main Internet outside Sweden is carried on two links (40Gb/s and 10Gb/s) between SUNET and NorduNet. The data used in this study was collected on the 10Gb/s link, which according to SNMP statistics carried 50% of all inbound but only 15% of the outbound traffic volume.

We collected data by using optical splitters attached to two Endance DAG6.2-SE cards (i.e., one measurement card for each direction). We configured the DAG cards to capture the first 120 bytes of each frame to ensure that the network and transport headers were preserved. After recording the traces, the IP-addresses were anonymized using the prefix preserving CryptoPAN software [14] and the remaining payload, beyond the transport layer, was stripped to ensure privacy and confidentiality. During data collection, the DAG cards were synchronized with each other using Endace's DUCK time synchronization [15], allowing us to merge the data into well-aligned bi-directional packet-header traces. Further details on the data collection and pre-processing can be found in [16].

To process and analyze general traffic properties we used available tools like CAIDA's CoralReef [17], as well as own specialized tools for additional sanity checking and result processing. For packets of special interest, the corresponding TCP flows were extracted and manually analyzed. For detection and classification of OOS TCP segments we used Tstat 2.0 [18]. Tstat's OOS detection and classification method is described in [13], and is a refinement of the methodology used by Jaiswal et al. [19]. We further describe this method in Section IV.

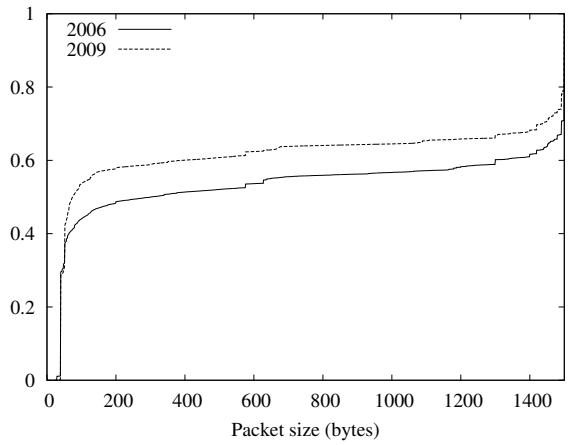


Figure 1. CDF for IP packet sizes for the 2006 and the 2009 measurements.

III. General Results

The 70 packet traces consist mainly of IPv4 packets (99.98% and 99.99% of all frames observed in 2006 and 2009, respectively). The remainder of the traffic is mainly IPv6 routing traffic (BGP). In the rest of this paper, only IPv4 traffic is considered. Note that the 2006 data is a subset of the data set analyzed in [1].

A. IP Traffic Characteristics

1) IP Packet Size Distribution: Earlier Internet measurements, conducted between 1997 and 2002 [2]–[4], [20], reported of cumulative IPv4 packet distributions being trimodal, with major modes at about 40 bytes (TCP acknowledgments), 576 bytes (the default datagram size [21]) and 1500 bytes (the Ethernet MTU). Default datagram sizes represented about $10 - 40\%$ of all packets. However, later measurements have reported of a much smaller fraction of default datagram sizes (e.g., 3.8% in 2004 [5]).

Figure 1 shows the cumulative distribution function (CDF) of packet sizes in our measurements. The packet size distributions are bimodal with major modes at around 40 bytes and 1500 bytes. The percentage of packets having the default datagram size is about 1% in both measurements, not even being among the three largest modes anymore. As we already reported earlier [1], this can be explained by the common use of PathMTU Discovery. We can also see that the fraction of small packets has increased significantly.

2) Transport Protocols: Table II shows the fractions of packets/bytes carried by each protocol compared to total IPv4 traffic. The figures confirm the domination of TCP as transport protocol, but also indicates an increasing trend in UDP traffic. The percentage of UDP packets has increased from 8.2% to 16.27%, and the amount of bytes from 3.4% to 8.53%. This is in line with other measurements that also have reported increased UDP traffic, especially in terms of flow numbers [22], [23]. The reason for this has been

 ISBN: 978-1-61208-133-5

Table I
SUMMARY OF THE DATA SETS.

Dataset	Collection Period	#Traces	Trace Dur.	Total Volume	Total #Packets
GigaSUNET	Apr.-Nov. 2006	35	10 min	2.3 TB	3.3×10^9
OptoSUNET	Jan.-Jun. 2009	35	10 min	4.6 TB	7.9×10^9

Table II
PROTOCOLS AT TRANSPORT LEVEL (IN %).

	GigaSUNET 2006		OptoSUNET 2009	
	Pkts	Data	Pkts	Data
TCP	91.50	96.50	82.90	90.40
UDP	8.20	3.40	16.27	8.53
ICMP	0.17	0.02	0.23	0.04
ESP	0.13	0.06	0.47	0.93

reported to be an increase in P2P signaling traffic, generating large numbers of small flows over UDP.

Also other protocols have become more prevalent. Especially the use of Encapsulating Security Payload (ESP) [24], used to enhance security and confidentiality of data transfers, has increased from 0.06% to 0.93%. Although ESP is only responsible for about 1% of the data, the increase is substantial. We speculate that this could be caused by a rising popularity of IPsec tunnels as a reaction to the new IPRED law in Sweden[1].

B. TCP Characteristics

In order to analyze TCP characteristics, we aggregated packets into bi-directional TCP flows. In this paper, a TCP flow is the same as a TCP connection. Thus, a flow is identified by a connection initiation (3-way handshake) and a termination (by FIN/RST signaling). Flows without an observed, complete handshake have been excluded from our analysis. This strict flow definition was used to allow more accurate results.

1) Flow Lengths: Even if our measurements do not contain flows longer than 10 minutes we investigated TCP flow size distributions. When considering TCP flows, the classical assumption is that TCP traffic is heavy-tailed, i.e., it consists of a large number of small flows (mice) and a small number of large flows (elephants) [26], [27]. The TCP flow size distributions for our measurements are plotted in Figure 2. The graph shows the CDF of TCP flow sizes in bytes.

Consistent with the classical assumption, the distribution of flow sizes appear to be heavy-tailed. About half of all flow sizes are around 1000 bytes only, but very large flows are not negligible and are responsible for a large fraction of the total traffic volume. On average, it appears as flow lengths have increased slightly from 2006 to 2009, but no significant differences are apparent. This confirms recent results by Qian et al. [8], reporting about no qualitative differences in TCP flow sizes when comparing AT&T backbone data from 2001 and 2008.

[1]On April 1, 2009, an anti-piracy law based on the European directive on the enforcement of intellectual property rights (IPRED) [25] came into effect in Sweden.

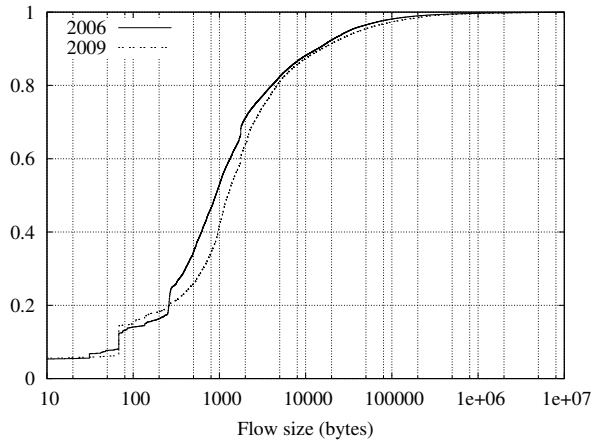


Figure 2. CDF for TCP flow lengths (in bytes).

2) TCP Options: Earlier measurements have shown a rather significant deployment of TCP options, such as Selective Acknowledgments (SACKs), Window Scaling (WS), Timestamps (TS), and Maximum Segment Size (MSS). Allman [6], for instance, reported that about 20% of all hosts allowed the WS and TS options. SACK was shown to be more commonly deployed, about 40%. In a recent study by Maier et al. [7], WS was reported to be advertised by at least one endpoint in $32-35\%$ of connections carrying data, and effectively used by $28-31\%$, TS was advertised in $11-12\%$ and used in 8%, and SACK was advertised in 97% and used in 82% of the connections.

We have previously shown that the use of the WS, TS, SACK, and MSS options is rather widespread [1], [28]. The MSS option, for instance, was advertised in about 99% of all the TCP SYN segments in [1]. In this paper, however, we only focus on established TCP connections. Thus, we only deal with connections that have had a proper SYN-SYN/ACK exchange. Table III shows the percentage of TCP option advertisement/usage for the connections that were established during our measurements. The advertised column shows if at least one party of the connection tried to use the option, while the used column shows if the corresponding option was actually used in the connection.

As indicated by the table, TCP options are used to a

 ISBN: 978-1-61208-133-5

Table III
TCP OPTION USAGE IN SUCCESSFUL THREE-WAY HANDSHAKES (IN %).

TCP	GigaSUNET 2006		OptoSUNET 2009	
Option	Advertised	Used	Advertised	Used
MSS	99.99	99.40	99.99	99.21
WS	21.20	18.60	44.22	37.64
SACK	96.34	80.36	98.43	86.83
TS	17.27	14.25	23.93	19.63

significant extent, and are even more commonly employed than in the DSL connections of the large European ISP studied in 2009 [7]. In our data, the MSS option is used by nearly all connections ($\approx 99\%$) in both 2006 and 2009. The use of WS have almost doubled, from about 19% to 38%. The use of SACK and TS have also increased, from 80% to 87% and from 14% to 20%, respectively.

IV. TCP OOS DELIVERIES

This section presents and compares OOS segments found in the measurements. As mentioned in the introduction, an OOS segment is a segment that arrives unexpectedly at the TCP receiver. In our definition of OOS we also include segments that already have been received, i.e., duplicated segment arrival. We start by describing the methodology used for identifying and classifying OOS segments. We then give an overview of OOS segments observed in our measurements. Finally, we provide an extended analysis of reordered TCP segments.

A. Methodology

The methodology used was originally developed in [19] and later extended in [13]. The methodology both identifies OOS segments and classifies them. For example, if a segment is lost and is retransmitted using a retransmission timeout the methodology will identify retransmission timeout as the underlying cause.

The basic detection of an OOS segment is straightforward; given a bi-directional traffic trace the sequence and acknowledgment numbers can be used to infer if segments are arriving in-order or OOS. To further classify OOS segments is a more complicated task. We chose to use the Tstat tool [18] which implements the methodology in [13]. A short description of the classification follows below, while the exact details of the classification process can be found in [13].

If the observed segment has both the same IP identifier and the same sequence number as an already observed segment it is due to network duplication (NetDup). If the observed segment is unacknowledged and the loss recovery of the sender is likely to have triggered the transmission of the corresponding segment it might be due to a retransmission. Triggering of the loss recovery mechanisms are assumed if the estimated loss recovery time is greater than the expected RTO, or if three duplicate acknowledgments have been observed (Fast Retransmit (FR)). If the loss recovery of the sender is likely to have triggered a retransmission of the segment although both the segment and its acknowledgment have been observed already the retransmission is unnecessary (Un.RTO, Un.FR). If the segment has been observed and acknowledged previously, and TCP window probing is conducted, it is classified as flow control (FC). If the segment has not been observed yet, and it is unlikely that the sender has invoked any of its retransmission mechanisms, it is reordered (Reo).

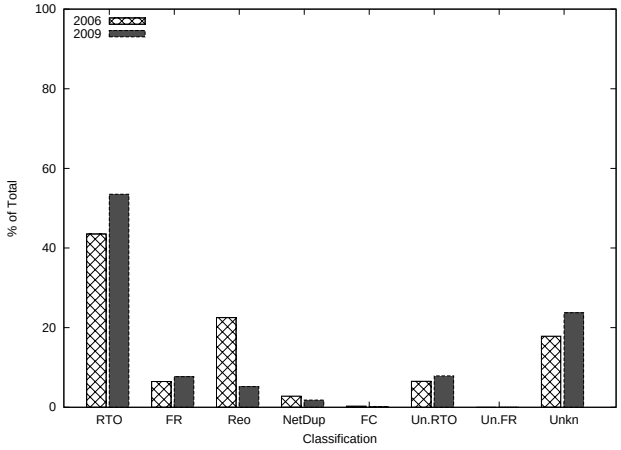


Figure 3. OOS classification of TCP segments in 2006 and 2009. The significant decrease in reordered segments, from 2006 to 2009, is the main difference between the data sets.

In addition to these categories, Tstat may also classify segments as "unknown" (Unkn). This classification is used whenever Tstat is unable to infer the exact cause of an OOS segment. As Tstat's OOS segment classifier uses IETF's TCP standards when calculating e.g., the RTO of a connection, segments can be OOS for no obvious reason. This happens as there are variations between different TCP implementations, both in logic and in configuration.

B. OOS Overview

Figure 3 shows the classification of all OOS segments in our measurements. In total, 1.6% of all segments were OOS, for both the 2006 and the 2009 data. This is about the same as, or slightly lower than, figures reported in related measurements: in [12] $0.9\% - 7.1\%$ of all packets in seven different traces were OOS, and in [13] an average of $5\% - 8\%$ were reported. The differences between our and related measurements can be of many reasons. For instance, [12] only considered connections with at least 10 segments and [13] reported large variations between different measurement points.

Although the amount of OOS seems to be stable between our measurements, the OOS distributions vary. Three distinctive differences can be observed. First, the amount of OOS segments due to RTOs are more common in 2009 (53.5%)

 ISBN: 978-1-61208-133-5

Table IV
TRAFFIC VOLUME AND OOS BREAKDOWN FOR 2006 TRAFFIC, ALL FIELDS IN % OF TOTAL.

Length	Pkts	Flows	OOS	RTO	FR	Reo	Dups	FC	Un.RTO	Un.FR	Unkn
SHORT	5.78	64.55	21.90	25.12	1.42	18.02	87.52	13.73	38.37	0.17	10.22
MEDIUM	10.48	27.46	17.94	24.75	2.25	11.19	7.18	1.88	37.68	0.67	10.28
LONG	83.74	8.00	60.16	50.13	96.33	70.80	5.30	84.39	23.94	99.17	79.51
BREAKDOWN				43.50	6.48	22.52	2.79	0.30	6.54	0.01	17.85

Table V
TRAFFIC VOLUME AND OOS BREAKDOWN FOR 2009 TRAFFIC, ALL FIELDS IN % OF TOTAL.

Length	Pkts	Flows	OOS	RTO	FR	Reo	Dups	FC	Un.RTO	Un.FR	Unkn
SHORT	5.29	60.10	17.86	20.43	0.98	17.21	64.83	17.57	35.15	0.00	8.37
MEDIUM	10.64	30.64	26.57	33.25	3.08	12.28	21.15	1.95	44.18	5.38	17.00
LONG	84.07	9.25	55.56	46.32	95.94	70.50	14.02	80.47	20.67	94.62	74.63
BREAKDOWN				53.51	7.68	5.19	1.81	0.18	7.89	0.00	23.74

than in 2006 (43.5%). Second, segments classified as OOS due to unknown reasons have increased from 2006 (17.9%) to 2009 (23.7%). The unknown classification is for instance used when a packet seems to be retransmitted but the number of dupACKs are less than three or the estimated RTO has not yet expired. The increase of OOS segments in this category might be related to the increase in TCP implementations that use more aggressive loss recovery mechanisms than the ones standardized in [29], [30]. Linux, for example, uses a minimum allowed RTO of 200 ms, while the standard is 1 s. Furthermore, Windows XP allows segments to be retransmitted after only 2 dupACKs. Finally, the amount of OOS due to packet reordering dropped significantly between the 2006 (22.5%) and 2009 (5.2%) measurements. We will analyze this more thoroughly in Section IV-D.

Although the measured backbone traffic is highly aggregated, it might be misleading to draw conclusions based solely on packet-level observations. A few flows could skew the statistics, such as long-lived high-volume elephant flows where every other packet is OOS. Therefore, OOS segments were also classified on a per-flow basis (see Figure 4). For the measurements in 2006, 17.3% of all flows had at least one OOS segment, and in 2009 16.4% of the flows had at least one OOS segment. The OOS distribution on flow-level is also about the same as on packet-level. Thus, it is not likely that the observed trends are due to a small number of misbehaving flows.

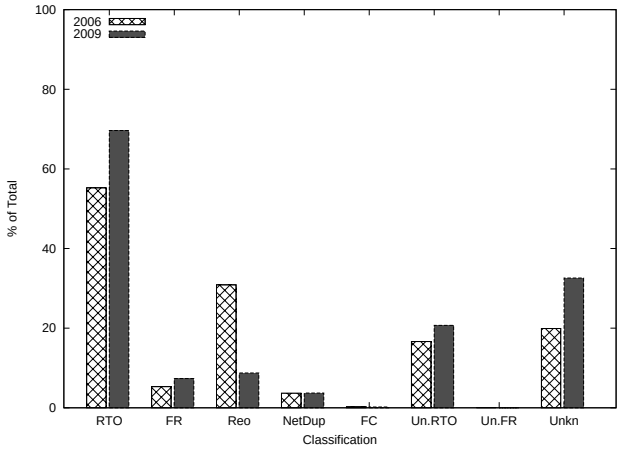


Figure 4. OOS classification of TCP flows in 2006 and 2009. The significant decrease in reordered segments, from 2006 to 2009, is the main difference between the data sets.

C. OOS Details

Tables IV and V show the 2006 and 2009 OOS distributions for different flow size classes. We present the results in a similar way as Mellia et al. [13]. The figures in the tables correspond to the ratio between the number of specific OOS events occurring in a given flow class over the total number of such OOS events. For instance, 17.21% of all reordered segments in the 2009 measurements (Table V) occurred in short flows.

Three different flow size classes are used in the tables. Short flows are flows including not more than five data segments (1-5). Medium sized flows have a minimum of six data segments and a maximum of 20 data segments (6-20). Long flows have a payload that is larger than 20 data segments (>20). The bottom lines of the two tables show the total occurrence of the different OOS classes. Thus, the bottom lines convey the same information as Figure 3.

The flow size distributions are quite similar in 2006 and 2009, moving slightly towards longer flows in 2009. About 60% of all flows are short (containing $5-6\%$ of all packets), $27-30\%$ of the flows are medium sized (containing 10% of all packets) and $8-9\%$ of the flows are long (containing $83-84\%$ of all packets). The amount of OOS segments in the different flow size classes have also shifted slightly, making medium sized flows subjected to more OOS in 2009 (from 17.94% to 26.57%), while the OOS in short and long flows has decreased somewhat. This is also visible when inspecting the different OOS categories where the medium sized flows now have a larger portion of almost every OOS

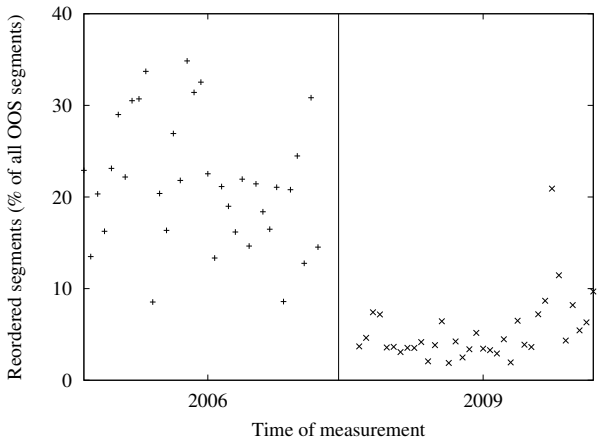


Figure 5. Reordered segments as a fraction of all OOS segments, per measurement.

category. In general, it is evident that the relatively small amount of packets belonging to short flows account for a large portion of all OOS segments ($17 - 22\%$). The large flows, which represent a vast majority of all the traffic do only account for $55-60\%$ of the OOS segments. Thus, short flows seem to be more subjected to OOS than medium size and long flows.

When comparing 2006 to 2009, the most interesting difference is the total distribution of OOS segments (the bottom line in the tables). For instance, the portion of RTOs and unknown events has increased while reordering has decreased significantly (from 22.52% to 5.19%). While the portion of reordering in 2006 are comparable with previous studies, e.g., Mellia et al. [13] found reordering to be responsible for 28.12% of all OOS segments in traces from 2004, the 2009 results displays a surprisingly low amount of reordering. It is very hard to find specific reasons to why these OOS categories have changed so significantly between 2006 and 2009. The increase in RTOs and unknown events might be a consequence of more recent, and aggressive, retransmissions schemes; RTO timers that expire before fast retransmit can be invoked; fast retransmit algorithms that requires less than three dupACKs. For the decrease in reordering, it is even harder to speculate about causes.

D. Packet Reordering

Packet reordering can occur for a number of reasons including e.g., multi-path routing and parallelism within routers [9], [10]. As mentioned in the introduction, novel networking technologies that now are being deployed use some of these techniques and are thus believed to create packet reordering in their mode of operation [31]. To mitigate the negative effects of reordering, a number of reordering robust TCP versions have been developed during the last years (e.g., [31]–[33]). Most of these proposals do, however, not inhibit the actual reordering but merely the negative effects reordering poses on TCP performance. It would therefore be intuitive that reordering might have increased the last years. Our measurements, on the other hand, indicate a significant *decrease* in packet reordering. It is, however, important to remember that novel networking technologies that might lead to increased reordering often are employed in specialized networks, and that such deployments do not affect backbone traffic that much.

Figure 5 shows the fraction of reordered segments (in $\%$ of all OOS segments) for our 70 collected traces in 2006 and 2009. As shown in the graph, the amount of reordering, and also the variance between the traces, is much smaller in the 2009 traces. For the 2006 traces the fraction of reordered segments goes from approximately 10% to 35%. The 2009 traces display a rather stable amount of reordering around $1 - 8\%$, except for a few outliers.

To rule out that reordering events are induced by a few specific networks (or a few specific routers), we looked at the amount of reordering events per /16 and /24 network prefix[2] (class B and class C networks). In 2006, the average number of reordering events per /16 network was 177, with an 95% confidence interval of $[139, 215]$. In 2009, the corresponding figure was only 45 with an 95% confidence interval of $[32, 57]$. In 2006, the average number of reordering events in /24 networks was 45, with a 95% confidence interval of $[33, 58]$. In 2009, the average number of reordering events per /24 network was only 15, with a corresponding confidence interval of $[12, 19]$. The mean values together with the confidence intervals lead us to the conclusion that the significant decrease in reordering events, between 2006 and 2009, was not caused by a few misbehaving routers/networks.

To further rule out temporal events as the cause for the decrease in packet reordering, we investigated the temporal distributions. Figure 6 shows OOS events during one measurement in 2009. The y-axis shows the number of OOS segments during each millisecond of the measurement (x-axis). The white bottom part of the graph shows the fraction of reordered segments. The frequency of OOS (and reordering) events does not follow any particular mode, but rather appear as noise. Although the graph only shows one trace, this type of distribution of OOS events is representative for all traces. Since we are not able to attribute the decrease in packet reordering to any specific network event, we speculate that modern networking equipment has been more carefully designed not to introduce packet reordering, e.g., by taking routing decisions on flow or IP-pair level rather than on individual packet level.

V. Conclusions

To reveal recent trends in TCP packet-level characteristics we have measured and compared highly aggregated back-

[2]Note that the applied prefix preserving IP address anonymization allows this type of analysis.

 ISBN: 978-1-61208-133-5

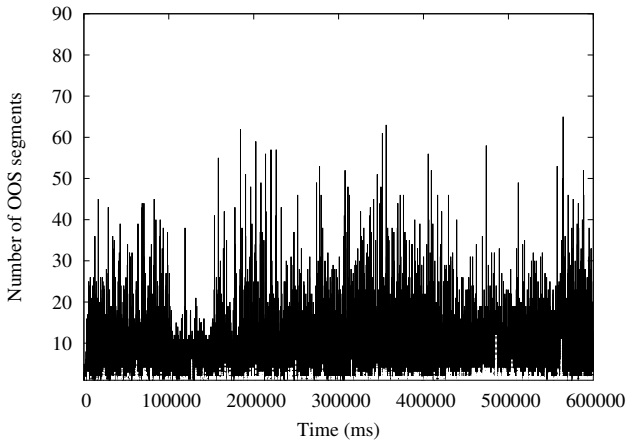


Figure 6. OOS segments during one of the 10 minute measurements in 2009.

bone traffic from 35 traffic traces collected in 2006 with 35 corresponding traffic traces collected in 2009. The analysis shows that although TCP is still the dominating transport protocol, the use of UDP has increased significantly from 2006 to 2009. Furthermore, the use of TCP options like WS, SACK and TS have also continued to increase over these years. The most frequently used TCP option is the MSS option, which is used in over 99% of all TCP connections.

We have also looked at TCP OOS deliveries and found that although the relative amount of OOS deliveries is stable, OOS caused by packet reordering has decreased significantly from 2006 to 2009. The change does not seem to be due to a few misbehaving hosts/routers or due to any major temporal event, but rather a general development. This is an interesting result, as many researchers have speculated in the increase of packet reordering due to novel networking technologies that create packet reordering as a side effect. These novel technologies are, however, mostly deployed in specialized networks and maybe therefore not prone to affect Internet backbones significantly. It is, however, important to continue considering the levels of packet reordering in backbones, as specialized networks and their mechanisms will be incorporated into the regular Internet.

ACKNOWLEDGEMENTS

This work was supported by SUNET, the Swedish University Computer Network.

REFERENCES

[1] W. John and S. Tafvelin, "Analysis of internet backbone traffic and header anomalies observed," in *IMC '07: Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurements*, San Diego, CA, USA, October 2007, pp. 111–116.

[2] K. Thompson, G. Miller, and R. Wilder, "Wide-area internet traffic patterns and characteristics," *IEEE Network*, vol. 11, no. 6, pp. 10–23, 1997.

[3] S. McCreary and K. Claffy, "Trends in wide area IP traffic patterns - a view from ames internet exchange," CAIDA, SDSC, UCSD, Technical Report, 2000.

[4] C. Fraleigh, S. Moon, B. Lyles, C. Cotton, M. Khan, D. Moll, R. Rockell, T. Seely, and C. Diot, "Packet-level traffic measurements from the sprint IP backbone," *IEEE Network*, vol. 17, no. 6, pp. 6–16, 2003.

[5] K. Pentikousis and H. Badr, "Quantifying the deployment of TCP options - a comparative study," *IEEE Communications Letters*, vol. 8, no. 10, pp. 647–649, October 2004.

[6] M. Allman, "A web server's view of the transport layer," *SIGCOMM Comput. Commun. Rev.*, vol. 30, no. 5, pp. 10–20, 2000.

[7] G. Maier, A. Feldmann, V. Paxson, and M. Allman, "On dominant characteristics of residential broadband internet traffic," in *IMC'09: Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurements*, 2009.

[8] F. Qian, A. Gerber, Z. M. Mao, S. Sen, O. Spatscheck, and W. Willinger, "TCP revisited: a fresh look at TCP in the wild," in *IMC'09: Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurements*, 2009.

[9] V. Paxson, "End-to-end internet packet dynamics," *IEEE/ACM Trans. Netw.*, vol. 7, no. 3, pp. 277–292, 1999.

[10] J. C. R. Bennett, C. Partridge, and N. Shectman, "Packet reordering is not pathological network behavior," *IEEE/ACM Trans. Netw.*, vol. 7, no. 6, pp. 789–798, 1999.

[11] L. Gharai, C. Perkins, and T. Lehman, "Packet reordering, high speed networks and transport protocol performance," in *13th International Conference on Computer Communications and Networks (ICCCN'04)*, Chicago, IL, USA, October 2004.

[12] S. Rewaskar, J. Kaur, and D. F. Smith, "A passive state-machine approach for accurate analysis of TCP out-of-sequence segments," *SIGCOMM Comput. Commun. Rev.*, vol. 36, no. 3, pp. 51–64, 2006.

[13] M. Mellia, M. Meo, L. Muscariello, and D. Rossi, "Passive analysis of TCP anomalies," *Computer Networks*, vol. 52, no. 14, pp. 2663–2676, 2008.

[14] J. Xu, J. Fan, M. Ammar, and S. B. Moon, "On the design and performance of prefix-preserving IP traffic trace anonymization," in *IMW '01: 1st ACM SIGCOMM Workshop on Internet Measurement*, San Francisco, CA, USA, 2001, pp. 263–266.

[15] S. Donnelly, "Endace DAG Timestamping Whitepaper," 2007, Endace Technology Limited.

[16] W. John, S. Tafvelin, and T. Olovsson, "Passive internet measurement: Overview and guidelines based on experiences," *Computer Communications*, vol. 33, no. 5, pp. 533–550, 2010.

[17] K. Keys, D. Moore, R. Koga, E. Lagache, M. Tesch, and kc claffy, "The Architecture of CoralReef: an Internet Traffic Monitoring Software Suite," in *A workshop on Passive and Active Measurements*, 2001.

 ISBN: 978-1-61208-133-5

[18] Politecnico di Torino, http://tstat.tlc.polito.it/index.shtml, October 2008.

[19] S. Jaiswal, G. Iannaccone, C. Diot, J. Kurose, and D. Towsley, "Measurement and classification of out-of-sequence packets in a tier-1 IP backbone," *IEEE/ACM Trans. Netw.*, vol. 15, no. 1, pp. 54–66, 2007.

[20] C. Shannon, D. Moore, and K. Claffy, "Beyond folklore: Observations on fragmented traffic," *IEEE/ACM Trans. Netw.*, vol. 10, no. 6, pp. 709–720, 2002.

[21] J. Postel, "The TCP maximum segment size and related topics," *Internet RFCs, ISSN 2070-1721*, vol. RFC 879, November 1983.

[22] M. Zhang, M. Dusi, W. John, and C. Chen, "Analysis of UDP Traffic Usage on Internet Backbone Links," in *Ninth Annual International Symposium on Applications and the Internet (SAINT)*, Seattle, USA, 2009, pp. 280–281.

[23] W. John, "Characterization and Classification of Internet Backbone Traffic," Department of Computer Science and Engineering, Chalmers University of Technology, Göteborg, SE, Tech. Rep., 2010, Doctoral Thesis, ISBN 978-91-7385-363-7.

[24] S. Kent, "IP encapsulating security payload (ESP)," *Internet RFCs, ISSN 2070-1721*, vol. RFC 4303, December 2005.

[25] European Parliament, "Directive 2004/48/EC of the European Parliament and of the Council," 2004, http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2004:157:0045:0086:EN:PDF (accessed 2010.01.18).

[26] Y. Zhang, L. Breslau, V. Paxson, and S. Shenker, "On the characteristics and origins of internet flow rates," *SIGCOMM Comput. Commun. Rev.*, vol. 32, no. 4, pp. 309–322, 2002.

[27] K. Lan and J. Heidemann, "A measurement study of correlations of internet flow characteristics," *Computer Networks*, vol. 50, no. 1, pp. 46–62, 2006.

[28] W. John, S. Tafvelin, and T. Olovsson, "Trends and Differences in Connection-behavior within Classes of Internet Backbone Traffic," *Passive and Active Network Measurement Conference (PAM)*, pp. 192–201, 2008.

[29] V. Paxson and M. Allman, "Computing TCP's retransmission timer," *Internet RFCs, ISSN 2070-1721*, vol. RFC 2988, November 2000.

[30] M. Allman, V. Paxson, and E. Blanton, "TCP congestion control," *Internet RFCs, ISSN 2070-1721*, vol. RFC 5681, September 2009. [Online]. Available: http://www.rfc-editor.org/rfc/rfc5681.txt

[31] S. Bohacek, J. P. Hespanha, J. Lee, C. Lim, and K. Obraczka, "A new TCP for persistent packet reordering," *IEEE/ACM Trans. Netw.*, vol. 14, no. 2, pp. 369–382, 2006.

[32] S. Bhandarkar, A. Reddy, M. Allman, and E. Blanton, "Improving the robustness of TCP to non-congestion events," *Internet RFCs, ISSN 2070-1721*, vol. RFC 4653, August 2006.

[33] A. Sathiaseelan and T. Radzik, "Reorder notifying TCP (RN-TCP) with explicit packet drop notification (EPDN)," *International Journal of Communication Systems*, vol. 19, no. 6, pp. 659–678, 2005.

 ISBN: 978-1-61208-133-5

Weighted Fair Resource Sharing Without Queuing Delay

Benedek Kovács
Budapest University of Technology and Economics
H-1111, Egry József u. 1
Budapest, Hungary
kovacsbe@math.bme.hu

***Abstract*—This paper presents a new method that provides weighted fair sharing without queues and delay while outperforming traditional traffic throttling mechanisms. A mathematical description and model is presented to justify the findings and to provide better knowledge about traffic throttling characteristics. Simulation, numerical results and statistical discussion are also presented to underpin the findings especially where exact mathematical results are not or just partly available.**

***Keywords*-Network management; Service level agreement; Overload control; Traffic shaping; Fair sharing; Maximal throughput**

I. INTRODUCTION

There are many overload and congestion control and load sharing problems to be solved in telecommunication networks of various kinds e.g., in the Internet Multimedia Subsystem. Considering any type of network and signalling protocol a protocol operation flow consists of *messages*. The network nodes are entities receiving these messages and they process them using their resources such as CPU capacity or memory. If they lack the resource to process an offered message then the node is *overloaded*. Similarly, in packet switched networks, congestion control and fair resource (link bandwidth) sharing aims to keep the system utilizing its resources at its rated capacity while providing satisfactory service for the users and quality of service for service classes.

Many overload control mechanisms are developed to avoid overload and congestion situations. In some of them the target node (the critical resource) itself can deny to serve the request/forward a packet or sometimes reject/drop (send a negative reply message using minimal resources/ignore) it (see Figure 2). Another solution is that the sender (*source*) does not send out the message if it knows for some reason that the *target* will not be able to serve it. This information can be hard coded or can be gained from measurements by the source itself or gathered via special messages from peers (see Figure 1). In all cases i.e. in all overload control mechanisms there is a decision point and logic that decides to reject or admit/send out the request. This entity is called the *throttle* which is in the center of our interest. We suppose that the control information and the requirements on the behavior are available and up to date for the throttle. (Such information is to be gained with measurement or with some closed/open loop control mechanism. It has a large literature and is out of the scope of this paper to examine the characteristics of such mechanisms.) The throttle itself can be physically located in the target as Figure 2 shows, in the source or might also be distributed in multiple sources as Figure 1 shows.

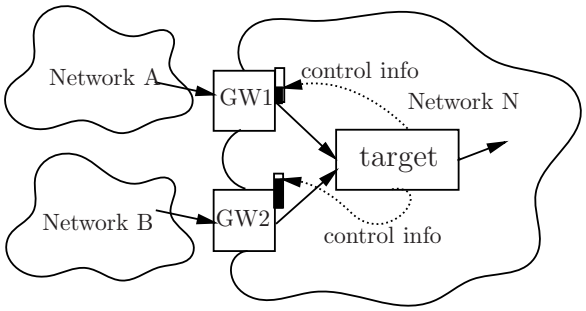


Fig. 1. A scenario where two operators A and B share a common core network N. The Gateways associated to each network implement distributed throttling to prevent the central nodes (target) to get overloaded. The central node—or any overload control agent— distributes control information to the Gateway Nodes to ensure fair service of requests from networks A and B while keeping maximum utilization of Network N.

The *throttle* entity discussed here is one very important and well defined part of overload control systems of any type as it has the role to reject (or drop) an *offer* or to let it go through i.e., admit it. When a *throttle* realizes a call gapping mechanism it makes the decision based only on previous offers thus no offers in the future are examined. (Call gapping means that calls are not admitted for a given period of time based on some measurements and collected information.) In our case the non-anticipative *throttle* is not allowed to delay an *offer* and only one *offer* arrives at a time i.e. the call gapping mechanism cannot buffer the *offer* and *admit* it later than it has arrived. This makes a fundamental difference from Weighted Fair Queuing and mechanisms like those in [4], [3]. Weighted fair resource sharing with no delay and maximum utilization is the main advantage of the method introduced and discussed in this paper.

For the sake of clarity, definition of terms and word usage is presented here:

 ISBN: 978-1-61208-133-5

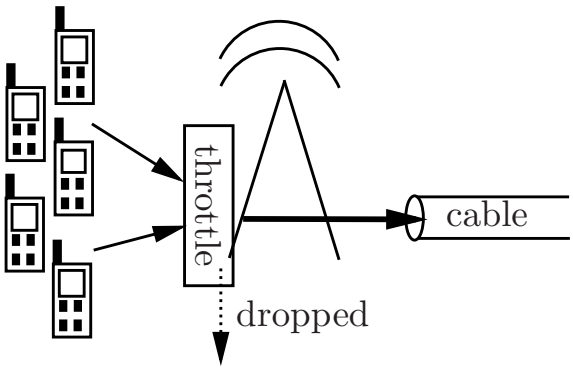


Fig. 2. The throttle is implemented in the Base Station and applied to avoid congestion on links when multiple users are sharing media in a small cell. Users not sharing media are also given a minimum share to be able to communicate. The throttle at the base station should select wisely which packet to be dropped and which to forward.

- The *throttle* decision function is a function mapping from the offer load point process to the set {*admission*, *rejection*}. (Each *throttle* is uniquely assigned a function γ that transforms the intensity process $\rho(t)$ of the income process to an intensity process of the admissions $\gamma(\rho(t)) = a(t)$. [9])
- An *offer* is the event for which the *throttle* has to decide on *admission* or *rejection*. If an offer is admitted it cannot be rejected (dropped) and vice versa, and there is no third possibility. An offer has properties: *arrival time, priority level and class* that can be measured.
- The *traffic class* and the *priority level* sets have finite elements.
- The *offered traffic* (or *offer load*) is the flow of *offers* modeled with a progressively measurable not necessarily stationary point process marked with the marks from the mark space that is the direct product of the set of priorities and classes. (This implies that the probability of two *offer* events occurring at the same time is zero.)
- The *admitted traffic* (or *throughput*) is the flow (i.e. the point process) of *admitted offers* (*offers* for which the *throttle* yields admission). The flow of admitted offers can be conditioned upon the whole history (past) of the offer load flow and upon the throttle parameters and undoubtedly on the decision strategy.

The above assumptions and definitions are natural and obvious and also necessary to make the discussion clear.

The typical verbal definitions of the requirements we investigate the different throttles against are given here preliminary. They are not precise and many contradict and can have multiple exact (i.e. mathematical) definitions with different results. In this paper we give exact definitions of these requirements:

- **Requirement-A** *Maximal throughput with bound*: No *offer* should be rejected if there is enough available capacity in the system to serve it, but no *offer* should be admitted if there is not enough available capacity to serve it in the system.
- **Requirement-B** *Priority levels*: Each *offer* may be assigned a priority level and the *offer* with higher priority shall be admitted in favor of the one with the lower priority level.
- **Requirement-C** *Throughput share for traffic classes*: The offers can be classified and for the traffic class i the s_i portion of the capacity of the target shall be provided.

The three requirements are associated with three behavior aspects. We propose a so-called *rate based call gapping* mechanism and compare it with Token Bucket [7] against **Requirement-A** and **Requirement-B**.

In Section II, we introduce traffic throttling mechanisms, namely the Token Bucket and the rate based call gapping. Then in Section III, the new call gapping mechanism is proposed and it is shown how it meets the requirements. We also show how it can be extended with the original Token Bucket concept, achieving similar throughput characteristics to the original algorithm while keeping the good properties of the introduced one. In Section IV we present our simulations and some figures about the offer and admission traffic flows with the three mechanisms. Using statistics we show how each mechanism meets **Requirement-B**.

II. Classical requirements in a queue-less environment

A. Traffic throttling, priority handling and weighted fair sharing

Many call gapping mechanisms have been developed for different purposes with different characteristics. One of the most important call gapping algorithms is the Crawford algorithm [5] but one of the most widely implemented solutions is Token Bucket call gapping mechanism defined in the Overload Control Standard H.248.11 [13].

1) The Token Bucket with parameters (r, W): The Token Bucket call gapping mechanism is the following: there is a bucket of available tokens representing available resources (free capacity) of the system. Requests are offered to the system and each of them is assigned a number of tokens needed i.e., the amount of resources it requires to be served. Once there are enough tokens in the bucket the request is admitted or dropped. (Thus no queues are applied and no delay is present in the system because of the Token Bucket call gapping algorithm.)

By the definition of the original Token Bucket the tokens are generated into the bucket with exponential distribution and the offers arrive with a Poisson process in most of the models that means, the time interval between the consecutive arrivals is also exponentially distributed. We analyze and describe such a variant.

Firstly we mention that decision about serving a request are often implemented differently. The most important difference is in the interpretation: rather than consuming the tokens the bucket fill b is increased when a request arrives. The token generation is then realized with decreasing the bucket fill. The maximum fill is the watermark W that cannot be exceeded and

 ISBN: 978-1-61208-133-5

also the bucket fill can not be lower than 0. This concept is equivalent to the original algorithm.

Secondly we consider deterministic token generation instead of the exponential one that is used in most cases (e.g., [13]), because it is much easier to implement and sometimes to analyze, as well.

Then the Token Bucket mechanism we discuss works as follows: When a new request arrives at t_n than the needed bucket fill is calculated: $b(t_n)$ as if the request was served. This is done with calculating the expected number of tokens that would have been generated from the time when the former service was served (t_{n-1}) then multiply it with the throughput capacity of the bucket i.e., the Token Bucket rate at t_n: $r(t_n)$ and subtract it from the former bucket size at $b(t_{n-1})$. Then it tests it against the preset constant watermark: W.

Definition 1: Token Bucket call gapping strategy $\gamma_t(r, W)$.

$$b(t_n) = \max\{\chi(t), b(t_{n-1}) - r(t_{n-1})(t_n - t_{n-1}) + \chi(t)\}, \quad (1)$$

where $\chi(t) = 1$ iff there is an offer. Admit if $b(t_n) \leq W$. If the offer is admitted, the above definition is used for the next value of the bucket fill b. If the offer is rejected, then $b(t_n)$ is recalculated with $\chi(t) = 0$.

In many solutions the offers for the bucket can be of different types with different resource needs thus $S_i\chi_i(t)$ is used for update, where S_i is the so-called "splash amount" i.e., the expected number of tokens needed to serve the request of type i. From now on we suppose that $S_i = 1$, since the calculations would be much more difficult without any qualitatively different result with respect to the requirements we consider now.

2) Rate based call gapping: One obvious solution to regulate traffic is to limit the admitted offers in a given time frame. The Crawford algorithm worked as follows: for each time period T when the number of offers reached a certain number c, the capacity of the target, no other offers are admitted in the given period. This introduced bursts in traffic thus not preferred in most cases.

The idea and our proposal is to maintain an estimate of the traffic intensity $\hat{\rho}$ updated at each t_n an offer arrives. The provisional admission rate estimate $\hat{a}(t_n)$ is also calculated and then compared to some goal function $g(t_n)$, say the actual maximal throughput rate $c(t_n)$. Iff $\hat{a}(t_n) \leq g(t_n)$ then the offer is admitted.

The essence of such a mechanism lies in the way how the estimates are calculated and how $g(t_n)$ is determined. The first statistics that can be used and examined as an estimate for the intensity of a point process is the simplest first moment type estimate:

Definition 2 (Periodic intensity estimate):

$$\bar{\lambda}(t_n; T) := \begin{cases} \dfrac{N(t_n) - N(t_n - T)}{T} & \text{if } t_n \geq T \\ \dfrac{N(t_n)}{t_n} & \text{if } t_n < T \\ \bar{\lambda}(0; T) := 0 & \end{cases} \quad (2)$$

To calculate the value of the periodic intensity estimate one has to maintain all $t < t_i \leq t - T$ event times and it is often too resource consuming thus not feasible in real time systems. An adaptive estimate is often preferred and introduced by patent [14] to control network traffic. The definition is the following:

Definition 3 (Dynamic intensity estimate):

$$\begin{aligned} \hat{\lambda}(t_n; T) &:= \max\{\tfrac{1}{T}, \tfrac{T\hat{\lambda}(t_{n-1};T) - (t_n - t_{n-1})\hat{\lambda}(t_{n-1};T) + 1}{T}\}, \\ \hat{\lambda}(t_0; T) &:= 0. \end{aligned} \quad (3)$$

We do not want to go into details of intensity estimation of inhomogeneous point processes in this paper.

B. Priority handling with Token Bucket and Rate-based call gapping

At first we would like to clarify that once the offered traffic is modeled with a point process and the throttle meets **Requirement-A** we cannot provide priority between the offers. Why? Suppose that we have an offer in the system and we have to decide if we should admit it or not. **Requirement-A** tells us to admit the offer if we have the capacity to serve it. Suppose that this is the case and see that if the throttle would not admit the current offer to reserve this capacity for offers of higher priority then it might happen that there will be no higher priority offer in the future and the throttle would suffer a loss of workload.

1) Priority handling with Token Bucket: However, giving up the maximal throughput requirement with turning it into an efficient-enough requirement, some priority handling naturally can be done. In the Token Bucket concept different watermarks are assigned to each priority level. The offers of lower priority are checked with a lower watermark. This means reserving a set of tokens (system resources) to the higher priority traffic. This method violates **Requirement-A** whenever $b(t)$ declines to 0 before rejecting an offer. Whenever this event has a low probability, using different watermarks for different priority levels is a good solution to meet **Requirement-B** with a Token Bucket throttle.

This is a strict priority handling in the sense that when possible, resources are always reserved for higher priority offers. The probability of rejecting an offer on a given priority level is given by a formula which is, together with other regulation properties, discussed in [10].

2) Priority handling with rate based call gapping only: For rate based call gapping mechanisms the above solution cannot be applied but less strict priority handling can be achieved using the dynamic intensity estimate 3 with different settings of parameter T.

The dynamic estimate is asymptotically unbiased with low variance proving its usability and also, the dynamic one often performs better than the most commonly used periodic estimate 2. Such discussion is pure mathematics thus out of the scope of this article.

It can be also proved that the expected bias of the dynamic estimate 3 is lower than the real intensity and this property is

 ISBN: 978-1-61208-133-5

monotonous in T. More exactly, at each t_n an offer arrives if $T_1 < T_2 < \infty$ then $\mathbf{E}[\hat{\lambda}(t_n;T_1)] < \mathbf{E}[\hat{\lambda}(t_n;T_2)] < \mathbf{E}[\hat{\lambda}(t_n;\infty)] = \lambda(t_n)$. The load is thus slightly underestimated, but the more likely the load is underestimated the more likely the offer is admitted i.e., higher priority offers should have lower T parameter and **Requirement-A** is met.

C. Weighted fair sharing

Fair sharing is about the differentiation between offers on the same priority level. (Note that traffic classes are not the same as priority levels.) Suppose a scenario like Figure 1 where the resources of Network N should be equally split between the two contractors while emergency (priority) calls still have to be served. In our other example minimal share for different data types should be achieved on limited bandwidth while it is allowed that there are some priority e.g., time critical packets such as voice, video, media, etc.

The solution in an environment without queues proposed in this paper is rather straightforward: one should measure the incoming rates for each class and check each offer against online calculated goal levels. The goal throughput rates can be adjusted dynamically according to the weighted fair sharing agreements and also according to the maximal throughput requirements.

The problem with the solution using Token Bucket only is that it does not measure the incoming traffic rates. (Using the bucket fill instead of measurement gives false result.) However, using any kind of rate based call gapping method the problem is solved. One can also measure the incoming rates for each class with the estimates presented above and set the bucket fill-up rate r according to the desired goal levels but this is already a combination of the methods. A solution for the network level case is proposed in the GOCAP standard [11].

III. Weighted fair sharing with no delay

In this Section our new method, the proposed fair sharing method with no delay is presented. At first we introduce the complete proposed procedure clearly. Then we discuss and prove how it satisfies all the requirements and what possible extensions, modifications or other solutions might result in a similar good algorithm. At the end of the discussion we present the relationship between the new method and the original Token Bucket algorithm.

A. The new call gapping algorithm $\gamma_g(c,T,g,s)$

Suppose that the consecutive offers arrive to the *throttle* at $... < t_{n-1} < t_n < t_{n+1} < ...$ time instants respectively. Each *offer* has a well defined *priority level* $j, j \in 1..J$ and *traffic class* $i \in 1..I$. Each *priority level* j has a priority parameter T_j assigned ($T_j \geq T_k$) if the *offer* with priority k has the higher priority and each *traffic class* has a pre-configured weight $i \mapsto s_i \in (0,1) \subset R$, (where $\sum s_i = 1$). For each i the algorithm maintains an estimation of the incoming *offer rate* $\hat{\rho}_i(t)$, a *provisional admission* rate $\hat{\alpha}_i(t)$ from which it calculates a *bounding rate* $g_i(t)$ and then according to the decision it estimates an *admission rate* $\hat{a}_i(t)$.

We suppose that the rate of the throttle varies with the following function: $c(t)$. (This value is determined and given for the algorithm and represents the capacity of the *throttle* and might be different from $r(t)$).

Definition 4: Define the proposed throttle decision strategy γ_g in the following way. Let us suppose that at t_n an *offer* arrives and the system is in state $\{t_{n-1}, \hat{\rho}_i(t_{n-1}), \hat{a}_i(t_{n-1})\}$ and $c(t_n)$:

1) Determine priority constants i.e., calculate T_j;
2) Update the incoming rates estimated for all i classes: $\hat{\rho}(t_n)$ with $\chi_k(t_n) = 1$ iff $i = k$, 0 otherwise;
3) Calculate a provisional admission rate for all i: $\hat{a}(t_n)$ with $\chi_k(t_n) = 1$ iff $i = k$, 0 otherwise;
4) Calculate the bounding rate for class i only: $g_i(t_n)$;
5) If $\hat{\alpha}_i \leq g_i$ then *admit* the offer and $a(t_n) := \alpha(t_n)$ else *reject* the *offer* and update $\hat{\alpha}(t_n)$ with $\chi_k(t) = 0, \forall k(!)$;
6) (Continue with 1. for the next event).

We propose to update $\hat{\rho}_i, \hat{\alpha}_i, \hat{a}_i$ according to the following equation:

$$\hat{\lambda}(t_n) := \frac{\chi(t_n)}{T_j} + \max\{0, \frac{T_j\hat{\lambda}(t_{n-1}) - (t_n - t_{n-1})\hat{\lambda}(t_{n-1})}{T_j}\}, \tag{4}$$

where $\hat{\lambda}$ is an estimator asymptotically unbiased for the $\lambda(t)$ real intensity of a point process thus to be replaced by $\hat{\rho}_i, \hat{\alpha}_i, \hat{a}_i$ and indicator $\chi_i(t_{n-1}) = 1$ iff the *offer* is of type i and 0 otherwise (or further specified like in step 5). Note that the time parameter T_j changes in time as well according to the priority level and that the former always has to be remembered.

To calculate the bound rate we introduce $u(t)$ the provisional used capacity according to **Requirement-B**:

$$\begin{aligned} u(t) &:= \sum_{\forall i} \min\{s_i c(t), \hat{\rho}_i(t)\} \\ &= \sum_{\hat{\rho}_i(t) \leq s_i c(t)} \hat{\rho}_i(t) + \sum_{s_i c(t) < \hat{\rho}_i(t)} s_i c(t) \end{aligned} \tag{5}$$

Thus the remaining (unused) capacity in the system is $c(t) - u(t)$. This has to be split between traffic classes with higher incoming rate then the agreed share $\hat{\rho}_i(t) > s_i c(t)$. Then

$$g_i(t) := \min\{\hat{\rho}_i(t), s_i c(t) + (\hat{\rho}_i(t) - s_i c(t))\frac{c(t) - u(t)}{\rho - u(t)}\}. \tag{6}$$

It is important to see that our method is capable to handle other class-wise throughput criteria than fair sharing and maximal throughput. Giving upper or lower bounds for g one can implement fairly complex throttle mechanisms.

As one can see the new method is more complex than the original token bucket mechanism. However, the processing cost of updating the few variables introduced is significantly smaller than processing the offers, thus it does not count even in case of overload.

B. γ_g meets all the requirements

Now that the strategy is introduced we prove that it meets all the requirements. We define each requirement mathematically

 ISBN: 978-1-61208-133-5

then we show how they are satisfied. We introduce some notation to make the discussion clear.

- $c(t)$ represents the true capacity of the system expressed in rate, deterministic and coming from an external input source.
- $\rho(t)$ is the real intensity of the offered traffic and $\hat{\rho}(t)$ is its estimate with (4).
- $a(t)$ is the real intensity of the admitted traffic and $\hat{a}(t)$ is the estimation of the intensity with (4).
- $\alpha(t_n)$ is the preliminary admitted traffic intensity for which the following stands: $\alpha(t) = a(t), \forall t < t_n$, and $\alpha(t_n)$ is the intensity $a(t_n)$ would have if the offer was admitted at time t_n, and its estimate is $\hat{\alpha}(t)$ accordingly.

1) Requirement-A : This requirement consists of two parts. Firstly, it says that there exists an upper bound for the system that should not be exceeded i.e., it limits the admission rate to avoid overload. Secondly, it tells us that once the limit is not exceeded then all the offers should be admitted to maximize the utilization. However, in theory the words capacity and bound can have many different definitions depending on the model we use for the target node.

The target node is often modeled with an inverse Token Bucket, i.e., a server with deterministic serving rate s and a queue of maximal length Q. It is very easy to see that the Token Bucket throttle $\gamma_t(s, Q)$ can perfectly meet the requirement in this case. (Note that this is only true supposing that there is no delay in the system between the throttle and the protected entity while $s(t) = r(t)$ is satisfied.)

Another approach is to assume that the target can handle requests on a maximal call rate c that is used as the bound at the throttle.

Both models have benefits and drawbacks while a mixture of them is used in practice. Speaking about the capacity of a node in Next Generation Networks engineers often refer to the call rate value in industrial contracts and Service Level Agreements. It is very important to note that the feedback driven overload control mechanisms work with call rate information too (see [13]). On the other hand a server with queue is a common model in the academic literature for the CPU capacity and Token Bucket (or versions of it) is proposed in many standards (e.g., [13] again) and implemented into nodes.

As a consequence we say that although it is rather difficult to give an exact definition for **Requirement-A** we can give a certain definition grabbing a few properties depending on the method we use.

Definition 5: Call rate bound. **Requirement-A** is met if $\sum E[a_i(t_n)] \leq c(t_n)$ (the throughput rate is bounded in expected value).

Theorem 1: The throttle with strategy γ_g meets the **call rate bound** requirement.

Proof: The proof relies on the fact that the estimator is asymptotically unbiased i.e., $\lim_{T\to\ +\infty} E[\hat{a}; T_i] = E[a_i]$ with negative bias if $T \geq 1/a_i$ (thus $E[\hat{a}_i] < E[a_i]$). The proposed strategy γ_g limits a_i in a way that $a_i \leq g_i$, thus showing $g(t) := \sum g_i(t) = c(t)$ competes the proof.

Define $u_1(t) := \sum_{i:\hat{\rho}_i(t)<s_ic(t)} \hat{\rho}_i(t)$ and $u_2(t) := \sum_{i:s_ic(t)<\hat{\rho}_i(t)} s_ic(t)$ thus $u = u_1 + u_2$ and then $g_i = \min\{\hat{\rho}_i, s_ic + (\hat{\rho}_i(t) - s_ic(t))\frac{c-u}{\rho-u}\}$. Although the system is non-stationary it is homogenous in time so $f(t) = \text{const.}$ for all functions. Now calculate $g(t)$:

$$\begin{aligned} g &= \sum g_i = \sum \min\{\hat{\rho}_i, s_ic + (\hat{\rho}_i - s_ic)\frac{c-u}{\rho-u}\} = \\ &= \sum_{i:\hat{\rho}_i<s_ic} \hat{\rho}_i + \sum_{i:s_ic<\hat{\rho}_i} s_ic + (\hat{\rho}_i - s_ic)\frac{c-u}{\rho-u} = \\ g &= u_1 + u_2 + (\rho - u_1 - u_2)\frac{c-u_1-u_2}{\rho-u_1-u_2} = c. \end{aligned} \quad (7)$$

■

Corollary 1: The following calculation of g can also be used:

$$g_i(t) = \min\{\hat{\rho}_i, s_ic(t) + (\hat{\rho}_i(t) - s_ic(t))\frac{c(t)-u(t)}{\rho-u(t)}\}, \quad (8)$$

where $u(t) = \sum_{i:\hat{\rho}_i(t)<s_ic(t)} \hat{\rho}_i(t) = \alpha(t)$. Then (7) becomes:

$$g' = u_1 + u_2 + (\rho - u_1 - u_2)\frac{c-u_1-u_2}{\rho-u_1-u_2} = c. \quad (9)$$

The difference between the two strategies is that in case of g the remaining capacity is split between the classes with higher offer rates proportionally to their weights while using g' the remaining capacity is split proportionally to the remaining offer rates. Both satisfies **Requirement-A** and **Requirement-C**as we will show. From now on g means either g or g' and the results will be obviously the same.

2) Requirement-B: As pointed out before, the priority requirement for call gapping is the most complex one in a way, since in the gapping algorithms it is supposing that we make decisions using measures on the past and the present offer. No future events can be used, thus **Requirement-B** is always satisfied. There is always one offer in the system and the throttle can admit or reject it according to **Requirement-A** and **Requirement-B**.

In the case of the Token Bucket call gapping, different watermarks W_j are introduced for each priority level j. One interpretation is that the bucket allows larger peaks for traffics with higher priority, thus $W_j < W_k$ whenever k represents the higher priority level. Doing this, the bucket implicitly reduces the throughput for lower priority traffics (the extra peak in the bucket has to be refilled with tokens i.e., $b(t)$ has to decline below the low watermarks to admit low priority traffic). Note that the different watermark levels have no effect if the offer rate is low with small peaks thus the rejection probability is small i.e., if there is no overload. Supposed that the true bound is $W = \max\{W_j\}$, this system preserves capacity for high priority traffic.

We give a similar solution for the problem through the timer parameter of the estimators: T. As defined above, we introduce a function of $T : j \mapsto T_j$ where $T_k \leq T_j$ if k represents the higher priority. (Note that it is the other way round for W_js.) The interpretation is that the estimator forgets the high offer

 ISBN: 978-1-61208-133-5

rates faster for the traffic of the higher priority. Let $T_m = \min\{T_j\}$ the true bound on the throttle using different T_js. This means that for low priority traffic it remembers the high peaks for a longer period, thus reserves capacity for the higher priorities similarly to the Token Bucket.

The two methods have different characteristics, but one thing is common. Both reserve capacity for higher priority traffic. Now we say that to meet **Requirement-B** the system has to have this ability and define it in the following way.

Definition 6: **Requirement-B**. Suppose that the throttle has rejected an offer at time t_{n-1}. Let $t_{n;j}$ be the closest time the throttle is able to admit an offer of priority level j. **Requirement-B** is met iff $\forall k, l(t_{n;k} \leq t_{n;l}) \Leftrightarrow (k \geq l)$ (k represents a higher priority).

The exact proof of this statement is not ready yet. The simulation results show that the proposed strategy satisfies **Requirement-B**. We discuss the statement in the Numerical Results Section.

3) Requirement-C: This is referred to as the throughput share requirement and tells us that there should be at least an s_i portion of the capacity dedicated to traffic class i.

Definition 7: **Requirement-C**. The **Minimum share requirement** is met if $\forall i : (\rho_i(t_n) \leq s_i c(t)) \Rightarrow E[a_i(t_n)] = \rho_i(t_n)$ i.e., if the offer rate of a traffic class is less than the agreed share, it should be fully admitted.

Theorem 2: The throttle with strategy γ_g meets **Requirement-C** in expected value.

Proof: We have the asymptotical unbiasedness for our estimators, thus $\lim_{T\to\ +\infty} E[\hat{a;T_i}] = E[a_i]$, meaning that the proof is true for the expected value of a_i.

Statement $\hat{a}_i(t_n) = \hat{\rho}_i(t_n)$ whenever $\forall i \hat{\rho}_i(t_n) \leq s_i c(t)$ is equivalent to the statement ($g_i(t_n) \geq \hat{\alpha}_i(t_n)$ thus) $g_i(t_n) \geq \hat{a}_i(t_n)$ whenever $\hat{\rho}_i(t_n) \leq s_i c(t)$. According to strategy γ_g: $g_i(t_n) = \hat{\rho}_i(t_n)$ whenever $\hat{\rho}_i(t_n) \leq s_i c(t)$ and since $\hat{\alpha}_i(t_n) \leq \hat{\rho}_i(t_n)$ because $\hat{a}_i(t_{n-1}) \leq \hat{\rho}_i(t_{n-1})$, it is true that $\hat{\alpha}_i(t_n) \leq g_i(t_n)$ thus the offer is admitted (and also $\hat{a}_i(t_n) \leq g_i(t_n)$). ■

C. Rate model for Token Bucket and a joint algorithm merging the methods

In this section we introduce a model for Token Bucket that is equivalent to the definition in Section II but makes calculations easier.

Definition 8: Token Bucket Rate Model Strategy: $\tilde{\gamma}_t(r, W)$ Let us define $T(t) = W/r(t)$ and use the following equation for updating the bucket rate variable:

$$\tilde{a}(t_n) = \frac{\chi(t_n)}{T} + \max\{0, \frac{T\tilde{a}(t_{n-1}) - (t_n - t_{n-1})r(t_n)+}{T}\}$$

where $\chi(t) = 1$ iff there is an offer at time t. Admit the offer iff $\tilde{a}(t_n) \leq r(t_n)$. If the offer is admitted then the above definition is the used for the next value of the bucket rate variable $\tilde{a}(t)$. If the offer is rejected then $\tilde{a}(t_n)$ is recalculated with $\chi(t) = 0$.

Theorem 3: The Token Bucket and the Token Bucket Rate Model Strategy are the same: $\gamma_t = \tilde{\gamma}_t$.

Proof: It is easy to show that $b(t_{n-1}) = \tilde{a}(t_{n-1})T \Rightarrow b(t_n) = \tilde{a}(t_n)T$ and the decision is $b = T\tilde{a}(t) \leq Tr(t) = W$ also trivial. ■

If one extends the Token Bucket for traffic class handling with some role like in the proposed mechanism it will not provide traffic class fairness. The reason is hidden in the fact that unlike $\tilde{\rho}, \tilde{\alpha}, \tilde{a}, \tilde{\beta}$ and all such estimators is not asymptotically unbiased i.e., $E[\tilde{\lambda}] = \lambda$ as $t \to \ +\infty$ is not true for the estimators defined with:

$$\tilde{\lambda}(t_n) = \frac{\chi(t_n)}{T_j} + \max\{0, \frac{T\tilde{\lambda}(t_{n-1}) - (t_n - t_{n-1})r(t_n)}{T}\}. \quad (10)$$

The bucket fill does not represent at all the used capacity in the system it only measures the peakedness of the traffic but these peaks can happen on low offer rates too.

On the other hand, the proposed method does not allow such big transient peaks in the traffic. Now we aim to make the proposed new call gapping to behave like Token Bucket. We define the following strategy that is a mixed architecture.

Definition 9: Rate Based Call Gapping with Bucket-type Aggregate Characteristics: γ_x Take all the definition from the new call gapping mechanism γ_g for $\hat{\rho}, \hat{\alpha}, \hat{a}, u, g_i$ and define $T_j(t) = W_j/r(t)$. Take W_j and the bucket fill change definition b from the original token bucket γ_t. Perform all the steps like in γ_g but decide using the following constraint equation: $\frac{b(t_n)}{W_j}\hat{a}_i(t_n) \leq g_i(t_n)$.

We will show numerically that the mixed algorithm behaves like Token Bucket on aggregate level and meets all the requirements. The source of the idea comes from the fact that $\hat{a}(t)$ places a strict bound on the rate thus $\hat{a}(t) \leq r(t)$ is always true as required. However we decrease the value of $\hat{a}$ thus allow peaks in the traffic like Token Bucket does. (See that Token Bucket γ_t allows temporary bounding violation rate-wise unlike γ_g but like γ_x. The bucket size related to the whole bucket is a kind of measure of this violation.)

1) $\gamma_{t'}$ and γ_x and Requirement-A : Here we discuss how the different algorithms meet the maximal throughput requirement. It is obvious that Token Bucket cannot meet **Requirement-A** in the way it was defined before since that definition assumed that the target has an infinite queue.

We do not aim to give an exact definition to **Requirement-A** but we derive relations between the bucket and the estimator based throughput characteristics. The number of admitted offers i.e., the probability of admission is in the center of our interest.

The probability of admission for token bucket depends on the offer rate with the following formula: $1 - \text{Erlang}[\rho, r]$. Thus the probability of losing calls is only defined at given values of ρ.

For rate based call gapping, since the estimator always over-estimates the rate ($\lambda < \hat{\lambda}$) and cuts the traffic strictly with c the admission rate is always below the target. But for the same reason it is possible that the offer is rejected although it could have been accepted according to the bound. The probability of this is the probability of estimating higher rate than c while the

 ISBN: 978-1-61208-133-5

true offer rate is lower: $P[\hat{\alpha} > c|\alpha < c] = 1 - \frac{P[\alpha<c-B[T]]}{P[\alpha<c]}$, where $B[T] = 1/(T(1-F[T]) + E[\Delta t|t < T]) - \alpha$ is the bias. (Knowing the exact bias if constant intensity is supposed for the offer rate, the bound can be modified to have maximal throughput and strict bound at the same time.)

The two methods can only be compared at a given value of the intensity. For all those values when the value of the intensity is not between $c - B[T]$ and c the γ_g strategy works perfectly. The Token Bucket drops a call with positive probability for any value of the offer rate and also might admit when the intensity is higher than allowed. This means that we cannot tell which method is better or has the higher throughput since it depends very much on the offer rate.

Theorem 4: The mixed strategy γ_x meets **Requirement-A** with appropriate watermark settings.

Proof: It is shown in Theorem 1 that $\sum g_i(t) = g(t) = c(t)$ and since the definition of g was not changed we should only examine what means to compare g_i to $\frac{b}{W_j}\hat{a}$ rather than to $\hat{a}$.

When we admit a request then $1 \leq b(t_n) \leq W_j \leq W_{\max}$ thus $\frac{1}{W_{\max}}\hat{a}_i \leq \frac{1}{W_j}\hat{a}_i \leq \frac{b(t_n)}{W_j}\hat{a}_i \leq \hat{a}_i$. This tells us that γ_x lets through more messages than γ_g since $E[\frac{b(t_n)}{W_j}\hat{a}_i] \leq E[\hat{a}_i]$. Fortunately the maximal watermark limits this overflow error $\frac{1}{W_{\max}}E[\hat{a}_i] \leq E[\frac{b(t_n)}{W_j}\hat{a}_i]$. It tells us that there is a setting of watermarks that guarantees bounding. (It is obvious that if $W_{\max} \to \ +\infty$ then $\frac{b}{W_{\max}}\hat{a}$ becomes very small and we always admit the request thus the theorem cannot be proved for any watermark settings.) ■

2) γ_x and Requirement-B: Some simple theorems are proved to show that the mixed strategy meets the priority and the throughput share requirements.

Theorem 5: Token Bucket strategy γ_t meets **Requirement-B**.

Proof: Obviously, the time to accept the next offer of priority level j is the time when the bucket level declines sufficiently to $b(t) \leq W_j$. For all levels $k > j$, $W_k > W_j$ i.e., $b(t)$ declines under the lower threshold later in time and the requirement is met. ■

Again it is rather hard to show that the mixed strategy γ_x meets **Requirement-B**. However, it seems to be trivial that γ_x satisfies **Requirement-B** more drastically than γ_t does. We have interesting simulation results presented about this property. We can see numerical results about this in Section IV.

3) γ_x and Requirement-C:

Theorem 6: The mixed strategy γ_x meets **Requirement-C**.

Proof: As pointed out γ_x admits at least all the offers γ_g does since $\forall i, \frac{b}{W}\hat{\alpha}_i \leq \hat{\alpha}_i$ is compared to $s_i c$ while a comparison of $\hat{\alpha}$ would be enough. This means that the mixed strategy provides minimum throughput share and fulfills **Requirement-C**. ■

IV. Numerical results and analysis

Although we have nice proofs on the good behavior of the proposed rate based call gapping mechanism the complete mathematical discussion about the differences and similarities with Token Bucket is not ready yet. It is also true that the requirements can be interpreted with definitions slightly different from those we gave. Therefore, we would like to present some simulation results and show that the findings are valid.

The simulation is written in *Mathematica* [15] and a *notebook* is available at *http://www.math.bme.hu/ kovacsbe/rbcg/BENEDEK-KOVACS-rate-based-call-gapping-PRELIMINARY-VERSION.nb* as an electronic appendix.

A. *Requirement-A*

The figure shows that all the mechanisms limit the admitted offer rate while trying to keep the highest throughput. In this scenario we examine the traffic on aggregate level i.e., there is only one traffic class for which the capacity of the throttle should be maximized and limited. The capacity is 1 offer/sec for the simple simulation case while the average number of offers per sec increases from 0.8 to 2 meaning that there is a 200% load on the node.

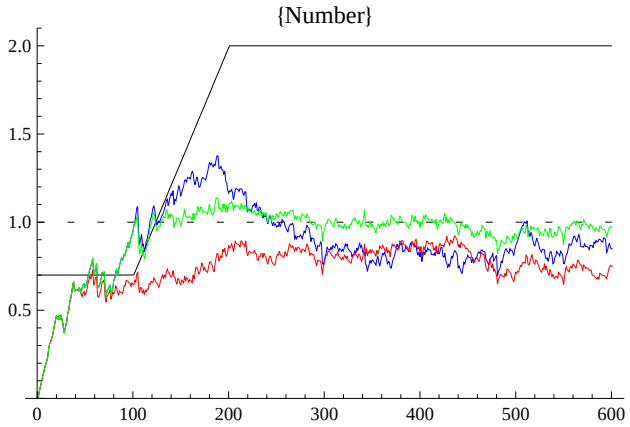


Fig. 3. The new algorithm (γ_g) on aggregate level. (Black: nominal offer rate, red: the token bucket's, blue: γ_g's, green: γ_x's throughput.)

As it can be seen in Figure 3 all three mechanisms limit the admitted traffic although Token Bucket allows considerable peak at the beginning. (The size of the peak depends on the parameters we set. Here the 1 offer/sec capacity is very small compared to the watermark which is set to 10.) On the other hand, rate based call gapping seems to under-utilize the system while the joint mechanism seems to have the smoothest and also maximal throughput.

After a total 600 offers from each traffic with the same exact trajectory the results show that $\gamma_t, \gamma_g, \gamma_x$ has admitted 415, 386, 404 number of calls respectively.

The problem with the mathematical discussion of maximal throughput is that the results depend very much severely on the value of the offer rate and capacity. It is only possible to compare the mechanisms at given rates, which is useless for real applications.

 ISBN: 978-1-61208-133-5

90	100	$W_H = 10$ $W_L = 15$	{0.,1.} [.002,.002]	{0.38,0.62} [.015,.015]	{0.01,0.99} [.006,.006]
150	100	$W_H = 10$ $W_L = 15$	{0.2,0.98} [.003,.003]	{0.4,0.6} [.007,.007]	{0.05,0.95} [.005,.005]
10	10	$W_H = 10$ $W_L = 20$	{0.,1.} [.000,.000]	{0.31,0.69} [.014,.014]	{0.,1.} [.000,.000]
10	10	$W_H = 10$ $W_L = 10$	{0.5,0.5} [.008,.008]	{0.5,0.5} [.009,.009]	{0.5,0.5} [.007,.007]

TABLE I
IN EACH ROW THE FOLLOWING QUANTITIES ARE PRESENTED RESPECTIVELY: TOTAL OFFER RATE: ρ, MAXIMAL THROUGHPUT: c, WATERMARK SETTINGS: $W_{\text{HIGH}}, W_{\text{LOW}}$ WHILE $T_j := W_j/c$. THEN PORTION IN REJECTED MESSAGES FOR TOKEN BUCKET, RATE BASED CALL GAPPING AND THE MIXED MECHANISM RESPECTIVELY.

B. *Requirement-B*

To discuss **Requirement-B** we provide the reader with some statistical results. The sample is generated with our simulation program. Generally there are two priority levels: normal and emergency calls. Each call is one of the two types with 1/2 probability. The means and the standard deviations are presented of 100 samples with 10 000 offers handled in each sample. The further setups for the simulation can be seen in Table IV-B.

It can be seen that all three methods reject less offer from those of higher priority but Token Bucket (γ_t) and the mixed mechanisms (γ_x) enforce a more strict priority handling than the simple proposal. Note that in case of sustained overload (row 2) almost all dropped offers are the lower priority ones.

C. *Requirement-C*

The results tell explicitly that unlike the new rate based call gapping proposal the original Token Bucket algorithm does not meet **Requirement-C**. We consider a scenario when there are two traffic classes Class A and Class B. The agreed share for Class A is the 20% of the total capacity of the node while the share for Class B is the remaining 80%. The offer rates set for the simulator are exactly the inverse of this for the two types of traffic.

The aggregate offer rate increases from 0.7 offers/sec to 2 offers/sec and reaches the scenario of 100% overload (the capacity of the node is 1 offer/sec while the offered rate is a 2 offers/sec on average). The offer rate of traffic Class B is 0.4 i.e., it is still under its provided share, thus all such calls are admitted. On the other hand, the whole remaining capacity should be granted to traffic Class A and it should be admitted on a higher level than the agreed share and only those exceeding the capacity limit are to be rejected.

Figure 4 shows the behavior of the rate based call gapping mechanism. With the proposed mechanism the minimum share is guaranteed for traffic Class B (the admission line is around the offered), while the requirement fails for Token Bucket. With the proposed method there is no rejected message of Class B, since it never offers on a higher rate than the agreed share. The throughput of the throttle is limited but also maximized, since Class A is granted all remaining capacity.

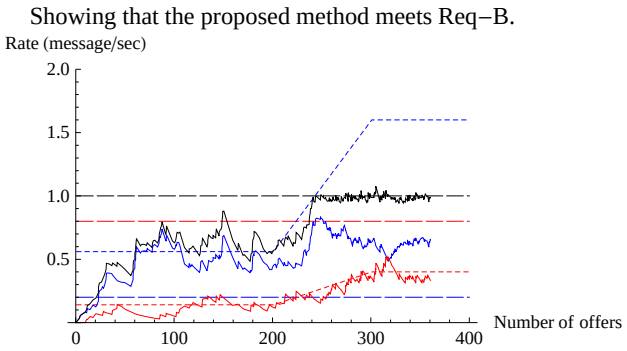


Fig. 4. The new algorithm (γ_g) with two traffic classes. Black: aggregate throughput rate, Red and Blue: the throughput rate of each traffic class, Solid: throughput, Dashed (large): nominal capacity, required minimal throughput, Dashed (small): nominal offer rates

V. CONCLUSION

We have presented a weighted fair sharing mechanism with no delay and its extension with the original Token Bucket to provide good network characteristics as well. These unique mechanisms meet the *maximal throughput with bound* requirement, handle priority messages and offers, and give minimum share for different traffic classes without using message buffers or queues. New ways of measuring traffic intensities are also proposed.

Examining the properties of the mechanisms we gave mathematical definitions of the three requirements and accompanied the mathematical model with several theorems. Still, the proof of priority handling is missing for the new methods, we presented statistical analysis instead. We used simulation that we implemented to underpin our proposal and findings.

Our rate based call gapping strategy can use different traffic intensity estimators. It is still remains an open task to find the optimal estimator and the optimal parameter setting of the estimators considering Poissonian input traffic with variable intensity or even a non-Poissonian (e.g., general renewal or Hawkes type) input process. The mathematical background on to prove the properties of the estimates of the intensity of a point process is to be published in the near future.

 ISBN: 978-1-61208-133-5

VI. APPENDIX

Notations:

$a, a(t)$	Real admission rate
$\hat{a}, \hat{\alpha}(t)$	Estimated admission rate
$b, b(t)$	Actual bucket fill
$c(t)$	Maximal capacity of the target (rate)
$g_i, g_i(t)$	Goal rate for traffic class i
$g, g(t)$	Sum of goal rates of all traffic classes
$r, r(t)$	Token Bucket's token generation rate
T	Parameter of the estimator
T_j	Parameter of the estimator for priority level j
$u, u(t)$	Used capacity according to **Requirement-B**
W	Watermark for Token Bucket
W_j	Watermark for offers of priority level j
$\hat{\alpha}, \hat{a}(t)$	Estimated preliminary admission rate
$\beta, \beta(t)$	Preliminary bucket size
γ_t	The token bucket throttle function
γ_g	The rate based call gapping throttle function
$\gamma_{g'}$	The variant of the rate based call gapping throttle function
γ_x	The rate based call gapping throttle function with Token Bucket extension
$\lambda, \lambda(t)$	Intensity (rate) of a Poisson process
$\hat{\lambda}, \hat{\lambda}(t)$	Estimated rate (intensity)
$\rho, \rho(t)$	Real offer rate
$\hat{\rho}, \hat{\rho}(t)$	Estimated offer rate

ACKNOWLEDGEMENTS

The research was motivated by Ericsson Research Hungary (Ericsson Telecommunications Hungary Ltd.) and the High Speed Network Laboratory, and was partially funded by the Hungarian Ministry of Culture and Education with reference number NK 63066 and the National Office for Research and Technology with reference number TS 49835. Special thanks to János Tóth (Budapest Univ. of Tech. and Eco., Dept. of Math. Analysis) for the careful reviews and support.

REFERENCES

[1] A. Demers, S. Keshav, and S. Shenker: Analysis and simulation of a weighted fair queuing algorithm, Journal of Interworking Research and Experience, pp. 3–26, 1990

[2] A. K. Erlang: "Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges", Post Office Electrical Engineer's Journal 10 (1917–18), pp. 189–197.

[3] A. K. Parekh and R. G. Gallager: "A generalized Processor Sharing Approach to Flow Control in Integrated Services Netwokrs: The Single-Node Case", IEEE/ACM Transactions On Networking, Vol. 1, No. 3, June 1993, pp. 344–357.

[4] J. Rexford, F. Bonomi, A. Greenberg, and A. Wong: Scalable Architectures of Integrated Traffic Shaping and Link Scheduling in High-Speed ATM Switches, IEEE Journal On Selected Areas in Communications, vol 15, issue 5, pp. 938–950, 1997.

[5] K. E. Crawford: "Method of controlling call traffic in a communication system", U.S. patent, no. 4224479, 1980.

[6] L. Takács: "Introduction to the theory of queues", Oxford University Press, New York, 1962, pp. 186–188.

[7] M. A. Gilfix: Method for Distributed Hierarchical Admission Control Across a Cluster, United States Patent, no. 0008090, 2006.

[8] P. P. Tang, Tsung-Yuan C. Tai: Network traffic characterization using token bucket model, INFOCOM'99 conference proceedings, vol 1. pp. 51–62

[9] Y. Ogata: "The asymptotic behavior of maximum likelihood estimators for stationary point processes", Annals of Institute of Satatistics and Mathematics, 30 (1978), Part A, pp. 243–261

[10] B. Kovács: "Mathematical remarks on Token Bucket", IEEE Conference on Software Telecommunications and Computer Networks, 24-26 Sept. 2009, pp. 151-155.

[11] M. Whitehead: "GOCAP – one standardised overload control for next generation networks", BT Technology Journal, Volume 23, Issue 1 (January 2005), pp. 147–153

[12] H.248 v2 protocol specification: *http://tools.ietf.org/html/draft-ietf-megaco-h248v2-04*, Last Access: 19 Mar. 2011.

[13] H.248.11 extension specification: ITU-T recommendation H.248.11

[14] "verload control in a quality-of-service aware telecommunications network", European Patent Office, The Hague, no. PCT/EP2008/059693, 2008, WO/2010/009764, App. no.: PCT/EP2008/059693, Publication: 28.01.2010, filing date: 24.07.2008

[15] *Mathematica*, Wolfram Research Inc. *http://www.wolfram.com*, Last Access: 19 Mar. 2011

 ISBN: 978-1-61208-133-5

From IPv4 to IPv6 – Data Security in the Transition Phase

Tomasz Bilski
Poznan University of Technology
Poznan, Poland
Tomasz.Bilski@put.poznan.pl

***Abstract*—IPv4 is a foundation of Internet communication. Designed many years ago the protocol is inadequate in modern networks. New sixth version is replacing the older one. It is often repeated that IPv6 was designated to solve some security problems. This statement is true only to some extent. IPv6 deployment for the future infrastructure IPv6DFI (especially in the transition phase) will have large impact, not always positive on many aspects of Internet services: network performance, data security, economy. As number of IPv6 networks grow, new threat awareness and understanding become more important. The paper attempts to present comprehensive survey on IPv6 security and to identify many issues of data security in the transition from IPv4 to IPv6 phase.**

Keywords: IPv4; IPv6; Internet security.

I. INTRODUCTION

Internet evolution from IPv4 to IPv6 is the biggest transformation in Internet infrastructure since its beginning. The process is (and will be for many years) very complex and resources (human, money) consuming. It must be expected that the transformation will have huge impact on many aspects of Internet services: network performance, data security, economy. In general, security issues in IPv6 are not better or worse than in IPv4, they are just different. There are risks related to all security features: confidentiality, integrity and availability. For many years we will live in a dual IPv4/IPv6 environment. The security issues could become complex to deal with in terms of implementation and configuration. In dual-stack architecture used in transition phase the problems resulting from IPv6 introduction may have unforeseen effects on IPv4 processing, affecting not only new services but also old services (based on IPv4).

The IP transition phase is an important research area of many teams (e.g., 6net [1], IPv6fix [2] in Japan, USGv6 [3] in the USA). There are some resources on different aspects of IPv6 security (e.g., [4] [5] [6]). Complete list of new threats and risks related to IPv6 is very long and it is very probable that we do not know all threats and risks. In the paper we try to present comprehensive survey on IP security issues with emphasis on the security in IPv4 to IPv6 transition phase.

Section II presents some remarks on solutions to the transition period problems. Main part of the paper (Section III) is dedicated to several issues related to security of IPv6 deployment and transition phase. General conclusions are given in the last part of the paper.

The research project is scheduled for years 2010-2013 and supported by scientific grant from the Polish Ministry of Education and Science.

II. FROM IPV4 TO IPV6

IPv6 had been proposed at IETF as the next generation of IP at early in the 1990's. IP transformation is long-term and complex process. Changes in software (operating systems, applications) as well as in hardware are needed on different TCP/IP layers. The process is not limited to IP version exchange. Many supplementary protocols, such as ICMP (Internet Control Message Protocol), DNS (Domain Name System), BGP (Border Gateway Protocol), OSPF (Open Shortest Path First), RIP (Routing Information Protocol) need to be modified or upgraded.

Unfortunately there is no single, commonly adopted solution to the IP transformation problem. We could not switch from IPv4 to IPv6 in one single point in time. We have to maintain existing IPv4 networks and slowly introduce IPv6 networks. So the two networks have to coexist and cooperate for most likely long time. It is expected that the phase will last for many years. Some reports (e.g., [7]) suggest the phase will reach 2025 year, with cost estimated in USA at $25 billion. And even in 2025 IPv6 global penetration will not achieve 100%. We have just started and so far some plans of IPv6 adoptions did not materialize. For example, according to Action Plan, published in 2008 for EU, Europe should widely implement IPv6 by 2010. It was predicted that by 2010 at least 25% of users will be able to connect to the IPv6 Internet and to access their most important content and service providers without noticing a major difference compared to IPv4 [8]. RIPE Survey [9] from mid 2009 leads to conclusion that achieving this 25% market penetration will be very difficult. There are many obstacles: cost, IT staff preparation (and also some form of human inertia), software implementations availability and maturity. The dangers of transition phase arise from many general reasons among them: a lack of information and a lack of documented operational experience, on which network administrators can draw. One of the important concerns in adopting IPv6 is security. Operators and enterprises are reluctant to deploy technology that may compromise security and eventually cause significant financial loss.

Most transition alternatives are a combination of dual-stack or dual-layer environments and packet tunnelling. Dual-stack means that both versions are running on the

 ISBN: 978-1-61208-133-5

network device [10]. IPv4 and IPv6 datagrams are transmitted over the same network links. Operating system and application decide, which version is used. Tunnelling means IPv6 datagrams are encapsulated in data field of IPv4 datagram or vice versa. Such tunnels may be configured automatically or manually between hosts, routers and networks. The transformation started, new IPv6 networks are working. It may be assumed that some users do not know their systems use IPv6 (e.g., MS Windows Vista default configuration has IPv6 turned on [11]).

III. SECURITY ISSUES

There are many security issues related to IPv6 deployment and transition phase. The security issues related to IP transformation phase may be divided into 2 groups. First group of security problems is associated with IPv6 features and implementations, such as cryptography tools, addressing scheme, security model, host mobility, software bugs. The second group is related to particular methods for IPv4/IPv6 cooperation. Security (especially resources availability) in a broad sense is affected by performance. Poor network performance could lead to availability threats. In order to boost performance security regulations may be loosen by administrators.

A. General Remarks

IPv6 was designed to improve data security. It introduced obligatory implementation (but its use is not required) of IPsec security mechanisms: AH (Authentication Header), ESP (Encapsulating Security Payload), IKE (Internet Key Exchange) Protocol. In general this makes protection (for higher layer protocols and applications) easier and more cost effective. The mechanisms are used to satisfy the requirements of access control, connectionless integrity, data origin authentication, confidentiality and protection against replay attacks [12]. It must be added that not all network devices are equipped with IPsec. For example, some printers, faxes, scanners do not use IPsec.

Furthermore such IPv6 features of the protocol as simplified header, greater number of available addresses have also impact on security. Simplified header makes routers more resistive to DoS attacks – packets are processed more rapidly. Greater number of addresses makes exhaustive host scanning (reconnaissance attack) of a typical /64 subnet unpractical.

IPsec is available for IPv4 but only as an option. Furthermore, IPsec means whole datagram protection so it is not attuned to work with NAT (Network Address Translation) used concurrently with IPv4. IPsec protects the whole datagram, so any modification of the header (NAT modifies addresses and port numbers) violates the security of the datagram. In the transition phase NAT is still used in some dual-stack solutions (e.g., [13]). Only in full IPv6 deployment phase NAT is not needed and full end-to-end IP security is deployable without those issues.

Common methods for preventing unwanted traffic from the Internet are firewalls and IPSes (Intrusion Prevention System). The security is based on the assumption that all the traffic is inspected at the edge of protected network. The problem is not all of the security devices and software are currently IPv6-capable (e.g., IPS may detect the traffic associated with common attacks and malicious behavior for IPv4 and at the same time might not be able to detect similar traffic when it is sent over IPv6) – in the consequence IPv6 may be used as a backdoor to the protected network.

IP transformation phase takes place at the moment of very dynamic Internet growth. Some forecasts estimate IP traffic will increase at a compound annual growth rate of 40% in 2008-2013 [14]. As a result current firewall systems that perform security screening through a common checkpoint will be increasingly degraded due to increasing number of datagrams to process.

IPv6 was designed to protect data. Unreasonably, its deployment may sometimes lead to decreasing security level, especially in the transformation period. The operating system vendors long ago started to support IPv6 in their products. Nevertheless interoperability and compatibility tests (e.g., [15]) of IPv6 implementations show some implementation problems. The problem occurs in network security systems that deal only with IPv4 datagrams. Operating system implementing both IP versions may use IPv6 without user explicit configuration – IPv6 datagrams are not screened and the protocol may be used to form a backdoor [6]. For example, Teredo, an IPv6 tunnelling tool developed by Microsoft is enabled by default in the Microsoft Windows Vista.

B. Cryptography Strength

IPsec uses various cryptographic techniques and tools: symmetric and asymmetric encryption algorithms, hash functions, pseudorandom number generators, key exchange protocols. For a given transmission process cryptography tools are negotiated with a use of IKE. It is assumed that the ultimate set of optionally available tools is changing. At the same time, in order to ensure interoperability, all IPv6 devices (and IPv4 with IPsec implementations) are required to employ some mandatory algorithms. The first problem is the requirements should be met in all devices, without regard to their processing power. This leads to some compromises in the algorithms strength. As for now according to [16] set of mandatory algorithms contains: AES-CBC with 128-bit keys, Triple DES-CBC and HMAC-SHA1-96.

Another problem is the existing algorithms and their implementations are continuously attacked (and sometimes broken) and will be attacked and broken in the future. This is more probable due to long period of IPv6 deployment. Strong algorithm may become weak. Well known examples are DES and MD5, included in the IPsec mandatory list in 1998 [17] but removed from the list in 2005 [18]. SHA1 hash function, which is mandatory at this moment is also known to have some weaknesses [19] and is a possible candidate for removal from the mandatory list.

Prospective replacement of the broken algorithms in all network devices will be very painful and resource consuming task. In some cases (e.g., devices with hardware implementations of cryptographic mechanisms) the task may be impossible to complete. It may be assumed that, in

 ISBN: 978-1-61208-133-5

outcome older, broken cryptography will be used for data protection.

C. End-to-End Security Model

There are two fundamental security models for communication protection: end-to-end and network-based. In end-to-end security model the end hosts provide the security services necessary to protect transmitted data. This model is used in banking applications. Network-based security refers to the practice of hardening the elements of a network to protect other devices. Both models have advantages and weaknesses. The two models may be integrated. Hybrid solutions can be used in the transition period.

IPv6 is integrated with IPsec transport mode dedicated to end-to-end security model. Switching on IPsec does not solve all security problems. First of all, it may not be assumed that communication endpoints can be trusted. Internal threats to data security occur more often than external. If an endpoint is not trusted then entire end-to-end security system could not be trusted to [20].

Additional weakness of the solution is caused by the fact that data are secured at the source and devices located inside the communication channel (gateways, firewalls, …) are not able to scrupulously analyze the traffic. For example, if inbound datagram is encrypted with ESP then it is possible to check IP address in a header but it is not possible to check if data field contains malicious load. For a firewall it is not possible to provide DoS (Denial of Service) prevention based on the expected TCP protocol behaviour – TCP segments inside IP data fields are encrypted so firewall could not check for example, the settings of particular TCP flags. In the same way outbound transmission analysis is also affected. DLP (Data Leakage Protection) is network sniffer, installed on gateway, looking for outbound transmission with predefined sensitive data that should not be transferred outside protected zone. DLP is called also Extrusion Prevention System. DLP systems could not perform their functions since they could not identify sensitive data in encrypted outbound datagrams.

There is a problem with fragmentation – since datagram defragmentation may be done only at the destination host firewall is not able to reassemble the datagram and eventually sanitize it.

Another problem is related to users' negligence. In the case of end-to-end security, users (not administrators) are responsible for data security. If users do not understand the security mechanisms they will not use them appropriately. For example, web browsers may display a warning about invalid server certificates, but users can override the warning and still make vulnerable connection.

A solution to the problem is proposed in the form of distributed firewall [21] [22]. The architecture consists of two firewalls: one at the network perimeter and the other integrated with end host. Firewall at the network perimeter performs general datagram filtering (e.g., based on source IP address) while host-based firewall inspects datagram more precisely. Only the end-host is able to decrypt datagram in order to check it thoroughly. The solution has negative impact on performance.

Additional problems are related to NAT and Quality of Service enforcement. Full end-to-end security is possible if NAT is not used, but some dual-stack solutions to the transformation phase use NAT for address conversion. Furthermore end-to-end security model makes QoS policy enforcement impossible.

D. ICMPv6 Issues

For a new version of IP a new version of ICMP is needed. The old version of ICMP dedicated to IPv4 may not be used with IPv6. ICMPv6 is more functional than its predecessor. It also replaces ARP (Address Resolution Protocol) – NDP (Neighbor Discovery Protocol) is a part of ICMPv6. Without secure network configuration ICMPv6 may lead to many new threats, for example, covert channel, NDP cache poisoning (similar to ARP cache poisoning) or DoS.

The functions added to ICMPv6 are the source of problems for firewall configuration. ICMP messages (used with IPv4) could be dropped by firewall without disturbing normal network operation. In the case of ICMPv6 firewall needs to allow some messages through the firewall and also messages to and from the firewall. Without the authorizations IPv6 procedures (e.g., neighbor discovery or stateless address configuration) could not work properly. As a result covert channel between firewall protected LAN and intruder may be established – for example, malicious IPv4 datagrams (in normal case rejected by firewall) could be hidden inside ICMPv6 messages (which are not rejected by firewall).

NDP is a new function added to TCP/IP stack for host IPv6 address autoconfiguration and address resolution. It has been shown [23] that NDP messages may be used to execute an attack on router resulting in network congestion and degradation of QoS.

Another DoS attack may be launched with a use of multicast transmission. IPv6 specifications forbid the generation of ICMPv6 datagrams in response to messages to global multicast addresses. But there are two exceptions (the *datagram too big* message and the *parameter problem* message). The attack uses these exceptions – error messages are returned as the responses when unprocessable (e.g., greater than Maximum Transmission Unit) datagrams are sent to multicast addresses. If datagram source address is spoofed (replaced with victim address) then many datagrams from multicast group are sent to a victim [5].

E. Host Authentication vs. User Anonymity

IPv6 addressing scheme is very complicated. There are many different types of addresses and many methods for address generation. For ease of configuration a 64-bit part identifying a particular host in the network may come from the interface identifier (e.g., Ethernet MAC address extended to 64 EUI (Extended Unique Identifier)). This ease of configuration leads to privacy problems. All communications of the given user can be linked together using constant interface identifier very easily. On the margin it may be added that IP addresses attributed to Internet users are personal data and are protected by EU Directives 95/46 and

 ISBN: 978-1-61208-133-5

97/66 [24]. In order to prevent such threats to privacy another method (pseudorandom) for address generation was established.

Among various methods for address acquirement we have stateless address autoconfiguration with pseudorandom host identifier selection [25]. The purpose is to change the interface identifier (and public address) from time to time. This way it is much harder for any eavesdropper to correlate Internet transactions to a specific network subject and the user anonymity is better protected. It must be added that a global routing prefix (usually /48), added to pseudorandom host identifier, is fixed for hosts in a given network and some privacy concerns remain.

Full anonymity protection is almost impossible. An attacker, who is on path, may be able to draw some conclusions with a use of: the payload contents of the transmitted datagrams and the characteristics of the datagrams such as datagram size and timing. Use of pseudorandom addresses will not prevent such payload-based correlation.

The same change in the address used for privacy protection could make it harder for a security administrator to define an address-based firewall policy access rule. Another problem is that such node behaviour with relatively high rate of address changes may be interpreted as DDoS (Distributed DoS), like SYN flood, attack and the transmission from the node may be blocked by firewall [5].

Address autoconfiguration has one more weakness. First 24 bits of MAC (Medium Access Control) addresses used in the process are related to specific vendor equipment. An attacker may scan network in order to discover specific vendor device reachability and facilitating an attack on known, specific for the given device security weaknesses.

Here we see typical development scenario: new feature in the protocol (address autoconfiguration) leads to a new security problem (threat to privacy), solution to the problem (pseudorandom address generation) leads to another difficulty (false DDoS alert).

Privacy is important feature of communication. Nevertheless it must be added that the feature may be used also to hide the source of illegal and nasty activities.

F. Software Bugs in IPv6 Implementations

Software implementing IPv6 is relatively new, less mature and has not been tested thoroughly. So far many bugs have been found in the software developed by all the major vendors of IT. It is very probable that numerous bugs will be found in the future. A question is if the bugs will be patched quickly? If not so, many new attack methods will emerge. Furthermore it may not be excluded that essential security features may be missing from early releases of software.

For example, Symantec discovered IPv6-related flaw in Vista. Fortunately it was patched by Microsoft (MS07-038 security update from July 2007). The flaw was related to Teredo tunnelling interface, which did not properly handled certain traffic, allowing to bypass firewall filtering and to obtain sensitive information with a use of IPv6 transmission. Similar problems occur in other operating systems. For example, bug number 6797796 in Solaris 10 may be used to execute DoS attack [26]. Old JUNOS (Juniper router operating system) versions (before May 2006) had a security bug, which could lead to router crash [27].

G. Mobile IP

Both Mobile IP and the increase of moveable IP devices will mean they will be in uncontrollable networks. In Mobile IP environment mobile host is reachable with a use of host routing protocol – normal route to a host is modified for a given recipient host. The method changes the way a datagram is sent to a host. In the effect it may be expected some forms of attacks will use the feature [28].

In the mobile IP environment new security threats will materialize in networks designated for use by foreign, visiting hosts. Such a network will have to loosen firewall rules. Mobile host needs to transmit some IP and ICMP packets (e.g., binding updates, datagrams with optional routing headers) necessary to maintain associations with home agent and other hosts. Firewall should be open for these datagrams – obviously this leads to a security risk.

Mobile IPv6 devices are often equipped with scarce resources and have low processing power. The resources and processing power may be not enough to protect the device: to filter incoming datagrams, to automatically update software (especially implementations of cryptography algorithms), to resist DoS attacks. Of course this problem is common and not related to a particular IP version but IPv6 users may mistakenly believe they are better protected.

H. Security of the Interoperability Methods

There are several means to operate in IPv4+IPv6 environment. The methods enable to transfer datagrams between hosts located in two generations of networks: IPv4 and IPv6. For example, datagram between hosts belonging to two separated IPv6 networks (islands) may be transferred (tunnelled) across IPv4 network – IPv6 datagram is encapsulated in data field of IPv4 datagram. In the future, as IPv6 deployment will spread and IPv4 use will diminish, the roles of IP versions will change. IPv4 datagrams will be encapsulated in data fields of IPv6 datagrams – new types of problems will emerge.

Some security problems are independent of tunnelling method, others are related to a particular tunnelling procedure. In the case of tunnelled datagrams devices enforcing security may inspect only the outer layer of the datagrams, which may be prepared by intruder to avoid filtering while malicious contents of the datagram remain unnoticed. If such datagrams reach a tunnel end-point inside the protected network they are decapsulated and from there can potentially be very harmful since within a network itself, defence levels are usually much lower.

It is obvious that in the case of tunnelling unencrypted IPv6 datagrams in IPv4 network all IPv4 security concerns influence data security. And this problem is regardless of tunnelling method.

There are many tunnelling/interoperability methods, for example: 6to4, Teredo, ISATAP and tunnel broker. Each method has individual impact on data security.

 ISBN: 978-1-61208-133-5

1) Teredo

Teredo uses UDP to tunnel IPv6 datagrams through IPv4 network. IPv6 datagrams are put into UDP segments, which are sent to the destination system via IPv4. Teredo requires a lot of datagram-sanity checks, which can prevent a number of attacks. The program also includes some decent anti-spoofing mechanisms. Nevertheless Teredo tunnelling may lead to new threats.

In Teredo architecture encapsulation/decapsulation is performed by end host. Any of internal (inside LAN protected by firewall) network's Teredo-enabled systems that can receive UDP datagrams can then act as an endpoint for IPv6 tunnels. It is much difficult to secure all such endpoints instead of a single firewall at the network boundary. In the case of a single firewall it is relatively easy for network administrator to control the traffic. But if malicious datagrams are hidden in Teredo tunnel then firewall is not able to discern them and block. Of course firewall could entirely block Teredo traffic (UDP predefined destination port 3544) but attack may be carried with a use of another UDP port. An attacker can send arbitrary IPv6 datagrams to a Teredo-enabled machine inside LAN. The machine may route the datagrams (with source routing mechanism of IPv6) to other host [29]. The problem is particularly related to MS Windows Vista, which in default configuration has both IPv6 and Teredo turned on [11].

2) 6to4

6to4 dual-stack is used to connect IPv6 networks across an IPv4 network. Unique 2002::/16 prefix is reserved for 6to4 systems. Network address with 2002::/prefix has IPv4 address 32 bits embedded immediately after the prefix. The IPv4 address indicates 6to4 router located between IPv6 and IPv4 network. All nodes inside IPv6 network have addresses with 48 bits prefix (2002 and 32 bits of IPv4 router address). If a 6to4 router receives an IPv6 datagram with 2002::/16 prefix then it sends it through IPv4 network, inside IPv4 datagram with IPv4 receiver address taken from 32 next (after 16 bits prefix) bits of IPv6 address.

It is assumed that IPv4 traffic from every address is accepted and decapsulated by 6to4 routers. The routers can be tricked to send spoofed datagrams anywhere. Anyone can send tunnelled spoofed traffic to a 6to4 router, and the router will believe that it is coming from a legal relay. There is no simple way to prevent such attacks, and longer-term solutions are needed in both IPv6 and IPv4 networks [30].

In addition it is suggested that 6to4 routers can be abused to carry DoS attack [31].

3) ISATAP

ISATAP is Intra-site Automatic Tunnel Addressing Protocol. ISATAP uses unusual form of IPv6 addresses. Address is made of 64-bit network prefix and interface identifier. Network prefix is received from ISATAP router while interface identifier contains an embedded IPv4 address (last 32 bits of IPv6 address). The IPv4 address is used in IPv4 headers when IPv6 traffic is tunnelled across an IPv4 network.

Risks related to ISATAP are similar to those related to 6to4.

4) Tunnel broker

Tunnel broker is based on third party servers (tunnel broker servers, tunnel servers) distributed in Internet. Tunnel broker provides tunnels for IPv6 datagrams in IPv4 networks. User/client has to register with the broker system, which sets up a tunnel to one of its tunnel servers. Client gets configuration settings from the broker and uses tunnel servers for communication. The problem is related to the requirement that all traffic passes through third party servers. Service availability as well as confidentiality are threatened. Exemplary DoS attack may be performed by malicious user demanding to establish such number of tunnels that exhausts the resources available in tunnel server [32].

I. Performance Issues

IPv6 was developed to improve network performance. There are many features that aim to fulfill this requirement: simplified header, better addressing scheme, ability to transfer very large datagrams, ban on fragmentation in routers. However, there are some issues of IPv6 with negative impact on performance. This indirectly may influence data security.

IPv6 allows transferring very large datagrams (up to gigabytes). On the other hand in the case of IPv6 tunneling in IPv4 network, the IPv6 path MTU for the destination is typically 20 bytes less than the IPv4 path MTU for the destination. IPv6 headers have fewer fields but are longer due to longer addresses. Minimum length of IPv4 header is 20 bytes while minimum length of IPv6 header is 40 bytes. This makes additional load to all the nodes (including routers, firewalls, bridges) in the communication channel. Of course these longer IP addresses are transferred not only in IPv6 headers but also in messages of many higher layers protocols (e.g., DNS, ICMP, BGP, OSPF) increasing network load.

In the transition phase routing may be done with two separate protocols and doubling the amount of processing in routers. This may lead to router's CPU overload and increase in routes convergence time.

It is obvious that the longer IP header the more time-consuming packet filtering process in firewalls. The question is if this will not force network administrator to switching off filtering in order to boost performance?

Dual-stack systems use tunneling for datagram transmission in heterogeneous environment. Datagram processing on hosts or routers sitting on tunnel ends adds extra time to total datagram delays. In the case of tunnel broker solution further delay is related to datagram transmission from source host to tunnel server, which may be topologically remote. For example, if hosts in Europe use American tunnel broker then transmission parameters (Round Trip Time, jitter, throughput, packet loss ratio) between two IPv6 hosts in Europe will be downgraded by intercontinental links.

 ISBN: 978-1-61208-133-5

Another set of problems is related to DNS. A performance problem is coupled with fallback process. In the IP transformation phase name servers will store two types of address resource records: A for IPv4 and AAAA for IPv6. It is assumed that no explicit information on address preference will be given to a client. The application may receive both IPv4 and IPv6 addresses for the same domain name. The application will have to try respectively addresses received from name server until the connection is successful. It is assumed that IPv6 addresses will be used initially. But if end system has no global IPv6 connectivity then the attempt to connect will be unsuccessful and host will switch to IPv4 address (this is known as fallback process). Due to TCP characteristics the fallback process may last up to about 190 s [33] – the effects on service access time are obvious. Increased number of AAAA queries sent to name servers is another source of performance deterioration. From IP transition point of view DNS is extraordinary service. Name servers and resolvers should be the first to be fully dual-stack capable.

In dual-stack architecture the IPv6 datagrams performance deterioration resulting from IPv6 processing may potentially have harmful unforeseen effects on IPv4 processing, affecting availability of services based on both protocols.

IPv6 allows to classify datagrams in order to diversify their processing by network devices. Some users (legal as well as hackers) may abuse the function and wrongly classify all sent datagrams as highest-priority. To enforce appropriate QoS policy the network device (e.g., gateway) needs to inspect headers and data fields of the datagrams. If the datagrams are encrypted such inspection (and QoS policy enforcement) will be impossible.

IV. CONCLUSION

Some general conclusions may be drown from IP evolution. The change is rather inevitable. New functions of IPv6 and ICMPv6 lead to new threats. IP transition period has (and will have for many years) great impact on Internet security, performance and economy.

Since all popular tunnelling methods (Teredo, 6to4, ISATAP, tunnel broker) use IPv4 networks, the security concerns related to IPv4 are still relevant. In popular dual-stack architecture the problems resulting from IPv6 introduction may potentially have unforeseen effects on IPv4 processing, affecting both services. There are many security issues related to IPv6 deployment. Complete list of new threats and risks related to IPv6 is very long. It is probable that the list will grow longer in the future. In general the security issues related to IP transition phase may be divided into 3 classes:

- related to IPv6 internal features,
- related to IPv6 implementations,
- related to IPv4 to IPv6 transition mechanisms.

A variety of risks and threats are results of the problems. In the previous sections we have described examples of threats from several categories:

- DoS attacks,
- covert channels through firewalls,
- privacy problems,
- extra complexity of management/security tasks,
- bugs in immature software,
- performance deterioration.

In the time of full IPv6 deployment IPv6 will be more than 30 years old. It is very unlikely that the protocol will be appropriate for Internet in for example, years 2020-2030.

Finally, it must be said that many attacks are targeted at the application layer. Since the attacks are unrelated to a particular IP version IPv6 deployment will not change the security level of the application layer.

REFERENCES

[1] 6net Large-Scale International IPv6 Pilot Network, 6NET Consortium, 2008, http://www.6net.org, (last access 7.03.2011).

[2] IPv6 Fix Official Homepage, WIDE Project, 2007, http://v6fix.net/index.html, (last access 7.03.2011).

[3] USGv6 Testing Program, NIST, 2010, http://www.antd.nist.gov/usgv6/testing.html, (last access 7.03.2011).

[4] S. Convery and D. Miller, IPv6 and IPv4 Threat Comparison and Best Practice Evaluation, 2004, Cisco Systems, http://www.cisco.com/security_services/ciag/documents/v6-v4-threats.pdf, (last access 30.11.2010).

[5] E. Davies, S. Krishnan, and P. Savola, IPv6 Transition/Coexistence Security Considerations, RFC 4942, IETF, 2007.

[6] S. Hogg and E. Vyncke, IPv6 Security, Addison Wesley, 2008.

[7] M.P. Gallaher and B. Rowe, IPv6 Economic Impact Assessment, NIST, October, 2005.

[8] Advancing the Internet. Action Plan for the deployment of Internet Protocol version 6 (IPv6) in Europe, Commission of the European Communities, Brussels, 2008, http://www.ipv6.eu/admin/bildbank/uploads/Documents/Commision/COM_.pdf, (last access 30.11.2010).

[9] M. Botterman, Towards IPv6 Deployment, RIPE 59 Lisbon, 2009, http://ripe59.ripe.net/presentations/botterman-towards-v6-deployment.pdf, (last access 7.03.2011).

[10] E. Nordmark, Basic Transition Mechanisms for IPv6 Hosts and Routers, RFC 4213, IETF, 2005.

[11] J. Hoagland, The Teredo Protocol: Tunneling Past Network Security and Other Security Implications, Symantec, http://www.symantec.com/avcenter/reference/Teredo_Security.pdf, 2007 (last acess 30.11.2010).

[12] E. Kent et al., Security Architecture for the Internet Protocol, RFC 4301, IETF, December 2005.

[13] A. Durand, Dual-stack lite broadband deployments post IPv4 exhaustion, Internet-draft, IETF, 2009.

[14] Cisco Visual Networking Index: Forecast and Methodology, 2008-2013, Cisco, San Jose, 2009.

[15] TAHI Project. Test and Verification for IPv6, http://www.tahi.org, (last access 7.03.2011).

[16] V. Manral, Cryptographic Algorithm Implementation Requirements for Encapsulating Security Payload (ESP) and Authentication Header (AH), RFC 4835, IETF, 2007.

[17] S. Kent and R. Atkinson, IP Encapsulating Security Payload (ESP), RFC 2406, IETF, 1998.

[18] D. Eastlake, Cryptographic Algorithm Implementation Requirements for Encapsulating Security Payload (ESP) and Authentication Header (AH), RFC 4305, IETF, 2005.

[19] V. Rijmen and E. Oswald, Update on SHA-1, Cryptology ePrint Archive Report 2005/010, http://eprint.iacr.org/2005/010, 2005 (last access 30.11.2010).

 ISBN: 978-1-61208-133-5

[20] M.H. Behringer, Why End-to-End Security is Necessary But Not Sufficient, Internet Protocol Journal, Cisco vol. 12. No. 2, Sep. 2009, pp. 20-26.

[21] S. Ioannidis, A. Keromytis, S. Bellovin, and J. Smith, Implementing a Distributed Firewall, Proceedings of Computer and Communications Security (CCS), November 2000.

[22] M. Kaeo, IPv6 Security Technology Paper, North American IPv6 Task Force (NAv6TF) Technology Report, NAv6TF, http://www.ipv6forum.com/dl/white/NAv6TF_Security_Report.pdf, 2006 (last access 30.11.2010).

[23] G. An and J. Nah, Effective Control of Abnormal Neighbor Discovery Congestion on IPv6 Local Area Network, LNCS, Volume 4159/2006, Springer Berlin/Heidelberg, 2006, pp. 966-976.

[24] Opinion 2/2002 on the use of unique identifiers in telecommunication terminal equipments: the example of IPv6, Data Protection Working Party, European Commission, http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2002/wp58_en.pdf, 2002 (last access 30.11.2010).

[25] T. Narten, R. Draves and S. Krishnan, Privacy Extensions for Stateless Address Autoconfiguration in IPv6, RFC 4941, IETF, 2007.

[26] Bug ID 6797926, Oracle Corporation, 2011, http://bugs.opensolaris.org/bugdatabase/view_bug.do?bug_id=6797926, (last access 7.03.2011).

[27] Juniper Security Advisory, UNIRAS, 2006, http://archive.cert.uni-stuttgart.de/uniras/2006/07/msg00013.html, (last access 7.03.2011).

[28] A. Mankin, Threat Models introduced by Mobile IPv6 and Requirements for Security in Mobile IPv6, Internet-Draft draft-team-mobileip-mipv6-sec-reqts-00, July 2001.

[29] P. Savola, Security of IPv6 Routing Header and Home Address Options, Internet draft, IETF, 2001.

[30] F. Ali, IP spoofing, Internet Protocol Journal, Cisco, volume 10, no 4, Dec. 2007, pp. 2-9.

[31] P. Savola and C. Patel, Security Considerations for 6to4, RFC 3964, IETF, 2004.

[32] A. Durand, P. Fasano, I. Guardini, and D. Lento, IPv6 Tunnel Broker, RFC 3053, IETF, 2001.

[33] T. Fujisaki et. al., Operational Problems in IPv6: Fallback and DNS issues, http://www.nttv6.net/~fujisaki/fallback.pdf, 2006 (last access 30.11.2010).

 ISBN: 978-1-61208-133-5

An Optimized Port Allocation Mechanism in the Context of A+P for Public IPv4 Address Sharing

Xiaohong Deng, Lan Wang, Daqing Gu
Orange Labs Beijing
France Telecom Group, Beijing, China
{xiaohong.deng; lan.wang; daqing.gu}@orange-ftgroup.com

***Abstract*—the IANA free pool of IPv4 addresses will be exhausted soon, how to use scarce IPv4 public addresses more efficiently while migrating to IPv6 is a challenge. A+P is recommended as a complementary method to Dual-stack Lite which aims at address public IPv4 address sharing problem in the context of IPv6 migration. Since A+P suffers from inflexible port allocation, this paper introduces an optimized A+P port allocation mechanism which allows customers negotiate IP-addresses of desired sharing ratios on their requirement. Moreover it enables A+P NAT using random source port selection algorithm which significantly improves security by preventing attacker's easy guessing the five-tuple. The test result shows that this mechanism enables great randomness of source ports selection behavior on A+P NAT.**

Keywords—IPv6 migration; Dual-stack Lite; A+P; Port randomization.

I. INTRODUCTION

The IANA pool for global public IPv4 address allocation is forecasted to exhaust by mid-2011.And IPv4-only legacies are ubiquitous crossing telecom infrastructure. Since IPv6 and IPv4 are incompatible protocols, IPv6 could not replace IPv4 in order to solve the public IPv4 exhaustion problem immediately. Instead, both protocols will co-exist for a long period of time. With public IPv4 address sharing solutions, higher utilization ratio of available public IPv4 addresses can be achieved. These solutions can be deployed before the majority of the Internet becomes IPv6-capable and most communications could be done through IPv6.

A. IPv4 Address Sharing Solutions

Several solutions have been proposed in IETF to address IPv4 address shortage while migrating to IPv6, such as NAT444 [1], A+P [2], Dual-stack Lite [3]. Recently, IETF reached a consensus on Dual-Stack Lite as a promising solution. It provides the broad band service provider a scalable and easy way to introduce IPv6 while keeping IPv4 reachability to its customers. In Dual-Stack Lite, IPv4 addresses among customers are shared using two technologies: IP in IP (IPv4-in-IPv6) and NAT. A+P is similar to Dual-Stack Lite, the difference is that A+P NAT locates at the customer premises, while Dual-Stack Lite NAT locates at the carrier network.

Many applications require ALG to work through a NAT. At the same time, more and more applications expect incoming connections, such as peer-to-peer ones. Making sure those subscriber-provided services working properly in a Dual-stack Lite environment is important. Unfortunately, service providers are not in the position of provisioning such applications and ALGs. In this case, in [3], A+P is recommended as a complementary method to Dual-stack Lite in order to deal with the subscriber-provided services' ALG issues, which would break some subscriber-provided services if ALG issues are not well treated. Reserving certain ports under the control of customers is one way to enable the Customer Premises Equipment (CPE) A+P NAT to process this kind of traffic. Figure 1 shows an example: Ports 5002-5004 are reserved for A+P NAT; Incoming packet that falls into the A+P ports range bypasses the Dual-stack Lite Carrier Grade NAT (CGN), and directly is sent to the tunnel endpoint, an A+P aware CPE, from the Address Family Transition Router (AFTR). CPE then locally NAT the packet to internal hosts, otherwise the incoming packet will traverse the AFTR's NAT.

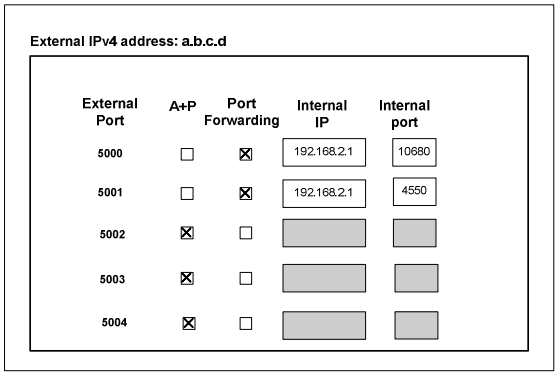


Figure 1. ISP portal address & port control table

B. Motivation and Objectives

Firstly, dynamic port assignment is used in Dual-stack Lite mode to maximize the address sharing ratio. On the other hand,

 ISBN: 978-1-61208-133-5

A+P mode allocates ports in a cookie-cutter fashion, i.e., a range of ports are pre-allocated to each CPE. Concerns have been raised when public IPv4 addresses are shared among a large mount of CPE, while only a limited N number of TCP or UDP port numbers are available per CPE in average. In fact, pre-allocating N ports is not encouraged according to several service providers' report: the average number of connections per customer is the single digit, while thousands or tens of thousands of ports could be used in a peak by any single customer browsing a number of AJAX/Web 2.0 sites. If a smaller number of ports per CPE (N in the hundreds) are allocated, it is expected that customer's applications could be broken in a random way over time. If a large number of ports per CPE are allocated (N in a few thousands), the address sharing ratio will be decreased. Furthermore, customers may require different amount of ports, for example, enterprise customers may expect more ports to support simultaneous sessions; some customers have many terminals connected to CPE which may require more ports.

Secondly, a number of "blind" attacks can be performed against the TCP and similar protocols by identifying the transport protocol instance, i.e., the five-tuple (Protocol, Source Address, Destination Address, Source Port, Port). Since A+P pre-allocates N (N<65,536, usually less than several thousands in order to achieve a large sharing ratio more than a single digit) ports to CPE, it is more easy to guess the five-tuple and in turn increase the probability of successful attacks.

This paper proposes a optimized port allocation mechanism for A+P which aims at addressing those two defects of A+P by two efforts, 1) providing customer oriented differentiated services for A+P to allow customer negotiate IP-addresses of desired sharing ratios based on their requirement; 2) providing a source port randomization algorithm to achieve better security by preventing attacker's easy guessing the five-tuple.

C. Orgnaization

The rest of this paper is organized as follows. Section II introduces related works; Section III presents our proposed optimized port allocation mechanism. The simulation results are discussed in Section IV. Finally, this paper is concluded in Section V.

II. RELATED WORK

Several methods have been proposed for allocating A+P parameters. In [4], DHCPv4 Options for allocating port restricted public IPv4 address and a range of ports are defined. Two IPCP Options, Port Range Value Option and Port Range Mask Option to convey one range of ports (either contiguous or not contiguous) pertaining to a given IP address have been discussed in [5]. These two IPCP Configuration Options provide a way to negotiate the Port Range to be used on the customer Premises. The sender can use the Configure-Request message to carry request which Port Range associated with a given IP address is desired, or to request the peer providing the configuration, the peer then can provide this information by NAKing the option, and returning a valid Port Range associated with an IP address.

Both of these methods propose A+P port allocation mechanism and negotiation process, either by using DHCP semantics or by PPP IPCP negotiation. Neither of them addresses the problem described in Section I.

III. OPTIMIZED A+P PORT ALLOCATION MECHANISM

Firstly, a mechanism was designed to allow service provider provisioning the differentiated qualities of service by allocating different sharing ratio IP-addresses to different customer. As customer gets better service with lower sharing ratio IP-address, the one demanding better service pays more for the lower sharing ratio IP-address. The operator could configure IP-addresses pools with different service levels depends on different sharing ratios. As illustrated in Table I, it shows an example seven service level IP-addresses pools was configured according to seven sharing ratios. With sharing ratios varies from 1 to 64, which means available ports for each customer varies from 65,536 to 64, the service level decrease from level 0 to level 6. Level 0 provides one unshared global IP-address to customer as nowadays operator does. Each customer could request different service level IP-address according to their requirements on quality of service, which depends on the sharing ratio, the lower sharing ratio, the higher quality of service with higher price.

TABLE I. AN EXAMPLE OF SERVICE LEVEL ADDRESS POOLS

Service level address pools	Sharing ratio	Available ports
Level 6 address pool	64	1024
Level 5 address pool	32	2048
Level 4 address pool	16	4096
Level 3 address pool	8	8192
Level 2 address pool	4	16384
Level 1 address pool	2	32768
Level 0 address pool	1	65,536

Secondly, to provide a way for customer to generate random ports while guaranteeing them in customer restricted port range, the core idea is simple: Choosing M bits to from a customer ID bits for a set of customers which sharing the same IP-address, and then identifying the customers in the same set by allocating unique M bits customer ID values. Hence the M bits customer ID could guarantee the ports inside the customer restricted port range. With regard to each shared IP-address, as its sharing ratio is given, the number of customer ID bits M is

 ISBN: 978-1-61208-133-5

decided by $\log_2^{Sharingratio}$. In order to facilitate the reuse of existing Port Randomization algorithms, two parameters are derived, customer ID pattern and customer ID value, customer ID pattern is derived from setting the customer ID bits to '1' and the left bits to '0', customer ID value is derived from setting the Customer ID bits to a unique value allocated from the operator side and the left bits to '1'. An example is shown in Figure 2, a Level 6 address pool of sharing ratio 64 for a sharing IP-address-A is selected, and then 6 bits are chosen as Customer ID bits, which are the 3rd, 4th, 7th, 9th, 10th, 12th bit. All the customers sharing IP-address-A get "0000101101001100" as customer ID pattern, and each of them get a unique customer ID value. As shown in Figure 3, take the customer#5 for example, for a random generated port, just take a bitwise AND operation with customer ID pattern and then take a bitwise OR operation with customer ID value of customer#5, the result port would be inside the customer#5 restricted port range. Hence it's easy for customer NAT reusing the port randomization algorithms referred in [6], it could be done by slight modification to the existing port randomization algorithms. Take the Algorithm 1: Simple port randomization algorithm [6] for example, the Figure 4 shows the A+P simple port randomization algorithm, only one line code was inserted to the original simple port randomization algorithm.

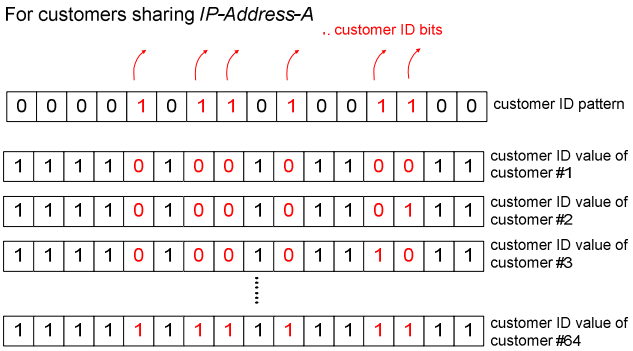


Figure 2. Customer ID pattern and customer ID value

To ensure the customer NAT could generate as random ports as possible to prevent attackers, three heuristic rules to choose M bits of Customer ID are proposed:

1) To avoid allocate a range of continuous ports to customer, the location of M bits for the Customer ID Pattern should not take place in the most significant bits.

2) To avoid allocating only even ports to customers with the least significant bit '0' or only odd ports to customers with the least significant bit '1', M bits Customer ID Pattern should not involve the least significant bit.

3) To avoid allocating a regular range of ports to customer, the location of M bits for the Customer ID Pattern should not take place in the continuous bits.

Pseudocode for Customer ID Pattern bits chosen is shown in the Figure 5.

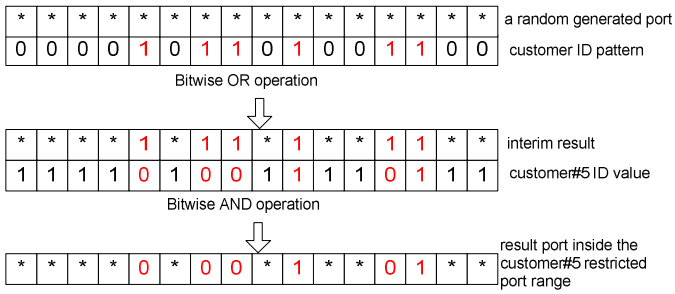


Figure 3. Calculate port inside the customer restricted port range

```
/* A+P Ephemeral port selection function */

num_ephemeral = max_ephemeral - min_ephemeral + 1;
next_ephemeral = min_ephemeral + (random() %
num_ephemeral);
next_ephemeral_in_range = next_ephemeral ||
customer_ID_pattern & customer_ID_value
count = num_ephemeral;
do {
if(five-tuple is unique)
return next_ephemeral_in_range;
if (next_ephemeral == max_ephemeral) {
next_ephemeral = min_ephemeral;
} else {
next_ephemeral++;
}
count--;
} while (count > 0);
```

Figure 4. A+P simple port randomization algorithm

Customer could get IP-address in desired sharing ratio by negotiating Customer ID pattern with operator, the negotiation could be done by several ways, either by negotiating DHCPv4 Option for allocating port restricted public IPv4 address defined in [4], or by negotiating PPP IPCP Options defined in [5] after Softwire [7] is established and IPCP reaches the Opened state. The negotiation process is out of scope.

Note that customers sharing the same IP address share the same customer ID pattern; however customers with different IP addresses, no matter if they are using the same sharing ratio, are using different customer ID patterns which is granted by random customer ID choosing algorithm.

 ISBN: 978-1-61208-133-5

```
/* Heuristic algorithm for Customer ID bits chosen */

M = log2(SHARING_RATIO);
Genbits:
/*Guarantee Heuristic rule 1 and Heuristic rule 2*/
for (i = 0; i < M;)
{
bits[i] = GenRandomBit();
if(bits[i] == 0 || bit[i] = =15); else i++;
}
/*Sort array in decreasing order in order to match the
condition for rule 3*/
Sort(bits);
/*Guarantee Heuristic rule 3*/
for (i = 1; i < M;)
{
if(bits[i-1] = = bits[i]+1) i++;
else break;
}
if(i= =M) goto Genbits;
return bits;
```

Figure 5. Customer ID choosing algorithm

IV. EVALUATION

A. TCP/UDP Source Port Considerations

Since a successful attack against the TCP or UDP requires the attacker to have knowledge of a valid five-tuple (protocol, source IP address, source port, destination IP address, and destination port). The protocol, source and destination IP addresses are obvious as they are the specific services the attacker will be spoofing. The destination port is also obvious, the attacking aims at well-known services, which announce well-known ports to public. The only difficult part of guessing this is the TCP/UDP source port, since it different for each new TCP/UDP session. As for TCP RESET attack, the attacker could assume the destination port of 179 for BGP, as for Domain Name System (DNS) cache poisoning attack, it could assume the destination port of 53 (the port number IANA has assigned for DNS). For an attacker, additional requirement of a correct source port would increase the difficulty of the attack by a factor or 16. Random source ports would increasing the numerical attack space "from 2^{32} to 2^{48}", hence increase the difficulty of an attack. Unfortunately, even if random source ports are supported by implementations, routers or gateway devices that perform network address translation (NAT), often rewrite source ports for tracking NAT session. When some NAT devices modify source ports without random source port selection algorithms, it will increase the risk of successful attacks.

The following section will evaluate port randomization of A+P NAT with or without our proposed mechanism and its impact on the DNS cache poisoning attacks.

B. DNS Cache Poisoning Attacks

DNS servers usually store results in a cache to speed further lookups for efficiency, which makes DNS server vulnerable to DNS cache poisoning attacks. DNS cache poisoning is a maliciously created that provides data to a caching name server that did not originate from authoritative DNS sources. Once a DNS server has received such non-authentic data and caches it, it is considered poisoned. The answers from a poisoned DNS server cannot be trusted. The clients may be redirected to malicious web sites that will try to steal clients' identity or infect clients' computers with malware.

DNS requests contain a 16-bit transaction IDs, used to identify the response associated with a given request. Unless the attacker can successfully predict the value of the transaction IDs and return a reply first, the server won't accept the attacker's response as valid. Even if it may be possible to guess these transaction ID values in advance, but as long as the server randomizes the source port of the request, the attack may become more difficult, since the fake response must be sent to the exactly same port that the request originated from. The essence of the problem is that DNS resolvers don't always use enough randomness in their transaction IDs and query source ports. Increasing the randomness of transaction IDs and query source ports may increase the difficulty of a successful poisoning attack. United States Computer Emergency Readiness Team (US-CERT)'s Vulnerability Note VU#800113 describes deficiencies in the DNS protocol and implementations that can facilitate cache poisoning attacks. Most implementations do NOT randomise the port number. In most cases, the same port number 53 was always used.

As stated above, some DNS implementations use a number of mechanisms to protect themselves from the attacks. First of all, queuing received request eliminates the possibility of birthday attack. Secondly, sending the request from the random dynamic UDP port and accepting only answers to this source port. In consequence the probability of successful attack is significantly decreased because the attacker has to guess two 16-bits numbers (both UDP port number and transaction ID). Because these numbers are independent the success probability would be in this case $P = \frac{1}{2^N}$, where N is in practice near 32 (IANA has reserved 0 through 1023 for The Well Known Ports, not all 65,536 could be Ephemeral Port). Even if the attacker uses a high speed connection, the probability of success is relatively small, because the attacker is not able to generate about 2^{32} fake answers in time when DNS is waiting for the reply from the authoritative DNS.

Therefore, upgrading DNS server and DNS resolver to implementations of good randomness is essential to defence against DNS Cache Poisoning Attack. However when DNS resolver behinds routers, firewalls, or other gateway devices

 ISBN: 978-1-61208-133-5

that perform network/port address translation (NAT), if the NAT device does not implement random ephemeral port selection algorithms, consequently it will remove source port randomness implemented by DNS server and stub resolvers. Experiments have been conducted to evaluate if A+P NAT bring deficiency of port randomness when a DNS resolver behinds a NAT.

C. *Comparison of DNS Port Randomness in Three NAT Scenarios*

There is a web-based DNS randomness test tool on the Domain Name System Operations Analysis and Research Centre (DNS-OARC) [8] to help user estimate if their name servers are vulnerable to DNS Cache Poisoning Attacks. This estimation was based on the randomness score of source port and Transaction ID Randomness of DNS resolver. We use the djbdns [9] which sends requests form various source ports and only accepts answers sent to source port, and put it behind NAT devices to test if there is potential randomness deficiency that A+P NAT may bring in.

Three scenarios are given for comparison: In scenario A, DNS resolver behinds a NAT implementing "simple port randomization" algorithm recommended in [6]; in scenario B, DNS resolver behinds a NAT implementing "A+P simple port randomization algorithm" due to our optimized A+P port allocation; in scenario C, DNS resolver behinds an A+P NAT implementing "port increment by 1" algorithm. The estimation result shows that our "A+P simple port randomization algorithm" inherits greater randomness than the algorithm1 recommended in [6]. As shown in Figure 6 and Figure 7, both of them are evaluated as GREAT source port randomness. On the contrary, the traditional A+P NAT "Increment by 1 algorithm" removed source port randomness implemented by the tested DNS server, and is evaluated as POOR source port randomness as shown in Figure 8.

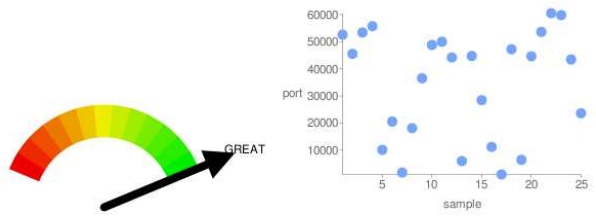


Figure 6. Evaluation DNS randomness in Scenario A

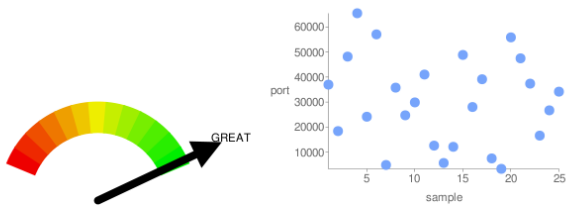


Figure 7. Evaluation DNS randomness in Scenario B

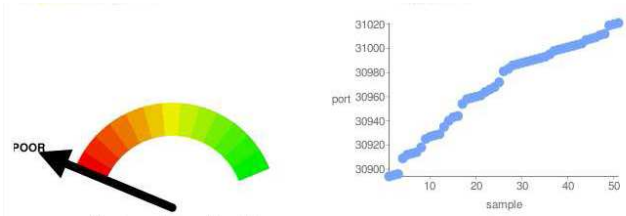


Figure 8. Evaluation DNS randomness in Scenario C

The scatter diagram of 25 sampled ports of scenario A, scenario B and scenario C are shown in Figure 9-11 respectively. The source ports generated by "Simple port randomization NAT" and "A+P simple port randomization NAT" are well distributed, varies from lower bound to upper bound; while "A+P port increment by 1 NAT" generates sequential ports in a very limited range from 30900 to 30925.

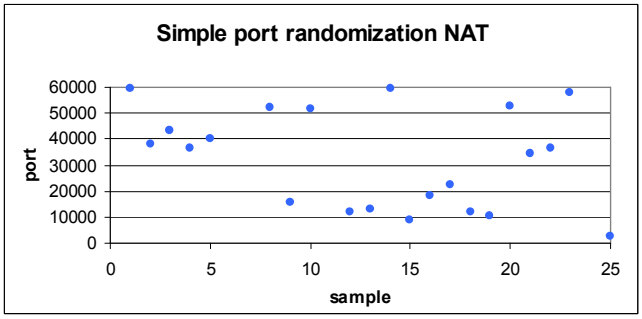


Figure 9. Sampled ports from DNS randomness test in Scenario A

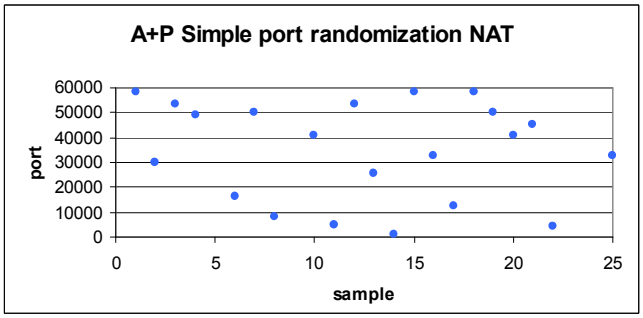


Figure 10. Sampled ports from DNS randomness test in Scenario B

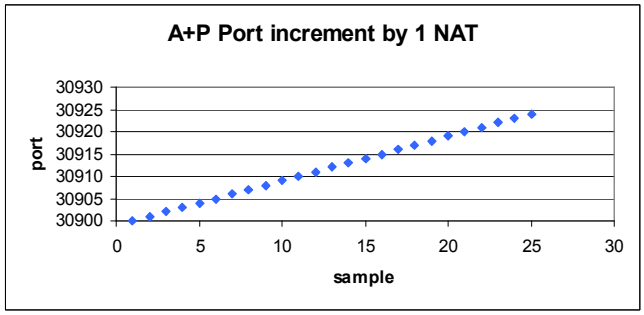


Figure 11. Sampled ports from DNS randomness test in Scenario C

Furthermore, more ports are sampled and Standard Deviation (STDDEV) is used as a measure of randomness as indicated in [8]. The port randomness in scenario A-C are tested. Table II shows the relationship of Port Randomness Score and STDDEV range. In each experiment, the same numbers of ports were sampled for scenario A-C, and the samplings repeat six times. Figure 12-15 shows the STDDEV comparison of the three scenarios when 100, 500, 1000, 5000 ports were sampled respectively, From these four figures, we can see that "Simple port randomization NAT" and "A+P simple port randomization NAT" almost have the same good performance on STEDEV, while "A+P port Increment by 1 NAT" is far underperformance.

TABLE II. PORT RANDOMNESS SCORE

Port Randomness Score	STDDEV Range
GREAT	3980 - 20,000+
GOOD	296 - 3980
POOR	0 - 296

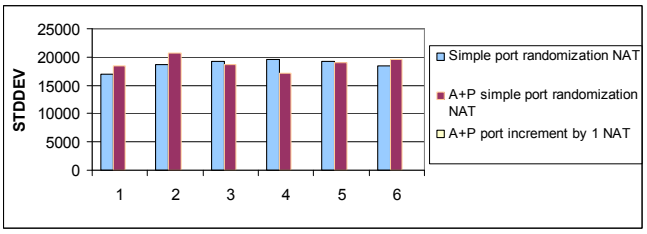


Figure 12. STDDEV Comparison of 100 ports

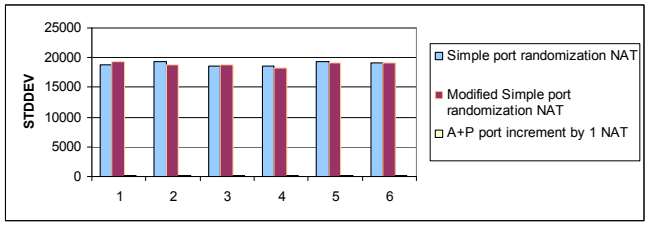


Figure 13. STDDEV Comparison of 500 ports

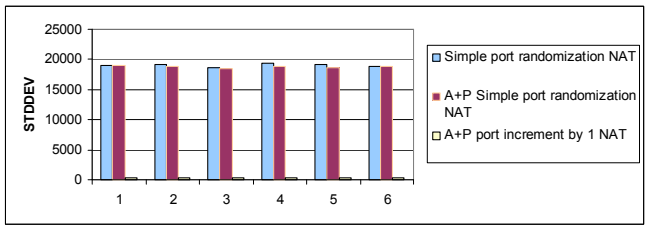


Figure 14. STDDEV Comparison of 1000 ports

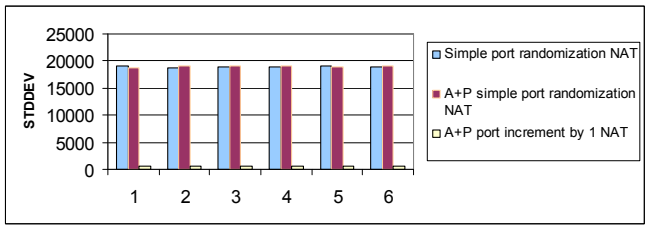


Figure 15. STDDEV Comparison of 2000 ports

V. CONCLUSION

This work introduces an optimized port allocation mechanism for A+P enhancement which has two key features: 1) it provides customer oriented differentiated services by allowing customer negotiates IP-addresses of desired sharing ratios based on their requirement; 2) it supports port randomization. The test result shows that, without our mechanism, "A+P port incremental by 1 NAT" brings randomness deficiency to DNS server and consequently makes DNS server vulnerable to DNS poisoning attacks, while our "A+P simple port randomization NAT" has as great randomness as " simple port randomization NAT " which doesn't bring randomness deficiency to DNS server. Hence this work mitigates A+P's two major defects coming along with allocating ports in cookie-cutter fashion.

REFERENCES

[1] J. Yamaguchi, Ed., "NAT444 addressing models", draft-shirasaki-nat444-isp-shared-addr-03(work in progress), March 8, 2010.

 ISBN: 978-1-61208-133-5

[2] Bush, R., "The A+P Approach to the IPv4 Address Shortage", draft-ymbk-aplusp-04 (work in progress), October 27, 2009.

[3] Durand, A., "Dual-Stack Lite Broadband Deployments Following IPv4 Exhaustion", draft-ietf-softwire-dual-stack-lite-02 (work in progress), March 8, 2010.

[4] Bajko, G. and T. Savolainen, "Port Restricted IP Address Assignment", draft-bajko-v6ops-port-restricted-ipaddr-assign-02 (work in progress), November 2008.

[5] M. Boucadair, Ed, "Port Range Configuration Options for PPP IPCP", draft-boucadair pppext-portrange-option-01(work in progress), July 2009.

[6] M. Larsen, Ed, " Transport Protocol Port Randomization Recommendations", draft-ietf-tsvwg-port-randomization-07 (work in progress), November 30, 2009.

[7] RFC5571, B. Storer., " Softwire Hub and Spoke Deployment Framework with Layer Two Tunneling Protocol Version 2 (L2TPv2)", June 2009.

[8] Web-based DNS Randomness Test, https://www.dns-oarc.net/oarc/services/dnsentropy, March, 2011

[9] Djbdns, http://cr.yp.to/djbdns.html, March, 2011.

 ISBN: 978-1-61208-133-5

Analysis of Security Vulnerability in Cooperative Communication Networks

Ki Hong Kim
The Attached Institute of ETRI
Daejeon, The Republic of Korea
Email: hong0612@ensec.re.kr

***Abstract*—Cooperative communication is a new and emerging wireless communication that exploits spatial diversity to improve wireless channel capacity. Cooperative medium access control (CoopMAC) protocol is a MAC protocol that involves an intermediate relay between a transmitter and a receiver in the cooperative network. In this paper, we identify various attacks against CoopMAC and analyze security vulnerabilities in CoopMAC. From our analytical results, it can be induced that there is a need for an efficient authentication procedure which provides reliability and security for normal CoopMAC communication. To our knowledge, this is the first comprehensive case study of security vulnerabilities caused by possible security attacks in CoopMAC. Our results can be used to design an efficient and secure communication mechanism for cooperative networks.**

***Keywords*-CoopMAC; cooperative communication; security vulnerability; authentication**

I. INTRODUCTION

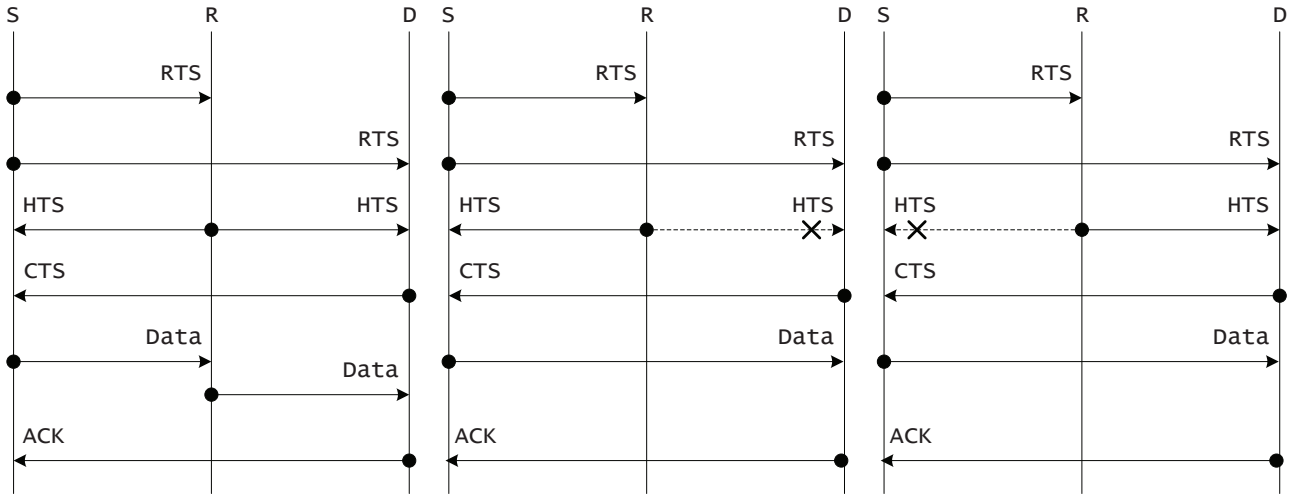
Cooperative communication is indispensable for making ubiquitous communication connectivity a reality. Cooperative communication is an innovative wireless communication mechanism that takes advantages of the open broadcast nature of the wireless communication channel and the spatial diversity to improve channel capacity, robustness, reliability, delay, and coverage. In the cooperative communication network, when the source node transmits data packet to the destination node, some nodes that are close to source node and destination node can serve as relay nodes by forwarding replicas of the source's data packet. Among the forwarding methods employed by the relay nodes, amplify-and-forward (AF), decode-and-forward (DF), and compress-and-forward (CF) are the most common methods. The destination node receives multiple data packet from the source node and the relay nodes and then combines them to improve the communication quality [1][2].

A MAC protocol called CoopMAC is designed to improve the performance of the IEEE 802.11 MAC protocol [3] with minimal modification. It is able to increase the transmission throughput and reduce the average data delay. It also utilizes the multiple transmission rate capability of IEEE 802.11b, 1 to 11Mbps, and allows the source node far away from the access point (AP) to transmit at a higher data rate by using a relay node [4][5].

Although cooperative communication has recently gained momentum in the research community, there has been a great deal of concern about cooperative communication mechanism and its security issues. There have been several previous related works regarding communication techniques and security issues for cooperative network. The work in [1][2] described wireless cooperative communication and presented several signaling schemes for cooperative communication. In [4][5], a new MAC protocol for the 802.11 wireless local area network (WLAN), namely CoopMAC, was proposed and its performance was also analyzed. The potential security issues that may arise in a CoopMAC were studied in [6], and various security issues introduced by cooperating in Synergy MAC were also addressed in [7]. The [8] suggested cross-layer malicious relay tracing method to detect signal garbling and to counter attack of signal garbling by compromised relay nodes, while the [9] presented the distributed trust-assisted cooperative transmission mechanism handle relays' misbehavior as well as channel estimation errors. Also, a performance of cooperative communication in the presence of a semi-malicious relay which does not adhere to strategies of cooperation at all time was analyzed in [10], and a statistical detection scheme to mitigate malicious relay behavior in DF cooperative environment was developed [11]. The examination of the physical consequences of a malicious user which exhibits cooperative behavior in a stochastic process was discussed in [12]. The [13] described a security framework for leveraging the security in cognitive radio cooperative networks. However, most of the works on cooperative communication is focused on efficient and reliable transmission schemes using the relay and identification of general security issues caused by the malicious relay node. No work has been done on the analysis of denial of service (DoS) vulnerability caused by an attacked relay node in cooperative communication environments.

In this paper, a cast study of DoS attack in CoopMAC is presented for the first time. Security vulnerabilities at each protocol stage while attacking a cooperative communication, namely between a source node and a relay node, between a relay node and a destination node, and between a source node and a destination node, is analyzed and compared. This study differs from previous works in that it concentrates on one significant aspect of security vulnerability in the CoopMAC, namely DoS vulnerability of CoopMAC caused by the Dos attack of attacker node. This is believed to be the

 ISBN: 978-1-61208-133-5



(a) Cooperative Tx via R (b) Directive Tx by D's HTS receiving failure (c) Directive Tx by S's HTS receiving failure

Figure 1. Control packet exchange in CoopMAC protocol

first comprehensive analysis and comparison of the security vulnerability from possible DoS attack in CoopMAC. The analytical results can be used to design an efficient and secure communication mechanism for cooperative communication security.

The remainder of this paper is organized as follows. In section II, we describe the characteristics of CoopMAC. Next, in section III, we identify some possible security attacks against CoopMAC and then analyze the security vulnerabilities at each protocol stage of CoopMAC. Finally, in section IV, we review our conclusions and detail plans for future work.

II. Overview of CoopMAC Protocol

CoopMAC is a MAC protocol based on the IEEE 802.11. It employs request-to-send (RTS) and clear-to-send (CTS) control packet to establish communication, which are overheard by other nodes besides the source node and the destination node. The CoopMAC is totally compatible with the legacy 802.11 protocol. It shows a communication throughput increase and also reduces the transmission delay experienced each data packet [4][5].

The exchange of control packets for CoopMAC is shown in Fig. 1. First, source node S senses the communication channel condition, busy or idle. If the channel is idle, source node S sends the RTS packet, reserving the channel for network allocation vector (NAV) duration. If not, source node S should wait the channel is idle and then send the RTS packet. When the relay node R receives the RTS packet and decodes it successfully, it responds with a helper ready-to-send (HTS) packet to the source S and the destination node D. After receiving the RTS packet followed by HTS packet, destination node D sends CTS packet to reserve the channel for cooperative communication via the relay node R. Even if destination node D does not receive the HTS from the relay node R, it sends the CTS packet to the source node S. But in this case, it reserves the channel for direct transmission between the source node S and the destination node D (Fig. 1(b)).

Once source node S receives the HTS packet from the relay node R and the CTS packet from the destination node D respectively, the cooperative transmission between source node S and destination node D via the relay node R starts (Fig. 1(a)). That is, source node S sends the data packet to relay node R and relay node R then forwards the packet received from source node S to destination node D. On the other hand, if source node S has not received the HTS packet from relay node R before the CTS packet is received from destination node D, it transmits the data packet directly to destination node D (Fig. 1(c)). After destination node D successfully receives the data packet from source node S, it sends an acknowledgment (ACK) packet to source node S. Otherwise, destination node D sends a negative acknowledgment (NACK), notifying source node S of the failure of transmission. In addition, if source node S receives no response from destination node D within a specific timeout period, it will also notice the failure of transmission to destination node D. For more complete details of CoopMAC protocol, please refers to [3][4].

III. Security Vulnerabilities in CoopMAC

Due to broadcast nature of the wireless channel and cooperative nature, cooperative communication suffers from various attacks.

For example, in Fig. 2, if source node S want to transmit data packet to destination node D using the relay node R, it first sends out the RTS, and the relay node R then reply with a HTS to source node S and destination node

 ISBN: 978-1-61208-133-5

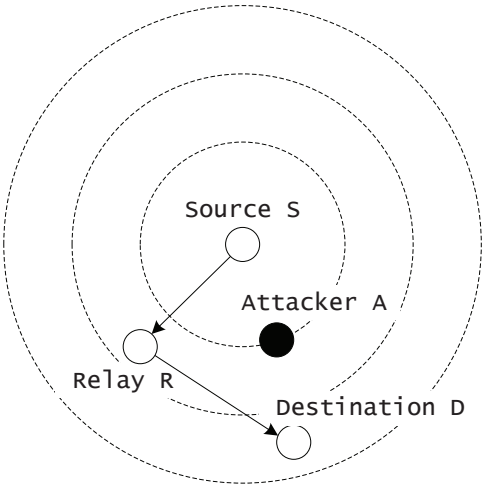


Figure 2. Example of DoS attack by neighboring node in CoopMAC

D. After receiving the RTS and the HTS, the destination node D finally sends a CTS to source node S. However, let's assume attacker node A is much closer to source node S than destination node D or it is between the source node S and the destination node D. In this case, attacker node A disguises itself as destination node D and responds with a CTS to source node S. There is no countermeasure to avoid this attack. That is, currently there is no suitable countermeasure mechanism to prevent a reply attack in the physical connection and authentication mechanism to authenticate destination node D. Therefore, an arbitrary attacker can respond with a CTS to neighboring nodes and thus it results in disruption of the normal cooperative transmission between nodes.

The goal of the attacker node is to obstruct the communication between source node and destination node. These attacker nodes would exploit the weakness in cooperation procedures, especially in the control packet exchange, and disguise themselves as legitimate relays. We will introduce some cases of attacks according to the control packet of CoopMAC next.

A. Attack on RTS Control Packet

In the CoopMAC as shown in Fig. 3(a), attacker node A sends the faked RTS to relay node R and destination node D, and then waits for the HTS from relay node R as well as CTS from destination node D. After the attacker node A receives the HTS and the CTS, it sends a fake data to the relay node R. Consequently, this attack results in a transmission disturbance in the RTS and the data packet from source node S. Accordingly, source node S can not start data transmission to relay node R.

On the other hand, as shown in Fig. 3(b), attacker node A intentionally sends the faked RTS to only destination node D. The legal RTS from source node S can be rejected by destination node D due to an illegal previous RTS received from attacker node A. Hence, CTS is sent from the destination node D to attacker node A, which causes source node S to continuously wait for the CTS from destination node D. As a result, normal cooperative communication between source node S and destination node D can not be guaranteed.

B. Attack on HTS Control Packet

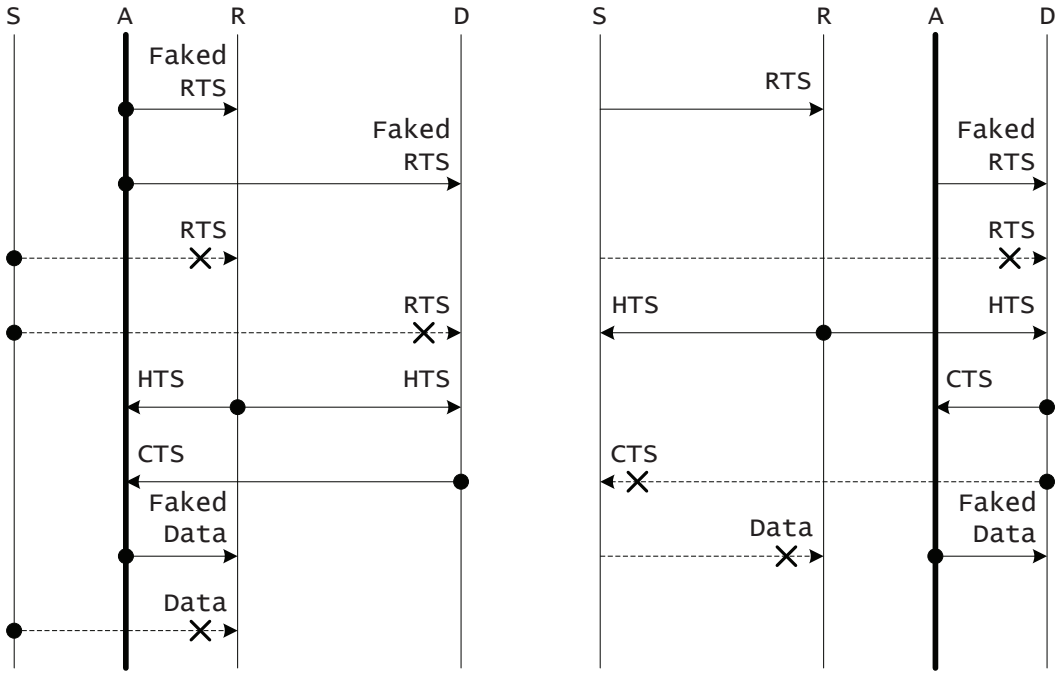
As shown in Fig. 4(a), the faked HTS is sent from attacker node A to source node S and destination node D. Accordingly, the legal HTS from relay node R is denied by source node A and destination node D. Then, destination node D sends CTS to source node A. After receiving the faked HTS and CTS, source node S starts data transmission to attacker node A, but relay node R. Due to this false transmission to the attacker node A, cooperative communication between source node S and destination node D via relay node R is not established.

The potential security vulnerability from faked HTS in the CoopMAC is also shown in Fig. 4(b). In the case of sending faked HTS to only destination node D, since the destination node is typically not come to know of this, although the legal HTS is sent from the relay node R to destination node D, it is denied by destination node D. Then, the destination node D sends a CTS to source node S in order to notify that it successfully receives the control packet. This also means that attacker node A is an intended legitimate relay node forwarding data packet. Therefore, if relay node R receives the data packet from source node S, it doesn't forward data packet to the destination node D, but forwards it the attacker node A. Finally, the attacker node A denies cooperative communication service to the source node S by simply dropping the data packet it receives. It also spoofs an ACK, causing the source node S to wrongly conclude a successful transmission.

C. Attack on CTS Control Packet

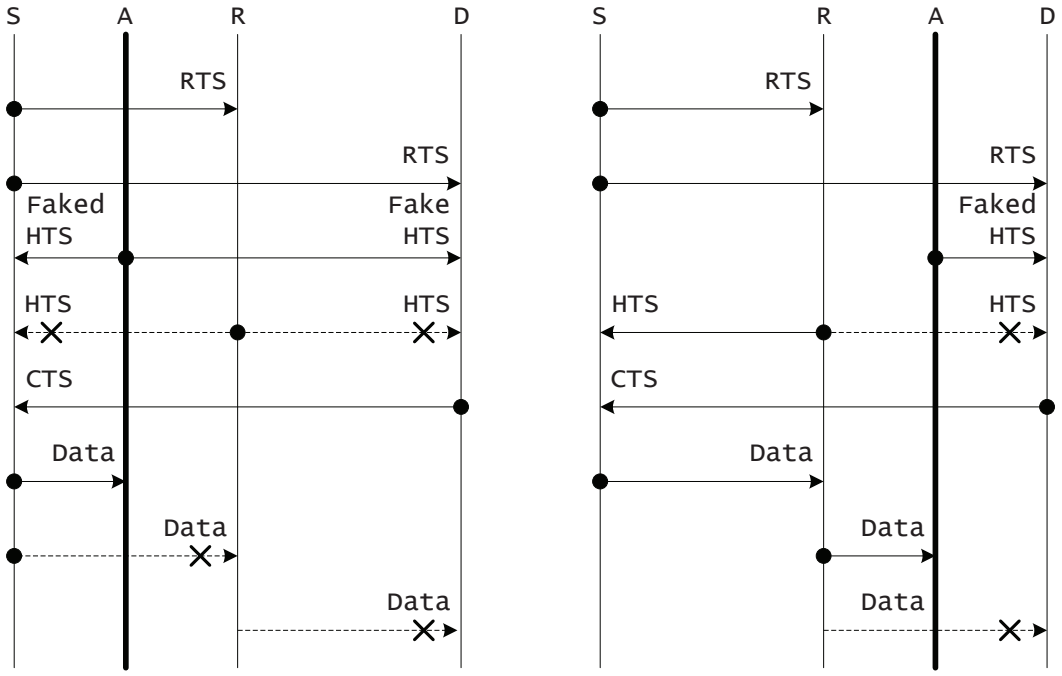
Fig. 5 shows a security vulnerability which caused by the faked CTS from attacker node A. In this case, the attacker node A sends a faked CTS to the source node S, informing the source node S that it is an intended recipient of future data packet. And, since the authentication is not applied to CTS packet, the legal CTS from destination D can be rejected by source node S due to a previous illegal CTS from attacker node A. Just after receiving the CTS from attacker node A, source node S transmits data packet to relay node R. Subsequently, the relay node R receives the data packet and then forwards received data packet to attacker node A. The attacker node A may try to deny communication service to the source by deliberately not forwarding data packet received from the relay node R. Consequently, cooperative communication between source node S and destination node D is not established.

 ISBN: 978-1-61208-133-5



(a) Faked RTS to R and D (b) Faked RTS to D

Figure 3. Security vulnerability by RTS packet attack



(a) Faked HTS to S and D (b) Faked HTS to D

Figure 4. Security vulnerability by HTS packet attack

 ISBN: 978-1-61208-133-5

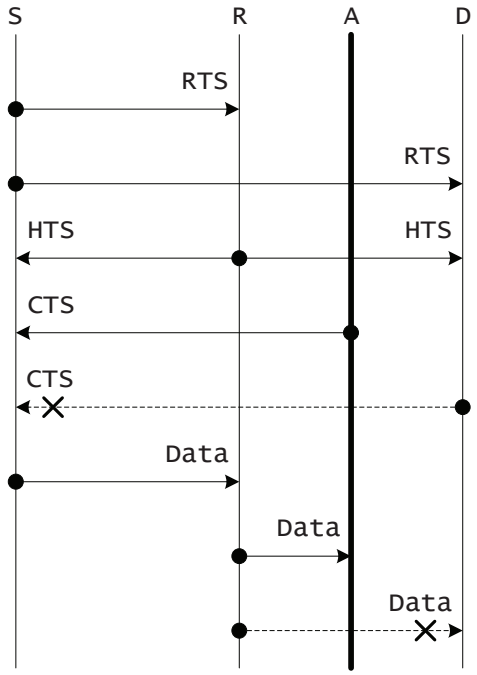


Figure 5. Security vulnerability by CTS packet attack

IV. Conclusion and Future Works

Security is a principal issue that must be resolved in order for the potential of cooperative communication networks to be fully exploited. However, security issues related to the design of cooperative network have largely not been considered.

CoopMAC is one such extension to the MAC sublayer. It was proposed to take advantage of cooperation, while remaining backward compatible with legacy IEEE 802.11. This paper presented the first case study of DoS attack in the CoopMAC. It also analyzed security vulnerabilities at each protocol stage while attacking a control packet exchanged among nodes. This work is the first comprehensive analysis of security vulnerability caused by DoS attack in CoopMAC. It can be significant in the use of designs of efficient authentication mechanism for secure CoopMAC. Moreover, our analytical results can be applied not only to cooperative network security, but also wireless sensor network (WSN) security design in general.

In the future, the authors will attempt to design and implement power-efficient authentication mechanism suitable for cooperative network. The plan is then to examine the effect that the proposed authentication mechanism has on the performance and efficiency of the cooperative transmission.

References

[1] A. Nosratinia, T. E. Hunter, and A. Hedayat, "Cooperative Communication in Wireless Networks," *IEEE Communication Magazine*, Vol. 42, Issue. 10, pp. 74-80, 2004.

[2] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative Diversity in Wireless Networks: Efficient Protocols and Outage Behavior," *IEEE Trans. on Information Theory*, Vol. 50, No. 12, pp. 3062-3080, 2004.

[3] Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, *ANSI/IEEE Std 802.11*, 1999 Edition (R2003), 2003.

[4] P. Liu, Z. Tao, and S. Panwar, "A Cooperative MAC Protocol for Wireless Local Area Networks," *IEEE ICC '05*, pp. 2962-2968, 2005.

[5] T. Korakis. Z. Tao, S. Makda, and B. Gitelman, "It is Better to Give Than to Receive - Implications of Cooperation in a Real Environment," *Springer LNCS 4479*, pp. 427-438, 2007.

[6] S. Makda, A. Choudhary, N. Raman, T. Korakis, Z. Tao, and S. Panwar, "Security Implications of Cooperative Communications in Wireless Networks," *IEEE Sarnoff Symposium*, pp. 1-6, 2008.

[7] S. Kulkarni and P. Agrawal, "Safeguarding Cooperation in Synergy MAC," *IEEE SSST '10*, pp. 156-160, 2010.

[8] Y. Mao and M. Wu, "Tracing Malicious Relays in Cooperative Wireless Communications," *IEEE Trans. on Information Forensics and Security*, Vol. 2, No. 2, pp. 198-207, 2007.

[9] Z. Han and Y. L. Sun, "Securing Cooperative Transmission in Wireless Communications," *IEEE MobiQuitous '07*, pp. 1-6, 2007.

[10] S. Dehnie, H. T. Sencar, and N. Memon, "Cooperative Diversity in the Presence of a Misbehaving Relay: Performance Analysis," *IEEE Sarnoff Symposium*, pp. 1-7, 2007.

[11] S. Dehnie, H. T. Sencar, and N. Memon, "Detecting Malicious Behavior in Cooperative Diversity," *IEEE CISS '07*, pp. 895-899, 2007.

[12] S. Dehnie and N. Memon, "A Stochastic Model for Misbehaving Relays in Cooperative Diversity,", *IEEE WCNS '08*, pp. 482-487, 2008.

[13] H. Marques, J. Ribeiro, P. Marques, A. Zuquete, and J. Rodriguez, "A Security Framework for Cognitive Radio IP Based Cooperative Protocols," *IEEE PIMRC '09*, pp. 2838-2842, 2009.

 ISBN: 978-1-61208-133-5

Analysis on IPv6 Transition Solutions and Service Tests

Xiaohong Deng, Lan Wang, Tao Zheng, Daqing Gu
France Telecom Group, Beijing, China
{xiaohong.deng; lan.wang; daqing.gu}@orange-ftgroup.com

Eric Burgey
France Telecom R&D Paris, France
eric.burgey@orange-ftgroup.com

***Abstract*—during the long IPv6 transition phase, multiple transition approaches may co-exist in the same network to enable v4-to-v4 communication and v6-to-v4 communication over IPv6 access. In this paper, we present our integrated platform with different transition solutions, such as AplusP, Dual-stack Lite and NAT64/DNS64, and analyze application behaviors and potential issues that may impact on the deployments. Particularly, results of application tests indicate that dual-stack application may break either because application layer is IP version dependent or because application has difficulties with NAT64 traversal if only NAT64/DNS64 is deployed; but same dual-stack application may work well in a Dual-stack Lite and NAT64/DNS64 mixing environment.**

Keywords—IPv6 migration; Dual-stack Lite; NAT64/DNS64; AplusP.

I. INTRODUCTION

The IANA pool for global public IPv4 address allocation is forecasted to exhaust by mid-2011. Yet IPv4-only legacies are ubiquitous crossing telecom infrastructure. As IPv6 and IPv4 are incompatible protocols, IPv6 could not replace IPv4 in order to solve the public IPv4 exhaustion problem immediately. Instead, both protocols will co-exist for a long period of time. The common thinking for more than 10 years has been that the transition to IPv6 will be based solely on the dual stack model until IPv6 takes over IPv4 before we ran out of IPv4. However, this has not happened. The IANA free pool of IPv4 addresses will be depleted soon, well before sufficient IPv6 deployment will exist. As a result, many IPv4 services have to continue to be provided even under severely limited address space. As a result, saving IPv4 address is one of concerns for IPv6 transition solutions. Dual-stack Lite [1], AplusP [2] and NAT64/DNS64 [3], described in this section, are transition approaches that address IPv6 introduction as well as IPv4 address sharing.

A. IPv6 transition Solutions

- Dual-stack Lite

The Dual-stack Lite technology [1] is intended for maintaining connectivity to legacy IPv4 devices and networks when service provider networks make a transition to IPv6-only deployments after the exhaustion of the IPv4 address space.

Dual-stack Lite enables a broadband service provider to share IPv4 addresses among customers while migrating to IPv6 by combining two well-known technologies: IP in IP tunnel and NAT. The principle is simple, 1) moving the current NAT performed on the Home Gateway (HGW) to Carrier Grade NAT; 2) IPv4 traffic are transported over IPv6 access network by IPv4-in-IPv6 softwires, an IP in IP tunnel defined in RFC5571[4]. Dual-stack Lite specification introduces two new terms: the DS-lite Basic Bridging Broad Band element (B4) and the DS-lite Address Family Transition Router element (AFTR). A B4 element is a function implemented on a dual-stack capable node, either a HGW or a directly connected device that creates a tunnel to an AFTR. An AFTR element is the combination of an IPv4-in-IPv6 tunnel end-point and an IPv4-IPv4 NAT implemented on the same node.

As illustrated in Figure 1, each HGW has only IPv6 access, yet many customers are still configured with RFC1918 [5] private addresses. Therefore the traffic generated by terminals is tunneled by the B4 element to the AFTR via an IPv4-in-IPv6 softwire, where B4 element is acting as Softwire Initiator (SI) and the AFTR is acting as Softwire Concentrator (SC). After de-capsulation, the AFTR then translate RFC1918 private addresses realm to public IPv4 address realm. Per subscribers tunnel endpoints ID are identified by AFTR to distinguish RFC1918 private address space per subscriber.

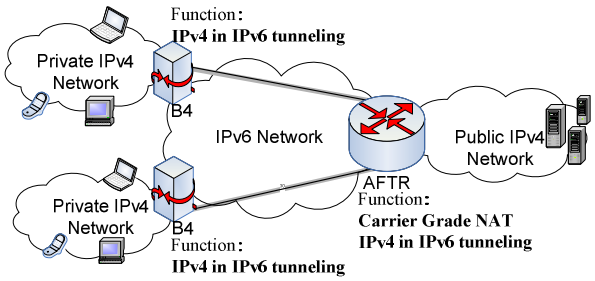


Figure 1. Dual Stack-Lite

- AplusP

The principle in AplusP [2] is straightforward; 16 bits stolen from TCP/UDP port field are attached to the IPv4 address to identify different customers that sharing the same public IP address. Hence, multiple HGW share a common global IPv4 address, but with separate, non-overlapping, port ranges. Each HGW can use the address as if it were its own public address, except that only a limited port range is available to be used. An IPv6 address derived from a pre-assigned public

 ISBN: 978-1-61208-133-5

IPv4 address plus a specified range of ports are allocated to each HGW; as a result, Port Range Router (PRR) can route incoming traffic to destination HGW according to a destination IPv6 address generated from destination IPv4 address and port fields of the IPv4 packet header.

As demonstrated in Figure 2, AplusP uses the same two technologies as DS-lite: IP in IP tunnel and NAT, but in different ways. First, the NAT is performed on the HGW rather than on Carrier Grade and the HGW NAT should ONLY use the restricted source port range. Second, although IPv4 traffic are also transported over IPv6 access network by IPv4-in-IPv6 tunnels as the DS-lite does, the AplusP HWG should pre-assigned an IPv6 address that is derived from a pre-assigned public IPv4 address plus a specified port range so that Port Range Router can route incoming traffic to proper HGW according to a destination IPv4 address and port.

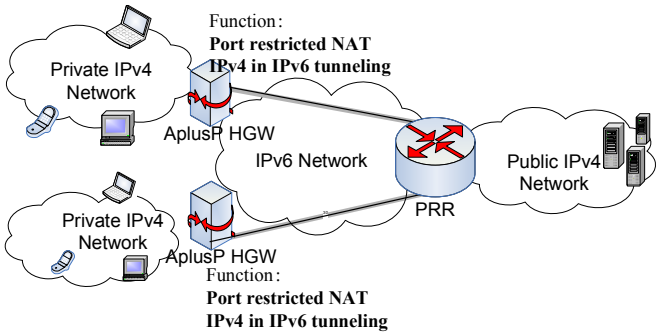


Figure 2. AplusP

- NAT64/DNS64

It has been reached a consensus that both networks coexist until IPv6 takes over IPv4. However, IPv6 growth has been much slower than anticipated. Therefore, IPv6-only deployments face a challenge of communicating with the predominantly IPv4-only rest of the world. Likewise, a similar problem is encountered when legacy IPv4-only devices need to reach the IPv6 Internet. One initial proposal was NAT-PT RFC2766 [6]. However, it has been declared obsolete in RFC4966 [7] due to its various issues. Still, to address this challenge, IPv4/IPv6 translation continued to be a major focus of interest at IETF. Recently, IETF BEHAVE working group has been working on IPv4/IPv6 translation solution and has resulted in several draft documents. The general framework for IPv4/IPv6 translation is described in [8], which also explains the background of the problem and some use cases. NAT64/DNS64 [3], describes a stateful IPv6-to-IPv4 NAT translation which allows IPv6-only clients to talk to IPv4 servers using unicast UDP, TCP, or ICMP. The public IPv4 address can be shared among several IPv6-only clients. Used in conjunction with DNS64, which is a mechanism for synthesizing AAAA resource records (RR) from A RR and NAT64's prefix, NAT64 requires no changes in the IPv6 client or the IPv4 server.

The Figure 3 illustrates NAT64/DNS64 principle and a home network use case of NAT64/DNS64. When a host e.g., Host-1 in IPv6 network wants to talk to a host e.g., Host-2 in IPv4 network, NAT64 together with DNS64, which works as DNS proxy and derive AAAA RR from A RR, translate IPv6 TCP, UDP and ICMP to IPv4. DNS64 can be either standalone device or embedded within NAT64. Dotted line describes a use scenario, where IPv6-only terminals in a home network need to contact with IPv4 peer. In this scenario, no changes are required on HWG and only IPv6 Router Advertisement (RA) or DHCPv6 are expected to offer automatically configuration for IPv6 devices.

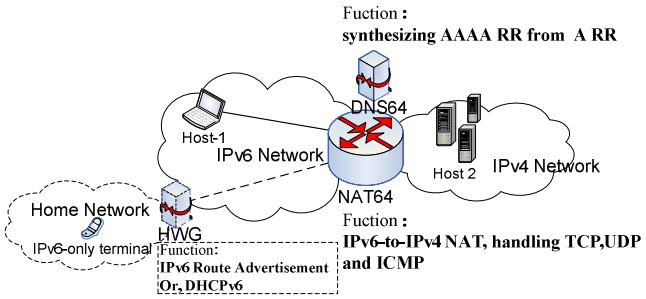


Figure 3. NAT64/DNS64 and a use case

IETF Softwire working group has submit Dual-stack Lite [1] to the IESG, which is responsible for technical management of IETF activities and the Internet standards process, to make it be considered as a Proposed Standard. While NAT64/DNS64 [3l] has been approved by IESG and been sent to RFC Ed Queue. Recently, many vendors have claimed that Dual-stack Lite and NAT64/DNS64 approaches have been included in their product roadmap; for example: Cisco's Carrier-Grade IPv6 (CGv6) Solution and HUAWEI's Carrier-grade NAT (CGN) Solution.

B. Motivation and Objectives

IPv4 address sharing is an important mechanism during transition phase. Application behaviors in terms of port consumption not only impact the deployment factor (i.e., port range size) for AplusP solution but also play an important role in determining the port usage per customer on AFTR for Dual-Stack Lite. During our study, a concern that NAT64 may break existing popular applications has been arising. Hence, it is of interest to test application's NAT64 compatibility. In addition, since solutions addressing different use cases which may co-exist in the same subscribers' network, it is essential for network providers to understand potential issues and how applications cope with multiple transition solutions. Therefore, we have implemented three transition solutions in an integrated platform on which application behaviors were tested.

C. Orgnaization

The rest of this paper is organized as follow. Section II introduces related works; Section III presents our platform in which three transition solutions are implemented and integrated. The test results are discussed in Section IV. Finally, this paper is concluded in Section V.

II. RELATED WORK

A. Opensource AFTR Implementation

Internet Systems consortium (ISC)'s AFTR [9] which is available for free download under the terms of the ISC License

 ISBN: 978-1-61208-133-5

(a BSD style license), implements a Dual-Stack Lite AFTR as described in [1].

A Dual-Stack Lite deployment includes at least one AFTR in the ISP's network core, and one B4, which is the IPv4 default router for all hosts behind it and customer-side tunnel initiator. For testing and demonstration purposes, despite that B4 functionality can also be built into general-purpose computers (e.g., in FreeBSD or Linux), ISC's AFTR has used a Linksys WRT54GL running OpenWrt [10], an embedded Linux distribution for home gateways, and released a WRT54G prebuilt images which is prebuilt with functionalities that make a B4 set up an IPv4-in-IPv6 tunnel with AFTR.

As Dual-stack Lite is under the standardization process in the softwire working group of the IETF and there may be changes to the specification before it is finalized as an RFC, ISC AFTR have been actively tracking the current specification. As such, ISC AFTR considers itself as a work in progress, for testing and experimentation only. It is an open source implementation meant to promote the development of open standards for IPv4-to-IPv6 transition technology.

B. Open source NAT64/DNS64 Implementation

Be funded by the NLnet Foundation and Viagénie, Ecdysis [11], has developed an open-source implementation of a NAT64 gateway to run on open-source operating systems such as Linux and BSD. The gateway is comprised of two separated modules: the DNS ALG and the IP translator. The DNS ALG is implemented in two DNS open-source server: Unbound and Bind. The IP translator is implemented in Linux as kernel module using Netfilter facilities and in openBSD as a modification of Packet Filter (PF).

C. AplusP Implementation

We had implemented both Aplusp HGW and Port Range Router (PRR) on a Linux platform [12].

For Aplusp HGW, using Netfilter framework, the IPv4 port restricted NAT operation performed by CPE was implemented by simply setting rules through iptables tool on Linux. After the NAT operation on the CPE, the NATed IPv4 packets were sent to a TUN interface which represented as a virtual network interface in Linux and enabled with IPv4-in-IPv6 encapsulation/decapsulation functions developed by us.

PRR, located in the interconnection point of the IPv6 network and IPv4 network, is implemented with two main functions: 1) IPv4-in-IPv6 encapsulation/decapsulation; 2) destination port based routing function, which is for the IPv4 traffic originated from the IPv4 Internet and destined to the shared IPv4 address realm of the operator. Likewise, TUN driver is also used in PRR to achieve function 1). Function 2) is realized by pre-assigning an IPv6 prefix that maps from IPv4 address and port range to each CPE, and generating IPv6 destination address according to IPv4 destination address and port. To facilitate test and experiment on AplusP solution, recently, we are considering release this AplusP implementation under open source license.

III. IPV6 TRANSITION SOLUTION IMPLEMENTATION

A. Overview of Implementation

Based on the related work stated in previous section, first, we integrated all of these three IPv6 transition solutions: AplusP, Dual-stack Lite and NAT64 into a realistic ADSL access environment, which consists of HGW, DSLAM, PPPoE server and DHCPv6 server. As illustrated in Figure 4, we customized two types of HGWs, running OpenWrt, on Linksys WRT54GL: a) AplusP enabled home gateway has been uploaded with our AplusP HGW functions and configured with PPPoE client and DHCPv6 client; b) Dual-stack Lite/NAT64 enabled home gateway, providing both Dual-stack Lite and NAT64 solution for the same subscriber, has been configured with IPv4-in-IPv6 tunneling, PPPoE client and DHCPv6 client for Dual-stack Lite; and IPv6 Route Advertisement (RA) for NAT64. The IPv6 provisioning to HGWs is via IPv6 over PPPoE provided by a PPPoE server, where a DHCPv6 server is co-located.

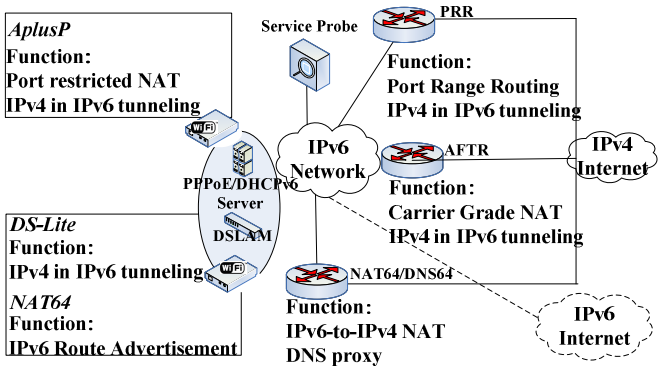


Figure 4. IPv6 transition solutions mock up

Currently, we only have IPv4 Internet access and in future we will also have access to the IPv6 Internet which is illustrated by dotted line in Figure 4.

Because new DHCPv6 options for AplusP and Dual-Stack Lite have not been standardized yet, we used user defined options for mockup purpose. The DHCPv6 server was configured to convey different set of DHCPv6 options, some of which are user defined, to AplusP HGW and Dual-stack Lite/NAT64 HGW separately. The user defined DHCPv6 options are shown in Figure 5.

```
#AplusP DHCPv6 options
option dhcp6.gateway code 54 = ip6-address;
option dhcp6.ipv4 code 55 = ip-address;
option dhcp6.port code 56 = unsigned integer 16;
option dhcp6.range code 57 = unsigned integer 16;

#Dual-Stack Lite DHCPv6 options
option dhcp6.softwire code 58 = ip6-address;
option dhcp6.name-servers code 59 = ip6-address;
option dhcp6.pubadd code 91 = ip6-address;
option dhcp6.radvd code 92 = ip6-address;
option dhcp6.defgateway code 90 = ip6-address;
```

Figure 5. DHCPv6 options for AplusP and Dual-Stack Lite

B. Service Probe in AplusP

Besides PRR, AFTR and NAT64/DNS64, we also developed and deployed a Service Probe in our IPv6 network, which use IPv6 TCP socket to ask AplusP HGW for NAT session usage, and store AplusP NAT statistics in a Mysql

database to further analyze application behaviors in terms of port and session consumptions. The detailed test results and analysis are presented in the next section.

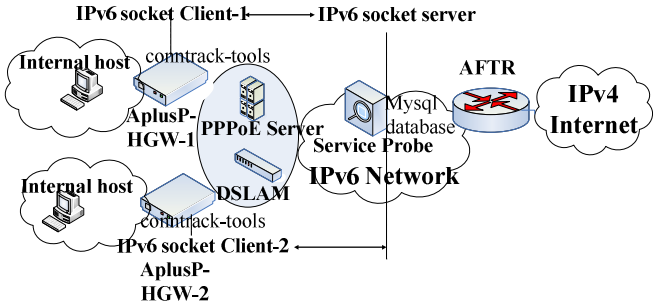


Figure 6. Service Probe and AplusP HGW

To implement service probe, we have configured conntrack-tools [13], which is the module that provides stateful packet inspection for iptables, on the AplusP HGW to collect statistics of the iptables NAT usage. All the statistics of AplusP HGW NAT are then sent via IPv6 socket to Service Probe which is responsible for further storage and analysis of application behaviors, more specifically, TCP/UDP ports and sessions consumption during communication. Service Probe socket is designed by I/O multiplexing approach so that it can monitor multiple AplusP HGWs at the same time, e.g., in our test bed there are two HGW as shown in the Figure 6. Database design on Service Probe is simple and shown in Figure 7 and Figure 8. The first table in the Mysql database stores key information of all NAT sessions, received from AplusP HGW, including (si, sp, di, dp, protocol, msi, msp, sla, ela, status, pr) where the end time which is N/A until the session is expired, and the status of the session which is either active when the session is valid or history when the session is expired. By scanning the first table, for per internal host, a second table is instanced to keep tracking and storing port numbers and session numbers that this internal client is using in an every second basis.

si	source address
sp	source port
di	destination address
dp	destination port
proto	the protocol
msi	mapped source address
msp	mapped source port
sla	the start time of a session
ela	the end time of a session
status	either active or history session
pr	port rage

Figure 7. Fields description of Service Probe's first table

ip	IP address of a client
pn	The port number of a client used
sn	The session number of a client used
ttime	The current time
pr	Port range

Figure 8. Fields description of Service Probe's second table

C. *Dual-stack Lite and NAT64 implementation*

Introducing NAT64 may bring impacts, for instance some applications may break due to incompatible with NAT64. Yet with both Dual-stack Lite and NAT64 enabled network, it was not clear that how application may behave in the multiple transition solution enabled network. For example, we were not certain whether apps choose IPv6 over IPv4 or if they choose both. Therefore we have tested apps' compatibility with NAT64 as well as apps' compatibility with two transition solution enabled network. Furthermore, we have observed apps' behaviors in terms of IP preference, more specifically, how applications, in a both Dual-stack Lite and NAT64 enabled subscriber network, deal with AAAA and A DNS record and which IP version protocol (IPv4 or IPv6) is preferred to initiate communication. To do so, a Dnsmasq (a DNS forwarder), whose upstream DNS server is configured with DNS64, was installed in the B4 element as shown in Figure 9. DNS64 returns both AAAA and A RRs to Dnsmasq which in turn forwards the responses to the host behind B4. Since we do not have native IPv6 access yet, AAAA RRs returned by DNS64 are generated from A RRs and NAT64's prefix instead of native AAAA RRs.

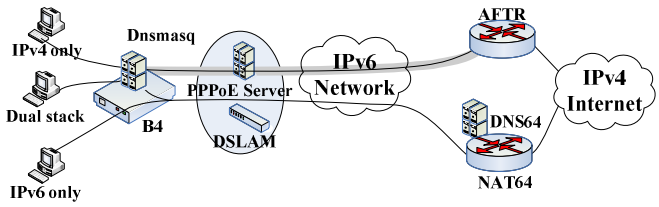


Figure 9. Dual-Stack and NAT64

IV. TEST RESULTS AND ANALYSIS

A. *Application behaviors in interms of port/session consumptions*

We tested popular applications, such as news website, video website, skype, BitTorrent and GoogleEarth, to investigate how many ports and sessions they are costing on the NAT mapping table dynamically from the first NAT bindings being established to the last one being destroyed. All the figures in this sub-section are derived from the Mysql database described in sub-section B of the previous section.

As illustrated in Figure 10, when open a news website (e.g., [15]) that often contains a number of images and flashes, it takes up to four minutes from the first NAT binding being established to the last one being destroyed. During the four minutes new NAT bindings are established while the old ones are expiring. For IE, the port consumption dramatically rose and reached the peak of 20 ports at the 18th seconds and then decreased gradually to zero at the 200th seconds. While for firefox, after opened a dozen of ports at the beginning, it then gradually increased to 25 until the 120th second, and finally dropped gradually to zero until the 240th second. It is evident that even if visit the same website, port consumption varies from web browser to web browser.

 ISBN: 978-1-61208-133-5

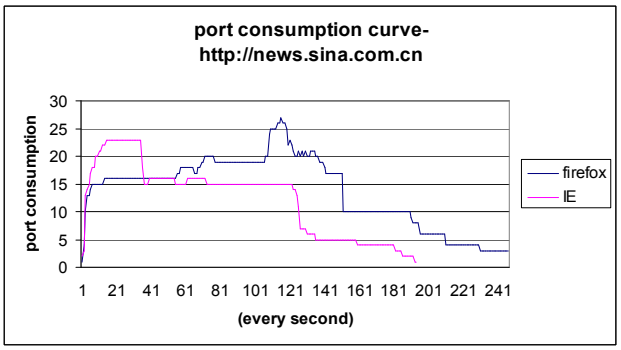


Figure 10. Comparision of port consumption between firefox and IE

Figure 11 shows same evidence that firefox consumes more ports than IE when open a video website which cost up to 80 ports during browsing its main page.

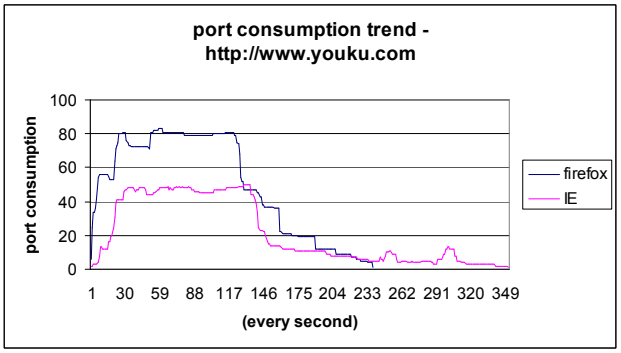


Figure 11. Sampled ports from DNS randomness test in Scenario B

Port consumption comparison among different applications, including BitTorrent, Google Earth, skype and using firefox to visit news website and video website, is demonstrated in the Figure 12. BitTorrent has constantly occupied hundreds of ports while downloading and dominated in the port consumption. Firefox has consumed dozens of ports for about two minutes and ranked second after BitTorrent, while others merely cost less than 25 ports during a whole communication process for each.

The session consumption comparison among the same set of applications are also illustrated in Figure 13. Unlike other apps which consume similar amount of sessions as ports, BitTorrent established five hundreds of sessions even though the port consumption was relatively low (under a hundred) in the first minute of the communication, because when BitTorrent initiates a downloading it first uses the same source port to connect to the different destinations (destination IP and port) therefore one source port multiplexing different sessions. Besides, Skype is another example that uses one source port to multiplex different sessions thereby saving source port consumptions on NAT.

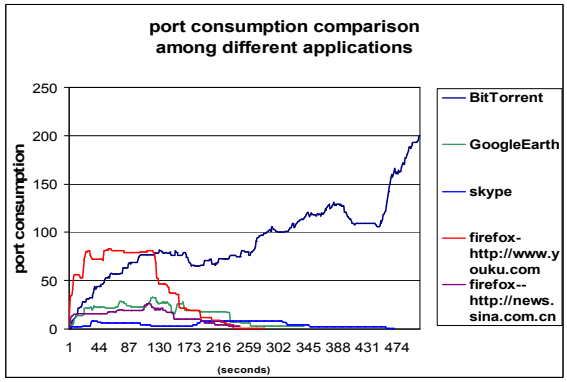


Figure 12. Port consumption comparison among different apps

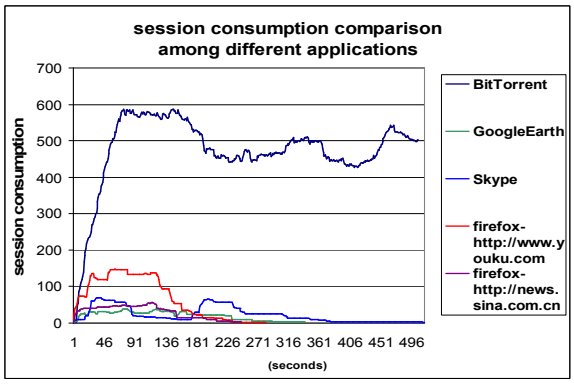


Figure 13. Session consumption comparison among different apps

Furthermore, we have tested and compared mobile apps and PC apps. The test results shown in Figure 14 and Figure 15 indicates that even the same app, either web-browser chrome or Google Earth, the mobile release - Android chrome and Android Google Earth consumed fewer ports than the PC release – Windows chrome and Windows Google Earth respectively.

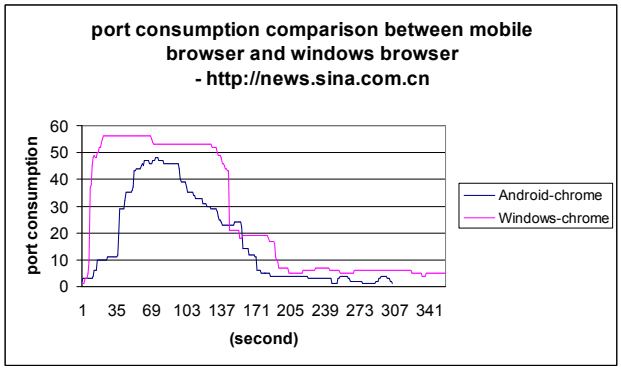


Figure 14. Comparison between mobile and Windows broswer

 ISBN: 978-1-61208-133-5

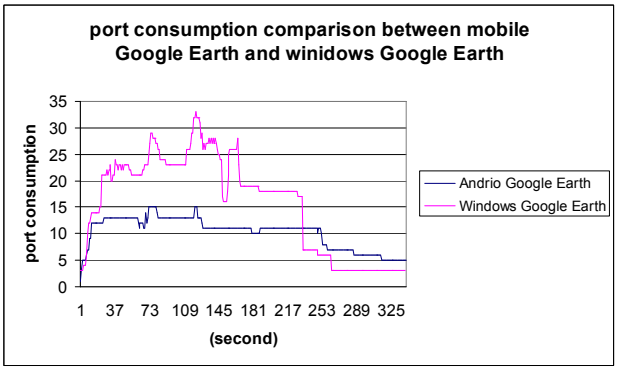


Figure 15. Comparison between mobile and Windows GoogleEarth

B. *Application compatibility with NAT64*

Tested Apps	Compatible	Non-compatible
Firefox (Non-vedio) v3.6.12	√	
video website		√
IE (Non-vedio) v6.0	√	
Skype v5.0	√	
Google Earth v5.2.1	√	
Live Messenger 2009		√
QQ 2010		√
uTorrent v2.2		√
BitComet v1.23		√

Figure 16. Apps compatibility with NAT64

Test results of apps' compatibility with NAT64 are listed in the Figure 16. Vesting video website, Live Messenger, QQ, uTorrent and BitComet break in a NAT64 only network, while firefox, IE, Skype and Google Earth work well with NAT64.

C. *Application compatibility with NAT64 and Dual-stack Lite mixing network*

Besides investigation on NAT64 solution, we have tested if the both Dual-stack Lite and NAT64 enabled network has influence on apps. In Figure 17, the test results show that all of them work well without configuration change, manual restart or other human interference.

Tested Apps	Compatible
Firefox (Non-vedio) v3.6.12	√
video website	√
IE (Non-vedio) v6.0	√
Skype v5.0	√
Google Earth v5.2.1	√
Live Messenger 2009	√
QQ 2010	√
uTorrent v2.2	√
BitComet v1.23	√

Figure 17. Apps compatibility with both NAT64 and DS-Lite enabled environment

D. *Application behaviors in terms of IP version preference*

Regarding both Dual-stack Lite and NAT64 enabled subscriber network, we further tested and analyzed whether IPv4 or IPv6 is preferred to initiate the communication. It has been shown that all the applications that we investigated issued both A and AAAA DNS query, and except QQ (an Instance Messenger) and BitComet (a BT Client) completely ignored AAAA RR, all other applications made use of both IPv6 and IPv4 to communicate with peers/servers. IPv6 usage portion depends on apps and use cases, some of which use all IPv6 while others used all IPv4; some of which use major in IPv6 while others used major in IPv4. As a result, as illustrated in Figure 18, we classified apps into five categories: 1) all IPv6, 2) major IPv6, 3) half/half, 4) major IPv4 and 5) all IPv4. Again, as we do not have native IPv6 access, all the IPv6 communication was talking to IPv4 servers/peers via NAT64.

- all IPv6

For web browsers, AAAA records have higher priority and when visit non-video website, both firefox and IE used IPv6 to talk (to IPv4 server) through NAT64.

- major IPv6

Skype and Google Earth used AAAA to initiate IPv6 connections (to IPv4 server) via NAT64. Yet, according to our captured packets (wireshark), we still found that there were a few IPv4 connections.

- half/half

When firefox and IE were used to visit a video website (e.g., [15]), the main pages were downloaded through IPv6 via NAT64 and after IPv6 video downloading failure due to NAT64 traversal failure, firefox and IE then requested IPv4 and vedio downloading was done via AFTR.

Tested Apps	All IPv6	Major IPv6	Half/ half	Major IPv4	All IPv4
Firefox (Non-vedio) v3.6.12	√				
video website			√		
IE (Non-vedio) v6.0	√				
Skype v5.0		√			
Google Earth v5.2.1		√			
Live Messenger 2009				√	
QQ 2010					√
uTorrent v2.2				√	
BitComet v1.23					√

Figure 18. Apps classified by IPv6 usage portion

 ISBN: 978-1-61208-133-5

- major IPv4

During login and authentication phase, Live Messenger were using IPv6, but after that, because the IPv6 client did not send the same application layer message as the IPv4 client, the IPv6 client (behind NAT64) failed to get reply from IPv4 server. Then, Live messenger automatically switched to IPv4 by which all the rest of communication were done. What we have learned from this case is that the application layer should be IP version agnostic in order to decrease impacts introduced by IPv6 transition solutions. uTorrent(a BT client) is another instance that used IPv6 for login/authentication but IPv4 for the data exchange, for IPv6 peer was not able to talk to IPv4 peer.

- All IPv4

Although QQ and BitComet issued both A and AAAA quires, they completely ignored AAAA RR and only used IPv4 for communication.

V. CONCLUSION

It is likely to have multiple transition approaches in the same subscriber network. Therefore, we have implemented AplusP, Dual-stack Lite and NAT64/DNS64 in an integrated platform and investigated application behaviors in this platform. Firstly, port/session consumption on NAT that impacts on the deployment factors for both AplusP and Dual-stack Lite has been tested. Secondly, application's NAT64 compatibility is presented. Results of application tests indicate that dual-stack application may break either due to IP version dependent of application layer or NAT64 traversal difficulties if only NAT64/DNS64 is deployed; yet same dual-stack application may work well in a Dual-stack Lite and NAT64 mixing environment.

REFERENCES

[1] Durand, A., "Dual-Stack Lite Broadband Deployments Following IPv4 Exhaustion", draft-ietf-softwire-dual-stack-lite-06 (work in progress), August 11, 2010.

[2] Bush, R., "The A+P Approach to the IPv4 Address Shortage", draft-ymbk-aplusp-08 (work in progress), January 5, 2011.

[3] M. Bagnulo and P. Matthews., " Stateful NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers", draft-ietf-behave-v6v4-xlate-stateful-12 (work in progress), July 10, 2010.

[4] RFC5571, B. Storer, "Softwire Hub and Spoke Deployment Framework with Layer Two Tunneling Protocol Version 2 (L2TPv2)", June 2009.

[5] RFC1918, Y. Rekhter,"Address Allocation for Private Internets", February 1996.

[6] RFC2766, G. Tsirtsis, "Network Address Translation - Protocol Translation (NAT-PT)", February 2000.

[7] RFC4966, C. Aoun, "Reasons to Move the Network Address Translator - Protocol Translator (NAT-PT) to Historic Status", July 2007.

[8] F. Baker, "Framework for IPv4/IPv6 Translation", draft-ietf-behave-v6v4-framework, August 17, 2010.

[9] Internet Systems Consortium, https://www.isc.org/software/aftr, March, 2011

[10] OpenWrt, a Linux distribution for embedded devices, http://openwrt.org/, March, 2011.

[11] Ecdysis: open-source implementation of a NAT64 gateway , http://ecdysis.viagenie.ca/, March, 2011.

[12] Z. Xiaoyu, X. DENG, "Implementing Public IPv4 Sharing in IPv6 Environment", April 2010.

[13] conntrack-tools, Connection tracking userspace tools for Linux, http://conntrack-tools.netfilter.org/, March, 2011.

[14] http://www.sina.com.cn, March, 2011.

[15] http://www.youku.com, March, 2011.

 ISBN: 978-1-61208-133-5

An Encryption Scheme for Color Images Based on Chaotic Maps and Genetic Operators

El-Sayed M. El-Alfy
College of Computer Sciences and Engineering
King Fahd University of Petroleum and Minerals
Dhahran 31261, Saudi Arabia
alfy@kfupm.edu.sa

Khaled A. Al-Utaibi
Information and Computer Science Department
University of Ha'il
Ha'il, Saudi Arabia
alutaibi@uoh.edu.sa

***Abstract*—Secure transmission and storage of color images is gaining growing importance in recent years due to the proliferation of multimedia network applications. In this paper, we propose a novel scheme based on chaotic maps and genetic operators for encrypting color images. The capability of the proposed approach to efficiently generate cipher images with very low correlation coefficients of adjacent pixels is demonstrated through some experimental results for several benchmark images. It is also shown that the approach is very sensitive to any slight changes in the secret key values.**

Keywords-image encryption; chaotic maps; genetic operator; information hiding; data security.

I. INTRODUCTION

Color images are being transmitted and stored heavily over the Internet and wireless networks taking advantage of rapid development in multimedia and network technologies. However, as there is always a potential risk of information security in such interconnected environments, protecting confidentiality of color images has become an increasingly important issue in many areas such as remote sensing and satellite imagery, astrophysics, seismology, agriculture, radiology, telemedicine, ecosystems, industrial processes, military communications, and image archiving. Several image encryption schemes have been suggested in the literature to meet this requirement [1][2]. However, due to the processing overhead resulting from the large data size of digital images and the high correlation among pixels, traditional encryption techniques, such as DES, AES and RAS, are found to be inefficient for image encryption [7][8][11].

The pseudorandom nature and other properties of chaotic systems, including sensitivity to initial conditions and non-periodicity, have made them attractive alternatives among the proposed approaches for image encryption [7]. The first chaotic based image encryption algorithm was proposed in 1989 [17]. Recently, there is a growing interest in this area and several approaches have been proposed in the literature [3]-[6]. In [7], Lin and Wang proposed an encryption algorithm based on chaos with PWL memristor in Chua's circuit. Their algorithm uses two main operations of image scrambling and pixel replacement. Fu and Zhu proposed another technique based on logistic maps with permutation and circular bit-shift methods for confusion and diffusion [8]. The method proposed by Yanling is based on logistic chaotic sequences and image mirror mapping [9]. A 3D image encryption scheme using logistic maps with bit permutation was presented in [10]. The scheme proposed by Kumar and Chandrasekaran is also a 3D image encryption, but with a different approach where Lorenz attractor is used directly for image encryption [11]. Lue *et al*. proposed an image encryption algorithm based on spatiotemporal chaos [12]. In their work, the plain image block data is masked by the values extracted from a spatiotemporal chaotic system, and then shuffled according to the maximum state value in the system. Wei-Bin and Xin proposed an algorithm that uses Arnold cat map to shuffle the pixels of the plain image and 1D Henon's chaotic system to change the shuffled pixels by XOR operation [13]. The algorithm proposed by Wang and Zhang is based on S-boxes in AES and chaotic sequences generated by logistic maps [14]. The algorithm presented by Flores-Carmona *et al*. in [15] is based on CML (Chaotic Map Lattice), which allows direct encryption and decryption of color digital images. A 3D Bakeer map encryption technique was proposed by Hongelei and Guang-Shou in [16]. Chong Fu *et al*. proposed an image encryption scheme based on 3D Lorenz system to improve the security and performance of the encryption system over conventional one dimension chaos based ones [18].

The previously mentioned algorithms are restricted to grayscale images. Though, some of them can be easily extended to handle color images, this extension comes with a cost of increased computation time as a result of additional information required to represent color components. Therefore, many color-image encryption techniques use block-based encryption which is usually faster than stream-based encryption although it may be less secure. One example of block-based encryption algorithm for colored images was proposed by Pareek *et al*. [19]. This algorithm uses an external key and two logistic maps. The first map is used to generate the initial conditions of the second map which is used to select the type of encryption operation among eight different encryption operations (*e.g*., NOT, XOR, etc.). In order to make the cipher robust against attacks, the external key is modified after encrypting each block of pixels. The color image encryption algorithm proposed by Shubo *et al*. in [20] uses two logistic maps coupled such that the first logistic map updates the parameter of the other. The encryption operation is performed by a simple XOR operation of the binary sequence of the plain-image with the keystream binary sequence generated by the second logistic map. Another color image encryption based on a modified logistic map and 4-dimensional hyper-chaotic maps was proposed in [21]. In this, paper, we propose an alternative scheme for encrypting color images based on

 ISBN: 978-1-61208-133-5

chaotic-maps and genetic operations as tools for confusion and diffusion. The simple and fast computation of crossover and mutation operations compared to regular confusion and diffusion operations allows the algorithm to implement stream-based encryption which usually provides better security than block-based encryption.

The rest of this paper is organized as follows. In Section II, we give a detailed description of the proposed image encryption algorithm. Experimental results in Section III demonstrate various performance and security measures of our algorithm. Section IV concludes the paper by summarizing the proposed work and the obtained results.

II. THE PROPOSED ALGORITHM

A. The General Structure of the Algorithm

The general structure of the proposed algorithm is shown in Figure 1. It consists of four units: logistic map, quantification, crossover, and mutation. The logistic map generates four chaotic sequences based on the given controlling parameters (μ_1, μ_2, μ_3, μ_4) and initial values (x^o_1, x^o_2, x^o_3, x^o_4) which represent shared keys used by the encryption and decryption algorithms. The quantification unit maps the four chaotic sequences to four key streams which are then used to control the crossover and mutation operations. The purpose of the crossover unit is to cause image confusion by scrambling the image pixels row-wise and then column-wise. The mutation unit is used to mask the intermediate image obtained by the crossover unit with a random image; thus causing image diffusion.

B. Logistic Map

Logistic map is widely used in chaotic cryptography for their simplicity and high sensitivity to initial conditions. It is defined by:

$$x_{n+1} = \mu x_n (1 - x_n), \tag{1}$$

where μ is a control parameter, x_n is a real number in the range [0,1] and x_0 is an initial condition. When $3.569955672 < \mu \le 4$, the system becomes chaotic [10].

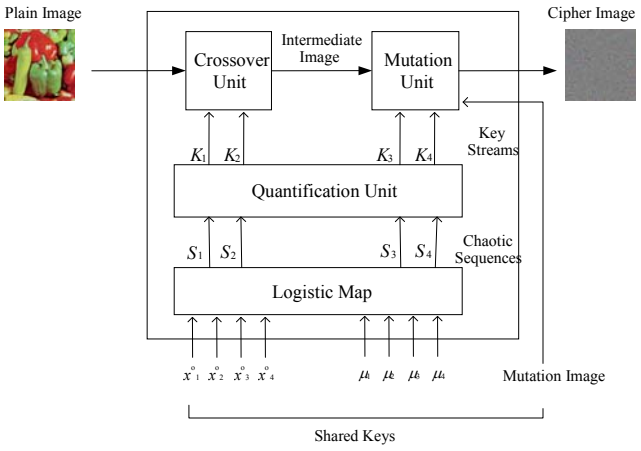


Figure 1. Layout of the proposed algorithm.

In the proposed algorithm, the logistic map is used in a similar manner as in [8] to generate four chaotic sequences (S_1, S_2, S_3, S_4). These sequences are generated based on some given controlling parameters (μ_1, μ_2, μ_3, μ_4) and initial values (x^o_1, x^o_2, x^o_3, x^o_4) which are considered as shared keys for encryption and decryption.

C. The Quantification Unit

Most of chaotic systems generate real-valued sequences which need to be mapped to integer/binary sequences (*i.e.*, key streams) which will be used to control the confusion and diffusion units. Basically, there are three techniques commonly used in the literatures: normalization, threshold level functions, and ordered chaotic sequence.

In normalization methods, a real value, X, in the chaotic sequence can be mapped to a digital value, D, in the key stream using the following relation:

$$D = \left\lceil \frac{(X - X_{\min}) \times D_{\max}}{X_{\max} - X_{\min}} \right\rceil, \tag{2}$$

where $X_{\min}$ and $X_{\max}$ are the minimum and maximum values in the chaotic sequence to be quantified, and $D_{\max}$ is the maximum required value of the key stream.

In the second method, each value, x_i, in the chaotic sequence is converted to a binary bit, b_i, using a single level threshold function defined as:

$$b_i = \begin{cases} 0 & x_i < 0.5 \\ 1 & x_i \ge 0.5 \end{cases}. \tag{3}$$

The third method as described in [10] is based on mapping the key stream to the element's positions in the sorted chaotic sequences. In this method, the elements in the chaotic sequence, X, are sorted in ascending order to form an ordered sequence X'. If the chaotic sequences are non-periodic, then each element in X has exactly one position in the sorted sequence X'. These positions are taken to be the values of the key stream. For example, suppose that X = {0.87, 0.34, 0.12, 0.75, 0.03, 0.88, 0.56, 0.04}, then the sorted sequence X' = {0.03, 0.04, 0.12, 0.34, 0.56, 0.75, 0.87, 0.88}. Since each element in X has exactly one position in X' (*e.g.*, 0.87 has position 7), the key stream is given by the sequence K = {7, 4, 3, 6, 1, 8, 5, 2}.

The first method is a simple and fast way to map the chaotic sequences to integer key streams, but it is subject to rounding errors. The threshold level function gives uniform distribution of the generated key streams. However, it is a lengthy process and requires long chaotic sequences (each bit requires one chaotic value). The third method is used in the proposed algorithm for its simplicity and short computation time.

The quantification unit in our algorithm receives four chaotic sequences (S_1, S_2, S_3, S_4) generated by the logistic map and convert them to four key streams (K_1, K_2, K_3, K_4) which will be used to control the operation of the crossover and mutation units. The length of the 1st and 3rd key streams is M, and the length of the 2nd and 4th key streams is N, where $M \times N$ is the size of the plain image in pixels.

D. The Crossover Unit

The crossover unit is used to change the order of the image pixels row-wise and column-wise by means of a multi-point crossover operation. The unit is controlled by the two key streams, K_1 and K_2, generated by the chaotic map and quantification units. The first key stream controls the crossover operation on the image rows whereas the other key controls the crossover operation on the image columns. Each two consecutive elements in the key stream select two rows/columns for the crossover operation and determine the positions of the cut points. The number of cut points in the crossover operation is a variable parameter that should be set by the user prior to encryption/decryption process. For example, this value can be set to $\lfloor M/2 \rfloor$ for row-crossover and $\lfloor N/2 \rfloor$ for column-crossover. The idea of selecting the two rows/columns and determining the positions of the cut points can be explained as follows. Assume that the two consecutive elements of the key stream are E_i and E_{i+1}, then rows/columns number E_i and E_{i+1} are selected for crossover operation. The positions of the cut points are computed as follows:

$$\begin{aligned} r_1' &= |E_i - E_{i+1}| \bmod L \\ r_2' &= (r_1' + |E_i - E_{i+1}|) \bmod L \\ &\vdots \\ r_P' &= (r_{P-1}' + |E_i - E_{i+1}|) \bmod L \\ (r_1, \ r_2, \ \ldots, \ r_P) &= sort(r_1', \ r_2', \ \ldots, \ r_P') \end{aligned} \quad (4)$$

where P is the number of cut points, (r_1, r_2, …, r_P) are their positions, and L is the length of the row (or column), *i.e.*, $L = M$ (or N). Note that the sort procedure rearranges the values of the temporary variables in ascending order. For example, assume that the number of cut points is 4 and two consecutive elements in the key stream K_1 are 5 and 8. Then, the 5th and 8th rows will be selected, and the positions of the cut points will be determined as shown in Figure 2.

The computation of the positions of the cut points can be optimized, if the set (r_1, r_2, …, r_P) is computed in advance based on all possible values of $|E_i - E_{i+1}|$ and store them in a lookup table referenced by $|E_i - E_{i+1}|$. After selecting two rows (or columns), i and j, and determining the positions of the cut points r_1, r_2, …, and r_P, the multi-point crossover operation is performed by swapping RGB of pixels in the even segments of the two rows (or columns) i and j as shown in Figure 3. Note that, it is possible to swap odd segments instead of even ones.

E. The Mutation Unit

The mutation unit is the last stage in the encryption process. To obtain the final cipher image, the mutation unit masks the intermediate image resulting from the crossover stage with a random image using XOR operation. For this purpose, the sender and receiver must first agree on some randomly generated image and keep it secret. Then, the mutation unit XORs every pixel in the intermediate image with pseudo-random pixel from the secrete image selected by the values of the two key streams K_3 and K_4. For instance, the $(i, j)^{th}$ pixel in the cipher image is obtained by XORing the corresponding pixel in the intermediate image with $(p_i, q_j)^{th}$ pixel of the secret image, where $p_i \in K_3$ and $q_i \in K_4$. This process is explained further by means of a simple example of 4×4 image as shown in Figure 4.

F. Operation of the Proposed Encryption Algorithm

Given an RGB color image, where each one of the three color components (*i.e.*, red, green and blue) is represented as an $M \times N$ matrix, the general operation of the proposed encryption algorithm is described as follows:

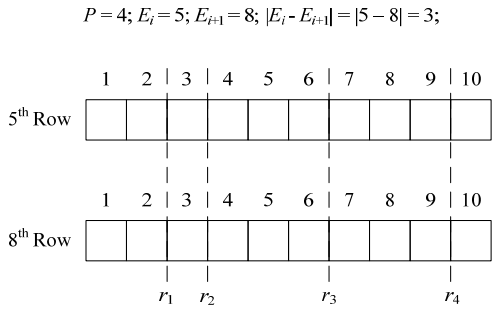
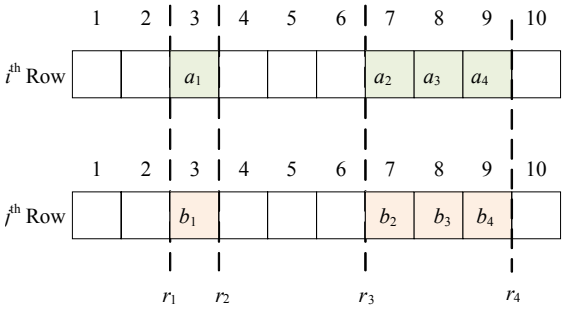
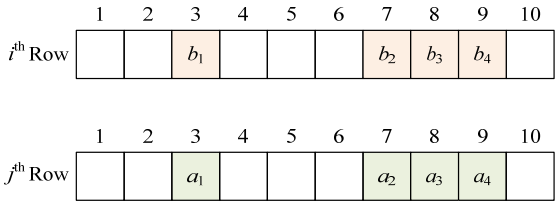


Figure 2. Example of determining the positions of cut points.



(a) Before Crossover



(b) After Crossover

Figure 3. Example of the crossover operation.

Step 1: Using the logistic map with key values (μ_1, x^o_1, μ_2, x^o_2, μ_3, x^o_3, μ_4, x^o_4) generate four chaotic sequences S_1, S_2, S_3 and S_4 where $|S_1| = |S_3| = M$ and $|S_2| = |S_4| = N$.

Step 2: Using sorted chaotic sequence method, obtain four key streams K_1, K_2, K_3 and K_4 where $|K_1| = |K_3| = M$ and $|K_2| = |K_4| = N$.

Step 3: Perform crossover operation row-wise on each individual $M \times N$ matrix using the key stream K_1.

Step 4: Perform crossover operation column-wise on each individual $M \times N$ matrix using the key stream K_2.

Step 5: Perform mutation operation on each individual

matrix by XORing each pixel in the intermediate matrices obtained by the crossover operation with a random pixel selected from corresponding matrix in the secret image based on the key streams K_3 and K_4. The secret masking image as mentioned previously is generated randomly and shared by the sender and the receiver.

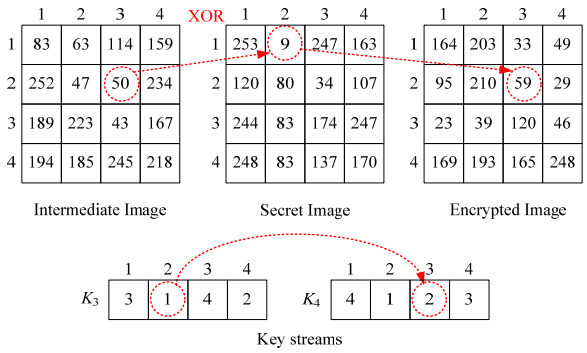


Figure 4. Example of the mutation operation of one color component of 4×4 block.

G. Decryption Algorithm

The decryption algorithm is identical to the encryption algorithm discussed above except that the order of the basic operations is reversed. That is, after generating the required key streams in steps 1 and 2, the decryption algorithm applies mutation operation first followed by column wise crossover operation then row-wise crossover operation.

III. EXPERIMENTAL RESULTS

To empirically assess the effectiveness of the proposed technique, we have carried out a number of experiments using MATLAB 7.7.0 (R2008b). These experiments include image encryption and decryption, histogram analysis of the plain and encrypted images, key space and sensitivity analysis and correlation coefficient analysis.

A. Image Encryption and Histogram Analysis

For this experiment, we have considered a 24-bit color image of size 256×256 pixels shown in Figure 5 (a), which is available at USC-SIPI image database in TIFF format [22].

This image is encrypted using the proposed technique with a key = {3.7158, 0.11, 3.89858, 0.25, 3.76158, 0.35, 3.8458, 0.552}. The resulting encrypted image is shown in Figure 5 (b). The histograms of red, green and blue channels of the plain and the encrypted images are shown in Figure 6. It is clear from this figure that the histograms of the encrypted image are uniform and significantly different from the histograms of the plain image. This result indicates that it is very difficult to use statistical analysis to attack the proposed encryption algorithm.

B. Key Space and Sensitivity Analysis

The secret key of the proposed technique is (μ_1, x^o_1, μ_2, x^o_2, μ_3, x^o_3, μ_4, x^o_4), where $\mu_i \in (3.569945672\ldots, 4]$ and $x^o_i \in (0,1)$, $i = 1,2\ 3,\ 4$, μ_i and x^o_i are both double precision. Since double precision can represent about 16 decimal digits, the key space of the proposed algorithm can be estimated as $(10^{14})^4 \times (10^{16})^4 = 10^{120} \approx 2^{398}$. Note that the range of μ_i is (3.569945672…, 4]; therefore a 14-digit precision is assumed. Thus, brute-force attacks on the key are computationally infeasible.

Figure 5. (a) original plain image. (b) encrypted image.

The brute-force attacks on the key streams, K_1, K_2, K_3, and K_4, generated by the quantification unit is also computationally infeasible as there are $L_i!$ combinations for each sequence, where L_i is the length of each sequence ($i = 1$, 2, 3, 4). Note that when these sequences are considered together to control the crossover and mutation operations, then the total possible combinations become $(L_1 \times L_2 \times L_3 \times L_4)! = (M^2 \times N^2)!$.

We have carried out a key sensitivity test using a key that is one digit different from the original key to decrypt the encrypted image. The resulting image is totally different from the original image as shown in Figure 7. This demonstrates that the proposed algorithm is very sensitive to any change in the secret key value.

C. Correlation of Two Adjacent Pixels

In this experiment, the correlation between two adjacent pixels in the plain image and encrypted image is tested. The following formula [19] has been used to calculate the correlation coefficients in horizontal and vertical directions:

$$C_r = \frac{N\sum_{j=1}^{N}(x_j \times y_j) - \sum_{j=1}^{N} x_j \times \sum_{j=1}^{N} y_j}{\sqrt{\left(N\sum_{j=1}^{N} x_j^{\,2}\right) - \left(\sum_{j=1}^{N} x_j\right)^2 \times \left(N\sum_{j=1}^{N} y_j^{\,2}\right) - \left(\sum_{j=1}^{N} y_j\right)^2}} \qquad (5)$$

where x and y are gray scale values of two adjacent pixels in the image, and N is the total number of pixels selected from the image for calculation. The experiment was performed by randomly selecting 4096 pairs of adjacent pixels from the plain image and the encrypted image shown in Figure 5, and then calculating the correlation coefficients using (5).

The results are shown in Figure 8. Frames (a) and (b) respectively show the distribution of two horizontally adjacent pixels in the original and encrypted images. Similarly, Frames (c) and (d) show respectively the distribution of two vertically adjacent pixels in the original and encrypted images. The correlation coefficients for the two adjacent pixels in the original and encrypted images are shown in Table I. These results show clearly that the distribution of two adjacent pixels in our results is more uniform than that reported in [20].

 ISBN: 978-1-61208-133-5

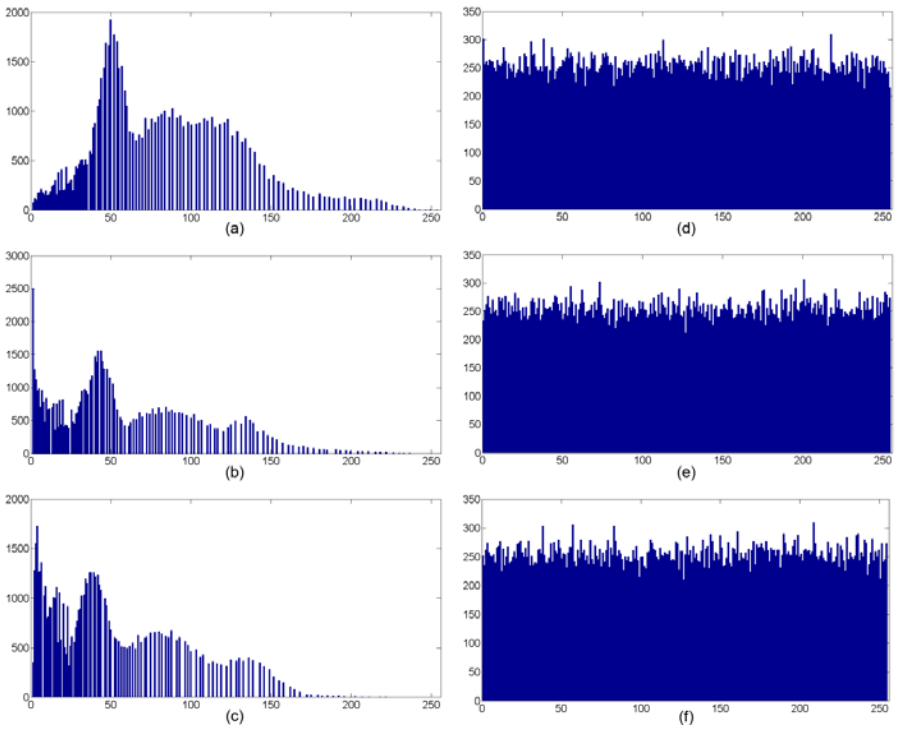


Figure 6. Histogram analysis: (a), (b), and (c) histograms of red, green and blue channels of the plain image shown in Figure 4 (a). (d), (e) and (f) histograms of red, green and blue channels of the encrypted image shown in Figure 3 (b).

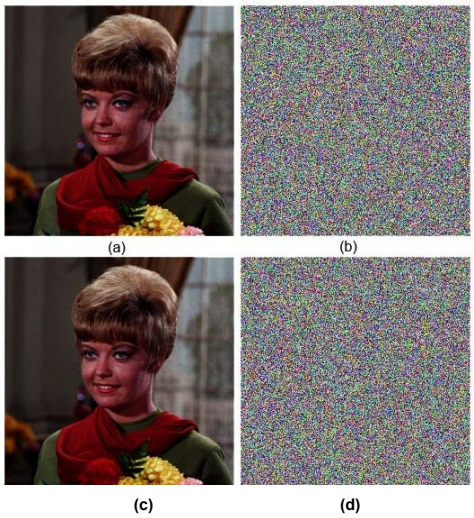


Figure 7. Key sensitivity: (a) plain image. (b) encrypted image. (c) decrypted image with key = {3.7158, 0.11, 3.89858, 0.25, 3.76158, 0.35, 3.8458, 0.552}. (d) decrypted image with key = {3.7159, 0.12, 3.89859, 0.26, 3.76159, 0.36, 3.8459, 0.553}.

TABLE I. CORRELATION COEFFICIENTS FOR TWO ADJACENT PIXELS IN PLAIN AND CIPHER IMAGES SHOWN IN FIGURE 4.

	Plain Image	Cipher Image
Horizontal	0.96784	0.00131
Vertical	0.95966	0.00012

In addition, we have carried out an extensive study of the correlation between plain image and its corresponding cipher image for several other images in the USC-SIPI image database. Results of this experiment are shown in Table II. It is clear that the correlation coefficients obtained by our proposed algorithm are very small which indicates that there is no correlation between the plain image and its corresponding encrypted image. Also, the correlation coefficients obtained by our algorithm are generally smaller than those obtained by the algorithm proposed in [20].

TABLE II. CORRELATION COEFFICIENTS BETWEEN SEVERAL PLAIN & CORRESPONDING ENCRYPTED IMAGES.

File Name	File Description	Size	Correlation Coefficient
4.1.01	Girl	256×256	-0.002601
4.1.02	Couple	256×256	-0.001354
4.1.03	Girl	256×256	0.005903
4.1.04	Girl	256×256	-0.005237
4.1.05	House	256×256	0.001596
4.1.06	Tree	256×256	-0.001793
4.1.07	Jelly beans	256×256	-0.001413
4.1.08	Jelly beans	256×256	0.002144
4.2.01	Splash	512×512	-0.000950
4.2.02	Girl (Tiffany)	512×512	-0.001311
4.2.03	Baboon	512×512	0.001832
4.2.04	Girl (Lenna)	512×512	0.000118
4.2.05	Airplane (F-16)	512×512	0.000396
4.2.06	Sailboat on lake	512×512	0.001111
4.2.07	Peppers	512×512	-0.001362
house	House	512×512	-0.000095

 ISBN: 978-1-61208-133-5

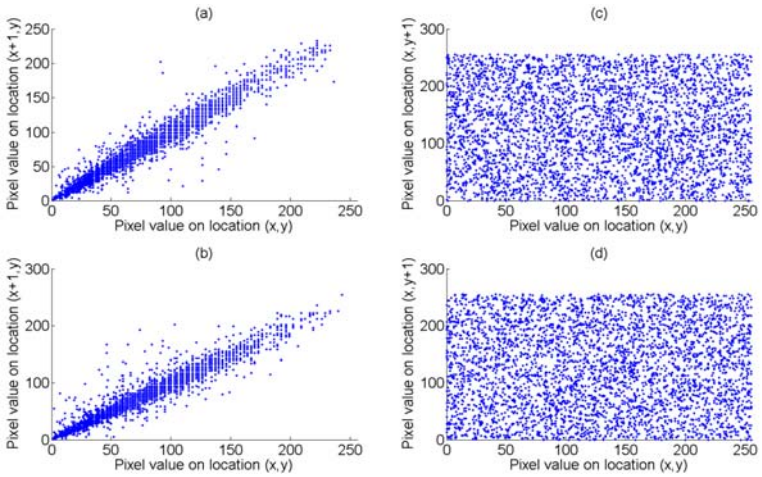


Figure 8. Correlation of two adjacent pixels: (a) and (b) distribution of two horizontally adjacent pixels in the plain and encrypted images presented in Figure 4. (c) and (d) distribution of two horizontally adjacent pixels in the same plain and encrypted images.

IV. Conclusion and Future Work

A novel approach based on chaos is presented in this paper for encrypting color images. The encryption/ decryption algorithms use a logistic map to generate four chaotic sequences which are converted to four key streams using sorted chaotic sequences method. The generated key streams are used to control multi-point crossover and mutation operations, which result in image confusion and diffusion respectively. Several experiments are conducted and the results show that the proposed approach is capable of generating encrypted images with uniform distribution of the pixel values and very low correlation coefficients of adjacent pixels. It is also very sensitive to any changes in the secret key values. We are now working on modifying the proposed approach to handle each color component independently and to consider the inter-color correlation for increasing the secrecy of the cipher image.

Acknowledgment

The authors would like to acknowledge the support of King Fahd University of Petroleum and Minerals (KFUPM) and the University of Ha'il, Saudi Arabia, during development of this work.

References

[1] M. Padmaja and S. Shameem, "Secure Image Transmission over Wireless Channels," in Proc. of the Int. Conf. on Compu. Int. and Multimedia Applications, (ICCIMA 2007), 2007, pp.44-48.

[2] B. Furht, D. Socek, and A. M. Eskicioglu, "Fundamentals of multimedia encryption techniques," in B. Furht and D. Kirovski, (eds.) Multimedia Security Handbook, CRC Press, Ch. 3, 2005.

[3] X. Li and D. Zhao, "Optical color image encryption with redefined fractional Hartley transform," Int. J. for Light and Electron Optics, vol. 121, no. 7, April 2010, pp. 673-677.

[4] C.J. Tay, C. Quan, W. Chen, and Y. Fu, "Color image encryption based on interference and virtual optics," Optics & Laser Technology, vol. 42, no. 2, March 2010, pp. 409-415.

[5] K. Martin, R. Lukac, and K. N. Plataniotis, "Efficient encryption of wavelet-based coded color images," Pattern Recognition, vol. 38, no. 7, July 2005, pp. 1111-1115.

[6] W. Chen, C. Quan, and C.J. Tay, "Optical color image encryption based on Arnold transform and interference method," Optics Communications, vol. 282, no. 18, Sept. 2009, pp. 3680-3685.

[7] Z. Lin and H. Wang, "Image encryption based on chaos with PWL memristor in Chua's circuit," in Proc. of the Int. Conf. on Commun., Circuits and Systems, July 2009, pp. 964-968.

[8] C. Fu and Z. Zhu, "A chaotic image encryption scheme based on circular bit shift method," in Proc. of the 9th Int. Conf. for Young Computer Scientists, (ICYCS 2008), Nov. 2008, pp. 3057-3061.

[9] W. Yanling, "Image scrambling method based on chaotic sequences and mapping," in Proc. of the 1st Int. Workshop on Education Tech. and Computer Science, (ETCS '09), March 2009.

[10] Y. Feng J. Li and X. Yang, "Discrete chaotic based 3D image encryption scheme," in Proc. of the Sympos. on Photonics and Optoelectronics, (SOPO 2009), Aug. 2009, pp. 1-4.

[11] G.M.B.S.S. Kumar and V. Chandrasekaran, "A novel image encryption scheme using Lorenz attractor," in Proc. of the 4th IEEE Conf. on Industrial Electronics and Applications, (ICIEA 2009), May, 2009, pp. 3662-3666.

[12] L. Luo, M. Du, B. He, F. Zhang and Y. Wang, "An image encryption algorithm based on spatiotemporal chaos," in Proc. of the 2nd Int. Congress on Image and Signal Proc., (CISP '09), Oct. 2009, pp. 1-5.

[13] C. Wei-Bin and Z. Xin, "Image encryption algorithm based on Henon chaotic system," in Proc. of the Int. Conf. on Image Analysis and Signal Proc., (IASP 2009), April 2009, pp. 94-97.

[14] D. Wang and Y. Zhang, "Image encryption algorithm based on s-boxes substitution and chaos random sequence," in Proc. of the Int. Conf. on Computer Modelling and Simulation, (ICCMS '09), Feb. 2009, pp. 110-113.

[15] N. J. Flores-Carmona, A. N. Pisarchik and M. Carpio-Valadez, "Encryption and decryption of images with chaotic map lattices," CHAOS Journal, vol. 16, pp. 1-6, 2006.

[16] Y. Honglei and W. Guang-Shou, "The compounded chaotic sequence research in image encryption algorithm," in Proc. of the WRI Global Congress on Intelligent Systems, vol. 3, May 2009, pp. 252-256.

[17] R. Matthews, "On the derivation of a chaotic encryption algorithm," Cryptologia, vol. 13, no. 1, Jan. 1989, pp. 29-41.

[18] Chong Fu, Zhen-chuan Zhang, Ying-yu Cao, "An improved image encryption algorithm based on chaotic maps," in Proc. of the 3rd Int. Conf. on Natural Computation, (ICNC 2007), 2007.

[19] N. Pareek, V. Patidar, K. Sud, Cryptography using multiple one-dimensional chaotic maps, Communications in Nonlinear Science and Numerical Simulation, vol. 10, no. 7, pp. 715–723, 2005.

[20] S. Liu, J. Sun, Z. Xu, "An Improved Image Encryption Algorithm based on Chaotic System", J. of Computers, Vol 4, No 11 (2009), 1091-1100, Nov. 2009.

[21] Y. Cao and Y. Fu, "Color image encryption based on hyper-chaos," in Proc. of the 2nd Int. Congress on Image and Signal Proc., (CISP'09), Oct. 2009.

[22] http://sipi.usc.edu/database/

 ISBN: 978-1-61208-133-5

The Impact of Corporate Culture in Security Policies – A Methodology

Edmo L. Filho[1], Gilberto T. Hashimoto[1], Pedro F. Rosa[1], Joao H. P. de Souza[2]
Universidade Federal de Uberlândia[1]
Uberlândia – Minas Gerais – Brazil
Universidade de São Paulo[2]
São Paulo – São Paulo – Brazil
edmo,gilbertot@algartelecom.com.br,
frosi@facom.ufu.br, joaohs@usp.br

Albene Teixeira Chaves
UNIMINAS/FACIMINAS – União Educacional Minas Gerais
Uberlândia – MG – Brazil
atchaves2@gmail.com

***Abstract*—Despite security policies, standards, awareness strategies and tools currently in place, employees are still being involved in risky behaviors that jeopardizes businesses. Meanwhile, although security policies are the cornerstone of well-designed security strategies, recent studies have demonstrated poor adherence or even negligence in accordance with the rules security policies specify. This observed behavior is related to the fact that business permeates different countries, cultures, and understanding human nature and culture is still a key success factor to information security not well-supported by established security policy development and deployment methodologies. As its outcome, this paper addresses a ubiquitous methodology to develop security policies considering the evaluation of culture and its impacts over security policy adherence.**

Keywords-security policy; awareness; culture, congruence model.

I. INTRODUCTION

As far as employees are using business networks to communicate, collaborate and access data, critical corporate information is being introduced into a broader environment that is more vulnerable and difficult to protect. Employees have available an increasing number of interactive applications and devices such as smart phones and handhelds. Besides that, individuals find it difficult to have a true boundary between work and home life [1] and they spend time sharing personal and business information on social networking sites [2]. As a result the frontiers between working inside or outside the company have completely disappeared and calling into question the traditional method to secure the perimeter.

The situation is further complicated by the increasing in the outsourcing activities. Although outsourcing can increase information security risks, in today's increasingly global competitive environment, most organizations have had to transform and outsourcing is a common strategy to reduce costs. In addition, this strategy can pose a company to different cultures in the same business process or a project.

The current scenario can lead to the extension and potential dilution of protection controls and an increase in the number of third parties given the same access rights and privileges as "natural" employees. Examples of common risks and mistakes [3]-[4]-[5] include (being not limited to): using unauthorized programs, misuse of corporate computers, unauthorized physical and network access, misuse of passwords and transfer sensitive information between work and personal computers.

Corporate culture is the total sum of the customs, values, traditions and meanings that make a company different from others. It is often called "the character of an organization" since it embodies the vision of the company's founders. The values of a corporate culture influence the ethical standards within a corporation, as well as managerial and security behavior.

Senior management may try to determine a corporate culture. They may wish to impose corporate values and standards of behavior that specifically reflect the objectives of the organization. Generally, these corporate values and standards of behavior are derived from the culture of the nation. As a consequence an obstacle to adherence of corporate culture naturally arises when companies extrapolate its frontiers by business expansion or acquisition of other companies. Regional and cultural differences will manifest themselves in a variety of security threats and business risks.

In addition, there will also be an extant internal culture within the workforce. Work-groups within the organization have their own behavioral quirks and interactions which, to an extent, affect the whole system. Roger Harrison's four-culture typology, and adapted by Charles Handy, suggests that unlike organizational culture, corporate culture can be imported. For example, computer technicians will have expertise, language and behaviors gained independently of the organization, but their presence can influence the culture of the organization as a whole.

Security Policies [8] are the cornerstone of a successfully information security architecture, because it provides clear instructions about information security and establishes management support. Policies are used as a reference point for a wide variety of information security activities including: designing controls into application systems and networks, establishing user access controls, conducting cybercrime investigations; and keep workers aware of punishment related to security violations.

 ISBN: 978-1-61208-133-5

However, in order to be effective security policies must be accompanied by an exhaustive and endless awareness program. The education and training helps minimize the cost of security incidents, and assure the consistent implementation of controls across an organization's information systems and business process.

Firewall [7]-[12] and Intrusion Prevention System (IPS) [6] are important building blocks of a security topology. Network and/or security administrators often rely on their services to protect against the majority of threats and to enforce security policies. However, security is not keeping up with technological and social changes in the workplace, there are ways to circumvent or ignore enforcement rules and, depending on the way security policies are deployed, people's culture has strong influence on adherence or not of these security policies.

Even in experienced international companies, many well-meaning universal applications of management theory ended up being a fiasco when these practices were faced with other cultures. It is not different when considering a security policy. What performs well in a country company may not in another country. Security controls need to be workable in a variety of environments and developed, implemented and supported with people's behavior in mind.

The goal of this paper is to propose and evaluate a method to develop and deploy security policies considering the diversity of culture that companies may confront. The groundwork of the methodology is built over an integrated and consistent approach. As far as we know, to evaluate the impacts of people's culture in the security policy development and deployment is a hard task. In Section II, III, and IV the necessary background to develop the methodology is presented. In Section V, the methodology is described and detailed. In Section VI, the most important results are shown. Finally, section VII presents some concluding remarks and suggestion for future work.

II. CORPORATE CULTURE BACKGROUND

Culture is a common system of meanings, which shows what people should pay attention, how should act and what to value. Strong culture is said to exist where staff respond to stimulus because of their alignment to organizational values. In such environments, strong cultures help companies operate with efficiency, cruising along with outstanding execution and perhaps minor tweaking of existing procedures.

Conversely, there is weak culture where there is little alignment with organizational values and control must be exercised through extensive procedures and bureaucracy. Considering security policies, this control is mainly exercised through application of enforcement rules by configuration and usage of security appliances.

Where culture is strong people do things because they believe it is the right thing to do, however there is a risk of another phenomenon, "group think". This is a state where people, even if they have different ideas, do not challenge organizational thought, and therefore there is a reduced capacity for innovative thoughts. This could occur, for example, where there is heavy reliance on a central charismatic figure in the organization, or where there is an evangelical belief in the organization's values, or also in groups where a friendly climate is at the base of their identity (avoidance of conflict). In fact group think is very common, it happens all the time, in almost every group. Members that are defiant are often turned down or seen as a negative influence by the rest of the group, because they bring conflict.

Of all the data losses reported by the UK Government after the nefarious case of the leaking of personal details of 25 million people in a single incident involving the UK Government's Revenues and Customs Department (HMRC), 95% is due to cultural factors or the behavior of people whilst only 5% is believed to be due to technology issues [13].

Every culture distinguishes itself from others by means of specific solutions to specific problems [14]. The categories of problems can be viewed under three aspects: problems that arise from people's relationship, passage of time and environment. Due to its main objective the article focus on people's relationship and the five guidelines to understand the ways humans relate to each other:

Universalism versus Particularism – In the universalism approach is possible to define what is good and what is bad and this criterion is always applicable. In the Particularism culture more attention is given to the obligations of the relationships and specific circumstances. For example, instead of assuming that a good law should always be followed, the Particularism reasoning is that friendship has special obligations and hence may be a priority.

Individualism versus Collectivism – People see themselves primarily as individuals or basically part of a group? Moreover, it is more important to concentrate on the individual so that they can contribute to the community, or is it more important to consider the community first?

Neutral or Emotional – The nature of our interactions should be objective and impartial or is it acceptable to express emotion? In several places the business relationships are generally tools for reaching an objective. Emotions are avoided in order not to compromise discussions. However, several cultures consider the manifestation of emotions a natural part of business.

Specific versus Diffuse – When the person is engaged in a business relationship, there is real and personal contact rather than the specific relationship recommended in the contract.

Achievement versus Attribution – Achievement means that the person is judged by his recent activities and history. Attribution means that the status is conferred by birth, kinship, gender or age, but also for their connections, who you know and professional training.

Innovative organizations need individuals who are prepared to challenge the status quo—be it groupthink or bureaucracy, and also need procedures to implement new ideas effectively.

Most organizations are facing some kind of transformation and traditional cultures are facing the impacts of globalizations and being rebuilt, including perceptions and behavior towards security. If not addressed clearly, cultural

 ISBN: 978-1-61208-133-5

changes can cause uncertainty and doubts in employees or third parties, impacting adherence to security policies.

The Congruence Model [15] is a methodology to address the cultural and business changes. The methodology deals with changes to both formal and informal cultures as well as the infrastructure and business processes. The congruence approach is already being applied by the security community [13].

III. SECURITY POLICY BACKGROUND

In spite of organization's size, their businesses, or the extent to which it uses technology, information security is an important matter that should be addressed by explicit policies. However, the settlement of security policies is itself based on a specific framework that requires methodology to write, structure, effective review, approval, enforcement and awareness process [8]-[9].

Security policies are high-level statements that provide guidance to those who must make present and future decisions. An information security policy document is vital for many reasons. Beyond the definition of roles and responsibilities for workers, partners, suppliers, a policy document sensitizes them to the potential threats, vulnerabilities and problems associated with modern information systems. A consistent awareness program is fundamental to achieve the security policy goals. Education and training helps minimize the cost of security incidents, and helps assure the consistent implementation of controls across an organization's information systems.

The well-known methodologies for developing security policies [8] do not address the issue of corporate culture in depth, only guidance to make policies compliance to corporate culture is provided. Many obstacles to compliance of security policies arise when these policies are deployed as canned goods in different cultures. For example, is difficult to understand some of the cultural, religious and societal pressures of the India's caste system and its implications: orders are expected to be obeyed and the rules required at work will always be less important than behavior deep-rooted over countless generations.

Figure 1 depicts the approach mainly used around the world to develop and deploy security policies.

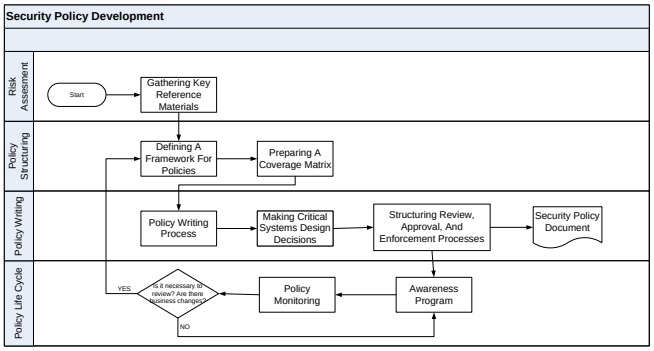


Figure 1. Security Policy development fluxogram.

Legal aspects are also an important aspect of security policy life cycle. Each country has its own legal system that must be evaluated before developing or deploying an imported policy.

Currently, the effectiveness of security policies considering data leakage is an important concern. Regardless of the type or mode of data leakage, recent research [9] reveals that one out of four companies does not even have a security policy and for businesses with policies, the findings reveal a significant gap between the beliefs of security staff regarding employee compliance and the actual behavior of them. The reasons why employees knowingly overlook or bypass security policies and put corporate data at risk are mainly a result of a failure to communicate security policies and create an awareness behavior in accordance with local culture.

The proposed methodology presents an adaptive strategy to security policy development and awareness program based on the analysis of the culture throughout the five guidelines to understand the ways humans relate to each other and the application of the Congruence Model.

IV. AWARENESS BACKGROUND

The data loss issue encompasses everything from confidential information about one customer being exposed, to strategic files of a company's product being sent to a competitor. Whether deliberate or accidental, data loss occur any time employees, third-party, or other insiders release sensitive data about customers, finances, intellectual property, or other confidential information in violation of company policies and regulatory requirements.

Beyond the methods used to educate about information security the approach needs to sensitize employees to the types of attacks that they might encounter. Employee thinking needs to be stimulated via real-world examples. The awareness program must include topics about the threats, and on how to secure "your own" environment. The awareness approach is a key success factor to the development of a security framework and shall be measured. Surveys are a traditional method of measuring awareness [11]. However, measuring attitudes and awareness have a poor correlation with behavior.

An adaptive strategy to the awareness program is based on marketing, psychology principles, and a qualitative information security awareness scorecard [10]. Blogs and social network forums integrated with monitoring process are used to energize employee involvement. The expected results may be evaluated through internal quizzes considering a rewarding process.

Repetition of information security policy ideas is essential. Repetition impresses users and other audiences with the importance that management places on information security. Education also prevents workers from saying "I never heard about that."

The channels used to express a policy will determine how the policy should be written. For example, if videotape will be used, then an abbreviated colloquial style should be employed. If a policy document will reside on an intranet web server, then a more graphic and hypertext-linked style is

 ISBN: 978-1-61208-133-5

appropriate. If policies will be issued through a series of paper memos, then short and concise text-oriented expressions will be required. The ways that the organization currently uses or intends to use information security policies should also be examined [8].

The education process must consider the third parties as they are now always present somewhere in the organizations. Effective education is difficult in multicultural an outsourced environments where suppliers are growing rapidly and hiring hundreds of new employees to support companies' requirements. The cost and time spent to education tends to prove its benefits in a short time.

V. PROPOSAL

Based on the analysis of the culture throughout the five guidelines to understand the ways humans relate to each other, it is possible to define four types of companies related to corporate culture: the Family, Eiffel Tower, Guided Missile and Incubator [14]. Figure 2 summarizes the relationship between the employees and their notion of company.

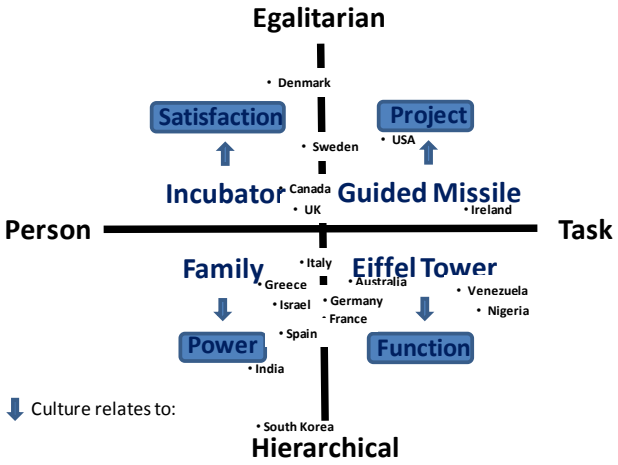


Figure 2. Cultures of the company.

A comprehensive study on the types of culture related to companies and how to determine the type through research can be found at [14] and Trompenaars' database. Examples from the 16 questions used to measure corporate culture include how to measure the hierarchy level and conflicts.

According to the type of corporate culture that will be generally derived from the culture of the nation, security policies will have more adherence or not depending on the strategy to develop and implement.

Another important study of how values in the workplace are influenced by culture can be found at [16]. For example, "*Brazil's highest Hofstede Dimension is Uncertainty Avoidance (UAI) is 76, indicating the society's low level of tolerance for uncertainty. In an effort to minimize or reduce this level of uncertainty, strict rules, laws, policies, and regulations are adopted and implemented. The ultimate goal of this population is to control everything in order to eliminate or avoid the unexpected. As a result of this high Uncertainty Avoidance characteristic, the society does not readily accept change and is very risk adverse*".

For the purpose of this article the factors thinking, learning and change (responses) should be evaluated in order to determine the impacts in the development and awareness process of security policies.

The Family culture deals more with the intuition than rational process. It focuses on the development of people over people's performance. The knowledge is based on trial and error and the individual is more important than the task. The change process is essentially political and top down. The mentors and managers are important actors in the learning process. Pattern examples of national corporate culture include France, Spain, India and Japan [14].

The Eiffel Tower considers that in order to perform his functions the professional must accumulate the necessary skills and always keep evolving. Human resources are evaluated like financial capital and cash. The change process in this culture is always slow, executed through rules of change and considering a formal process. This kind of culture doesn't adapt well to turbulent environments. Companies with this culture profile generally avoid and resist to changes. Examples include Germany, Holland and Denmark [14].

The Guided Missile culture reviews its objectives through a constant feedback process. Then it is a circular and not linear culture. It rarely changes its main objective and everything necessary is done to keep and achieve the objectives. The directions are corrective and conservative. The learning process includes the personal contact and interactions within a group. It has a practical approach instead of theoretical and focuses on the problems instead of discipline. Changes are fast in this kind of culture, as the objectives moves new groups of work are formed to support the new demands and the old groups are diluted. This culture tends to be individualist. Examples include Canada, USA and United Kingdom [14].

The Incubator culture is based on the idea that people's satisfaction is more important than the company itself. To tolerate the company the main people's objective is to serve the incubator for self-expression and self-satisfaction. Companies in this kind of culture often operate as an intense emotional environment, having a minimal hierarchical structure and the authority is strictly personal. When the members are in harmony the change process is usually fast and spontaneous. This culture is creative, although doesn't survive to changes on the demand patterns. Sweden is a common example of this culture [14].

Figure 3 depicts the relation between the types of culture and the level of difficult to develop and deploy security policies.

As the authors could observe during the process of development and deployment of security policies in the last 10 years considering wholesale, telecom, data center, agribusiness, transportation and real estate companies – some of them multinational – formal cultures tends to facilitate the whole process. However, exceptions can happen.

 ISBN: 978-1-61208-133-5

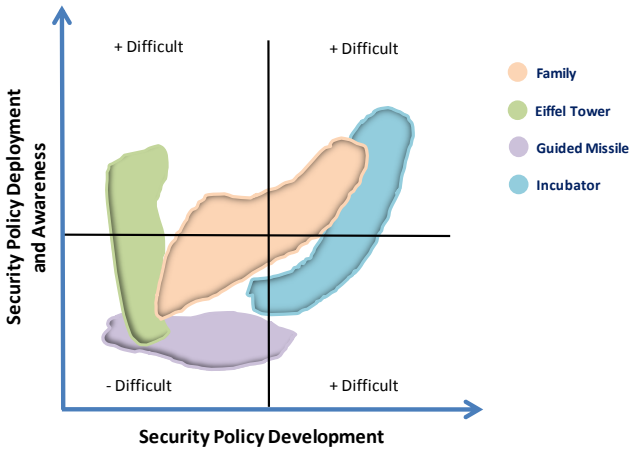


Figure 3. Impacts of culture in Security Policy process.

Table I depicts the relation between corporate culture, the security policy development, awareness process and the necessary steps to include in the traditional security policy development process after the analysis of the culture pattern predominant in the organization.

The pattern analysis must follow or be included in the Risk Assessment phase. The guidelines are an adaptation of Trompenaars' "how to manage and be managed by the corporate culture" to the Security Policy (SP) development and deployment phases .

TABLE I. CORPORATE CULTURE AND SECURITY POLICY

Culture	SP Development	SP Deployment	Awareness
Family	• Top Down approach[1]; • Conquer the corporate leaders[2]; • Evangelize the CEO and leaders[3]; • Abuse of risk examples to convince them[4]. • Make the leaders feel more powerful and with more control through SP's; • Show to the leaders the **impacts of errors** more than the advantages of SP's	• 1 and 2; • Make people feel as the "owner of the process[4]"	• 1 and 4; • Conquer the team's members; • Repetition is a must[6];
Eiffel Tower	• 1 and 2; • Show the leaders the **advantages** of SP's[5];	• Decentralized; • Usage of Project Management methodology.	• The awareness program should sell SP as status;
Guided Missile	• Top Down and decentralized approach. • 5; • Present the avoided risks in financial numbers;	• Usage of project management methodology.	• The awareness program must show results in financial numbers and impact over "salary".
Incubator	• Decentralized approach. • Discover the most influent individuals of the network and evangelize them.	• Make people feel SP as innovative and important to their objectives. • 6.	• The awareness program should create "challenge conditions"; • 6.

There are different methods to apply the guidelines that will depend on the available time and resources. Understanding corporate culture, security professionals will have strong likelihood to establish security policies integrated into the organization's culture.

VI. EVALUATION

In spite of the methodologies already in place and the proposal, developing and deploying security policies is not a easy project even in corporate cultures classified as guided missile. The authors' experience and observation through the last 10 years showed that evangelization is a good strategy. Being nearest the employee and third party bring results in a short time than being far from.

Security metrics is a nascent discipline with more questions than answers [17]. Moreover, the choice of security metrics may lead to a false sense of security or otherwise misdirect security efforts and strategy. Measurement does not guarantee safety as usually the metrics are related to past events.

The best way to measure the effectiveness of the proposal is to observe human behavior towards information security. The number of incidents per amount of employees is a suggested metric to be monitored. Then, it is necessary to put in place tools to monitor frauds and other incidents.

Table II presents examples of the most predominant type of culture [14], the difference between the number of end users and the number of decision makers who are aware of a policy regarding acceptable use of company resources [9].

Why is there a lack of connection between policy makers and the employees who must conform to policies every day? According to the survey results [9] one crucial reason is a lack of direct and consistent communication, and 11% of employees say that security policies were never communicated to them or that they were never educated about the policy.

TABLE II. THE DISCONNECT BETWEEN END USER AND SECURITY POLICY AWARENESS

Country	End User	Decision Makers	Culture
USA	45%	76%	Guided Missile
France	49%	74%	Eiffel Tower

 ISBN: 978-1-61208-133-5

Italy	46%	77%	Family
India	54%	77%	Family
UK	50%	71%	Incubator

Europe had the highest prevalence of this belief, where the United Kingdom – Incubator – (25%) and France – Eiffel Tower – (20%) far exceed the global average. Germany – Eiffel Tower – also has a high percentage of employees who claim that IT never communicates security policies to them (16%).

The survey results associated with the analysis of corporate culture support the author's proposal and the relation presented in Figure 3.

VII. CONCLUSION AND FUTURE WORK

The organization's performance rests upon the alignment of each of the components – the work, people, structure and culture – where the higher the congruence between them, higher will be the performance of the organization. Information security risks may damage the desired company results and security policies are the cornerstone of a security framework – the starting point to avoid or minimize damages.

However, the methodologies already in place to develop security policies didn't consider the impacts of culture in adherence to them. Information security risks are a critical issue for companies, as the number of incidents continues to increase. Whether it's a malicious attempt, or an inadvertent mistake, these risks can decrease a company's trademark, reduce shareholder value, and blemish the company's goodwill and reputation. Furthermore, applied security technology is not enough once the human factor is essentially the weakest part considering corporate culture.

The article proposes a starting point to discuss and evolve the impacts of culture in security policies adherence. The article also presents a methodology which, in a nutshell, is to include in the Risk Assessment phase the verification of the predominant pattern of culture in the organization, and after that follow the proposed guidelines to the specific culture in order to achieve success. The proposal is likely to progress in conjunction with further research in this area.

Future work is necessary to investigate an evolution of the corporate culture analysis considering automation of the process through the usage of OWL (Web Ontology Language). Ontologies have been used to knowledge management and organization [18], for example, in Artificial Intelligence (AI) ontologies have been used to explicitly declare the knowledge embedded in knowledge-based system and to facilitate knowledge share and re-use. Another challenge is to develop an effective method to evaluate employees' adherence and commitment to security policies considering the established corporate culture.

REFERENCES

[1] J. Walker, R. Samani, M. Henshaw, A.Yeomans, P. Wood, T. Holman, M. Westmacott, A. Davis, O. Ross, A. Sehmbi, L. Orans, and S. Janes, "CW Security Think Tank: How to prevent security breaches from personal devices in the workplace", 12/07/2010, http://www.computerweekly.com/Articles/2011/01/05/244377/CW-Security-Think-Tank-How-to-prevent-security-breaches-from-personal-devices-in-the.htm 03/24/2011.

[2] E. L. Filho, G. Hashimoto, P. Rosa, and J. Machado, "A Security Framework to Protect against Social Networks Services Threats" Proc. IARIA Symp. The Fifth International Conference on Systems and Networks Communications (ICSNC) 2010, IEEE Press, Aug. 2010, pp. 189-194, doi: 10.1109/ICSNC.2010.36.

[3] W. Willinger, R. Rijaie, M. Torkjazi, M. Valafar, and M. Maggioni, "Research on online social networks: time to face the real challenges," ACM SIGMETRICS Performance Evaluation Review, Dec. 2009, pp. 49-54, doi: 10.1145/1710115.1710125.

[4] CISCO (2008), "Data Leakage Worldwide: The High Cost of Insider Threats", http://www.cisco.com/en/US/solutions/collateral/ns170/ns896/ns895/white_paper_c11-506224.html 03/23/2011.

[5] CISCO (2008), "Data Leakage Worldwide: Common Risks and Mistakes Employees Make", http://www.cisco.com/en/US/solutions/collateral/ns170/ns896/ns895/white_paper_c11-499060.html 03/23/2011.

[6] M. Rash, A. Orebaugh, G. Clark, B. Pinkard, and J. Babbin, "Intrusion Prevention and Active Response: Deploying Network and Host IPS", SYNGRESS PUBLISHING, p. 4-20, 2005.

[7] E. L. Filho, "Arquitetura de Alta Disponibilidade para Firewall e IPS Baseada em SCTP", Department od Computer Science, Federal University of Uberlândia, p. 50-59, 2008.

[8] C. Wood, "Information Security Policies Made Easy" 10th Edition, INFORMATION SHIELD, Inc., 2005.

[9] CISCO (2008), "Data Leakage Worldwide: The Effectiveness of Security Policies" http://www.cisco.com/en/US/solutions/collateral/ns170/ns896/ns895/white_paper_c11-503131.html 03/23/2011.

[10] G. Stewart, "Maximising the Effectiveness of Information Security Awareness Using Marketing and Psychology Principles" Department of Mathematics, Royal Holloway, University of London, RHUL-MA-2009-2, Feb. 2009.

[11] C. Roper, J. Grau, and L. Fischer, "Security Education, Awareness and Training, SEAT from Theory to Practice". ELSEVIER BUTTERWORTH-HEINEMANN, 2006.

[12] E. L. Filho, G. Hashimoto, and P. Rosa, "A High Availability Firewall Model Based on SCTP Protocol," Proc. IARIA Symp. Systems and Networks Communications (ICSNC 08), IEEE Press, Dec. 2008, pp. 202-207, doi: 10.1109/ICSNC.2008.63.

[13] C. Colwill, "Human Factors in Information Security: The Insider Threat – Who Can You Trust These Days?", ELSEVIER, Inform. Secur. Tech. Rep. (2010), doi: 10.1016/j.isrt.2010.04.004.

[14] F. Trompenaars and C. Hampden-Turner, "Riding The Waves of Culture: Understanding Diversity in Global Business", 2th Edition, MCGRAW-HILL, 1998.

[15] O. Wyman, "Congruence Model: A Roadmap for Understanding Organizational Performance", http://www.oliverwyman.com/ow/pdf_files/Congruence_Model_INS.pdf 03/23/2011.

[16] G. Hofstede, "Cultural Dimensions", http://www.geert-hofstede.com/ 03/23/2011.

[17] C. Nelson, "Security Metrics An Overview", ISSA Journal, August 2010, pp. 12-18.

[18] E-S. Abou-Zeid and J. Molson, "Towards a Cultural Ontology for Interorganizational Knowledge Processes", IEEE Press, Jan. 2003, vol. 1, pp. 8c, Proceedings of the 36th Annual Hawaii International Conference on System Sciences (HICSS'03), doi: 10.1109/HICSS.2003.1173645.

 ISBN: 978-1-61208-133-5

Future Architectures for Public Warning Systems

Michelle Wetterwald, Christian Bonnet, Daniel Camara
EURECOM, Sophia Antipolis, France
firstname.lastname@eurecom.fr

Sebastien Grazzini
Eutelsat, Paris, France
sgrazzini@eutelsat.fr

Jérome Fenwick
Groupe SYNOX, BALMA, France
jfenwick@groupe-synox.com

Xavier Ladjointe, Jean-Louis Fondere
Thales Alenia Space, Cannes, France
[firststname.lastname]@thalesaleniaspace.com

***Abstract*— Natural disasters have often made the headlines in the past years. As a consequence, many actions have been started by the public authorities to reduce the damages and the number of casualties. In that objective, the French project RATCOM aims at developing an alert system in case of coastal tsunami. Its downstream components will propose reliable and efficient communication systems to relay the alert. In parallel to the integration of the existing technologies in the project demonstrator, a survey analysis has been performed to identify the communications technologies and networks which are in preparation but not yet operational, and which will increase the efficiency and quantity of individuals reachable by the future population alert networks. Each of these technologies is not sufficient by itself, but their combination will improve drastically the efficiency of the alerting global system.**

Keywords-tsunamis; alerting; public warning system; broadcasting networks.

I. INTRODUCTION

Natural disasters and the thousand of casualties they usually cause raise a major concern at public authorities' level. A milestone event in this field was the Indian Ocean tsunami that happened in December 2004. This event raised the question of how to improve the protection of the population and prevent so many deaths. In fact, the main answer relies in the fast distribution of the information: information about the best behaviour to adopt in case of a disaster, and more importantly, information about the imminent arrival of a disaster.

The South East part of the French Mediterranean coast has been identified by the experts as the potential location for small-sized tsunamis. These could be caused by major landslides in the underwater area, few kilometres away from the coast. One of these tsunamis occurred in 1979 in front of Nice and made several million euros' worth of damage. As a prevention tool, the RATCOM project [1], started in 2009, aims at developing an alerting system towards the public safety professionals on one hand and the citizens on the other hand. The project is organized around two major components: the upstream component and the downstream component. The upstream component is responsible to monitor the events occurring at the sea and report the risk level to a control centre. The control centre then makes the decision to generate an alert and forwards it to the downstream component which is responsible to disseminate the warning within the shortest time frame possible. The best-known method to broadcast this type of information is by triggering the operation of alert sirens. However, more modern technologies exist nowadays that can help reaching a larger quantity of people. The RATCOM downstream component aims at identifying and setup a network linking these technologies into a single framework. Some of these technologies are currently operational and will be included in the final project demonstration. An additional survey paper activity has been conducted to identify other technologies that are not ready at this point in time, but in the future may become relevant to our warning system. The suitability of these techniques to be included in the project global downstream component has been analyzed. The final objective of this study is to draw up an inventory of the technologies and networks that are not yet operational, but are relevant in the context of a future public warning system.

This paper is organized as follows. The second section considers systems of communication close to their deployment phase, with a probable delay of less than three years, and having the ability to be connected to a warning system in the medium term. Derived from digital broadcast systems, the DVB-SH (Digital Video Broadcasting for Handheld Satellite) uses the coverage capabilities of satellite networks. Satellites offer also the possibility to provide redundant connections and improve the strength of the whole system. WiMAX (Worldwide Interoperability for Microwave Access) is a new technology which can provide service to larger areas than Wi-Fi (or IEEE 802.11), whose concept is somewhat similar. The new capabilities and possibilities of connecting current and upcoming mobile cellular networks are discussed. In the third part are presented prospective technologies that are currently being defined and standardized, but which will be effectively operational in a period longer than five years. They are essentially the Public Warning System integrated in mobile phone networks, broadcast technology in these global

 ISBN: 978-1-61208-133-5

networks and Vehicular Networks. Finally, we draw our conclusion to this study in the last section.

II. USING STATE OF THE ART TECHNOLOGIES

In this chapter some technologies that are still in their early phase of deployment or will be in the next two or three years are described. The section focuses only on the technologies which seem to be relevant for the broadcast of warning messages to the public. These technologies (satellite systems, WiMAX and CBS / LTE) have the ability to quickly reach a greater proportion of the population, including people on the move. They constitute a part of the population who could not be informed by more traditional methods such as the television. For each of the technologies is presented a fast description then an analysis of why it is relevant for the public warning and for the interconnection with our alert distribution system is performed.

A. Satellite systems

Satellite systems can be used in two different manners: first manner consists in broadcasting directly information to handheld devices. This is the DVB-SH technology. The second manner consists in strengthening the whole system by operating connections redundantly with terrestrial links which may be at risk.

The DVB-SH is a standard derived from the DVB-H standard to distribute broadcast video, audio and data to mobile devices such as mobile phones. Mobile TV is definitely set to become the next major media market of tomorrow. The publication in November 2004 of the DVB-H standard, seen by analysts as a possible solution for providing mobile television, was the starting point of a series of work on this new mode of television programs consumption. While DVB-H is designed primarily for use in the UHF terrestrial broadcasting only, the DVB-SH tries to exploit the S band, where there are opportunities for Mobile Satellite Services (MSS). Thus, this standard, created specifically for distributing content via satellite in mobility situation, makes a major innovation in the telecommunications world by satellite: it enables the addition of a network of terrestrial repeaters, called CGC (Complementary Ground Component) to complement the satellite coverage.

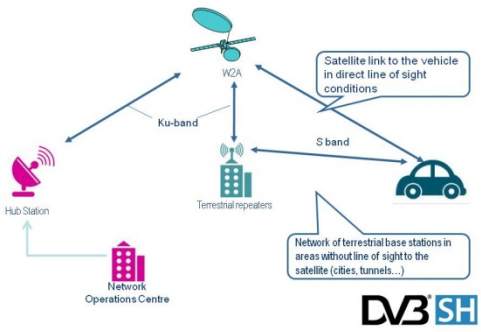


Figure 1. DVB-SH broadcast network architecture

One of the major problems in terms of warning systems is to quickly reach a large number of people, whether they are in a mobility situation or not and, if possible, at a reduced cost. The DVB-SH broadcast network meets these criteria through the variety of devices able to receive the signal (mobile phones, vehicular terminals, etc.) as well as through the possibility of sharing the same flow between a large number of people via the satellite. Accordingly, it becomes quite interesting to interface our alerting network and demonstrate the potential offered by hybrid broadcast architecture. Three warning systems are considered in the framework of this study: first, the broadcast of video / audio warning on TV / Radio mobile satellite devices, second, the broadcast of a detailed report about the alert to the TV / Radio mobile satellite devices for interested people and finally the triggering via the satellite of fully autonomous and easily installable alerting peripherals (e.g. on beaches).

The first two cases are closely related; they actually consist in stopping the Radio / TV programs to replace them with the tsunami warning. The procedure is very similar to what exists for the abduction alerts on TV but would be applied to mobile TV. The major innovation lies in the fact that, simultaneously with the program stop, an alert is sent as data traffic and the user can view this bulletin with the same device. This bulletin can be updated to indicate, for example, the end of the alert. The third case is the satellite triggering of alerting fully autonomous and of easy maintenance devices. Indeed, with DVB-SH, it is possible to receive the signal with a small omnidirectional antenna and one can imagine devices (sirens, billboards...) independent of terrestrial communications networks that can be triggered remotely via a satellite signal. This new type of installation would benefit from reduced costs because no wired connection would have to be planned and its assembly and disassembly in urban areas would be simplified. The positioning of the devices would be defined only by taking into account the risk factor and not the availability of a terrestrial network. This freedom enables an improvement of the efficiency of the devices. Moreover, such a warning system would benefit from a complete independence from terrestrial communications networks which can be damaged by natural disasters.

As the second manner to use this technology, the Ku band satellite connection systems or VSAT (Very Small Aperture Terminal) serve redundant network nodes or quickly connect fixed subscribers or isolated alert networks. These systems make use of satellite dishes with a diameter less than 3 meters and terminals (or modem) that allow bidirectional communications. They provide the following intrinsic advantages: a minimum ground infrastructure, an immediate area covering several alert networks from one or several countries at the same time and a simple and rapid deployment. With a satellite link, it is possible to connect either a comprehensive warning system, in which case we preferably connect via the satellite the control node responsible for the warning broadcast on this network, or a specific node of the warning network, such as a siren, a VMS (Variable-Message Sign) or any other equipment that would require redundancy or that just needs to be connected to the

 ISBN: 978-1-61208-133-5

network. Such a node can be a warning system sharing the same satellite link or a single important subscriber connected to the satellite endpoint.

The choice of the satellite as the transmission system component on the downstream component is justified by the desire, first, to avoid congestion or interruption of the terrestrial networks which can become harmful in case of a tsunami, and, secondly, to be able to quickly connect a warning system or a single isolated but important subscriber. In this case, the satellite will thus be used for the redundancy of critical network nodes (connected to a warning system or a subscriber of critical importance in the decision process), to connect the system to an existing warning network, or just to quickly connect a siren or isolated warning sign.

B. WiMAX networks

The WiMAX technology is standardized by the IEEE (Institute of Electrical and Electronics Engineers) under IEEE 802.16 and addresses several objectives: fixed mobile convergence, higher flow rates, compliance with quality of service constraints, etc. Compared with the architecture of conventional cellular systems such as EDGE (Enhanced Data rates for GSM Evolution) or UMTS (Universal Mobile Telecommunications System), the architecture of a WiMAX network is based on components that are intended to remain close to the Internet standards, as pictured in Figure 2.

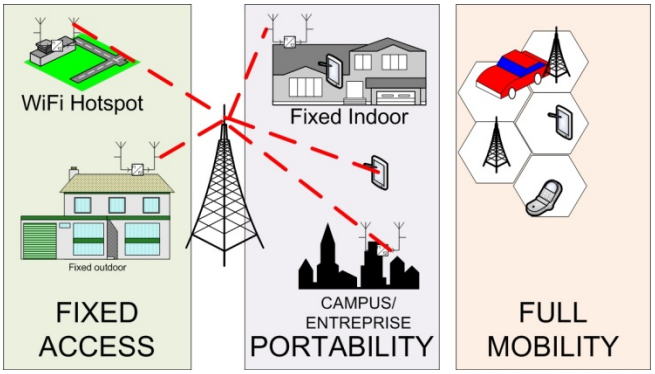


Figure 2. WiMAX Mobile Environment

The standard provisions various types of communications. For point to point transmission, it aims to link transmission points separated by a few dozens kilometres for the multiplexing of IP traffic with the support of differentiation and service guarantee. This type of application is similar to radio-relay transmission while it provides the spectral efficiency and intelligent management of IP traffic. It comes in support of network deployments that would not be economically viable if done in wire line technologies. The systems for point to multipoint transmissions without mobility provide the Internet IP traffic from a connection point of the wired network to a group of buildings or homes through the radio interface. User equipments within the buildings are basically PCs that receive a service equivalent to an ADSL (Asymmetric Digital Subscriber Line) access. This standard thus targets to address the so-called "white areas" in which a typical deployment of ADSL based on a wired infrastructure would be too expensive to setup. In the point to multipoint transmission with mobility version, the WiMAX radio signal terminates directly on the terminal of the final user. This system can accommodate the wireless ADSL users, but also PC terminals (usually laptops) for a mobile Internet access.

The WiMAX offers a continuous connection for the transfer of IP packets. Accordingly, it can support any type of warning system based on data transmission. An interesting feature is its ability to support a Broadcast / Multicast mode called MCBCS (Multicast and Broadcast Services). In the same perspective as the 3GPP MBMS (Multimedia Broadcast/Multicast Service) technology, the WiMAX plans to provide broadcast services in geographical areas managed by the system. In a Multi-BS (MBS) system, several BSs located in the same geographical area, called MBS_ZONE, can transmit the same broadcast / multicast messages simultaneously on a single radio channel. It should be noted that a BS may belong to several MBS_ZONEs. A mobile terminal that registers for an MBS service can receive information from all the BSs of the MBS_ZONE without having to register with a specific BS of the area. In addition, it can receive the MBS signals from several BSs simultaneously for an improved reception quality. This broadcast service enables the usage of the WiMAX technology as a potential support for public alerting messages.

C. 3G and LTE cellular networks

The CBS technology allows sending through the GSM (Global System for Mobile communications) network one or more small messages to all the mobile phones located within a specific area covered by one or several broadcasting Base Transceiver Stations (BTS). The information can be broadcast over several channels, possibly one per language used for the broadcast message. The user must first select the channels to which he wants to subscribe. This technology allows broadcasting a mass message without network performance problems. However, setting the terminal requires an adequate communication plan to the population associated with a technical support team able to assume the setup on heterogeneous consumer devices, in the best case when they are compatible, since the CBS feature has been removed from many terminals in favour of more vending features. In any case, this technology has been selected by the standards to carry the messages of the Public Warning System (PWS), as will be explained in Section IIIB.

The LTE is a project led by the 3GPP standards body for the publication of the technical standards of the future fourth generation mobile telephony. It enables data transfer at very high speed, with a longer range, a higher number of calls per cell and lower latency. For the operators, the LTE involves changing the core network and the radio transmitting stations. New compliant mobile terminals must also be developed. Considering the limitations of the current solutions in terms of deployment and performance, the LTE generation allows, with continuous connections, to be able to alert all the terminals almost simultaneously in a specific area, using dedicated short messages. The question of the

 ISBN: 978-1-61208-133-5

penetration rate of terminals with 4G subscriptions is an important element in the relevance of the solution for an alerting system. The number of users accessing the 3G services has been increasing sharply since the latest developments of devices such as the iPhone, Android, BlackBerry or Windows phones and the commercialization of unlimited flat rate packages. The population currently reached with 3G mobile subscriptions will probably evolve to the upcoming 4G systems rapidly due to the effect of device renewal.

III. USING ENHANCED UPCOMING TECHNOLOGIES

This analysis has been completed a prospective study of networks currently in the phase of definition and standardization, and which are of interest for the future population warning systems. In a first step are introduced the future vehicular networks, whose deployment is planned for the second half of the decade. The advantage of such networks is that, in addition to being able to reach the drivers, they operate in a cooperative mode. As a result, these networks are resilient to the possible destruction of the communications infrastructure. In a second step are presented the future developments of broadcast technology for mobile cellular networks (CBS and MBMS) and their integration in terms of standards into warning systems. The presented techniques were initially developed for a tsunami warning network in Japan and subsequently generalized to a more comprehensive Public Warning System (PWS). The CBS technology is used here again. This standard is part of the GSM, UMTS-3G and future LTE operational standards. Its advantage is that it allows the global broadcast of short messages (SMS-type) and thus overcomes the limitations due to network overload when targeting a large population. It also contains features that allow to "wake up" idle mobile phones and select the geographical coverage for the broadcast, making it particularly suitable for a connection to a global alerting system. However, it is somehow questioned since its deployment differs according to operators and countries.

A. Vehicular Communications

This new mode of communication from vehicle to vehicle is based on the new standards for Intelligent Transport Systems (ITS). Here we introduce the ETSI TS 102 636-3 standard [3], which is under development for ITS and GeoNetworking. The most interesting feature of this standard is the definition of a set of methods to distribute, and route messages in specific geographical areas. For example, in the advent of an emergency situation, the message is sent to the vehicles concerned by this emergency in the destination area. It would, in this way, reach only the concerned vehicles, not disturbing drivers outside the target region. The communication among entities may be between Vehicle to Vehicle (V2V), Infrastructure to Vehicle (I2V), Vehicle to Infrastructure (V2I), Infrastructure to Infrastructure (I2I) and all the concatenation of these basic scenarios. In the GeoBroadcast communications, the same message is forwarded to all the ITS stations in the defined area, as shown in Figure 3.

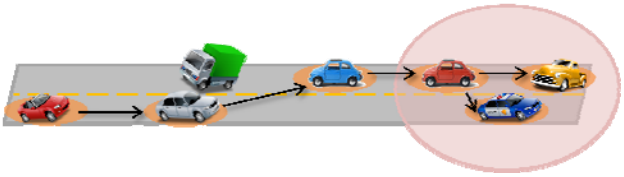


Figure 3. GeoBroadcast type of V2V communication

Some additional research has been conducted on the propagation of public safety warning messages using GeoBroadcast and Delay Tolerant Networks (DTN) techniques [4]. The main purpose of such work is to increase the coverage of the existing network to reach more people in a faster way. People in vehicles usually do not watch TV and may not be listening to the radio. In the future, cars will be equipped with devices helping to increase road safety that will be constantly active to provide the drivers with the information about the road conditions. The work described in [4] proposes that the vehicles act as virtual roadside units (vRSUs) and help on the spreading of the warning messages in case of an emergency. The intention is to decrease the “last mile” information access problem. The evaluations show that the mechanism is robust and efficient even over different disaster scenarios. Thus the use of vRSUs is an effective way to distribute warning messages to vehicles in a region.

B. Cellular Communications

Some new technologies and actions have recently been introduced in the 3GPP standardization for cellular systems which are relevant to public warning systems. The first part describes the two candidate technologies that can comply with the broadcasting requirements in case of a major event. Both technologies offer a global broadcast capability, which means that a message is sent only once and received at once by all the target terminals. The CBS has been part of the standards since the early GSM, even if not always deployed by operators, so it is technically compliant with all the existing enabled mobiles in the market. It permits to broadcast unacknowledged messages to all the receivers within some particular defined geographical areas known as cell broadcast areas. A CBS page is comprised of 93 characters and up to fifteen pages may be concatenated to form a message. Messages are broadcast cyclically at a frequency and for a duration agreed with the information provider. Mobiles can selectively display only the messages chosen by the Mobile user. In addition, a message that has been formerly successfully received is not displayed a second time. The second technology, the MBMS is an enhancement of the 3G systems which provides a point-to-multipoint capability for Broadcast and Multicast Services [5], allowing resources to be shared in the network. Since it is more recent, it has more constraints, but it also brings the capability to disseminate multimedia information (video, audio, pictures) in addition to the text messages. As the LTE is enhancing the capacity and efficiency of the cellular networks, the MBMS is evolving and adapted to benefit from these improvements.

 ISBN: 978-1-61208-133-5

Some actions have recently been taken in the 3GPP standardization groups to implement public notification warnings. Japan launched the first step with the ETWS (Earthquake and Tsunami Warning System), delivering Warning Notifications specific to Earthquake and Tsunami simultaneously to many mobile users located in Warning Notification Areas who should evacuate from an approaching Earthquake or Tsunami. An ETWS warning may be required in a very urgent timeframe (down to 4 seconds for the primary notification or initial alert) and is characterized by the capability to provide a very short notification period. A secondary notification can be delivered afterwards, carrying a larger amount of information such as text, audio or graphics to instruct what to do and where to get help, or a valid route from present position to an evacuation site. In a further release, this system has been generalized into the PWS (Public Warning System) [6] which targets worldwide objective, including the CMAS (Commercial Mobile Alert System) in the USA or the support of European requirements.

Some early technical studies considered both some variants of the CBS and MBMS broadcasting technologies for the PWS. Since the CBS is more mature from a standardization point of view, it is the solution that has been adopted. However, because the MBMS will be part of the future LTE systems and can convey larger amount of data, it is still an interesting candidate to support future alerting systems. One of its drawbacks, though, is that it lacks the geo-localization feature of the CBS system. An enhanced system has been proposed in [7] to extend the MBMS by developing cross-layer cooperation where the networking protocol and the cellular system collaborate to improve the efficiency of the geographical radio coverage. It enables a more precise and efficient delivery of the broadcast information, taking advantage of the comprehensive knowledge of the infrastructure and network topology by the mobile operator. Only the base stations located in the target zone participate in the distribution of the message, as shown in Figure 4.

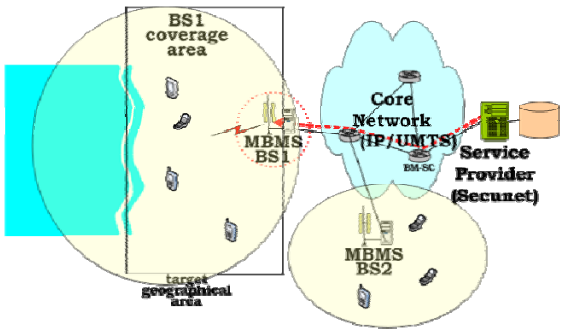


Figure 4. RATCOM Application Scenario with MBMS

Users located outside of their coverage do not have to filter out the un-necessary information, increasing the efficiency of the system.

IV. CONCLUSION

In this paper were presented several technologies to be deployed in the medium to long term. Some are completely new (WiMAX, DVB-SH or vehicular networks), others are the future evolution of existing communication networks (Ku band satellite networks, cellular mobile telephony).

Each of these technologies offers specific characteristics and particular interest for the broadcast of warning messages. The DVB-SH satellite broadcast network reaches a large number of users by stopping the radio and TV programs received on fixed or mobile devices, and replace them by the alert bulletin. It does not require the availability of a terrestrial network and therefore does not run the risk of being damaged by a natural disaster. The Ku band satellite network connections can also serve as redundancy to the existing network nodes in the case of failure due to a major problem, enabling the safe operation of critical network nodes or connecting a subnet that was isolated. The WiMAX technology, which is in its early deployment, is based on features close to the Internet. It offers the ability to support a Broadcast / Multicast mode and thus to provide broadcast services in geographical areas that cannot be easily connected with a legacy wired network. The CBS is based on existing cellular networks and deployable at medium term. In countries like Japan, CBS is used for mass message broadcast, even if its setting is somehow problematic. It will be advantageously replaced by the LTE that will achieve permanent connections towards all the terminals with a 4G subscription. Vehicular Networks will allow the broadcast of information from car to car in a specific geographical area. The advantage of this technology lies in the fact that it requires no infrastructure and can reach people while they are travelling. The future evolution of cellular networks is still being defined. With CBS and MBMS technologies, it is possible to broadcast a single message to many users, so at a lower cost from the point of view of radio resources, while capitalizing on a network and a massive penetration rate. The PWS systems take advantage of these features to provide a comprehensive model of early warning network. A proposal to extend the geographical feature of MBMS and increase its efficiency has also been introduced.

All these technologies can reach in a very limited time a significant number of users and are particularly relevant to a potential connection to the downstream component of a future public warning system. The availability of these technologies in the near or longer future depends mainly on their commercial success, according to business models and the return on investment expected from their deployment. Nevertheless, it is the administrative authorities who ultimately may decide on the development and promote the implementation of the functionalities needed to connect them to a global safety system.

 ISBN: 978-1-61208-133-5

REFERENCES

[1] RATCOM project http://ratcom.org, last accessed on 15.01 2011.

[2] ETSI TS 102 585 V1.1.2 : "System Specifications for Satellite services to Handheld devices (SH) below 3 GHz"

[3] ETSI TS 102 636-3: "Intelligent Transportation System (ITS); Vehicular Communications; GeoNetworking; Part 3: Network architecture".

[4] D. Camara, C. Bonnet, and F. Filali, Propagation of Public Safety Warning Messages, IEEE WCNC 2010, pp 1-6, Sydney, Australia

[5] 3GPP TS 23.246, “MBMS; ARCHITECTURE AND FUNCTIONAL DESCRIPTION”, V8.3.0 (03-2009)

[6] 3GPP TR 22.268; “Public Warning System (PWS) Requirements”; V9.2.1 (06-2009)

[7] M. Wetterwald, A case for using MBMS in geographical networking, ITST 2009, pp 309-313, October 2009, Lille, France.

 ISBN: 978-1-61208-133-5

Virtual Use Method of CGI by DACS Web Service Based on the Next Generation PBNM Scheme Called DACS Scheme

Kazuya Odagiri
Advanced Institute of Industrial Technology
Tokyo, Japan
odagiri@aiit.ac.jp, kazuodagiri@yahoo.co.jp

Syogo Shimizu
Advanced Institute of Industrial Technology
Tokyo, Japan
shimizu-syogo@aiit.ac.jp

Naohiro Ishii
Aichi Institute of Technology
Aichi, Japan
ishii@aitech.ac.jp

***Abstract*—as a work for managing a whole network effectively without a limited purpose, there is the work of a PBNM (Policy-based network management). The PBNM has two structural problems such as communication concentration from many clients to a communication control mechanism called PEP (Policy Enhancement Point) and the necessity of the network system updating at the time of introducing the PBNM into LAN. Moreover, user support problems in campus-like computer networks such as troublesome user support in updating a client's setups and coping with annoying communication cannot be improved by the PBNM. To improve these problems, we have been studied a next generation PBNM, which overcomes theses problems and has the function that does not exist in the existing PBNM, and called it a DACS (Destination Addressing Control System) Scheme. By the DACS Scheme, communication concentration from many clients to the PEP is solved, and system updating becomes unnecessary. Moreover, user support at updating the client's setups and coping with annoying communication by the DACS Scheme becomes very effective. In this study, to raise the effectiveness of this scheme, we show a virtual use method of CGI (Common Gateway Interface) by using the DACS Web Service, which is the Web Service realized by the DACS Scheme that we have been proposed before.**

Keywords- CGI; DACS Scheme; PBNM; destination NAT; packet filtering

I. INTRODUCTION

In computer networks where the usage policies are well defined, the network management is relatively easy. This is the case of enterprise computer networks, where security policies and access control lists are well defined. On the other hand, in campus-like computer networks, the management is quite complicated. Because the computer management section manages only a small portion of the wide needs of the campus network, there are some user support problems as follows. For example, when the mail boxes on one server are relocated to different server machines, an update of user machine's setups is necessary. Most of computer network users in a campus are students. Since students do not check frequently the e-mail, a usual operation is to make them aware of the settings update. This administrative operation is executed by means of web pages and/or posters. For the system administration, individual technical support is a stiff part of the network management.

As the work of network management, there are various kind of works such as the server load distribution technology [1][2][3], VPN (Virtual Private Network) [4][5]. However, these works are performed forward the specified different goal, and don't have the purpose of effective whole network management. As the work for managing a whole network, there is the work of Opengate [6][7], which controls Web accesses from LAN (Local Area Network) to internet. This work has the limited purpose of controlling Web access to internet. As the work for managing a whole network effectively without the limited purpose, there is the work of a PBNM (Policy-based network management) [8][9][10][11] in IETF (Internet Engineering Task Force). However, the PBNM has two structural problems such as communication concentration from many clients to a communication control mechanism called Policy Enforcement Point (PEP) and the necessity of the network updating at the time of introducing the PBNM into LAN. Moreover, it is often difficult for the PBNM to improve the user support problems in campus-like computer networks explained above.

To improve these problems of the PBNM, we show a next generation PBNM, which overcomes theses problems and has the function, which does not exist in the existing PBNM, and called it DACS (Destination Addressing Control System) Scheme. As the works of DACS Scheme, we showed the basic principle of the DACS Scheme [12], and security function [13]. In addition, we showed new user support realized by use of the DACS Scheme [14]. The past work of the DACS Scheme's mechanism was executed as a network management scheme for campus-like computer networks. In this paper, to raise the effectiveness of this scheme, we show the virtual use method of CGI (Common Gateway Interface) by using the DACS Web Service. The DACS Web Service is the Web Service realized by the DACS Scheme that we have been proposed before. The rest of paper is organized as follows. Section II shows motivation of this research. In Section III, we describe the content of the DACS scheme. Then, in Section IV, the content of DACS Web Service is explained. In Section V, virtual use of the CGI program is shown.

 ISBN: 978-1-61208-133-5

II. MOTIVATION

In the world of Internet, programs as the CGI [15] are often disclosed so that many users can use them without any charge. Because they are developed at an individual level, it is often impossible to use them in a company and university practically. However, when they are developed by a skilled developer, it is possible to use them practically. For example, in the software such as a bulletin board and groupware, they are not always referred only form same group members in each user's group. In many cases, when they are used in multiple groups, they are placed by being multi-copied. Because they are accessed from users in other group, there is the possibility of data leak. To be concrete, when they don't have an authentication mechanism, it becomes possible for users in other group to access the program. When they can acquire the URL for the program, the data of them is referred through the program by using the URL.

Therefore, in this study, the virtual usage method of the CGI program is shown. To be concrete, by using the DACS Web Service that the authors have been studied, it is realized. The DACS Web Service is the service that is realized on the network introducing the DACS Scheme, which is a scheme of Policy Based Network Management (PBNM). Because it is a service limited to Local Area Network (LAN) at this time, the method is also limited to the usage on the LAN.

III. THE DACS SCHEME

In this section, the content of the DACS Scheme is described.

A. Existing PBNM

As the works on existing network management, there are various works such as authentication [16][17], the server load distribution technology [1][2][3], VPN [4][5] and quarantine network [18][19]. However, these works are performed forward the specified different goal. Realization of effective management for a whole network is not a purpose. These works are performed for the specific purpose, and don't have the purpose of managing a whole network. As the work for managing a whole network, there is the work of Opengate [6][7], which controls Web accesses from LAN to internet. However, this work has the limited purpose of controlling Web access to internet. As the work for managing a whole network effectively without the limited purpose, there is the work of the PBNM [8][9][10][11] in IETF. The content of the PBNM is described in Figure 1.

To be concrete, in the point called PDP (Policy Decision Point), judgment such as permission and non-permission for communication pass is performed based on policy information. The judgment is notified and transmitted to the point called the PEP, which is the mechanism such as VPN mechanism, router and firewall located on the network path among hosts such as servers and clients. Based on that judgment, the control is added for the communication that is going to pass by.

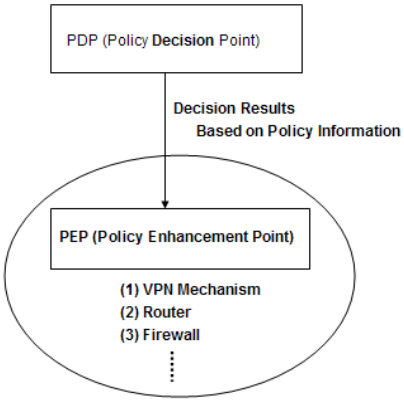


Figure 1. PBNM in IETF

B. Basic Principle of the DACS Scheme

Figure 2 shows the basic principle of the network services by the DACS Scheme. At the timing of the (a) or (b) as shown in the following, the DACS rules (rules defined by the user unit) are distributed from the DACS Server to the DACS Client.

(a) At the time of a user logging in the client.
(b) At the time of a delivery indication from the system administrator.

According to the distributed DACS rules, the DACS Client performs (1) or (2) operation as shown in the following. Then, communication control of the client is performed for every login user.

(1) Destination information on IP Packet, which is sent from application program, is changed.
(2) IP Packet from the client, which is sent from the application program to the outside of the client, is blocked.

An example of the case (1) is shown in Figure 2. In Figure 2, the system administrator can distribute a communication of the login user to the specified server among servers A, B or C. Moreover, the case (2) is described. For example, when the system administrator wants to forbid an user to use MUA (Mail User Agent), it will be performed by blocking IP Packet with the specific destination information.

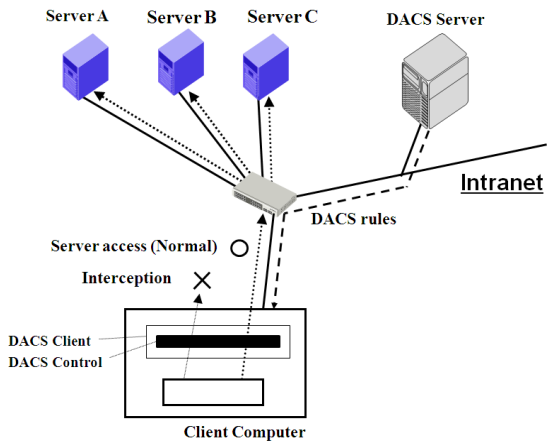


Figure 2. Basic Principle of the DACS Scheme

In order to realize the DACS Scheme, the operation is done by a DACS Protocol as shown in Figure 3. As shown by (1) in Figure 3, the distribution of the DACS rules is performed on communication between the DACS Server and the DACS Client, which is arranged at the application layer. The application of the DACS rules to the DACS Control is shown by (2) in Figure 3. The steady communication control, such as a modification of the destination information or the communication blocking is performed at the network layer as shown by (3) in Figure 3.

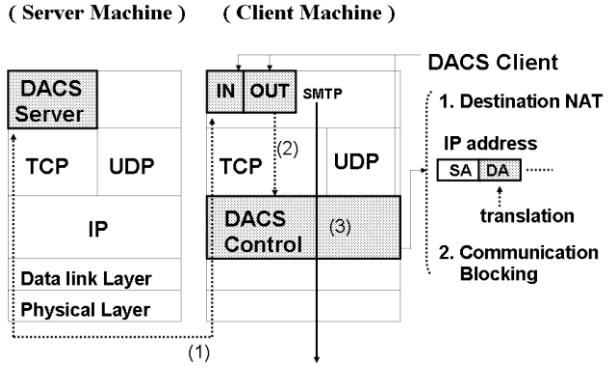


Figure3. Layer Setting of the DACS Scheme

C. Communication Control on Client

The communication control on every user was given. However, it may be better to perform communication control on every client instead of every user. For example, it is the case where many and unspecified users use a computer room, which is controlled. In this section, the method of communication control on every client is described, and the coexistence method with the communication control on every user is considered.

When a user logs in to a client, the IP address of the client is transmitted to the DACS Server from the DACS Client. Then, if the DACS rules corresponding to IP address, is registered into the DACS Server side, it is transmitted to the DACS Client. Then, communication control for every client can be realized by applying to the DACS Control. In this case, it is a premise that a client uses a fixed IP address. However, when using DHCP service, it is possible to carry out the same control to all the clients linked to the whole network or its subnetwork for example.

When using communication control on every user and every client, communication control may conflict. In that case, a priority needs to be given. The judgment is performed in the DACS Server side as shown in Figure 4. Although not necessarily stipulated, the network policy or security policy exists in the organization such as a university (1). The priority is decided according to the policy (2). In (a), priority is given for the user's rule to control communication by the user unit. In (b), priority is given for the client's rule to control communication by the client unit. In (c), the user's rule is the same as the client's rule. As the result of comparing the conflict rules, one rule is determined respectively. Those rules and other rules not overlapping are gathered, and the DACS rules are created (3). The DACS rules are transmitted to the DACS Client. In the DACS Client side, the DACS rules are applied to the DACS Control. The difference between the user's rule and the client's rule is not distinguished.

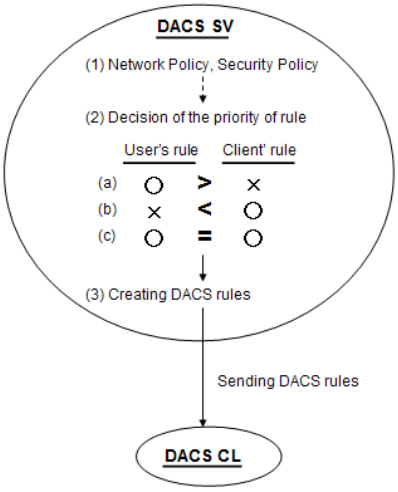


Figure 4. Creating the DACS rules in the DACS Server side

D. Security Mechanism of the DACS Scheme

In this section, the security function of the DACS Scheme is described. The communication is tunneled and encrypted by use of SSH. By using the function of port forwarding of SSH, it is realized to tunnel and encrypt the communication between the network server and the, which DACS Client is installed in. Normally, to communicate from a client application to a network server by using the function of port forwarding of SSH, local host (127.0.0.1) needs to be indicated on that client application as a communicating server. The transparent use of a client, which is a characteristic of the DACS Scheme, is failed. The transparent use of a client means that a client can be used continuously without changing setups when the network system is updated. The function that doesn't fail the transparent use of a client is needed. The mechanism of that function is shown in Figure 5. The changed point on network server side is shown as follows in comparison with the existing DACS Scheme. SSH Server is located and activated, and communication except SSH is blocked. In Figure 5, the DACS rules are sent from the DACS Server to the DACS Client (a). By the DACS Client that accepts the DACS rules, the DACS rules are applied to the DACS Control in the DACS Client (b). The movement to here is same as the existing DACS Scheme. After functional extension, as shown in (c) of Figure 5, the DACS rules are applied to the DACS SControl. Communication control is performed in the DACS SControl with the function of SSH. By adding the extended function, selecting the tunneled and encrypted or not tunneled and encrypted communication is done for each network service. When communication is not tunneled and encrypted, communication control is performed by the DACS Control as shown in (d) of Figure 5. When communication is tunneled and encrypted, destination of the communication is changed by the DACS Control to localhost

 ISBN: 978-1-61208-133-5

as shown in (e) of Figure 5. After that, by the DACS STCL, the communicating server is changed to the network server and tunneled and encrypted communication is sent as shown in (g) of Figure 5, which are realized by the function of port forwarding of SSH. In the DACS rules applied to the DACS Control, localhost is indicated as the destination of communication. In the DACS rules applied to the DACS SControl, the network server is indicated as the destination of communication. As the functional extension explained in the above, the function of tunneling and encrypting communication is realized in the state of being suitable for the DACS Scheme, that is, with the transparent use of a client. Then, by changing the content of the DACS rules applied to the DACS Control and the DACS SControl, it is realized to distinguish the control in the case of tunneling and encrypting or not tunneling and encrypting by a user unit. By tunneling and encrypting the communication for one network service from all users, and blocking the untunneled and decrypted communication for that network service, the function of preventing the communication for one network service from the client, which DACS Client is not installed in is realized. Moreover, even if the communication to the network server from the client, which DACS Client is not installed in is permitted, each user can select whether the communication is tunneled and encrypted or not. The function of preventing information interception is realized.

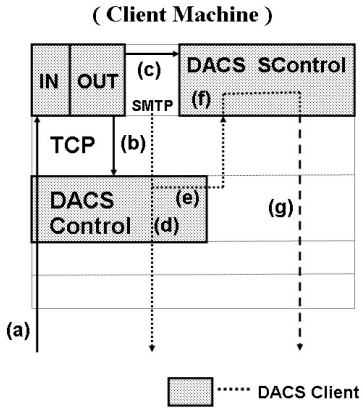


Figure 5. Extend Security Function

E. Effectiveness of the DACS Scheme

(a) Effective User Support at Changing Setups of Client with the DACS Scheme

When network system is updated, user support by the DACS Scheme is compared with user support by the Non-DACS Scheme, and an advantage of user support by the DACS Scheme is described. User support processes after updating the network system are described in Figure 6.

When the DACS Scheme is not introduced, notification for changing setups is sent to a user in a laboratory (2) after updating the network system (1). It is sent by E-mail and a homepage or a document. The user who accepts that notification updates a client's setups (3). If there is no problem in changing setups of the client, it is enabled to start the operating (4). When it is not possible to update setups by some causes, the user inquires to the network management section (5). In the network section, investigation by hearing comprehension for the user or investigation in the field is done (6). If a cause is specified, the coping way are considered, and carried out (7). It is a burden for a system administrator to support each user for every inquiry. When the DACS Scheme is introduced, a system administrator has only to change the DACS rules (8) at the time of updating the network system. After changing the DACS rules, communication control corresponding to new network system is started at a point in time when the user logs in to a client again (4). Because the system administrator with understanding the policy for using a laboratory network sets the DACS rules, a trouble by a cause except an artificial factor such as missing setups of the DACS rules does not occur. This process of user support is largely simplified in comparison with the process of user support by the Non-DACS Scheme.

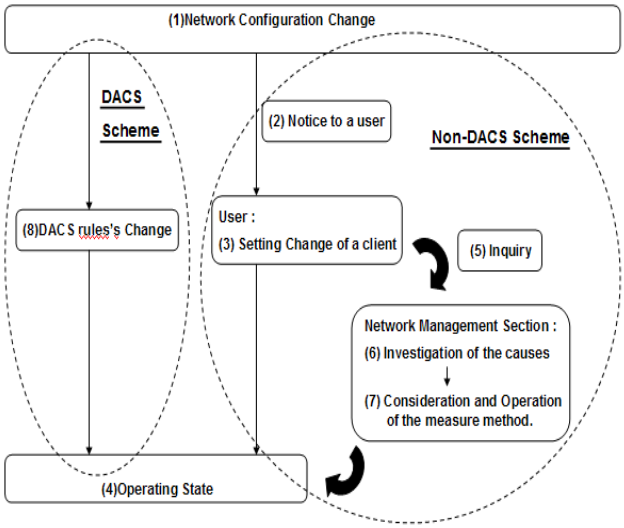


Figure 6. Process introducing the DACS Scheme

(b) Effective Coping with Annoying Communication by the DACS Scheme

To cope with the communication from a virus infection client and the communication with annoyance to other user such as streaming of moving and sound [20], a system administrator needs to specify, which user or client is transmitting the communication to. For example, when there is a direct cause in the client itself such as virus infection, the client must be specified. A user must be specified, when there is a direct cause in user oneself. When the IP address is managed dynamically by DHCP service, much time and effort is spent to specify the client or user. The coping process for annoying communication is described as shown in Figure 7 and explained with an example of the user support for a laboratory.

At first, annoying communication for other users is captured by communication detection through the mechanism such as F/W or IDS (1). Next, a source IP address of the annoying communication is acquired (2). To here, it is the same thing when the DACS Scheme is introduced or not introduced. When the DACS Scheme is not introduced, the process of user support is described in

 ISBN: 978-1-61208-133-5

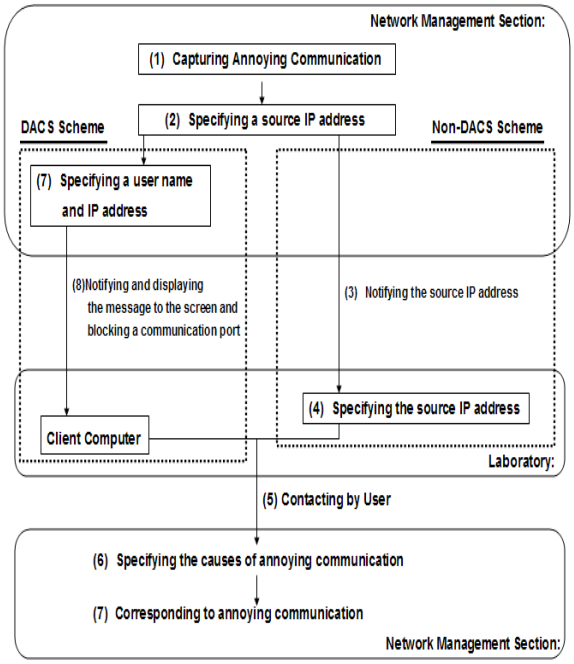


Figure 7. Change of User Support

the following. Under using DHCP Service, if a whole network is divided into multiple subnetworks, and each subnetwork is assigned to each laboratory, a system administrator can manage scope of the IP address used in a laboratory. If not so, the system administrator cannot manage it. In the case of the former, the IP address is notified to the laboratory (3), and the client transmitting the communication is specified (4). In the laboratory, because it is impossible to manage t the IP address that the client uses, the client is specified after investigating the network setups information of each client. It takes trouble very much. In the case of the latter, it is difficult to specify the client. This is because the system administrator cannot know the laboratory using the IP address. Even if the system administrator can know it, because it is needed to investigate the network setups information of each client, it takes trouble very much. After the client is specified, the user of the laboratory contacts a network management section (5). In the situation that a laboratory cooperates with a network management section, the cause specification of annoying communication and coping with it are done (6). On the other hand, when the DACS Scheme is introduced, source IP address of the annoying communication needs to be acquired (2) to specify the client first. When a user needs to be specified, a user name is specified from the IP address (7). When a user has a direct cause such as streaming of the moving picture and the sound, the message to notify abnormality is transmitted to the IP address of the client, which a user logs in. If a client has a direct cause such as infection by virus, the message to notify abnormality is transmitted to the IP address of the client. The message is displayed in the screen of the client. At the same time, the used port by annoying communication is blocked (8). The user sees the message of the screen, and contacts the network management section (5). In the situation that a laboratory cooperates with a network management section, specification of annoying communication and coping with it are done (6). It is shown that the DACS Scheme is effective at the following two points. The first point is that the client that transmits annoying communication is specified simply. The client that has a problem is specified by seeing the message of a screen at a glance. The second point is shown as follows. Because the influence to others is prevented by blocking a communication port of the client, time margin for the cause specification of annoying communication and the coping with it is generated effectively. When the urgent degree such as virus infection is high, the DACS Scheme is particularly effective.

IV. SYNOPSIS OF THE DACS WEB SERVICE

In this section, the synopsis of the DACS Web Service is described.

A. Two Kinds of Functions of Web Service Based on DACS Scheme

Two kinds of functions of Web Services based on DACS Scheme are described, here.

At first, the function to use data from database is developed. To realize this function, DACS Scheme needs to be extended, and the program on Web Server needs to be implemented in correspondence to the extended DACS Scheme as shown in Figure 8.

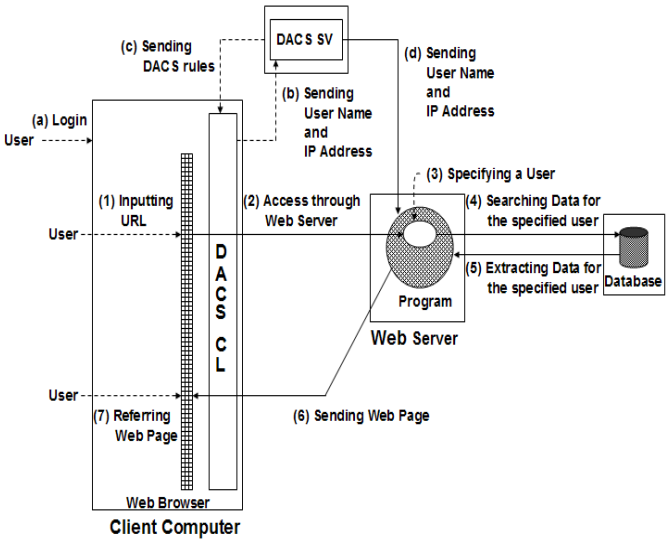


Figure 8. Function Using Data from Database

In the network with DACS Scheme, after a user's logging in a client (a), user name and IP address are sent to DACS SV (b). Then, DACS rules are sent back to DACS CL (c). Moreover, user name and IP address are sent to the program on Web Server. Then, the server side program on Web Server can identify the user by checking the login information and the source IP address from the client, and

can change the processing of the program every user. When each different user accesses the program with same URL, different information for each user can be searched and extracted from database, and be displayed on Web Browser. Through the processing from (1) to (7), this new function is performed.

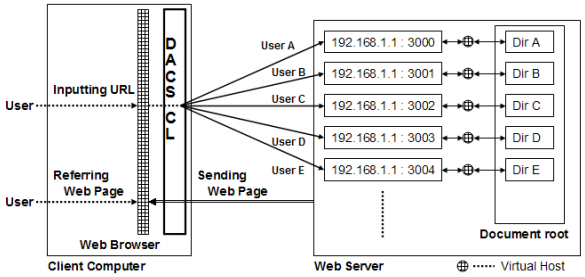


Figure 9. Function Using Data from Document Medium

Next, the function to use data from document medium is developed for the respective user. In the network with DACS scheme, different IP address and TCP port can be assigned for one host name by a user unit. Therefore, different document medium with same file name on different Web Server can be referred for each user by inputting same URL to Web Browser. When this principle is combined with the function of virtual host that is equipped as Web Server, it is possible to use Web Server as shown in Figure 9.

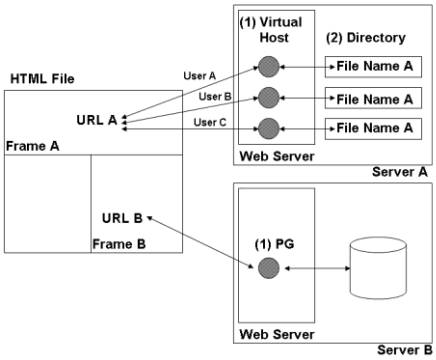


Figure 10. Web Service by two functions

By the function of virtual host, multiple groups of socket (IP address and TCP port) can be assigned for one Web Server. The referred document can be changed every socket. First, in Document root of Web Server in Figure 9, directories (Dir A,B,C,D….) are prepared for each user. By the function of virtual host, each directory is connected to each socket as one pare. By changing TCP port number (3000,3001,3002….) for one IP address (192.168.1.1), sockets corresponding to each directory are prepared. Next, movement on this mechanism is described. One user inputs one URL to Web Browser. When the URL is inputted by User A, the file in Dir A that is connected to the socket (192.168.1.1:3000) is referred. Equally, when by User B, the file in Dir B that is connected to the socket (192.168.1.1:3001) is referred. When by User C, the file in Dir C that is connected to the socket (192.168.1.1:3002) is referred. When the document medium with same name exists in each directory (Dir A,B,C….), each user can see different contents by inputting same URL to Web Browser. For information sender, because it is possible to notify information to the specific user by uploading document medium to the predetermined directory, information usage becomes largely wide. Because information sender can describe the content of document medium easily and freely, it is possible to communicate the information with much expressive power and impact.

As the result, by letting both functions coexist as shown in Figure 10, the Web Service that a user can use information on the network regardless of information storage form is realized.

B. Contents of the DACS Web Service

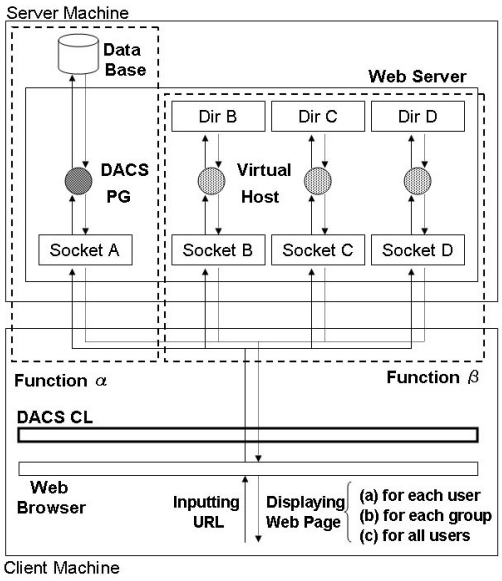


Figure 11. DACS Web Service

In Figure 11, synopsis of DACS Web Service is shown. The function to use data from database of information system such as a system managing results for a student, is shown as Function α. The function to use data from document medium such as a simple text file and a PDF file, is shown as Function β. After a user's inputting URL into Web Browser, communication control by DACS CL (DACS CTL) is performed. As the result, function α or Function β is used. Because the function of either is automatically selected every each URL according to DACS rules, a user can use data from information system or document medium dispersing on the network without being conscious of that function is used. In other words, a user can use information regardless of storage form and storage place of data freely and easily, if a user knows URL and the kind of information acquired by that URL. Even if whichever of Function α or Function β is used, data is displayed on Web Browser after inputting URL. Three kinds of data, which are sent by a user unit (a) and by a group unit (b) and by all users unit, are displayed.

Here, details of Function α are shown. After extension, the functions of retrieving data for each group (Function α2)

 ISBN: 978-1-61208-133-5

or for all users (Function α3) can be used. There are differences among Function α1, Function α2, and Function α3 in the program extracting data from a database for a request from a Web browser.

In the program of Function α1, data is extracted for each user, as shown by (1). In the program of Function α2, data is extracted for each group, as shown by (2). In the program of Function α3, data is extracted for all users, as shown by (3). In the existing function to retrieve data from a database, as shown by (1), it is possible to specify which user is sending communication through the Web browser. Therefore, the function is extended to set a correspondence list of the user and group name in the DACS Server and send that correspondence list from the DACS Server to the program of Function α2. As a result, because the program of Function α2 can recognize the group to which a user belongs, it is possible to extract information for each group. Even if a user belongs to multiple groups, it is possible to extract the data of all groups. In addition, it is possible to extract the data of a specific group by sending its group name as a parameter of the URL. In the program of Function α3, data is extracted for all of these users. Because it is the function of a normal Web Service that does not introduce DACS Scheme, it is generally realized without a technical problem.

Next, details of Function β are shown. Function β1 displays data of the document medium dynamically for each user. By use of this function, the function for each group (Function β2) and the function for all users (Function β3) are realized. Function β2 relates the URL for each group (Group URL1, Group URL2….) to each document, which is stored in each directory for each group. Function β3 relates the URL for all users (All Users URL) to the document, which is stored in the directory for all users. To send information, only uploading a file as a document medium into the predetermined directory (directory for each user, directory for each group, and directory for all users) is necessary. Information for each group can be recieved by the specific URL for each group. In addition, the users not belonging to each group can not access it by using the URL. Information for all users can be recieved by use of the URL for all users. By using the DACS Web Service, not only information for each user but also information for each group and for all users, can be used from the document medium.

V. VIRTUAL USAGE METHOD OF THE CGI

In this paper, the method that is realized by the Function α1 is proposed. By using this function, programs of the CGI is accessed virtually through same URL from users in each group. To be concrete, this method is realized by the following procedure.

(Step1) First setting of the CGI programs

First, CGI is set by a normal procedure. For example, the program files as the CGI are placed on the Web Server, and initial setting is performed. For example, the setting of initial parameter of the CGI and permission of the program files. As the result, users can use the programs of the CGI by inputting one URL into a Web Browser.

(Setp2) Setting for virtual use of the CGI programs

After copying the directory that stores the programs, it is pasted as another directory with another name. By repeating a similar operation, multiple directories for each group are prepared. At the same time, the content of the DACS rules is changed. As the result, users that belong to same group become possible to access the programs of same directory by use of a URL. On the other hand, users that belong to other group become impossible to access the above programs by use of same URL.

By these procedures, in the form of using same URL, users in each group can access the programs of the directory in each group, and can not access the programs of other group. Virtual use of the CGI programs is realized without a special mechanism.

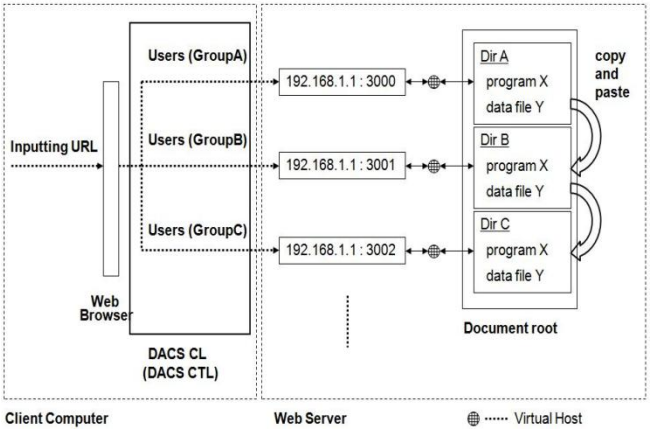


Figure 12. Virtual Usage of the CGI program

A concrete example of it is shown in Figure 12. As first step, the program X as the CGI program and other files such as data file are placed in directory A (Dir A in Figure 12), and initial setting of it is performed. As the result, users can access and use it. Next, second step is as follows. At first, Dir A is copied and pasted with another name. In Figure 12, Dir B and Dir C are the pasted directories. Each directory is named with the regularity. Though each socket is connected to each directory through the virtual host by the system setting, each name is allocated to be easy to automate the setting. At the same time, by changing the DACS rules, the host name in URL and the communication port is converted to each socket every group. In Figure 12, when users in Group A inputs one URL into a Web Browser, they access the program X in Dir A through by way of 192.168.1.1:3000. In the case of users in Group B, they access the program X in Dir B through by way of 192.168.1.1:3001. In the case of users in Group C, they access the program X in Dir C by way of 192.168.1.1:3002. Then, users in Group A con not access the program in Dir B and Dir C. Users in Group B con not access the program in Dir A and Dir C. Users in Group C con not access the

 ISBN: 978-1-61208-133-5

program in Dir A and Dir B. In this way, virtual use of the CGI program is realized simply.

REFERENCES

[1] S.K. Das, D.J. Harvey, and R. Biswas,"Parallel processing of adaptive meshes with load balancing," IEEE Tran.on Parallel and Distributed Systems, vol. 12, No. 12, pp. 1269-1280, Dec. 2002.

[2] M.E. Soklic,"Simulation of load balancing algorithms: a comparative study," ACM SIGCSE Bulletin, vol. 34, No. 4, pp. 138-141, Dec. 2002.

[3] J. Aweya, M. Ouellette, D.Y. Montuno, B. Doray, and K. Felske,"An adaptive load balancing scheme for web servers," Int.,J.of Network Management., vol. 12, No. 1, pp. 3-39, Jan/Feb. 2002.

[4] C. Metz,"The latest in virtual private networks: part I," IEEE Internet Computing, Vol. 7, No. 1, pp. 87-91, 2003.

[5] C. Metz,"The latest in VPNs: part II," IEEE Internet Computing, Vol. 8, No. 3, pp. 60-65, 2004.

[6] Y. Watanabe, K. Watanabe, E. Hirofumi, and S. Tadaki, "A User Authentication Gateway System with Simple User Interface, Low Administration Cost and Wide Applicability," IPSJ Journal, Vol. 42, No. 12, pp. 2802-2809, 2001.

[7] S. Tadaki, E. Hirofumi, K. Watanabe, and Y. Watanabe,"Implementation and Operation of Large Scale Network for User' Mobile Computer by Opengate," IPSJ Journal , Vol. 46, No. 4, pp. 922-929, 2005.

[8] S. Jha and M. Hassan,"Java implementation of policy-based bandwidth management," Int. J. Network management, John Wiley&Sons, Vol. 13, isuue. 4, pp. 249-258, July, 2003.

[9] G.M. Prerez, F.G. Skarmeta, S. Zeber, and T. Symchych,"Dynamic Policy-Based Network Management for a Secure Coalition Environment," IEEE Communications Magazine, Vol. 44, issue. 11, pp. 58-64, November, 2006.

[10] D.C. Verma,"Simplifying Network Administration Using Policy-Based Management," IEEE Network, Vol. 16, issue. 2, pp. 20-26, March-April, 2002.

[11] M. Sugano, S. Tanaka, Y. Sakata, K. Oguma, and N. Shiratori,"Application and Implementation of Policy Control Method "PolicyComputing" in Computer Networks," IPSJ Journal, Vol. 42, No. 2, 2001.

[12] K. Odagiri , R. Yaegashi , M. Tadauchi, and N. Ishii, " Efficient Network Management System with DACS Scheme : Management with communication control," Int. J. of Computer Science and Network Security, Vol. 6, No. 1, pp. 30-36, January, 2006.

[13] K. Odagiri , R. Yaegashi , M. Tadauchi, and N. Ishii, " Secure DACS Scheme, " Journal of Network and Computer Applications," Elsevier, Vol. 31, Issue. 4, pp. 851-861, November, 2008.

[14] K. Odagiri, R. Yaegashi, M. Tadauchi, and N. Ishii, "New User Support in the University Network with DACS Scheme," Int. J. of Interactive Technology and Smart Education.

[15] D. Robinson,"The WWW Common Gateway Interface Version 1.1,Internet Draft," 1995

[16] K. Wakayama, Y. Decchi, J. Leng, and A. Iwata,"A Remote User Authentication Method Using Fingerprint Matching," IPSJ Journal, Vol. 44, No. 2, pp. 401-404, 2003.

[17] S. Seno, Y. Koui, T. Sadakane, N. Nakayama, Y. Baba, and T. Shikama,"A Network Authentication System by Multiple Biometrics," IPSJ Journal, Vol. 44, No. 4, pp. 1111-1120, 2000.

[18] http://www.nec.co.jp/qxseries/solution/04.html 10.3.2011

[19] http://www.ntt.co.jp/journal/0512/files/jn200512049.pdf 10.3.2011

[20] H. Hu, J. Kashio, Y. Honda, and H. Suzuki,"Rate Control Method for Real Time Protocol (RTP) Enabling the Coexistence with TCP," IEICE Tran. on Communications, Vol. J84-B, No. 11, pp. 1994-2004, 2001.

 ISBN: 978-1-61208-133-5

Routing optimization in the transmission network

Mary Luz Mouronte, María Luisa Vargas, Paloma Martínez
Optical Network Provisioning
Ericsson
Madrid, Spain
mary.luz.mouronte.lopez@ericsson.com.

Abstract— **This paper describes a method to optimize the route lookup in a multivendor data transmission network built on Synchronous Digital Hierarchy (SDH) and Wavelength Division Multiplexing (WDM) technologies. The optimization is carried out in terms of time reduction, thereby having a wide field of application in network operation: restoring service after a failure, traffic redistribution scenarios, etc.**

The algorithm encompasses different aspects such as: number of routes to lookup, network structure, technological characteristics of routes, and situation of elements, all of them related to technical objectives. It also aims at integrating other features embodying other objectives apart from the aforementioned technical ones.

Keywords-component; route lookup optimization; SDH, WDM; transmission network features; transmission network elements; boundary conditions.

I. INTRODUCTION

The method to select a suitable path into a network for sending traffic is called routing. This process is carried out in different networks, particularly in data networks (data packet routing) and telephone networks (phone calls routed through numbering plans).

In **data networks**, the routing information is included into the final address codification. Routing tables are used at each node, and they contain adjacent node addresses, so it is not necessary to know the whole network topology.

Small networks can use manually configured routing tables, but when networks increase their size, manual management is unfeasible. This problem is intended to be resolved through automatic routing table creation. The process is based on the information included in the routing protocols, which allow the network to work in an autonomous way.

There are different kinds of algorithms to automatically populate the routing tables:

- Distance vector algorithms, based on the Bellman-Ford method [1]: each node assigns a cost to the distance to the known destine, and transmits it to the adjacent nodes. Each receptor builds its tables based on the received and stored information. If there is a node failure, the information is rebuilt in each one of the surviving nodes.
- Link state algorithm: each node shares its link information so all the nodes build a network map. This map is modified according to the link state. The route is obtained by means of basic short path algorithms, which assign cost to each link (Dijkstra algorithm, for example).
- Path vector protocol: a special class of distance vector algorithms, useful in very big networks.

All these algorithms (special mention to [2], [3], [4], and [5]) require features which make possible to establish a hierarchy among routes, such as bandwidth, delay, hop number and so on.

In **telephone networks**, routing tables are previously calculated in a centralized way. The main parameters for their elaboration are the network topology; the numbering plan and the traffic data analysis (see [6]).

Due to geographical and hierarchical structures, which are the base of the numbering plans, almost all telephone calls are routed using only the first numbers (prefix). The tables are designed specifically for each switchboard, so the maximum number analysis is defined for the geographical zone managed by the switchboard.

There is no dynamic route selection when there is a failure within the network. All the alternative routes are selected at the same time than the main ones.

As we can see in all the previous cases, the techniques used in each network are based on their specific characteristics, which are not useful in the **data transmission network**:

- In data networks, the routing table design, regardless of the selected method, is based on parameters that make possible to assign a cost to every links. In transmission networks, there is no measurable parameter that can be used to prioritize the paths. The only specific requirement at this point is to avoid loops in the final route when the algorithm selects a path in a node among different possibilities. In other words, an already visited node at the end of a possible path discards that path.
- In telephone networks, the routing table design is static, based on the network structure and the information included in the telephone numbers, with no dynamic selection. In transmission networks, there are static paths too, but if the network fails, the new path is selected dynamically. In this situation, there is no more static information than the network topology, so there is no hierarchy that can be used to select a

 ISBN: 978-1-61208-133-5

path. All the routes with available bandwidth have the same priority; there are no criteria to discard.

There are some recent studies aimed at route lookup in transmission networks, as it is a key point in the deployment of the future high capacity networks, all-based in optical links. There are some novel approaches to this problem, as it is developed in [9], applied to power networks but useful for optical networks too. There are some works too for calculating routes after deployment, in a general way as it is mentioned in [10] or oriented to a specific kind of traffic, such as the one analyzed in [11].

However, almost all of these works are focused on static calculation or pre-calculated routes, while our work is focused on real-time lookup without pre-evaluated routes. They are oriented to isolated networks (in terms of technology), not mixed ones, each island with different restrictions and conditions.

After the evaluation of all these points, the first step in the work will be the selection of the characteristics of the data transmission network that will facilitate the optimum route selection in a dynamic way according to network status. Subsequently, the work will describe the modifications in traditional algorithms to use all the selected features in the lookup.

II. DATA TRANSMISISION NETWORK CHARACTERISTICS

The procurement of a customer service between two points in the transmission network requires a route selection and setting between them, usually involving various nodes (equipments) aside from the access equipments. In normal service procurement, the route selection and setting is carried out at the circuit-provisioning step.

However, an additional situation requires a route selection. If there is a network failure, it will be necessary to turn aside the traffic in order to ensure the service. In that case, there is no previous selected route, so a new one must be selected and set at that moment. What is more, main concerns must be: the speed, a minimized traffic lost, and network status, to avoid failed equipments or connections.

The algorithms for data networks outlined before are very generic, and no efficient in particular ones, as it is showed by the modifications in the telephone network routing.

This work describes the adjustments of the previous algorithms to the data network characteristics, so a best efficiency can be achieved. The bandwidth required for a route within the data transmission network is translated into a signal type, and the viable route is defined through the relationship between different paths according to this signal type. The work takes into account different features, which at the end define the network in terms of signal types:

- **Hierarchy**: The network is hierarchically organized according to an increasing bandwidth. Paths with a specific signal type (i.e., a required bandwidth) are joined into another one with a higher one.
- **Safeguard**: At some hierarchic levels, there are path groups defining structures for traffic safeguard. These path groups must be taken into account for the route definition (e.g., path rings).
- **Technology**: Data transmission networks include a wide range of technologies. Within them, there are easier ways to switch routes (e.g., if the lambda switching can be remotely done in an optical network, this routing will be faster and more efficient than if it is necessary to make the switch in the equipment), directly affecting the delay in traffic restoration.
- **Network deployment**: The network capacity changes dynamically, so the information about the network topology must be properly updated for an efficient route lookup. The network can mainly change in three different ways: deployment (i.e. adding new equipments or paths resulting in an increased capacity); network variations due to failures, (decreasing capacity) or path redesign (changing relationships) and finally service provisioning, which decreases the network capacity.

III. ALGORITHM

This section describes the information included in the algorithm in order to use the specific selected features of the data transmission network to obtain optimized results. The algorithm is applicable to a multivendor transmission network.

The features outlined in the previous section are thoroughly explained below:

- Location (buildings): sites where the network equipment is placed.
- Transmission equipments within each location.
- Ports/points in each equipment: the equipment usually consists of a variable number of cards (depending of the equipment configuration) and each card includes a number of ports (depending on the card configuration). All of these are physical elements, but there is a logical element that must be taken into account, the termination point [7]. A physical port can be logically divided in different points, related to a specific signal type. A path can finish in a port or a point depending on its signal type, so the algorithm must differentiate and consider both of them.
- Links (paths) connecting equipments: The information handled by the algorithm is their free capability in terms of the number of paths and signal type of each one. This free capability can be modified according to the rules defined by the final points in the equipments, because there are specific configurations related to the equipment type and the network definition itself [7].
- Protection rings in the network: In a data transmission network, there are protection structures that ensure the traffic in some specific points, usually in big capability links. In these structures, there are usually two branches: one of them works in a normal way, and the

 ISBN: 978-1-61208-133-5

other is free (safeguard) until a failure turns up. At that moment, an automatic switch is performed and the traffic begins to flow through the free branch, keeping the service on. These structures are called “rings” and, in this work, they are described in terms of the equipments and links belonging to them.

All the necessary information in the algorithm must be updated dynamically, according to the variations carried out in the network topology mentioned before. These variations are translated on the assembly and disassembly of locations, equipments, ports, links and rings; configuration changes that modify the path that can be included into a link; and service procurement modifications (they entail link use, so the link capability is modified).

All the above information is related to the network structure. Nevertheless, there is additional information, regarding to the route itself which must be handled by the algorithm, so the lookup is optimized. The additional information is translated into the input parameters that the algorithm must receive, so it can obtain a route with the desired conditions:

- The **signal type** and **number** of routes the algorithm must lookup.
- The **termination points** or **ports** that are the ends of the desired route. As it is mentioned above, the route must use one or the other in both ends depending on the signal type.
- Required **route features**: These features can be expressed in terms of capability (a full link or only the needed bandwidth); possibility of network operations (links already have the desired capability or it can be achieved through configuration modifications); end situation (they must be in the same equipments or it is enough if they are in the same locations) and maximum number of links. The end situation must be taken into account because in certain conditions and technologies, it is not possible to reach the equipments and some hardware modifications are required. The maximum number of links is needed because sometimes the network is heavily messed [8] and the number of possibilities is dramatically increased, so this condition is a termination one.
- **Boundary conditions** as a route lookup limitation. These conditions include elements which must or must not be used (equipments or links) and rings where the route must be included. The first condition allows to avoid failed equipments or links, as well as to use other ones according to service requirements. The last one is necessary because, in some cases, the initial protection schemes must be kept.
- **Resource special features**. There are certain network conditions that can be useful to restrict the lookup, so it will be optimized. Usually they are related to the network status: to use failed resources if there is not other available route; to use temporary resources deployed only for a single scenario or another similar one. However, there is a special case related to network nature: to use only a specific kind of network (SDH or WDM, for example). This is necessary because, in some cases, there is no chance to send staff to manually switch routes, and the obtained route must allow automatic and remote operation.

The way the adapted algorithm works to find a route is outlined below.

The first step is to select the end equipment and the required resources according to all the input data. The next task is the lookup itself. There are three mutually exclusive lookups, depending on input data:

- Lookup into a ring.
- Lookup among locations.
- Lookup among equipments.

The first kind of lookup is the **lookup into a ring**. In this case, the first step is to get the equipments that constitute the ring. Next, the algorithm tries to locate a route defined only through links that belong to the ring (between those equipments).

For this lookup to get a valid route, it is necessary that the end equipments are included in the ring too. Although this condition could make think that it is not a very useful lookup, it can be valid if we are trying to restore the service in a specific point, and we define the set of links into the ring as the route that needs to be modified.

The algorithm can be formalized according to the next pseudocode:

```
GET EQUIPMENTS IN THE RING
IF (EqOri AND EqDes) IN THE RING
   SET EqIni = EqOri
   SET EqAnt = EqOri
   SET Route = ()
   LOOKUP:
       GET LINK RING (EqIni, Link1,
       EqDes1, Link2, EqDes2)
       IF (EqDes1 = EqDes) THEN
         ADD (Link1, Route)
         EXIT (OK)
       ELSE IF (EqDes2 = EqDes) THEN
         ADD (Link2, Route)
         EXIT (OK)
       ELSE IF UltEqu (EqDes1, EqDes2)
       THEN
         EXIT (NO_ROUTE)
       ELSE
         AsignEqu (EqAnt, EqIni, EqDes1,
         EqDes)
         GOTO LOOKUP
       END IF
```

The second type is the **lookup among locations**. This lookup is used, as mentioned before, when there is no certainty of obtaining a valid route totally defined through a sequence of

 ISBN: 978-1-61208-133-5

links and equipments, but we think that there is a sequence of locations and links, so in this case the network modifications are minor ones.

The lookup begins at the initial location. From that point, links are selected and the end location of each one is explored in the same way. When the lookup arrives to a location where there is no available link or the route length reaches the maximum number of links and it is not the end location, it goes back until the last one where there are other available links and tries to go to a different location.

In this case, the first and last locations are obtained through the end equipments, but the rest of the locations are obtained through the name of the links, not the equipments where the links end. The link selection takes into account the actual capability or the possibilities, which can be achieved through some network configurations, according to the input parameters. It also takes into account the special resource features: in a normal way, only permanent and working resources are selected, and there is no network restriction. These conditions change if there is any special feature, so the selection is widened (failed and temporary resources) or restricted (only a special network).

The algorithm can be formalized according to the next pseudocode:

```
GET_Loc_Equ (EqOri, LocOri)
GET_Loc_Equ (EqDes, LocDes)
SET Route = ()

LOOKUP_LOC (LocOri, LocDes, Route, Res) ⇒
SET LocIni = LocOri
LOOKUP:
   SET Res = NOK
   FOR Link IN GET_LINK_AVAI (LocIni,
   ContCond, EspFeat)
       GET_Loc_Link (Link, LocIni, LocEnd)
       IF (LocEnd = LocDes) THEN
         ADD (Link, Route)
         SET Res = OK
         EXIT (OK)
       ELSE IF (Long (Route) = NumMax – 1)
       THEN
         Next (Link)
       ELSE
         LOOKUP_LOC (LocIni, LocDes, Route,
         Res)
         IF (Res = OK) THEN
              EXIT (OK)
         ELSE
              NEXT (Link)
         END
       END
   END
   EXIT (Res)
```

The last kind of lookup is the most useful of them, the **lookup among equipments**. In this case, the route ends are the equipments themselves, not the locations, and the links must connect directly these equipments.

The lookup begins at the initial equipment. From that point, links are selected and the end equipment of each one is explored in the same way. When the lookup arrives to an equipment where there is no available link or the route length reaches the maximum number of links and it is not the end equipment, it goes back until the last one with available links and tries to go to a different one. The link selection has the same considerations described in the previous point.

The algorithm can be formalized according to the next pseudocode:

```
SET Route = ()

LOOKUP_EQU (EquOri, EquDes, Route, Res) ⇒
SET EquIni = EquOri
LOOKUP:
   SET Res = NOK
   FOR Link IN GET_LINK_AVAI (EquIni,
   ContCond, EspFeat)
       GET_Equ_Link (Link, EquIni, EquEnd)
       IF (EquEnd = EquDes) THEN
         ADD (Link, Route)
         SET Res = OK
         EXIT (OK)
       ELSE IF (Long (Route) = NumMax – 1)
       THEN
         Next (Link)
       ELSE
         LOOKUP_LOC (EquIni, EquDes, Route,
         Res)
         IF (Res = OK) THEN
              EXIT (OK)
         ELSE
              NEXT (Link)
         END
       END
   END
   EXIT (Res)
```

The last step, after any type of lookup obtains results, is the boundary check, so the routes that do not fill the boundaries are discarded. This check does not include the condition related to ring use because it is translated into a type of lookup.

The algorithm receives the number of routes that must select, but time is an important matter that has to be considered, because the lookup is useful only if it obtains routes in a reasonable period. Therefore, there is a general timer, and when it ends the algorithms gives as an output all the routes selected until that moment, although the number is smaller than the required one.

 ISBN: 978-1-61208-133-5

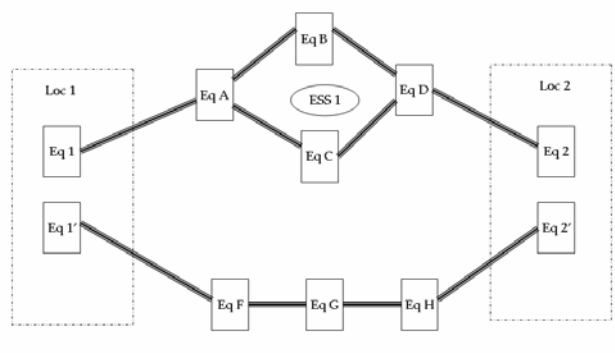


Figure 1. Topologic network view for a practical lookup

As the last step, we are going to explain the algorithm work with an example. In Fig. 1, we have a simplified network topology, where we can describe the expected behavior in each one of the lookups explained before.

In this topology, we have all the basic elements for the algorithm:

- Locations (Loc) and equipments (Eq), used for the route lookup.
- Rings (ESS), which define the protection schemes into the network.
- Links between the equipments.

The desired route will begin at Eq. 1 and will finish at Eq. 2. For a simplified explanation, we are supposing that all the resources are permanent and there is no failure into the network (the failed point will be outside this network). We are not describing the network type either, although the results are obtained mainly in an optical network, so we can consider that the ring is an optical ring.

The results obtained according to the different kinds of lookup will be:

- Lookup into a ring: In this topology, there would be no route between Eq. 1 and Eq. 2. The lookup should be between two equipments included into the ring: Eq A, Eq B, Eq C, Eq D, to obtain a valid result.
- Lookup among locations: in this case, the valid route would be between any equipment included in Loc 1 (Eq 1 and Eq 1') and any equipment included in Loc 2 (Eq 2 and Eq 2'), so there are two possible routes: crossing the ESS 1 or through equipments Eq F, Eq G and Eq H.
- Lookup among equipments: in this case, the only valid route would be between equipments Eq 1 and Eq 2, crossing the ring 1.

IV. RESULTS

The method has been applied over a huge network (more than 50,000 equipments) and to a wide range of cases varying the aforementioned parameters (route features and resource state). It has been verified that even the worst results in terms of lookup time mean a significant improvement compared to the non-optimized lookup. Telecommunication operators could efficiently solve most situations by means of this procedure.

The method has been applied over a huge network (more than 50,000 Equipments) and to a wide range of cases, defined by different signal types. In any signal type, we have varied:

- The route features;
- The boundary conditions and
- The resource special features.

It has been verified that even the worst results in terms of lookup time mean a significant improvement compared to the non-optimized lookup.

Each test has been defined following the process below:

- An origin and destination for the desired route are selected.
- A non-optimized lookup is carried out over the network.
- A signal type (route features) for the desired route is selected.
- An optimized lookup (taking into account only the route features) is processed.
- A set of variations is applied to all the boundary conditions for each origin and destination, showing different outcomes according to these variations.
- Start and finish time are stored for each lookup.

In Table 1, a summary of the results is shown to quantify the improvement. These results are obtained with six links as the maximum route length using as a 100-percent reference the time consumption for the non-optimized lookup, quantifying each result as a fraction of that time.

TABLE I. RESULT SUMMARY

Cases	Result
Maximum	63,46%
Minimum	28,13%
Average	52,63%

According to the input conditions, the lookup has been applied into the optical networks and through different networks. Results show that with different networks involved, the algorithm does not adapt so well to network features, but it improves the result too.

When the route length increases the improvement decreases very fast, and the lookup time is almost the same than in a general case, because in both of them, it grows exponentially, and the lookup is no useful in a practical way.

V. CONCLUSSIONS

The results show that taking into account not only geographical considerations as locations, but also data transmission network features, basically route continuity

 ISBN: 978-1-61208-133-5

through equipments, the lookup times are significantly minor than the original ones.

The explained lookup, using the algorithm with the proposed restrictions, is efficient if the route length is moderated. This supposition is reasonable in many scenarios within the telecommunication operators' transmission network. When the necessary route length increases over the average (around six hops), the lookup time tends to pair the time it would be obtained if the lookup used an algorithm without restrictions.

As mentioned before, after the lookup obtains a route there is a validation where the algorithm checks if the route fulfills the special network features and boundary conditions. This step offers a good point to introduce new considerations into the algorithms, so that the route optimization can embody other objectives apart from the technical ones (for instance: economic costs or energy consumption).

REFERENCES

[1] L. R. Ford Jr. and D. R. Fulkerson (1962). "Flows in networks". Princeton University Press.

[2] D. Medhi and K. Ramasamy (2007). "Network routing: algorithms, protocols, and architectures". Morgan Kaufmann.

[3] J. Doyle and J. Carroll (2005). "Routing TCP/IP", Included in: *Volume I, Second Ed. Cisco Press.*

[4] N. Spring, R. Mahajan, and T. Anderson (2003). "Quantifying the causes of path inflation". Included in: *Proc. SIGCOMM 2003*.

[5] Gerald Ash (1997). "Dynamic routing in telecommunication networks". McGraw-Hill.

[6] M., A. Wainwright (2003). "Small road network". Included in: *I. Kennedy, Teletraffic Lecture Notes, School of Electrical and Information Engineering, University of the Witwatersrand, 2003*.

[7] Recommendation G.709/Y.1331: Interfaces for the optical transport network (OTN). ITU.

[8] M.L. Mouronte, R.M. Benito, J.P. Cárdenas, A. Santiago, V. Feliú, P. van Wijngaarden, and. L.G. Moyano (2009). "Complexity in spanish optical fiber and SDH transport networks". Included in: *Computer Physics Communications, Volume 180, Issue 4, April 2009.*

[9] N. Leeprechanon, P. Limsakul, and S. Pothiya (2010). "Optimal Transmission Expansion Planning Using Ant Colony Optimization". Included in: *Journal of Sustainable Energy & Environment 1 (2010)*.

[10] S. Bohacek, J. Hespanha, J. Lee, C. Lim, and K. Obraczka (2007). "Game Theoretic Stochastic Routing for Fault Tolerance and Security in Computer Networks". Included in: *IEEE Transactions on Parallel and Distributed Systems, Vol. 18, Issue 9, September 2007.*

[11] C. Fang, C. Feng, and X. Chen (2010). "A heuristic algorithm for minimum cost multicast routing in OTN network". Included in: *IEEE Xplore May 2010*.

 ISBN: 978-1-61208-133-5

One Approach to Improve Bandwidth Allocation Fairness in IP/MPLS Networks Using Adaptive Treatment of the Traffic Demands

Tarik Čaršimamović
Directorate for Information Technologies
BH Telecom JSC
Sarajevo, Bosnia and Herzegovina
tarik.carsimamovic@bhtelecom.ba

Enio Kaljić, Mesud Hadžialić
Faculty of Electrical Engineering
University of Sarajevo
Sarajevo, Bosnia and Herzegovina
enio.kaljic@etf.unsa.ba, mesud.hadzialic@etf.unsa.ba

***Abstract*—In this paper, we propose an algorithm for adaptation layer in order to improve fairness in bandwidth allocation among different traffic classes in IP/MPLS networks under heavy traffic load. We propose a definition of the blocking frequency of traffic flows at the entry of autonomous network domain and proportional-priority coefficient per traffic class and use them as the input parameters of the adaptation mechanism. In order to evaluate the validity of the proposed algorithm, we need proper simulation tools and for that purposes we extend OPNET Modeler 14.5 with the modules for adaptation process. We also propose a development methodology for the design of modules for adaptation process within network simulators. The simulations results give us the proof of our hypothesis that with a proper adaptation layer we can improve the fairness of bandwidth allocation among different traffic classes under heavy network load and at the same time keep the required QoS conditions in the preferred boundaries.**

Keywords - Adaptation layer; algorithm; bandwidth allocation; blocking frequency; dynamic adaptation structure; development methodology; fairness; LSP; mpls; NGN; proportional-priority coefficient; RSVP.

I. INTRODUCTION

One of the key characteristics of the new generation network (NGN) environment is a necessity that the network is capable of handling an ever-increasing demand uncertainty, both in volume and time. In such environment, very often the total traffic demands exceed the available network capacity, and the traffic classes with higher priority could occupy the entire network capacity leaving no space for traffic flows with lower priority. Proper adaptation mechanisms could give acceptable results in adequate bandwidth allocation to the traffic variation and in fair treatment of all traffic classes. The fairness in the resource allocation among traffic flows depends on algorithm used during the process of adaptation to the real conditions of traffic load. The fairness of adaptation algorithm represents the ability of the model to distribute available resources in such manner that the probability of traffic blocking for any particular traffic class is the same as the overall blocking probability. We can use ratio P_i of the allocated resources G_i to the requested resources B_i of the particular traffic flow demand $(P_i = \frac{G_i}{B_i} x100)$ as a measure of the algorithm fairness. Three types of fairness index are possible [15]: balanced fairness, max-min fairness and proportional fairness. According to the results shown in this reference, proportional type of fairness is the most suitable type in the case when the network resources are distributed among different traffic classes and when the adaptive method of resource allocation is used. In the same paper, fairness index J for the proportional type of fairness among n traffic classes is proposed as such:

$$J = \frac{(\sum_{i=1}^{n} P_i)^2}{n\sum_{i=1}^{n} P_i^2} \quad (1)$$

If the value of the fairness index J is higher than 0.9, or in an extreme situation higher than 0.8, one can say that the resource allocation mechanism is fair [3]. Otherwise, variations in resource distribution are significant and blocking percentage of the lower-priority traffic classes is outside of the acceptable margins.

NGN is a packet-oriented network supporting Quality of Service (QoS) based on different type of transport technologies. The most preferred protocol in NGN is IP. There are different approaches for the QoS provisioning in IP based networks: Integrated Services (IntServ), Differentiated Services (DiffServ), combined IntServ/DiffServ, Multiprotocol Label Switching (MPLS), etc. MPLS is a popular transport technology that uses labels which are imbedded between layer two and layer three headers in order to forward packets. Packets are forwarded by switching packets on the basis of labels and not by routing packet based on IP header. One of the major advantages of MPLS networks is the inherent support to traffic engineering. We can also use a combination of MPLS and DiffServ and treat packets of the same Forward Equivalence Class (FEC) in accordance with the DiffServ procedure. Using MPLS Traffic Engineering (MPLS-TE) based on the network state detection we can balance traffic load among different Label Switched Paths (LSPs), but we cannot dynamically change allocated bandwidth to the LSPs.

 ISBN: 978-1-61208-133-5

In order to adapt to dynamics of traffic demands and to allocate sufficient bandwidth to the LSPs, as well as to improve fairness in the resources allocation among traffic flows, we introduce adaptation layer, working in two regimes:

- fuzzy controller regime, when the overall traffic demand is elastic and in average less than network capacity. In this case the adaptation process is realized by the means of fuzzy logic [19],
- proportional-priority regime, when the overall traffic demand is higher than the network capacity. In this case the adaptation process allocates bandwidth among traffic classes in such a manner that minimal bandwidth is guaranteed to each traffic class and the rest of network capacity is shared on the proportional basis among traffic classes, (equation 3).

The adaptation layer supports dynamic exchange between fuzzy controller regime and proportional-priority regime depending on the ratio between traffic load and the network capacity C. If the ratio is $\sum_{i=1}^{n} B_i \Big/ C \geq 1$ we switch the adaptation layer algorithm to proportional-priority regime, and when it decreases below 1, we switch adaptation algorithm to the fuzzy controller again.

In order to prove validity of our adaptation layer concept and sustainability of the fairness improvement concept of bandwidth allocation among traffic flows, we need proper simulation tools. Because there are no network simulators supporting the proposed adaptation layer algorithm and dynamics of this algorithm, we established a methodology for development of adaptation layer within network simulators and we also developed the adaptation layer code in C++ within OPNET core structure of node model (Label Edge Router - LER) and within core structure of process model of the Resource Reservation Protocol (RSVP-TE) used in the OPNET modeler.

II. MECHANISMS AND ARCHITECTURES FOR ADAPTIVE TREATMENT OF TRAFFIC DEMANDS

The goals of adaptive treatment of traffic demand in NGN are to:

- fulfill QoS requests of any traffic class,
- eliminate drops of any traffic flows,
- decrease congestion within network,
- Rise efficiency of network capacity.

In order to successfully achieve those goals, appropriate mechanisms for the bandwidth allocation, for the routing optimization and for reaction to the failure conditions are needed. During the research project COST 257 [4], several types of reactive and preventive approaches for network control are investigated:

- the flow control scheme (fluid flow model, discrete-time Markov model or control theory model) for reactive approach,
- the admission control method (Measurement Based Admission Control - MBAC, Traffic Description Based Admission Control - TDBAC, Experience Based Admission Control - EBAC or End-point Admission Control - EAC) for preventive approach,
- the active queue management or fuzzy congestion control as a new control trends.

Preventive controls usually try to limit the number of connections or to enforce connection to use only a limited amount of resources. In IP networks, the specifications of protocols such as RSVP or MPLS make admission control possible. MPLS traffic engineering is aimed at optimizing the network path to ensure efficient allocation of network resources, thereby avoiding the occurrence of congestion on the links. [1] [2]. During the research project Traffic Engineering for QoS in Internet at Large Scale (Tequila) [6], it was shown that a combination of MPLS and DiffServ could be acceptable solution for load balancing in IP networks when multi-path routing is used. Also, during the research project COST 239 [5] it was shown that, in case of large traffic load, the highest efficiency of the resource usage is in the networks which use border-to-border budget based network admission control (BBB NAC) as a budget-oriented method for allocation of virtual bandwidth. BBB NAC could be realized using RSVP extension for LSP tunnels establishing explicit LSPs with guaranteed bandwidth.

Several research projects investigate possible architecture of dynamic provisioning of QoS to the particular traffic flow. The basic result of KING (Key Components for Internet of the Next Generation) project [18] includes development of adaptive architecture in which, by continuous monitoring of network conditions, the network parameters could be adapted to traffic demand.

We combined the admission control method (MBAC – BBB NAC) with policy based control of adaptation layer to dynamically adapt budget of the NAC in order to decrease blocking frequency and to raise fairness of bandwidth allocation among traffic classes.

III. DESIGN OF MODEL FOR ADAPTATION PROCESS

A. Looking for a Simulation Tool for Validation of Designed Model

The adaptation processes are capable of a continuous monitoring of the network parameters and performing their adaptation in accordance to the ever-changing traffic demands. Possible solution for such structure of adaptation process could be an active resource allocation based on the dynamic monitoring of their availability and of their sustainability to effectively transfer traffic.

The efficiency of such structures should be evaluated and an adequate simulation model which can adequately represent mechanisms and architecture for adaptive treatment of traffic demand is needed. Therefore, in this chapter we investigated possibilities of existing network simulators to support structure of adaptation algorithm we used in this paper. The analysis is based on a review of scientific studies in this field and documentation available for the network simulators.

In [8], a scheme for an adaptive bandwidth reservation in wireless multimedia networks was examined. For the

 ISBN: 978-1-61208-133-5

purpose of validation of the proposed solution, necessary module for network simulator OPNET (Modeler 8.0) was developed. However, examination of the latest available version of the network simulator OPNET (Modeler 14.5) showed that these modules are not supported by the simulator manufacturers and as such is not included in the set of available modules. In [9], a solution for the adaptive bandwidth allocation in MPLS networks using a control with one-way feedback was given. The proposed solution was tested in a network simulator ns-2.27. But this solution is dedicated for particular problem and only in ns 2-27. Because of that it is inflexible for usage in general.

Elwalid et al. [10] discussed adaptive traffic engineering in MPLS networks and their effort to develop their own simulator is the significant sign that there is a poor support for the dynamic adaptation structures in the available network simulators.

Reviewing the documentation about the available network simulators we determined that none of those network simulators have built-in support for the dynamic adaptation structures.

As the available simulators have no appropriate support for dynamic adaptation structures, and the same is necessary to test proposed structures, one of the objectives of this paper is to establish the methodological approach to development of adaptation layer in the network simulator. This methodology will be used for development of the adaptive layer modules within a chosen network simulator.

B. One Possible Solution of Adaptation Layer

In [11][12][13][14], an active queue management as network control mechanism is proposed. This approach requires execution of adaptation layer processes at every node in the network, so making it unsuitable for MPLS based networks.

In this paper we use a different approach for solution of the adaptation layer algorithm in order to increase the resource usage efficiency, to provide proper QoS to any traffic class and to improve fairness in bandwidth allocation among traffic flows within MPLS based networks. The following will give a brief overview of the solution. A full description, that includes a detailed explanation of the algorithm, is given in [16]. In this paper, MPLS is used as a transport technology on the network layer. Network capabilities to provide sufficient resources at a given time for a given traffic class are controlled at ingress node. Instead of assigning bandwidth to a particular link, provisioning of QoS requirements is done by assigning a virtual budget at ingress node for all relations between ingress and egress nodes using BBB NAC. BBB NAC in MPLS for LSP with a guaranteed bandwidth can be established by RSVP extension for LSP tunnels and then managing the right to network access can be made for each stream. Adaptation layer collects statistical data from the network layer and uses that information to generate a global view of the current state in the network. Detection of the current state in the proposed architecture is based on the utilization of the budgets allocated to the NACs, the frequency of blocked reservation (e.g. blocking frequency) and on the utilization of links' capacity. Adaptation process includes adjustment of the amount of available bandwidth for each traffic class separately and optimization of internal routing.

Adaptation layer follows its own internal strategy and optimization algorithms in order to adapt network performance to the traffic load variations. Adaptation layer has two regimes:

- Fuzzy controller regime, which is realized by means of fuzzy logic, based on the blocking frequency of traffic flows. The adaptation process is executed in discrete cycles. Blocking frequency (FB) measured in particular cycle and difference in blocking frequency ((ΔFB) between two cycles are the input variables of triangle fuzzy membership function. This membership function and fuzzy rules, given in [16], are used for determination of the value of proportional coefficient (n_{ij}) to the bandwidth increment ΔB_i of i-th traffic class within j-th cycle. The bandwidth increment ΔB_i is given in advance for every traffic class. To adjust amount of allocated bandwidth G_{ij} of the i-th traffic class in j-th cycle to the actual traffic demand, adaptation algorithm changes allocated bandwidth of i-th traffic class in j-1st cycle with the value of $n_{ij}\,\Delta B_i$ in accordance with the following formula:
$$G_{ij} = G_{ij-1} + n_{ij}\Delta B_i \qquad (2)$$
- Proportional-priority regime, which is based on minimum bandwidth allocated to the i-th traffic class (min_{pi}) and proportional-priority coefficient δ_{ij} of the i-th traffic class in j-th cycle, performs its functions in accordance with the following formula:
$$G_{ij} = \min_{pi} + \delta_{ij}\left(C - \sum_{i=1}^{n} \min_{pi}\right) \qquad (3)$$

The criterion for switching between the two regimes is fulfilled when the sum of requested capacity of traffic classes B_i is bigger than network capacity (C) $\sum_{i=1}^{n} B_i > C$.

In the previous studies [17], behavior of the overall system, which performs its control functions automatically, autonomously and in an adaptive manner, is usually described by means of the following parameters:

- blocking frequency of traffic flows (FB),
- fairness of the allocation of resources to the traffic flows (P, J),
- utilization of network capacity (γ).

Blocking frequency of traffic flows (FB), we used as a key parameter for adaptation process, is defined as the total number of the rejected resource reservation (b_{miTN}) in all n classes of traffic within the determined time interval k. Measurement of rejected traffic flow is performed at the NAC any time new traffic flow ($c(f^{new}_{v,w})$ added to the existing traffic flows (c(f)) requests capacity which is higher than the available capacity (C(BBB)). of the given resources

between nodes v and w. While the frequency of blocking can be defined as the maximum blocking probability or a relative ratio of blocked and offered traffic, this definition of blocking frequency, which treats all traffic classes simultaneously and only at the input node, is simple to measure and easy to calculate:

$$FB = \sum_{i=1}^{n} FB_i, \quad FB_i = \sum_{T_{N=1}}^{k} b_{miT_N} \tag{4}$$

$$b_{miT_N} = countif\left\{\left[c(f^{new}{}_{v,w}) + \sum c(f)\right] > C(BBB)\right\}$$

Fairness of resource allocation between traffic classes depends on the resource (bandwidth) allocation algorithm used during the process of adaptation to the actual traffic demands. Consideration of fairness makes sense only if the total amounts of requested resources exceed the capacity of available network resources. Otherwise, the problem boils down to utilization of network resources and to load balancing in order to assess the cost of depreciation and to even utilization of network resources. Fairness of the adaptive algorithm is the ability of the model to distribute the available resources in such a way that any traffic class does not give preference outside of the defined priority mechanism. The main goal of equitable allocation of resources assessment includes quantification of differences in distribution of resources between traffic classes by measuring variations in the ratio of allocated resources. We used equation 1 to evaluate fairness of proposed adaptation algorithm. We also compare the same fairness index achieved in the network architectures operating in adaptation mode and in the network architecture operating in non-adaptation mode to evaluate improvement in fairness.

Simulation model of network architecture we used consists of the adaptation and network layers. The adaptation layer performs a calculation of the new bandwidth budget for network admission controllers (NAC) and a calculation of new metrics on the links in order to adapt network performance to the actual traffic conditions. Measurements of blocking frequency (FB), accepted traffic (c_{ai}), rejected traffic (c_{ri}) and LSP load are performed in the regular time intervals at the ingress node in order to provide input data for operation of adaptation layer. The network layer autonomously executes forwarding functions of the packets and ensures QoS requirements using the capabilities of existing technologies and protocols. This type of network architecture represents the optimal set of available technologies with flexible topological landscape. Admission control functions, based on timely adjusted allocated bandwidth and load balancing functions by means of MPLS traffic engineering capabilities, are executed only at the ingress node of the autonomous network domain.

IV. DEVELOPMENT METHODOLOGY OF AN ADAPTATION LAYER WITHIN NETWORK SIMULATOR

In order to develop an adaptation layer which is independent of the adaptation mechanism of the used technology or of the network simulator, it is necessary to define a development methodology. We established a development methodology of an adaptation layer within network simulator which has eight steps shown in Fig. 1. Each of those steps will be explained in this chapter.

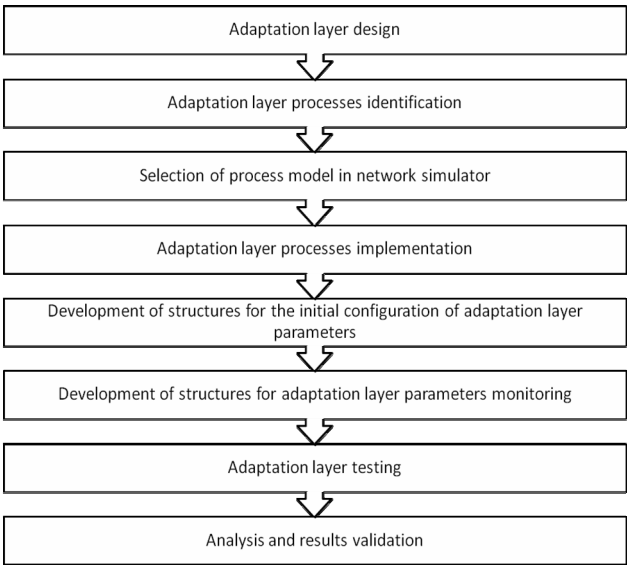


Figure 1. Development methodology of adaptation layer

A. Design of the Adaptation Layer

In the chapter III we explained the basic functions and principles of our adaptation layer. The detailed design of adaptation layer with adaptation algorithm, input and output variables, decision criteria and pseudo code of the adaptation layer components are given in [16].

B. Adaptation Layer Processes Identification

For the purpose of execution of adaptation layer functions we identify next three processes:

- measurement of input variables,
- adaptation, and
- output parameters control.

The first process is deterministic and it is performed in regular time intervals (T_N). The task of this process is measurement of flow intensity of each traffic class and a measurement of blocking frequency at the entry into the MPLS domain.

The process of adaptation is also deterministic, and it is performed in regular time intervals determined by the duration of discrete cycle of adaptation. The task of this process is to calculate a new budget based on input variables.

The last process is a stochastic process and its execution is caused by the decision results of adaptation process. The task of this process is to allocate a new budget to the network admission controller.

C. Selection of Process Model in Network Simulator

A number of network simulators are available today. Some of them, used in scientific researches, are ns-2/3, OPNET, Omnet + +, GloMoSim, Nets, etc. Selection of adequate network simulator should be based on the

characteristics of the simulators corresponding to the needs of adaptation layer developed as the target platform. We choose OPNET Modeler 14.5 as a proper network simulator considering next properties of the chosen simulator:

- simulation is based on the discrete network states (FSM-based approach),
- it supports traffic profile we intend to use during a simulation,
- it supports the network technologies and protocols we selected for a simulation model,
- it is suitable for prototype research such as this simulation model,
- it is easy to configure,
- it has relatively good documentation and support,
- it can be extended for adaptation layer (supports C++).

Since the proposed adaptive layer is a prototype of generic adaptation layer and as such does not exist in the selected network simulator, the whole adaptation layer should be developed based on the pseudo code given in [16], taking into account the constraints of simulator architecture. The architecture of the network simulator OPNET Modeler 14.5, extended with necessary modules for adaptation layer, is presented in Fig. 2 below.

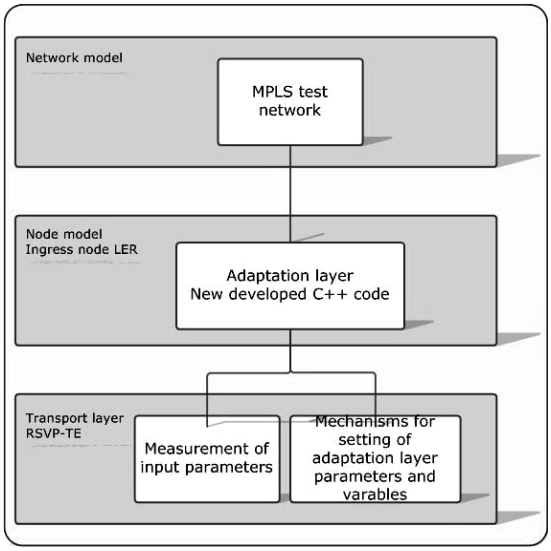


Figure 2. Hierarchical architecture of OPNET simulator

Network simulator OPNET Modeler 14.5 is hierarchically organized. A network model is located at the highest level of hierarchy. The network model is composed of nodes and links connecting the nodes. Each node is defined by the node model (workstation, switch, router, server, etc.). Node model consists of processors that are described in process models. Process models are described in FSM's (Finite State Machine) and transfer functions written in C++ programming language. Transfer functions rely on the core functions of the simulator. The core simulator consists of pre-compiled libraries, whose source code is not available. The kernel is based on discrete event simulation.

Node at which network admission control functions are performed is the ingress LER of the MPLS domain. Bandwidth control at the entry of the network is ensured by establishing an explicit path with guaranteed bandwidth. The protocol that is responsible for setting up LSPs is RSVP-TE. Therefore, the logical choice of process model within the simulator is RSVP process model. We used the network model which consists of five traffic sources nodes, the ingress edge router, four core MPLS routers, the egress edge router and five traffic destination nodes.

D. Adaptation Layer Processes Implementation

Ingress LER node model (Cisco 7600) consists of processors and queues associated with packet or statistic wires. We introduce the new statistical flows between the MAC (Media Access Control) queues and RSVP process model in order to take periodical measurements of the input variables, such as the mean intensity of flows, number of the rejected reservations, etc.

Process models in the OPNET network simulator are based on FSM. Passing from one state to another is initiated by different types of interruptions (packet arrival, arrival of new statistics value, user-defined stop, etc.).

During the transition stage different functions could be called. Besides the FSMs, the main components of a process model are the state variables, temporary variables, function block with headers, block functions, block for debugging and scheduling process block. Each process model also has attributes, interfaces, local and global statistics. Attributes and statistics can be promoted to a higher level, i.e. at the level of the node model.

E. Initial Configuration of Adaptation Layer Parameters

The initial parameters of adaptation layer, such as initial bandwidth per each traffic class (B_i), minimal bandwidth per traffic class ($min_{pi,}$), proportional-priority coefficient (δ_i), bandwidth increment (ΔB_i), are defined in [16]. Those initial parameters are subject to changes during the exploitation period (if the traffic environment changes dramatically) or during the simulation process (to be able to perform different simulation scenarios). For this purpose we need a proper structure within a simulator which offers changeability of the initial configuration settings and changeable setting of its parameters. The development process of that structure has the following steps:

- definition of the adaptation layer attributes within the set of the existing process model attributes,
- promotion of the attributes from a process model level to the level of the node model,
- coding the input function in C++ to retrieve attributes when the simulation starts.

F. Monitoring of Adaptation Layer Parameters

In Sections I and III, we defined parameters which should be monitored such as blocking frequency (FB),), fairness index (J), difference of blocking frequency (ΔFB) used as input variable for fuzzy membership function, etc. Those parameters should be measurable and monitored in order to qualify adaptation process execution and to use them as the

 ISBN: 978-1-61208-133-5

input parameters to the adaptation layer. For this purpose we need a structure within simulator which offers a possibility to measure and monitor the values of the adaptation layer parameters. The development process of that structure has the following steps:

- definition of the local statistics in the process model,
- promotion of statistics on the level of the node model,
- coding of the function in C++ to record statistics.

G. *Testing, Analysis and Result Validation*

Those two steps of the development methodology are explained in Section V.

V. THE SIMULATION RESULTS

The simulation model created for testing purposes of adaptation layer is shown in Fig. 3 below. All nodes in the access part of the network are connected using 10 Gbps links, while the core routers are connected using 1 Gbps links. OSPF protocol is used as an IGP, and RSVP-TE protocol is used for establishment of LSPs.

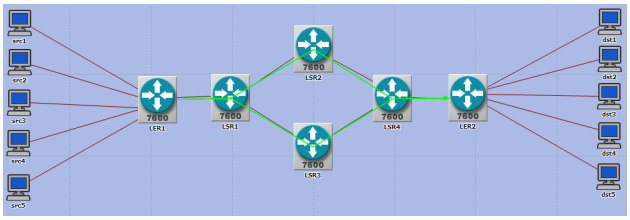


Figure 3. Simulation model for adaptation layer testing

The detailed dynamics of traffic demand of traffic classes, presented in Table 1 and used during a simulation process, are given in the thesis [16]. The traffic generators are configured so that the average network load is 80%. The network capacity is 2 Gbps. The peak load of the network is about 2.5 Gbps. The initial values for those traffic classes, in the case that average network load is 80% of the network capacity, are given in Table 1 below.

TABLE I. INITIAL VALUES OF TRAFFIC SOURCES

Traffic class	Initial BW kbps	Minimal BW kbps	Maximal BW kbps	BW incr. ΔBi	Proportional-priority coefficient δi
EF	8,270	5,990	14,000	125	$1.2\frac{B_{ij}}{\sum B_{ij}}$
AF1	465,110	319,760	700,000	10,000	$\frac{B_{ij}}{\sum B_{ij}}$
AF2	586,390	403,140	800,000	22,000	$0.9\frac{B_{ij}}{\sum B_{ij}}$
AF3	5,690	3,850	14,000	200	$0.9\frac{B_{ij}}{\sum B_{ij}}$
BE	431,310	286,528	700,000	6,000	$0.8\frac{B_{ij}}{\sum B_{ij}}$

During the simulation process we observe a distribution of requested bandwidth per each traffic class B_i and distribution of allocated bandwidth per each traffic class G_i in the same time window. We perform those observations in the adaptation mode of network architecture and in non-adaptation mode of the same network architecture in order to validate the accurate of the adaptation layer processes and to evaluate improvement in resource utilization as well as in QoS satisfaction of the requests of any traffic class.

We also observe values and distribution of fairness index (J) in both modes of network operation and values and distribution of ratio of the allocated resources to the requested bandwidth per each traffic class, in order to evaluate improvement of fairness in adaptation mode of network operation compared to the non-adaptation mode of operation. To justify the results of simulation process we repeated the simulation and the same measurements and observations in the case that the average network load is 100% of the network capacity. During the simulation process we take measures every 10 seconds and average those measurement values in time window of one minute, using those average values to calculate parameters which are needed for adaptation process of our adaptation algorithm.

By means of Figures 4 to 6 below, as a part of simulation results, we will show the outcomes of proposed adaptation algorithm, as well as of the extension of the OPNET structure. The whole scope of simulation results, from which we proof our entire concept, can be seen in [16].

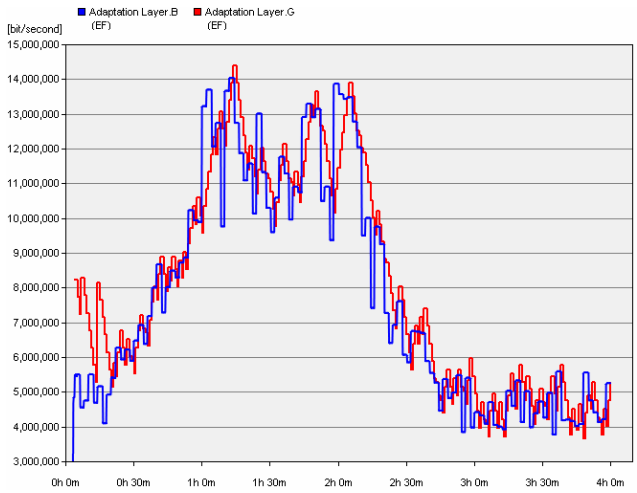


Figure 4. Requested and allocated bandwidth for EF traffic class

From the Fig. 4 we can see that allocated bandwidth G for EF traffic class pretty well follows the required bandwidth B. This confirms that the adaptation layer functions properly and accurately. We can see the same results for other traffic classes and for average load of 100%.

Fig. 5 shows us that introduction of adaptation layer improves the fairness of resource allocation in the network. During the non-adaptive mode, the ratio (P) of the allocated resources to the requested resources for EF class was unstable and goes up to 200%, while during the adaptive mode of network operation this percentage was stabilized

and dropped to 100%, as is the preferred value for all traffic classes.

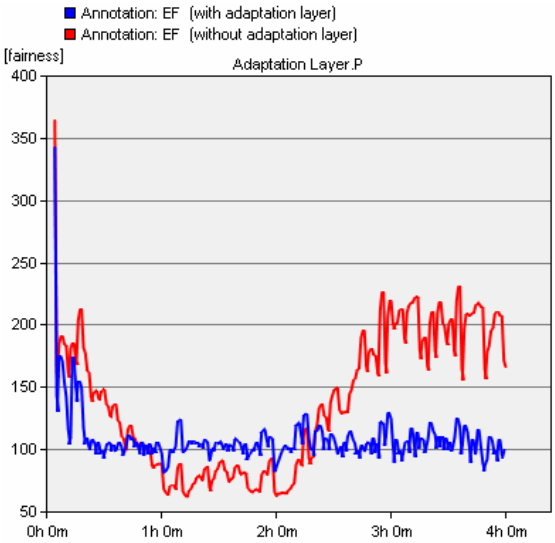


Figure 5. Ratio of the allocated resources for the EF traffic class

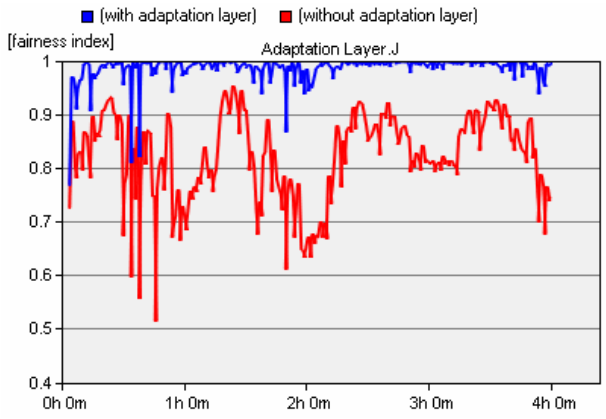


Figure 6. Fairness index

Fairness index (Fig. 6) is, in the adaptation mode of network operation, maintained above 0.96 with brief outages of up to 0.8, while the same index, in non-adaptation mode, is very unstable and drops up to 0.5. These results give us the proof of our hypothesis that with a proper adaptation layer we can improve the fairness of bandwidth allocation among different traffic classes under heavy network load and at the same time keep the required QoS conditions in the preferred boundaries. We can also conclude that the proposed adaptation algorithm behaves properly.

VI. CONCLUSION

The simulation results have shown that the proposed adaptation algorithm can significantly improve the fairness of bandwidth allocation among different traffic classes under a heavy traffic load in IP/MPLS networks, while keeping the required QoS conditions to any traffic class within the boundaries as preferred. The bandwidth allocated to any traffic class follows the required one, and in the case of a sufficient bandwidth, the QoS requests are guaranteed. We see future work in researching the impact of different fuzzy algorithms and membership functions in the adaptation layer.

REFERENCES

[1] A. Farrel, "Network Quality of Service: Know It All. Burlington," Morgan Kaufmann Publishers, USA, 2009.

[2] A. Stavdas, "Core and Metro Networks," John Wiley & Sons Ltd., 2010.

[3] C.A. Kamienski, "The Case of Inter-Domain Dynamic QoS based Service Negotiation in the Internet," Computer Communications, vol. 27, pp. 622-637, 2007.

[4] COST-257, Final Report, "Impact of New Services on the Architecture and Performance of Broadband Networks", 2000.

[5] C. Hoodgendoorn, "KING Research Project Overview," 2005.

[6] D. Godens, "Functional Architecture Definition and Top Level Design," Tequila Project, 2000.

[7] G. Hasslinger and J. Mende, "Measurement and Characteristics of Aggregated Traffic in Broadband Access Networks," in Proceedings of ITC 20, pp. 998-1010, Ottawa, 2007.

[8] X. Chen and Y. Fang, "An adaptive bandwidth reservation scheme in multimedia wireless networks," in IEEE Globecom, San Francisco, pp. 2830-2834, 2003.

[9] T. Yokoyama, K. Iida, H. Koga, and S. Yamaguchi, "Proposal for Adaptive Bandwidth Allocation Using One-Way Feedback Control for MPLS Networks," IEICE TRANS. COMMUN., vol. E90-B, no. 12, pp. 3530-3540, Dec. 2007.

[10] A. Elwalid, C. Jin, S. Low, and I. Widjaja, "MATE: MPLS Adaptive Traffic Engineering," in IEEE INFOCOM, Anchorage, pp. 1300-1309, 2001.

[11] W.-S. Lai, C.-H. Lin, J.-C. Liu, and H.-C. Huang, "Using Adaptive Bandwidth Allocation Approach to Defend DDoS Attacks," International Journal of Software Engineering and Its Applications, vol. 2, no. 4, pp. 61-72, Oct. 2008.

[12] A. Kamra, H. Saran, S. Sen, and R. Shorey, "Fair adaptive bandwidth allocation: a rate control based active queue management discipline," Computer Networks, no. 44, pp. 135-152, 2004.

[13] Y. Zheng, M. Lu, and Z. Feng, "Performance Evaluation of Adaptive AQM Algorithms in a Variable Bandwidth Network," IEICE TRANS. COMMUN., vol. E86-B, no. 6, pp. 2060-2067, 2003.

[14] R. Wang, M. Valla, M. Y. Sanadidi, and M. Gerla, "Using Adaptive Rate Estimation to Provide Enhanced and Robust Transport over Heterogeneous Networks," in Proceedings of the 10th IEEE International Conference on Network Protocols, Washington, DC, pp. 206-215, 2002.

[15] R. Jain, "The Art of Computer System performance Analysis," Wiley, New York, 1991.

[16] T. Čaršimamović, "Selection of parameters for the adaptive treatment of traffic in next generation networks (NGN)," PhD Thesis, University of Sarajevo, Sarajevo, 2010.

[17] T. Engel, E. Nikolouzou, A. Ricicato, "Analysis of Adaptive Resource Distribution Algorithms in the Framework of Dynamic DiffServ IP Network," Crete, 2001.

[18] U. Walter, M. Zitterbart, "Architecture of a Network Control Server for Autonomous and Effective Operation of Next Generation Network," Institut fur Telematik, Karlsruhe, 2006.

[19] Runtong Zhang, Yannis A. Phillis, Vassilis S. Kouikoglou, "Fuzzy Control of Queuing Systems," Springer, 2005.

 ISBN: 978-1-61208-133-5

Network Interface Grouping in the Linux Kernel

Vlad Dogaru, Octavian Purdilă, Nicolae Ţăpuş
Automatic Control and Computers Faculty
Politehnica University of Bucharest
Emails: {vlad.dogaru,tavi,nicolae.tapus}@cs.pub.ro

***Abstract*—The Linux kernel is a prime field for implementing experiments, many of which are related to networking. In this work we tackle a part of the network scalability issues of the kernel. More specifically, we devise a means of efficiently manipulating large numbers of network interfaces. Currently, the only approach to handling multiple interfaces is through repeated userspace calls, which add significant overhead. We propose the concept of network device groups, which are completely transparent to the majority of the networking subsystem. These serve for simple manipulation of large number of interfaces through a single userspace command, which greatly improves responsiveness. Changes are proposed both in kernel and userspace, using the iproute2 software package as support. Improvements in speed are visible, with changing parameters of thousands of interfaces taking less than a second, as opposed to over 10 seconds using a conventional approach. Implementation is simple, unintrusive and, most importantly, user-defined. This leaves room for future improvements which use the group infrastructure, some of which have already been proposed by third parties.**

***Index Terms*—Linux, kernel, network, device, scalability, grouping, iproute2**

I. Introduction

Computer networking poses interesting problems, both in the design of new protocols and techniques and in improving the scalability of existing concepts. It is particularly interesting to see what happens when the quantity of networking equipment, rather than the number of clients, grows. More specifically, we study the performance of the Linux kernel when dealing with thousands or tens of thousands of network interfaces. Because these cannot be physically fitted into a single machine, virtual interfaces are used.

Linux already offers a kernel module that emulates network interfaces. Before we can measure performance of these interfaces, a means of efficiently manipulating them is needed. The simple way to do this would be to repeatedly use an administrative tool for every interface. However, when dealing with a large number of interfaces, this proves inefficient. The creation of a new process for each of the interfaces offsets the useful work which is done. Moreover, even if the entire work could be done in a single process, communicating to the kernel that each interface is to be modified becomes a bottleneck. What is needed is an kernel API for specifying that a number of interfaces need to be modified in an identical manner. This includes activating or deactivating the interfaces, setting their MTU and other link-level parameters.

We propose a solution that introduces the concept of network interface group. A group is nothing but a simple integer tag; no relationship is implied between the devices in the same group. Their parameters can be modified either individually or collectively, thus no flexibility is lost. Because there is no intrinsic relationship between members of a group, group membership and policy is entirely defined by the administrator, not forced by the kernel. In many ways, the concept of an interface group is similar to that of a packet mark. It is used exclusively from userspace and its meaning is flexible.

By introducing network interface grouping, both of the factors slowing down interface manipulation are addressed. The administrator can use a single command to modify the parameters of a large number of interfaces; thus, scheduler overhead is eliminated. Further, userspace can address all the members of a group using a single request to the kernel, thus minimizing message transfers and system calls.

The solution described has been fully implemented using Linux and the iproute2 software package as support. The changes to existing code are unintrusive, with respect to both performance and code complexity. The code has undergone several rounds of review from the community and is, at the time of this writing, pending inclusion in the upstream kernel. This enables continued development in the area of network interface grouping and collective configuration. It also provides guaranteed maintenance from the community.

Performance tests feature a virtual machine setup. This enables harmless recovery from kernel errors and contained testing. Moreover, it speeds up the testing process because rebooting is low-cost. KVM was chosen for the task because, except for the live boot media, it demands no additional files or daemons. Additionally, KVM incurs little performance penalty on the host system (provided the host hardware supports the Intel VT-x or AMD-V extensions) in comparison to other virtual machines or emulators. Userspace utilities are provided by the busybox suite; this enables a full array of Linux utilities, with minimal dependencies.

Tests have been made featuring the `dummy` module of the Linux kernel. This module enables the user to add as many interfaces as they like by providing a parameter when inserting the module into the kernel. Because these interfaces have no physical meaning and are not critical to virtual machine operation, they are easily the target of tests involving batch modifications. We have tested two setups: one in which the user repeatedly calls the iproute2 `ip` command for each interface (using a shell script); the other involves using our proposed solution and changing the parameters of an entire group of interfaces using a single command. Test results are consistently orders of magnitude in favor of the latter

 ISBN: 978-1-61208-133-5

technique, particularly for larger numbers of interfaces.

II. Related Work

It is interesting to note that, while much research has gone into Linux network scalability, there is little interest in *configuration-time* optimization. Most progress is made towards *runtime* performance, be it packet handling [4] or routing performance [1].

III. Architecture

Network interface grouping targets the Linux operating system and, as such, must adhere to the kernel API. Because the group infrastructure proposed is simple, no additional data structures or complex algorithms were needed. Changes are entirely in existing source code files, so no modification was brought to the build system.

Although Linux supports loadable modules, changes were deemed sufficiently unintrusive not to warrant the modularity of the interface grouping routines. The only core data structure that was modified is `struct net_device`, which is by no means a performance-critical element. Thus, cache performance was not a problem and the necessary modifications could be made without resorting to memory profiling.

Kernel-side, the contributions are located in the `net_device` structure and in the API that the kernel provides to manipulate it. Historically, there have been two choices for modifying network parameters. The first interface is based on the `ioctl` system call. This involves creating a special descriptor and using successive `ioctl` calls to modify kernel parameters. It is used by the `ifconfig` utility, but both the API and the userspace components are currently deprecated.

The alternative is Netlink [3], a socket mechanism for interprocess communication. Netlink can serve as a means of communication between userspace and the kernel, as well as between two processes. Unlike other sockets, however, Netlink only works for a single host. Netlink provides multiple socket families, corresponding to different operating system parameters ranging from the neighboring system (ARP) to firewalling (Netfilter) or IPSec. The one of most interest to us is `NETLINK_ROUTE`, which is used to query and change routing and link information. This socket family is used by routing protocol implementations such as Quagga, as well as the `iproute2`, which we chose as support for our userspace modifications. Figure 1 shows how Netlink is used by `iproute2`.

Meant as a modern network configuration tool, the `iproute2` utility is developed in close relation to the Linux kernel. The development teams are closely related and patches are sent to the same mailing list for both the kernel networking subsystem and `iproute2`. As such, it was the natural choice for implementing userspace support for network interface groups. `iproute2` is fairly modular in architecture, with components for link-level, IP addressing, routing, tunneling and IPSec. Network group logic was added to the lowest-level component, informally named `iplink`.

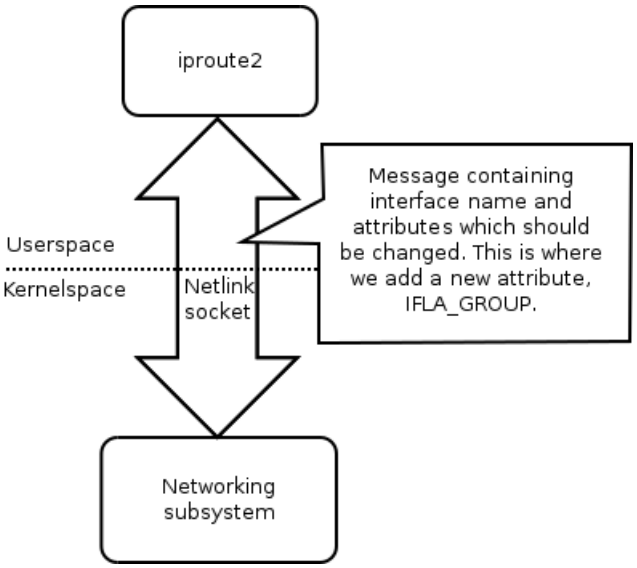


Figure 1. `iproute2`, a userspace tool, uses a Netlink socket to communicate with the kernel. We add new attributes to the message so that Netlink will be group-aware.

IV. Implementation

To have a properly functioning, albeit minimal, device group infrastructure, the following aspects need to be addresses:

1) A mechanism for grouping devices together.
2) A means of exporting group information to userspace, as well as a way to filter certain groups. The user can then view devices from a single group.
3) The possibility to change the group a device belongs to. We generally avoid referring to this action as *moving*, because no copying is required, as we shall see.
4) Finally, a way of specifying that a change affects an entire group and not just a single device.

The `net_device` structure has been modified to include group information in the form of an integer field. Thus, each network device belongs to a single group, there are no overlappings between groups. By default, all devices are created in group 0. The network device structure is not critical to performance (for instance, it is not used during routing), so modifying it does not raise cache efficiency problems. Nor is this structure part of intricate lists and other data, so it is fairly easy to understand the purpose of all its fields.

Group information is exported from kernel to userspace via route Netlink. The message format used in communication consists of a header followed by a variable number of attributes. Every time information is sent to the user, typically as a response to a `RTM_GETLINK` message, the group number is also packed in the message. Because of the flexible nature of Netlink messages, adding this information was trivial, in the form of a new attribute type. In turn, userspace unpacks the message information and interprets the attributes. The `list` operation in `iproute2` has been modified to accept a supplementary argument, the `devgroup` keyword, which

 ISBN: 978-1-61208-133-5

filters devices that belong to other groups. The user can now choose information they wish to see by filtering the relevant group.

The same Netlink attribute, called `IFLA_GROUP`, is used to set the group of a device. The sole difference is the direction of the message – user to kernel – and the type of message sent, typically `RTM_NEWLINK`. This type of message also contains the interface to operate on, specified either by the internal index number or the device name. Kernel-side modifications need concern with locking mechanisms, since the entire Netlink system is already protected by a lock. By ensuring that, kernel programmers have enabled future contributions that are less prone to bugs. Changing the group of a device means simply unpacking the message and assigning the requested value to the group field of the network device structure. No costly copying is involved.

Finally, the target of the whole infrastructure is being able to batch device parameter modifications. The aim is to be able, using a single Netlink message, to modify many network interfaces. Initially, another attribute was added, `IFLA_FILTERGROUP`. If this was specified in a user-originated message, the modifications described by the message were to be made on all the interfaces in a group, not just on a single one. However, the community expressed concern over adding two attributes for a relatively simple task, and a later iteration of the patch set uses a small hack and a single attribute. More precisely, if:

1) a userspace request has a negative interface identifier; normally, interface identifiers in Linux start at 1 and are all positive;
2) **and** it has no interface name specified;
3) **and** it contains a group attribute

then the modifications are to be made on the entire group, not just a single network device. By using the same attribute both for setting parameters and filtering devices, we eliminate the possibility of changing the group of an entire group of devices in a single request. But this was deemed an unnecessary corner case by the community in contrast to the API bloat it would introduce.

V. Experimental Setup

The initial tests we ran for the patch set have run into a tricky problem. We could not efficiently test our changes using many interfaces because of speed issues. More precisely, creating even only 1024 interfaces took around a minute. Because of the repetitive nature of testing our patch, we took the decision to test the setup without including `sysfs` in the kernel build. The `sysfs` filesystem is a modern way of configuring kernel parameters through a file-like interface, but, in our case, it was an important obstacle to performance. Without `sysfs`, new interface creation dropped from 55 seconds to 7 seconds, for 1024 interfaces. This is an important improvement for testing, not a real optimization.

Soon after, another performance obstacle had to be addressed, again not directly. VMWare, which is the de facto virtual machine in our environment, had a high overhead, not in terms of memory or processor consumption, but with respect to (re)booting and parameter modification. We migrated to the Kernel-Based Virtual Machine [2], which provides simple command-line manipulation, faster startup times, and generally behaves more like a simple program; for instance, when closing the KVM window, the virtual machine is stopped without further user queries. This might corrupt the machines disks, but we were not concerned about that.

Another aspect of the testing setup is the lack of a hard disk for the virtual machine. Because the kernel is the main part being tested and modified, storing it on a hard drive and copying it through the network to the virtual machine made little sense. Instead, we booted the machine from a live medium, generated by an in-house script. The script copies a few essential programs into an ISO image and bundles them with a kernel and an initial ramdisk. Aside from the iproute2 suite, scripts to create necessary entries in `/dev` and a minimal `passwd` file are provided. Manual creation of device files is necessary because the setup does not include Udev.

Basic tools are provided by Busybox [5]. The Busybox suite, dubbed *a swiss army knife for Linux*, is a collection of utilities designed to replace GNU coreutils, with a focus on small binary size and minimalism. Busybox consists of a single executable which encompasses the functionality of many utilities, from filesystem to process and text manipulation. It even contains a small version of `vi`. When launched, the Busybox executable looks at either its first parameter or the name it has at launch; this dictates its behaviour. So, launching `busybox ls` would trigger similar functionality to the `ls` command, while launching the Busybox executable using a link (symbolic or hard) named `vi` will have the effect of a `vi` clone. Busybox can be statically linked and it often is, but we chose not to do so. The live medium which we use to boot already contains the needed libraries. Finally, it should be noted that Busybox modularity is exemplary; it has a configuration menu in the spirit of the Linux kernel, from which the user can choose the modules which should be included. Configuration detail is surprisingly high – there are, for instance, options for more advanced `vi` features.

The final live image for the virtual machine is generated with `genisoimage`. The image is made bootable by including `isolinux`, a tool used for booting Linux from ISO images. The final image – containing a minimal kernel, an initial ramdisk, iproute2, Busybox, isolinux and the necessary libraries – amounts to little over 13 megabytes. Booting the virtual machine takes under 5 seconds, which is a boon to a generally error-prone testing cycle.

VI. Scenarios and Results

The sole method of testing the improvements is measuring the time needed to configure interfaces with and without using network device groups. We used an emerging technology called `perf` to profile the operations executed when bringing interfaces up, down, and when changing the Maximum Transfer Unit (MTU) of a device group. `perf` is a tool for manipulating the hardware performance counters of the

 ISBN: 978-1-61208-133-5

processors, similar in fashion to `oprofile`. The advantage of `perf` is that development is synchronized with the kernel, so communication between userspace and privileged mode is always up to date. Furthermore, user interface is much friendlier, with less complicated syntax and intuitive defaults.

Profiling has yielded expected results, but a bit disproportionate. It is clear that, by using device groups, modifying device parameters has a higher throughput, with proportionately less time spent in userspace and more in kernel mode. This is especially true because, when using device groups, we only create a single process; previously, a process was created for each interface. Even if the device manipulation was a straightforward action, it is important to note that creating a process, loading a new program into the address space, scheduling the process to run, and finally waiting for proper termination and cleanup are all very costly operations.

Indeed, operating system literature recognizes that performance critical applications should not create a large number of processes. This stems from the significant cost of a context switch. On systems which support virtual memory, which are virtually all systems nowadays, a specialized cache memory is employed in order to speed up virtual-to-physical address translation. The Translation Lookaside Buffer, or TLB, is an associative memory with pairs of virtual and physical addresses. When a context switch is triggered, the address spaces need to be changed, which in turn causes the flushing of all TLB entries.

Even factoring out userspace time spent, the total time spent in kernel mode is also significantly lower when using interface groups. Previously, a Netlink request was constructed for each interface which needed to be modified. This was an iterative process, constructing a message, then sending it, waiting for a reply, all this being done for each interface. Sending the request and receiving the response are translated into system calls, which are another recognized source of latency in the context of operating systems. Interface groups solve this problem by necessitating a single request for any number of interfaces, as long as they are in the same group. System call overhead is thus greatly reduced.

What is, in our view, remarkable about the proposed solution is that it actually scales better the more interfaces are involved. Without device grouping, the number of processes created and system calls is a multiple of the number of interface manipulated. Conversely, when employing interface grouping, a single process is created and a single Netlink request is made, regardless of the interface count. This leads to 50 times better performance with 1024 interfaces, but 65 times better with 2048 interfaces. We deem this an encouraging start into scalable network configuration.

The experiment results are shown in Table I. We timed the operation of changing the MTU of a number of interfaces, both traditionally (columns labeled *No group*) and using the new group interface (columns labeled *Group*). Because `sysfs` introduces a significant overhead, the experiments have been run on two different kernel configurations: one includes `sysfs` and the other does not.

	Without sysfs		With sysfs	
Interfaces	Group	No group	Group	No group
128	0.01	0.49	0.01	0.53
256	0.02	1.11	0.03	1.15
512	0.05	2.57	0.06	2.59
1024	0.17	7.02	0.20	7.51
2048	0.32	21.74	0.36	23.05

Table I
TIMING OF CHANGING INTERFACE MAXIMUM TRANSFER UNIT WITH AND WITHOUT INTERFACE GROUPING. ALL TIMES ARE IN SECONDS.

VII. CONCLUSION AND FURTHER WORK

The infrastructure implemented so far has proved to be scalable and generally accepted by the Linux kernel community. The patch set has already been through a feedback iteration, and is currently pending a second one. Valuable lessons in both design and communication with the kernel ecosystem have been learned in the process. Particularly, we have seen a surprising amount of suggestions for further improvement.

Although the patch, as it stands, is simple and generic, there have been suggestions to transform it into a hierarchical approach. The user would be able to define a tree-like structure, where a group can contain devices and other groups. Modifying a group would then have the semantic of modifying its devices and all the groups contained in it. A slight disadvantage of this would be little applicability, which is generally a sign of over-engineering. With respect to code, arborescent groups would require a separate group data structure, as opposed to a simple filed in the network device structure.

Another direction towards which the infrastructure can be developed concerns the handling of related devices, such as PPPoE (Point-to-Point Protocol over Ethernet) virtual devices and their supporting Ethernet device. It has been argued that group membership should be inherited, thus creating large groups without the need for explicit group changing. For instance, a PPP (Point-to-Point Protocol) server would add its relevant interface in group 42, then create hudreds or thousands of PPPoE interfaces having it as layer 2 support. These interfaces are implicitly created in group 42 and easily modifiable using a single command.

Because group membership is internally represented by a simple data type, an integer, there is no reason it should not be exported to userspace as such. The `sysfs` virtual filesystem is the current manner for the Linux kernel to expose information. A relatively simple addition would be the presence of the group tag in the corresponding `sysfs` directory of each device. An authorized user can then change the group of a device by using either `iproute2` or `sysfs`.

One particular critique of the grouping infrastructure was that it is not particularly user friendly, specifically that numbers have little meaning to users. Handling strings in kernel space is usually frowned upon when fixed-width tags can do better, so another solution is needed. Luckily, `iproute2` already has a convention in place. In the `/etc/iproute2` directory, the user can store associations between kernel numbers and

 ISBN: 978-1-61208-133-5

userspace significant names. We intend to add support for such an association. This remains within the initial self-imposed restriction of the implementation being simple in the kernel, but flexible in userspace.

Finally, the real scope of this endeavour is to improve the scalability of configurating network parameters. So far, we have treated part of the second level of the stack, but much work needs to be done for a fully functional interface. The most difficult problem we face is assigning relevant addresses to interfaces. That cannot be achieved with the current implementation, as identical MAC addresses would be next to useless for any setup. What needs to be implemented is incremental assigning of layer 2 and 3 addresses. One cannot efficiently do that in userspace, because it would bring about the problem of system call saturation discussed earlier. So the issue remains open, despite it being one of the key points of the original endeavour.

All in all, interface grouping is a simple, flexible interface that attempts to open the way to scalable network device configuration in the Linux kernel. Community feedback to it was positive, which means we can expect to see it in a future kernel release. Performance improvements are significant, even in the simple case of interface activating and deactivating. Improvement ideas abund, both in extending the existing structure and in improving the essential idea.

ACKNOWLEDGMENT

The authors would like to thank Jamal Hadi Salim for his initial implementation suggestion and constant feedback. Also, Răzvan Deaconescu and Laura Gheorghe provided valuable feedback for this article.

REFERENCES

[1] O. Hagsand, R. Olsson, and B. Gördén. Towards 10gb/s open-source routing. *Linux Kongress*, 2008.

[2] A. Kivity, Y. Kamay, D. Laor, U. Lublin, and A. Liguori. kvm: The linux virtual machine monitor. In *Proceedings of the Linux Symposium*, 2007.

[3] J. Salim, H. Khosravi, A. Kleen, and A. Kuznetsov. Linux Netlink as an IP Services Protocol. RFC 3549 (Informational), July 2003.

[4] J. H. Salim. When NAPI comes to town. Technical report, 2005.

[5] N. Wells. Busybox: a swiss army knife for Linux. *Linux Journal*, October 2000.

 ISBN: 978-1-61208-133-5

Unified Language for Network Security Policy Implementation

Dmitry Chernyavskiy
Information Security Faculty
National Research Nuclear University MEPhI
Moscow, Russia
milnat2004@yahoo.co.uk

Natalia Miloslavskaya
Information Security Faculty
National Research Nuclear University MEPhI
Moscow, Russia
NGMiloslavskaya@mephi.ru

***Abstract*—The problem of diversity of the languages on network security appliances' interfaces is discussed. Idea of the unified language for network security policy (ULNSP) implementation is proposed. A basic approach to the ULNSP formalization is considered. ULNSP Grammar and syntax examples are given. Further research on UNLSP is briefly discussed.**

Keywords-Network Security Policy, Network Security Appliances, Formal Language, Syntax, Grammar, Translator

I. INTRODUCTION

Communication in the modern world cannot be imagined without such a concept as a language – an effective way of representation and information transfer. At present there is a large variety of different languages created by people for their own needs: a sign language, a body language, a languages of mathematical and chemistry formulas, graphics languages applied in plotting and designer activity and others. All formats of input/output data define a language for information and communication technologies. Very often program systems have really complex languages for their interfaces, including declarations, instructions, expressions and many other kinds of data sets. This is the payment for programs and devices functionality.

The main reason of this diversity of possible languages is a tendency to represent the information in a form as much as possible short and convenient for a particular task solution. Information security (IS) sphere has not avoided this issue. Input and output languages are used for management of the majority of network security appliances (NSA), having their own set of instructions for implementation of the corporate IS policy (ISP), containing many rules (ISPR). Hereafter, by different NSA, we mean such appliances, for which at least one ISPR can be specified so that its implementation in one appliance is available only with the set of commands different from those used for implementation of this rule in another appliance. While protecting systems by different manufacturers, their command languages can differ so significantly that at the first glance it can be seen that there is nothing common among them. Obviously, the diversity of input languages on interfaces of NSA creates problems for IS managers because it is necessary to "translate" their ISPR for each particular device into a specific set of instructions. Such "translation" takes much more time than in case if all NSA could detect the identical commands. Moreover, it can result in errors in ISPR implementation configuring devices, especially if some rules are ambiguously formulated.

There are some solutions on the market that try to make a universal interface for ISP implementation. One of them is Check Point SmartDashboard [1], but this software product supports only Check Point security devices. Cisco Security Manager software [2] also uses policy-based approach to management of routers, firewalls and IPS systems, but, obviously, it supports only Cisco products. There are a lot of different solutions designed in order to improve an existing interface of a particular network security solution and make it more convenient, for instance, Activeworx IDS Policy Manager [3] is an application that provides GUI for policy-based management of Snort IDS. While network security models play an important role in any system, most research effort related to this topic are based on limited concept and do not discuss all the richness of current and emerging NSA. A development of the unified language for network security policy (ULNSP) implementation may solve these problems. The language will allow implementation of network security policy rules (NSPR) in a convenient form without being dependent on a particular device. Obviously, the language itself doesn't have any practical application; therefore its translator should be developed as well. Inherently, ULNSP together with its translator will form a universal interface between the human and NSA and, as a result, will increase the efficiency of the network security management.

Section II of the paper presents requirements and the main idea of UNLSP. Basic approach for formalization of the language is described in Section III. Section IV provides examples of NSPR defined with ULNSP. Practical application of the language is considered in Section V. Conclusions are given in Section VI.

II. UNIFIED LANGUAGE FOR NETWORK SECURITY POLICY BASIS

For the effective implementation of any policy, its rules must meet the requirements of maximum simplicity, clarity and ability for updating. Any policy rule should be formulated so that it could not be interpreted ambiguously. Implementation of these requirements should reduce the

 ISBN: 978-1-61208-133-5

probability that any rule will be ignored or implemented incorrectly, as well as the probability of bypassing the rules.

NSPR are specified in the corporate ISP in a natural human language and are subsequently implemented as a configuration (settings) for NSA. So the problem arises while translating NSPR into NSA commands by human (system, network or IS administrator). But different NSA recognize different languages, making it necessary to translate the same NSPR in a different set of commands. Such kind of translation may lead to typographical errors; so, that the device cannot accept this command. In the worst case, if a wrong command is syntactically correct, it can be applied by device, and this can cause incorrect functioning of the network or appearance of security breaches.

Proposed ULNSP would help to avoid such problems and solve the problem of NSPR portability and consequently improve network security management efficiency. In fact NSPR defined by ULNSP language will be the intermediate between ISPR and commands of a particular NSA (Fig. 1). In this case, network administrator (or other person responsible for the network security management) has to translate the rules from the corporate ISP into the rules in ULNSP, but this operation for a specific ISP must be done only once without concentrating on particular NSA models.

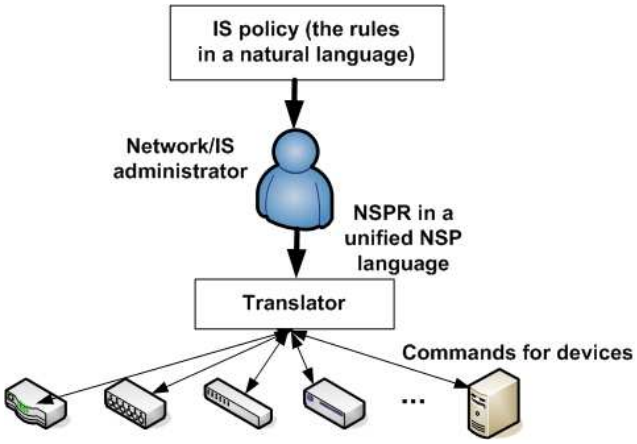


Figure 1. NSP Implementation using ULNSP.

Another important requirement is ULNSP extensibility – if we need to add a new type of devices or rules that the language should support, this process would be as simple as possible, and the changes wouldn't change the overall language structure.

III. ULNSP FORMALIZATION

For ULNSP formalization we use generative grammar G, which defines the rules for constructing sentences of the language $G = (T, N, S, R)$ [4], where T - a finite nonempty set - the terminal vocabulary (its elements are called terminal symbols TS); N - a finite nonempty set - non-terminal vocabulary (its elements are called non-terminal symbols); S - selected item of non-terminal vocabulary, so-called start symbol or axiom of the grammar; R - nonempty finite set of rules (productions), each of which has the form $\alpha \rightarrow \beta$, where α and β - chains on the dictionary $T \cup N$. In addition it is compulsory that $T \cap N = \varnothing$.

The chain β is directly derivable from the chain α in the grammar G (designated by $\alpha \Rightarrow_G \beta$), if the chain α can be represented as a concatenation of the three chains $\beta = \mu\xi\nu$ (some of them may be empty), chain β can be also represented as a concatenation of three chains $\beta = \mu\xi\nu$ and grammar G contains the production $\tau \rightarrow \xi$. Symbol $\Rightarrow_G$ denotes the binary relation on the set of all chains over the union of the vocabularies $T \cup N$. The chain β is directly derivable from the chain α in the grammar G (designated by $\alpha \Rightarrow_G^* \beta$), if in grammar G exists a finite set of strings $\pi_0, \pi_1, \ldots, \pi_n$, n>0 such that $\alpha = \pi_0$, $\pi_n = \beta$ for all $i = 1,\ldots, n$ holds $\pi_{i-1} \Rightarrow \pi_i$. Symbol $\Rightarrow_G^*$ denotes the reflexive transitive closure of $\Rightarrow_G$.

Formal language generated by G is a set of chains composed of TS of the grammar and the vocabulary derived from the grammar's start symbol:

$$L(G) = \{\alpha \in T^*: S \Rightarrow_G^* \alpha\}.$$

It is important to note that, in general, one language can be generated by different grammars.

Here an example of ULNSP grammar. The non-terminal N and the terminal T vocabularies should be defined:

N = {policy rule, identifiers, actions, functions, params, separators, permit action, deny action, traffic filtration, address translation, routing, interface, data link layer, network layer, transport layer, protocol ethernet, protocol IP, protocol ICMP, protocol TCP, protocol UDP, Ethernet params, IP params, ICMP params, TCP params, UDP params, address translation params, routing params, interface params, IPv4 addresses, IPv4 mask, Destination MAC, Source MAC, Type, MAC address, Version, IHL, Type of Service, Total Length, Identification, IP Flags, Fragment Offset, Time to Live, Protocol, Checksum, Source Address, Destination Address, Options, Source Port , Destination Port, Sequence Number, Acknowledgment Number, Data Offset, Reserved, TCP Flags, Window, Checksum, Urgent Pointer, Options, Length, Code, internal name, external name, local address, global address, interface ID, destination address, gateway, interface name, interface address, security level}

T={0,1,2,3,4,5,6,7,8,9,a,b,c,d,e,f,g,h,i,j,k,l,m,n,o,p,q,r,s,t,u,v, w,x,y,z,A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,W,X,Y, Z-,.,_,(,),},*

or in another words T consists of characters with ASCII codes from 33 to 126. In N a *policy rule* is an axiom of G.

Fig. 2 shows the scheme of ULNSP rules construction. This diagram is not a formalized method of representation of a grammar, but we use this figure in order to make an understanding of the language easier. Set of productions R can be found in Appendix.

 ISBN: 978-1-61208-133-5

The grammar of ULNSP is context-free. This grammar type is widespread in computer science and there are a lot of relatively efficient parsing algorithms that are applicable to detect chains of languages generated by context-free grammars.

IV. ULNSP SYNTAX AND SEMANTICS EXAMPLES

To block all traffic coming to the router's interface eth0 from the host with the address 192.168.1.1 to the host with the address 10.1.1.1 the rule in ULNSP will be as follows:

eth0 deny IP (192.168.1.1 10.1.1.1), or
eth0 deny IP 192.168.1.1 10.1.1.1

The following rule allows the transfer of TCP packets from the network 192.168.1.0/24 to the 80th port of the host with the address 10.1.1.2 on eth1 interface of the firewall:

eth1 permit IP (192.168.1.0/24 10.1.1.2) TCP (* 80), or
eth1 IP (192.168.1.0/24 10.1.1.2) TCP * 80

The following rule blocks all ICMP traffic via eth1:

eth1 deny ICMP

The following rule blocks TCP packets with FIN, URG and PSH flags to the host with the address 10.1.1.3 through the interface eth0 on NSA:

eth0 deny IP (* 10.1.1.3) TCP (* * * * * * FIN URG PSH * * *)

The rule for network address translation of the addresses 192.168.1.0/24 on the interface eth0 to addresses 10.1.1.0/24 on eth1 looks like that:

NAT (eth0 eth1 192.168.1.0/24 90.1.1.0/24), or
NAT eth0 eth1 192.168.1.0/24 90.1.1.0/24

For a static address translation 192.168.1.2 in the address 10.1.1.2 you can use the following rules:

NAT eth0 eth1 192.168.1.2 90.1.1.2

Translating of the address range 192.168.1.1 - 192.168.1.10 into the address 10.1.1.1 can be expressed by the following rule:

NAT eth0 eth1 192.168.1.1-192.168.1.10 10.1.1.1

As can be seen from the above examples of usage, ULNS has a convenient syntax and intuitively obvious semantics.

V. NLNSP TRANSLATOR

The main goal of NLNSP development is to create a universal interface between a human and NSA. It is obvious that the language itself has no practical applicability, because all ISPR, described in it, in fact, are a formal expression of rules defined in any natural language. In order to make the language applicable to the real devices a translation system is needed. The functions of this system include establishing a connection (telnet, ssh and so on) to the device; identifying the device, its version and functionality; translating NSPR from ULNSP into commands of a particular device; performing additional device configuration (if necessary).

Thus, in the ideal case, this system allows to use the same unified set of NSPR for all NSA with the required functionality making all required settings to adjust them in accordance with NSP. In reality it is almost impossible to cover all the existing diversity of devices of different manufacturers, types and versions, so the main challenge in this case is to support as many systems as possible.

The final result of the development is a software product, which allows to configure NSA by implementing NSP described in ULNSP. A security administrator will be able to set up a device in accordance with NSP, without being dependent on manufacturer of this device.

Let us consider the main convenience of the system. When designing a network security system for a corporation, the requirements come from its ISP. In fact it doesn't matter what kind of device is applied in a particular part of a network. It is important that this solution should have required functionality. Because the basic ULNSP concept is a function-based approach for formalization of the rules, it won't be a problem to formulate NSPR using the language. For configuring any NSA in accordance with the policy it is just necessary to connect to this device using the proposed system and input ISPR in the unified language. All necessary settings will be done automatically by the system.

VI. CONCLUSION

The main peculiarities of created ULNSP are context-free grammar and its extensibility. The function-based approach in rules formalization used allows to add new types of rules easily. Inherently, any policy rule formalized in ULNSP describes security function policy (defined in ISO/IEC 15408 [5]). The language also provides convenient syntax and intuitively obvious semantics. For example, all parameters for TCP traffic filtration rules follow RFC 793 [6].

ULNSP translator will help to automatically configure NSA in accordance with the corporate ISP by implementing the corresponding rules described with the language. For this purpose it is just necessary to know IP-address of the device and ISPR in ULNSP.

Today, ULNSP and its translator support a basic set of NSPR for firewalls and routers. The future challenge for the development is an extension of the set of rules that could be formalized with ULNSP such as rules for IDS/IPS systems, DLP systems, VPN rules and so on.

REFERENCES

[1] http://www.checkpoint.com/products/smartcenter/smartcenter_manag manag.html (last access date 16/03/2011)

[2] http://www.cisco.com/en/US/products/ps6498/index.html (last access date 16/03/2011)

[3] http://activeworx.org/programs/idspm/index.htm (last access date 16/03/2011)

[4] Karpov Y.G. Fundamentals of translators design. - St.Petersburg.: BHV, 2005 (In Russian).

[5] ISO/IEC 15408 Information technology - Security techniques - Evaluation criteria for IT security.

[6] RFC793 - Transmission Control Protocol.

 ISBN: 978-1-61208-133-5

APPENDIX

Set of productions R:

1) policy rule→ identifiers, separators, actions, separators, functions
2) identifiers→ α|ε
3) separators→ « »
4) actions→ permit action | deny action
5) permit action→ ***permit****|ε*
6) deny action → ***deny***
7) functions→ address translation | routing | interface | traffic filtration
8) address translation n→ ***NAT****,* ***(****, address translation params,* ***)***
address translation → ***NAT****, separators, address translation params*
9) address translation params → internal name, separators, external name, separators, local address, separators, global address
10) internal name→ α; external name→ α
11) local address→ IPv4 address, IPv4 mask
local address→ IPv4 address, -, IPv4 address
12) global address→ IPv4 address, IPv4 mask
global address→ IPv4 address, -, IPv4 address
13) IPv4 address→ $\beta_1.\beta_2.\beta_3.\beta_4$
14) Pv4 mask→ /ξ|ε
15) routing→ ***Route****,* ***(****, routing params,* ***)***
routing → Route, separators, routing params
16) routing params→ interface ID, separators, destination address, separators, gateway
17) interface identifier→ α
18) destination address→ IPv4 address, IPv4 mask
19) gateway→ IPv4 address
20) interface→ ***Interface****,* ***(****, interface params,* ***)***
interface → ***Interface****, separators, interface params*
21) interface params → interface ID, separators, interface name, separators, security level, separators, interface address
*22) interface name→ α|********
*23) security level→ τ|********
*24) interface address→ Адреса IPv4, Маска IPv4|**
25) traffic filtration→ data link layer | network layer | transport layer |
data link layer, separators, network layer |
data link layer, separators, transport layer |
network layer, separators, transport layer |
data link layer, separators, network layer, separators, transport layer
26) data link layer→ protocol Ethernet
27) network layer → protocol IP | protocol ICMP
28) transport layer→ protocol TCP | protocol UDP
29) protocol Ethernet→ ***Ethernet****,* ***(****,Etherne paramst,* ***)***
protocol Ethernet→ ***Ethernet****, separators, Ethernet params*
protocol Ethernet→ ***Ethernet***
30) Ethernet params→ Destination MAC, separators, Source MAC, Type
*31) Destination MAC→ MAC address|**
*32) Source MAC→ MAC address|********
33) MAC address→ $\sigma_1\sigma_2$**:**$\sigma_3\sigma_4$**:**$\sigma_5\sigma_6$**:**$\sigma_7\sigma_8$**:**$\sigma_9\sigma_{10}$**:**$\sigma_{11}\sigma_{12}$
34) Type→ ***0x****$\sigma_1\sigma_2|\beta_1|$********$|\varepsilon$
35) protocol IP→ ***IP****,* ***(****, IP params,* ***)***
protocol IP→ ***IP****, separators, IP params*
protocol IP→ ***IP***
36) IP params → Version, separators, IHL, separators, Type of Service, separators, Total Length, separators, Identification, separators, IP Flags, separators, Fragment Offset, separators, Time to Live, separators, Protocol, separators, Checksum, separators, Source Address, separators, Destination Address, separators, Options
IP params → Source Address, separators, Destination Address
37) Version→ ***0x****$\sigma_1|\varphi|$********
38) IHL→ ***0x****$\sigma_1|\varphi|$**
39) Type of Service→ ***0x****$\sigma_1\sigma_2|\beta_1|$********
40) Total Length→ ***0x****$\sigma_1\sigma_2\sigma_3\sigma_4|\eta|$********
41) Identification→ ***0x****$\sigma_1\sigma_2\sigma_3\sigma_4|\eta|$********
*42) IP Flags→ θ|********
*43) Fragment Offset→ η|********
*44) Time to Live→****0x****$\sigma_1\sigma_2|\beta_1|$********
*45) Protocol→****0x****$\sigma_1\sigma_2|\beta_1|$********
*46) Checksum→****0x****$\sigma_1\sigma_2\sigma_3\sigma_4|\eta|$**
*47) Source Address→ IPv4 address, IPv4 mask | IPv4 address, -, IPv4 adress|**
*48)Destination Address→ IPv4 address, IPv4 address|IPv4 address,-, IPv4 address|**
49) Options→ ***0x****$\sigma_1\sigma_2\sigma_3\sigma_4\sigma_5\sigma_6\sigma_7\sigma_8$, Options$|\varepsilon|$********
50) protocol TCP→ ***TCP****,* ***(****,TCP params,* ***)***
protocol TCP→ ***TCP****, separators, TCP params*
protocol TCP→ ***TCP***
51) TCP params → Source Port , separators, Destination Port, separators, Sequence Number, separators, Acknowledgment Number, separators, Data Offset, separators, Reserved, separators, TCP Flags, separators, Window, separators, Checksum, separators, Urgent Pointer, separators, Options
TCP params → Source Port , separators, Destination Port
52) Source Port→ $\eta|>\eta|<\eta|\eta_1$**-**$\eta_2|$********
53) Destination Port→ $\eta|>\eta|<\eta|\eta_1$**-**$\eta_2|$********
54) Sequence Number→ ***0x****$\sigma_1\sigma_2\sigma_3\sigma_4\sigma_5\sigma_6\sigma_7\sigma_8|\nu|$********
55) Acknowledgment Number→ ***0x****$\sigma_1\sigma_2\sigma_3\sigma_4\sigma_5\sigma_6\sigma_7\sigma_8|\nu|$**
56) Data Offset→ ***0x****$\sigma_1|\varphi|$********
*57) Reserved→ χ|********
*58) TCP Flags→¥| χ|********
59) Window→ ***0x****$\sigma_1\sigma_2\sigma_3\sigma_4|\eta|$********
60) Urgent Pointer→ ***0x****$\sigma_1\sigma_2\sigma_3\sigma_4|\eta|$********
61) protocol UDP→ ***UDP****,* ***(****, UDP params,* ***)***
protocol UDP→ ***UDP****, separators, UDP params*

 ISBN: 978-1-61208-133-5

protocol UDP→ ***UDP***
62) UDP params → Source Port, separators, Destination Port, separators, Length, separators, Checksum
UDP params → Source Port, separators, Destination Port
63) Length→ $\mathbf{0x}\sigma_1\sigma_2\sigma_3\sigma_4|\eta|^*$
64) protocol ICMP→ ***ICMP****, (, ICMP params,)*
protocol ICMP→ ***ICMP****, separators, ICMP params*
protocol ICMP→ ***ICMP***
65) ICMP params → Type separators, Code, separators, Checksum
66) Code→ $\mathbf{0x}\sigma_1\sigma_2|\beta_1|^*$
where α – a finite chain of TS on the dictionary *{0,1,2,3,4,5,6,7,8,9,a,b,c,d,e,f,g,h,i,j,k,l,m,n,o,p,q,r,s,t,u,v, w,x,y,z,-,_} ⊂T*;

βi – string of TS belonging to the set *{0,1,2,3,4,5,6,7,8,9, 10,11,12,…,253,254,255} ⊂T*;
ξ - a chain of TS belonging to *{1,2, …, 32}⊂T*;
τ - a chain of TS belonging to the set *{0,1,2, …, 100}⊂T*;
σ_i – a terminal symbol belonging to the set *{0,1,2, …, 9, A, B, C, D, E, F}⊂T;*
φ - a chain of TS belonging to *{1,2, …, 15}⊂T*;
η - a chain of TS belonging to *{1,2, …, 65535}⊂T*;
θ - a chain of TS belonging to the set *{1,2, …,7}⊂T*;
γ - a chain of TS belonging to the set *{0,1,2, …,31}⊂T*;
ν - a chain of TS belonging to the set *{0,1,2, …, 4294967295}⊂T;*
χ - a chain of TS belonging to the set {0, 1, 2, … ,63}*⊂T;*
¥ - a sequence consisting of space-separated distinct elements of *{URG, ACK, PSH, RST, SYN, FIN}.*

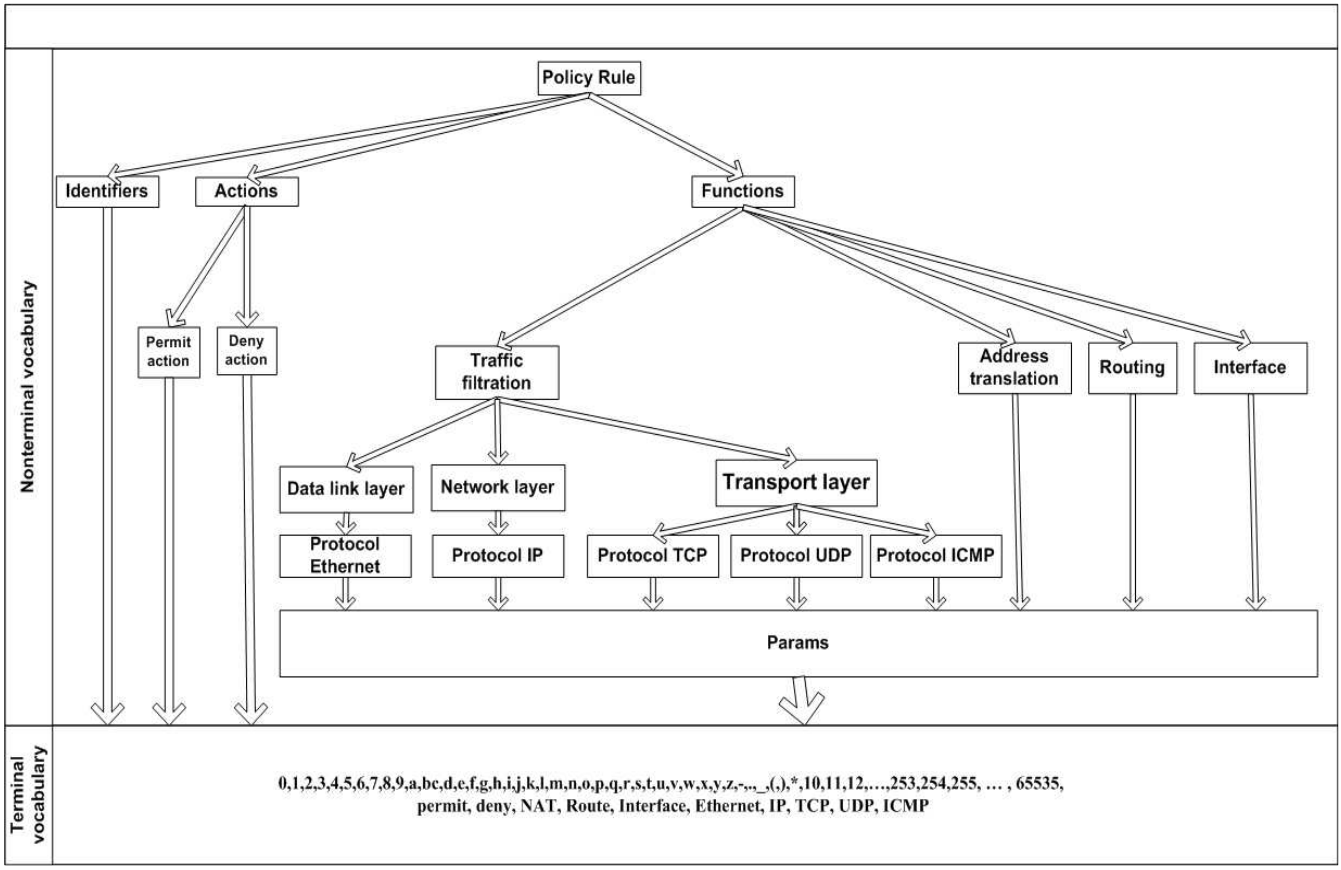


Figure 2. ULNSP rules construction.

Network Security Threats and Cloud Infrastructure Services Monitoring

Murat Mukhtarov
Information Security Faculty
National Research Nuclear
University MEPhI
Moscow, Russia
milnat2004@yahoo.co.uk

Natalia Miloslavskaya
Information Security Faculty
National Research Nuclear
University MEPhI
Moscow, Russia
NGMiloslavskaya@mephi.ru

Alexander Tolstoy
Information Security Faculty
National Research Nuclear
University MEPhI
Moscow, Russia
AITolstoj@mephi.ru

***Abstract*—Today Cloud Computing and virtual infrastructure are one of the most popular ways to deploy application hosting and web-farm platforms. Cloud Infrastructure services also known as "Infrastructure as a Service" (IaaS) are the way to deliver computer infrastructure, typically virtual environment as a service. Distributed nature of IaaS and likelihood that different customers can use the same server and network deliver new security threats. Security of open source platforms of Cloud Services is discussed. Threats that impact on availability components of platform and customer separation features are shown. The distributed way of network security monitoring of availability and integrity of IaaS is described.**

Keywords-Cloud computing, Infrastructure as a Service, Virtual Infrastructure, Application Hosting, Network Security

I. INTRODUCTION

Infrastructure as a Service (IaaS) is the next-generation way to provide customers with IT resources on demand principle. Customers can buy as much "Infrastructure" as they need, i.e. "pay per use" axiom. This is a way to reduce operational expenses on IT and shift some of risks to outsourcing companies. Such type of service is very convenient for small-business and medium-size companies to get access for the novel IT technologies and collaboration services, but there are some security threats which occur in the cloud. The first main threat may happen when some customer's virtual private servers (VPS) use the same shared hardware and network devices with others customer's VPS simultaneously. In this case configuration errors may sometimes occur, hence some unauthorized access accidents may happen. Up to 31% data breaches in Australia involved third parties such as Cloud Computing (CC) IaaS providers [1].

The second one is the availability issue: business critical data and applications are stored in one place (as we say "all eggs are put in a same basket"). Large-Scale botnets are able to deliver DDoS attack to the biggest ISP and Hosting providers (Such as Bitbucket, Amazon EC2), so there are lots of the related risks: failure of the hardware, hypervisor software, guest software, network channels, etc. as a result of successful DDoS attack or system-wide failure [2].

One of the ways of Cloud networks monitoring is to use network telemetry principal with such protocols as Cisco Netflow [3] or IPFIX [4]. Design and architecture of the cloud provide opportunity to use Netflow/IPFIX probes on the hypervisor without performance reduction for the sake of the kernel-acceleration technologies (such as PF-RING in Linux Kernel). Another way to monitor connections inside IaaS cloud infrastructure is introduced in the paper. IPFIX protocol is very similar to Cisco Netflow v9, but it is not proprietary, open-standard and has some improvements [4], which can be used on open source systems such as Linux or BSD-derivate systems (FreeBSD, OpenBSD). IPFIX is flexible, lightweight way for basic network security monitoring such as connection control and volume-based traffic estimation [5].

II. CLOUD SERVICE INFRASTRUCTURE TYPICAL ARCHITECTURE AND THREATS

IaaS expands CC services from web hosting and application hosting to end-user services (e.g. virtual desktop workplace). Supporting such a service becomes possible for the sake of several novel technologies and new license agreements which are provided by some software vendors such as Citrix and Microsoft. On the other hand development of open source desktop systems (KDE, GNOME, XFCE, etc.), designed to run popular Linux distributions (Ubuntu, OpenSuse, Debian, Redhat), makes possible to use such systems as desktop environment on desktop virtualization applications. Open source platforms of Cloud Services like Amazon and Bitbuket consist of Hypervisor system, as usual it is Xen-based or Kernel Virtual Machine (KVM)-based hypervisors, storage component based on Linux Volume Manage (LVM) and OpenISCSI – IP Storage Network (IP SAN), external Internet channels and intercommunication network. Each component has its own security threats that should be monitored and controlled. We focus on threats which impact on availability components of platform and customer separation features. Cloud Service provides rather more services than traditional datacenters but there are also rather more surfaces of attack, such as data separation issue, shared storage and availability of platform in common. Therefore securing such a platform is more difficult task than securing perimeter-based traditional datacenter and the problem of monitoring of IaaS platforms is very complex. Data storage, storage network and interconnection network are shared between all customers of IaaS, also external

 ISBN: 978-1-61208-133-5

network channels are common for all (Fig. 1). So attacker needs to compromise one of the components of IaaS platform, which are shared between customers to impact on the IaaS service in general. That is why it is important to use network security monitoring methods, which are to detect such impacts on transport network and shared network recourses in time.

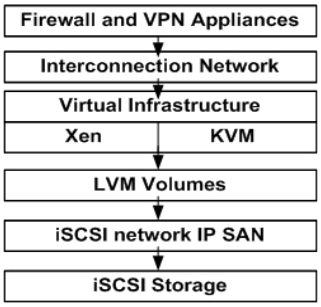


Figure 1. Architecture of open source software based IaaS platform.

A. External network channels

External network channels of nearly all datacenters including ISP's (such as Amazon EC) are vulnerable for the DDoS attacks, because attackers use large-scale bot networks. Network channels become point of failure as well for datacenter of Cloud infrastructure in general, as for individual customer, because each customer's network channel has finite bandwidth.

The second type of availability threat seems to be more difficult to detect and it requires distributed way of network security monitoring. Since such type of attacks is traffic volume based, the best way of lightweight monitoring of such type of attacks is using network "flow" protocols such as Cisco Netflow or IPFIX. Traffic streams from external network channels through access servers, usually going in VLAN, which is mapped to each customer, so the probe should be set on the enter point to the customers VLAN, for example on Broadband Remote Access Server (BRAS) or per Hypervisor. The second method is better for use since huge volume of flow data can impact on BRAS performance, but on the other hand using probes on each hypervisor machine can spread total load between virtual infrastructure servers.

Usually, external channels ISP's use traffic scrubbers (Cisco Guard, solutions like Cisco-Arbor Cleaning Pipes, etc.) for protection. They have capabilities allowing them to distinguish between "good" and "bad" traffic. They mitigate DDoS attacks by forwarding only good traffic and dropping attack traffic [6]. Before going to clean bad traffic from good one, a scrubber has to identify bad traffic. Cisco and Arbor use for that purpose several techniques, but all of them are based on Netflow v5/v9 analysis opposite to direct traffic intercept. So it is possible to use best practices and principles of commercial solutions with open source IaaS platforms. There are lots of open source implementations of flow-based traffic collectors (ipcad, flowtools, ntop, nprobe, ndsad, flowd, Vermont, etc.), which could be successfully used for network security monitoring purpose in Virtual Cloud Infrastructure (VCI). Their advantage is ability to install them on open source hypervisor platforms (Linux-based Xen and KVM), opaque for customer's software and without performance reduction.

B. Shared storage network

Shared storage network is a "point of failure" of whole IaaS infrastructure, also some iSCSI and volume mounting misconfiguration may impact on data separation between each customer and as a result some confidential data loss may occur. Usually open source Virtual Cloud is built on IP SAN (Storage Area Network) networks, because traditional FC SAN networks are rather expensive and it is not reasonable to use them in couple with open source software-based VCI. IP SAN network is based on iSCSI (Internal Small Computer Interface) protocol. iSCSI is an IP protocol that is a storage networking standard for linking data storage facilities. It is designed to carry out SCSI commands over IP networks, hence it could facilitate data transfers over local and external networks. Unlike traditional FC SAN, which requires special-purpose cabling, iSCSI can be run over long distance using existing network infrastructure. But using iSCSI is associated with several security threats: unauthorized accessing iSCSI Logical Unit Number that makes it possible to mount iSCSI running storage devices; authentication bypassing using some of attacks on CHAP protocol that is used to authenticate iSCSI peers; bypassing logical network isolation through VLAN misconfigurations or VLAN hopping attacks.

Taking that into account it can be concluded that customers cannot be sure that their sensitive data inside IaaS Cloud is safe. To improve data storage security, IaaS provider should monitor this threat by using some mechanisms, based on internal Linux/Unix system logging, such as syslog and mount table control scripts, and controlling VLAN separation Flow-based network measurements.

C. Shared internal network devices

Shared network devices also become one more point that needs to be controlled. Their main security risks are VLAN policy misconfiguration issues and VLAN hopping issues. As a result the separation between customers may be breached. Thus some customers may be able to have unauthorized access to essential data, stored on network resources on Virtual Service Infrastructure, Data Bases, Internal Web Portals and so on.

Another type of those threats is manipulation with Layer 2 functions of the switches, like an ARP poisoning, CAM table overflow etc. The result of such manipulations maybe unauthorized traffic interception and some sensitive data may be stolen. To avoid those risks some Layer 2 securing techniques such as "port-security", DHCP Option 82, port

 ISBN: 978-1-61208-133-5

authorization with 802.1x, virtual LAN with 802.1q are usually used. But sometimes configuration errors occur. For example there are several typical misconfigurations: native VLAN usage that equals 1; using 802.1q ports for customer link with native VLAN configured; allowing connections to one customer to VLAN's of others; 802.1x VLAN mapping errors – as a result of authorization process customer able to access prohibited VLANs.

The greater the size of the Virtual Infrastructure is, the more the likelihood of misconfigurations will be. Thus, the main tasks on network security monitoring of Virtual Infrastructure are to detect and to notify about separation failures. To control integrity of separation policy it is also convenient to use one of the flow-based monitoring protocols such as Netflow or IPFIX, but they should support "VLAN-ID" field in the flow template.

III. MONITORING NETWORK SECURITY AND POLICY INTEGRITY IN VIRTUAL SERVICE INFRASTRUCTURE

IaaS services's complex and tenant nature oblige service providers to use complex way of monitoring network security of their clients. In addition to traditional IDS, which have perfect present experience of known signatures' detection, the service provider must be able to detect availability threats such as DDoS attacks and anomaly network traffic flows, which may occur as a result of misconfiguration. In this view, it is very important to keep separation between customers' VPS and virtual networks.

There are several technologies, used in virtual infrastructure networks: separation of customers in own VLAN (802.1q VLAN) and isolating customers' services inside virtual appliance, controlled by hypervisor. Some of network vendors also support transport network technologies such as MPLS/VPLS network, MAC-in-MAC technology providing another separation methods for private networks. But such services are adapted to be opaque to an end customer. There are two main security threats - cloud availability (robustness against DDoS attacks) and shared network devices and hardware controlling. So we propose to monitor and detect such threats at an early stage, using IPFIX or Netflow v9 protocols, which are very similar.

A. Flow-based measurement

Netflow v9 or IPFIX provides useful information for security analysis such as IPv4/IPv6 headers, source IP, destination IP, source port, destination port, TCP flags, TOS, QOS, volume of traffic per flow, direction of the flow, interface, AS number and some additional ISP specific information: VLAN number, MAC address, MPLS labels. There are lots of techniques and software of flow analysis, based on analyzing Cisco Netflow v5/v9 data, namely ntop, nfsens, nprobe, flowd and some commercial products, e.g. Cisco MARS. However, it is not reasonable to use commercial implementations of Netflow collectors and security tools on open source cloud platforms.

One of the main IPFIX/Netflow v9 protocol advantages is its bidirectional flow (or bitflow), allowing tracking full connection opposite to Netflow v5. Trivial examples of biflow applications include initial round trip time (RTT) estimation, detection of connection establishment or other transactions for the purposes of an incident detection and response, and the separation of unanswered traffic for scan detection purposes [5].

Bidirectional flow measurement is very useful for a network security application, since it provides information about full connection that makes it possible to analyze each stage of the connection establishment for TCP protocol and track client responses for UDP protocol. For example, it is very useful to monitor and track HTTP and DNS connections and detect deviations in those connections, like scans or Flood attacks. In contrast to usage of unidirectional flow it provides information initiation and end of connection that enables to monitor and control integrity of this first initial dialog establishment success.

Bidirectional flow principle also reduces traffic that generates netflow/ipfix probe in a way as shown in Fig. 2.

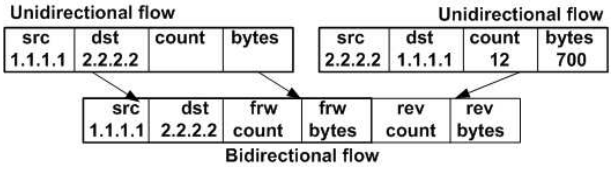


Figure 2. Unidirectional flow and Bidirectional flow.

Thus it is reasonable to use flow-based measurement for VCI monitoring problem.

B. Flow probes placement in Virtual Infrastructure

Flow-based measurement protocols are very convenient for classification and traffic volume analysis. Fig. 3 shows that netflow/ipfix probes can be placed in VCI network.

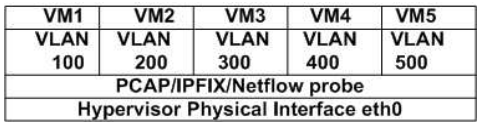


Figure 3. PCAP/Flow probe can be set on physical interface of Hypervisor machine.

Thus by using open source software IaaS providers are able to apply powerful tools to monitor network security (nfsen, ntop, flow-tools, Vermont, etc). So it is possible to use libpcap library compatible Netflow collector with virtual network interface card such as "tap" or "tun" Linux interfaces. Here it is an example with fprobe and nfdump on each virtual interface:

Linux# fprobe –itap0 –fip nfdump_host:9000
Linux# fprobe –itap1 –fip nfdump_host:9001
Linux# fprobe –itap2 –fip nfdump_host:9002

or on main physical interface:

Linux# fprobe –ieth0 –fip nfdump_host:9996

 ISBN: 978-1-61208-133-5

There are a lot of network devices vendors which support Cisco Netflow v9 or IPFIX protocols. It is possible to analyze flow-data that contains VLAN-ID field on the following Cisco network switches: Catalyst 4000/4500 and 6000/6500, but additional Netflow module is a prerequisite. On the other hand lots of vendors support IPFIX/Netflow v9 flow export out of box, such as Nortel, Extreme Networks, Juniper, etc. So it is not very difficult to check separation policy integrity with Netflow v9/IPFIX enabled on such a switching device. To make Cisco Router exports VLAN-ID field within flow-data typing IOS, cli command is needed:

Router(config)# ip flow-capture vlan-id

It is possible to export VLAN-ID field within IPFIX/Netflow v9 data on Linux host to use nProbe collector:

Linux# nprobe –n nfdumphost:9996 –i eth0 –T " %SRC_VLAN, %DST_VLAN, %IPV4_SRC_ADDR, %IPV4_DST_ADDR, %IN_SRC_MAC, %OUT_DST_MAC"

This is lightweight and chip way to monitor virtual interfaces inside Linux-based Cloud systems, which can be implemented in the current network architecture. On the other hand Netflow v9/IPFIX enables to monitor VLAN ID in traffic flows, which allows network administrator to control integrity of separation between IaaS provider's customers. VLAN ID monitoring using flow-based protocols makes it possible to detect and inform a security officer about network separation misconfigurations in time.

To provide excess coverage VLAN information travelling network it is important to use flow probes on a Hypervisor host as well as on network equipment. Each Hypervisor host has its own Flow probe that exports data to a collector, where VLAN information should be analyzed and compliance control should be performed.

It makes it possible to have information about whole VLANs in one place. It is no sense weather trunk interface or access VLAN interface using on Hypervisor host.

C. Flow analysis methods and tools

There are lots of statistical methods of volume-based raw traffic analysis, based on classification, abnormal behavior, baseline methods, detection of anomalies and deviations [7]. Most of them can be used to analyze Netflow/IPFIX data. Basically Netflow analyzing process is reduced to find one of several data sets: Top N and Baseline; Top N Session; Top N data; Pattern matching: port matching, IP address matching. TopN principle allows finding a source of activity that cause anomaly, worm attack, flood attack and it is based on volume deviations estimation. One of the lightweight flexible ways to implement IPFIX/Netflow v9 flow-data analyzer with its own analysis algorithm is to use Perl Flow.pm library [8].

It is better to use accomplished solution that could be built by means of combing several open source software. Open source tools such as nTop and nfsen provide functionalities to set threshold values of some traffic types. They provide information about volume (e.g. http, dns, Mircosoft-RPC traffic, etc). Increase of one traffic type in time can be easily monitored without drastic impact on performance of network equipment, virtual appliance or hypervisors software. Open source nfdump utility can be used for TopN analysis. There are several internal implementations of TopN with "-s statistics" option:

Linux@root# nfdump -M /netflow/directory -R file1:fileX –s srcip/dstport/pps/packets/bytes 'dst port 80' –O bytes

Obviously those output entries, which exceed regular values, may signify some network traffic inconsistency or network attack. Arguments of nfdump tool shown above enable it to detect DDoS attack against Web server. Centralized data management of flow-probes and IDS, like SNORT project, can be implemented using open source session-based network data correlation engine Prism++ [8].

In order to detect VLAN separation flow-data should be analyzed. It is possible to keep table of mapping customer's subnets and VLAN-ID's. Each incoming Flow should be aggregated by VLAN-ID field. Then it is possible to detect separation breach by means of comparing each aggregated flow with VLAN-ID – Subnet mapping table. If unauthorized network subnet in the given VLAN-ID is detected, comparator notifies about separation issue.

The described scheme of IPFIX/Netflow v9 data analysis provides opportunities for lightweight and efficient detection of network security issues, related to multicustomer VCI Servicesdiscussed above.

D. Impact on hypervisors perfomance

Flow collection is rather lightweight technique of network security monitoring. It achieves good performance results for several reasons: no need to intercept whole traffic traveling across the network and no need to analyze whole network packet – only headers information.

Flow analysis provides a network administrator or a network security officer with traffic volume-based quantitative evaluation.

Also Netflow sensor, implemented in Cisco routers and firewalls, also does not cause major impact on performance. For example, Cisco Systems provides following performance evaluation for 65000 flows Netflow v9 and 8903 packets per second :

Cisco 7200 Platform with NPE G1 CPU utilization	9 %
Cisco 7200 Platform with NPE G2 CPU utilization	8 %
Cisco 3845 Router	9 %
Cisco 2811 Router	53 %

Figure 4. Cisco Routers CPU utilization for 65000 Netflow v9 flows [10]

Here is an approach of evaluation performance impact on Hypervisor running 3 virtual machines with following initial data - 1 Virtual CPU, 256 RAM, 5Gb Virtual Device HDD, 100 mbp/s Virtual NIC, System Debian Lenny, also Apache is running.

Hypervisor configuration is one Intel DualCore E8400 Processor, with 2048mb RAM and 500Gb HDD without RAID.

 ISBN: 978-1-61208-133-5

For testing purpose we used file with size 1024mb, that was took from dd command:

Linux@root# dd if =/dev/zero of=/var/www/test_root/test.iso bs=1M count=1024

So we stressed Web server, trying to send GET requests to this file until Apache web-server forked enough childs (worker model) to take 80 % of CPU usage.

So we make comparison results with running and not running nProbe collector on Hypervisor system of CPU load Hypervisor System. Here are the tables for Hypervisor CPU Load without and with nProbe collector (fig. 5 and fig. 6 correspondently):

CPU Load	Hits per minute
22%	174 hits/minute
25%	243 hits/minute
34%	312 hits/minute
51%	362 hits/minute
74%	486 hits/minute

Figure 5. CPU Load of web server for hits per minute without nProbe running

CPU Load	Hits per minute
20%	171 hits/minute
26%	247 hits/minute
33%	311 hits/minute
52%	372 hits/minute
75%	492 hits/minute

Figure 6. CPU Load of web server for hits per minute with nProbe running

It seems that general impact on CPU is caused by Apache worker process. nProbe collector process in top –S output, always takes 0 % of CPU time.

To measure CPU load and Hits per minute we use Apache mod_status and net_snmp packages. For controlling we checked out CPU usage with top Unix-command and Nagios nrpe sensor.

Accuracy of results is not very high, we use rough estimates, but for evaluation performance of flow analysis that should be enough.

It is obvious that CPU usage impact will be noticeable only on huge amount of traffic – like thousands packets per second. Traffic rate is not very high in common web applications and low performance virtual platforms. Real CPU usage impact may occur only for flow analysis, performing on ISP equipment such as backbone routers.

IV. CONCLUSION

VCI services have several security issues and attack surfaces: customers use the same external network channels, shared network devices (separation is implemented via VLAN technologies), storage network and hardware. It is important to monitor and control availability of customers' virtual appliance and keep customers, separated in Virtual Infrastructure network. Flow-based measurement protocols such as Netflow v9/IPFIX are suggested to monitor separation of the customers, by means of controlling VLAN-ID in each flow and mapping it to the customer. Netflow v9/IPFIX flow-data analysis also provides opportunities for monitoring deviations of several types of traffic that may occur as a result of DDoS attacks or some network worms' activity inside or outside IaaS platform infrastructure. This way of monitoring network security of open source software, based VCI, is more productive and easy to implement in existing Virtual Clouds due to design and implementations of Netflow v9 and IPFIX protocols.

REFERENCES

[1] Tay L. and Kotadia M. Data breaches to cost more in the cloud http://www.securecomputing.net.au (2010). (23 March 2011)

[2] McNamara P. DDoS attack against Bitbucket darkens Amazon cloud http://www.networkworld.com (2009). (23 March 2011)

[3] Claise B. RFC 3954 Cisco Systems Netflow Services Export Version 9 (2004).

[4] Claise B. RFC 5101 Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of IP Traffic Flow Information (2008).

[5] Boschi E. and Trammell B. Bidirectional Flow Measurement, IPFIX, and Security Analysis pp. 8-10 (2006).

[6] Ramachadran V. and Nandi S. Bleeding Edge DDoS Mitigation Techniques for ISPs pp.8-9 (2004).

[7] Shanbhag S. and Tilman W. AnomBench: A Benchmark for Volume-Based Internet Anomaly Detection pp. 1-3 (2009).

[8] Kobayashi A. Net::Flow - decode and encode NetFlow/IPFIX datagrams http://search.cpan.org/~akoba/Net-Flow/ (2008). (24 March 2011)

[9] Dresslerand F. and Carle G. "HISTORY-HighSpeedNetworkMonitoring and Analysis," in 24th IEEE Conference on Computer Communications pp.2-4 (2005).

[9] Netflow Perfomance Analysis, Technical White Paper, Cisco Systems (2007).

 ISBN: 978-1-61208-133-5

Adaptive Scheduling Scheme for Multicast Service in Multiuser OFDM System

JooHyung Lee
Department of Electrical Engineering
KAIST
Daejeon, Republic of Korea
Joohyung08@kaist.ac.kr

Jong Min Lee
Data Transmission Tech. Development Team
SK Telecom
Seoul, Republic of Korea
jminlee@sk.com

Seong Gon Choi
Department of Electrical Engineering
Chungbuk National University (CBNU)
Cheongju, Korea
sgchoi@chungbuk.ac.kr

JunKyun Choi
Department of Electrical Engineering
KAIST
Daejeon, Republic of Korea
jkchoi@ee.kast.ac.kr

Abstract— **These Wireless communication systems have been developed to support users' various requirements. Multicast scheme is proposed for various types of service. Basically, the group Modulation and Coding Scheme (MCS) level for multicast transmission depends on the instantaneous worst channel user to provide reliable communication. However, this causes the low bandwidth efficiency for overall system. In order to overcome this problem, the proposed algorithm considers not only MCS efficiency of groups but also available overall system resources. The performance evaluation shows that proposed algorithm reduce the overall blocking probability and improve the throughput and revenue compared with traditional minimal and Proportional Fair (PF) based schemes.**

Keywords- Multicast, MCS efficiency, OFDM, Scheduling.

I. INTRODUCTION

Since wireless access technology and end user device, such as mobile, laptops, have been developed, user behavior is not restricted to using voice service by wireless device. As user who requests various multimedia broadcasting and streaming such as Internet Protocol Television (IPTV) increase, it is important to allocate resource efficiently [1]. Wireless multicast transmission can be a good solution to reduce the resource consumption for delivering the same contents to user who interested in certain group [2].

The major wireless multicast technologies used in various 3G/4G deployment models are Multicast Broadcast Service (MBS) [3] by WiMAX-The Worldwide Interoperability for Microwave Access, Multimedia Broadcast Multicast Service (MBMS) [4] by 3GPP, and Broadcast and Multicast Services (BCMCS) [5] by 3GPP2. These technologies commonly use Orthogonal Frequency Division Multiplexing (OFDM) technology with Adaptive Modulation Coding (AMC) in order to provide high bit rate and efficiently utilize the downlink bandwidth. By independently managing the each user, AMC can provide high bandwidth efficiency in unicast transmission. However in multicast transmission, it is not efficient since the multicast group MCS level is adjusted by only a user who has worst channel condition in a group [2]. Therefore, capacity saturation can be happened as the number of users increase because of depending on the instantaneous worst channel user in multicast transmission [6].

In order to cope with this problem, many researchers have proposed schemes especially considering throughput. Koh and Kim suggest the PF Scheduling for multicast service [7]. Kang and Cho suggest the dynamic packet scheduling for multicast [8]. Gopala and Gamal suggest the policy based scheduling for multicast [9]. Although these schemes can enhance system throughput by selecting maximal MCS level, it has low cell edge performance and causes high blocking probability. Therefore, there is no way to serve cell edge users and it will be a fatal problem if multicast is not provided to some static users. Xu Ning and Viver Guilame [10] concentrate to guarantee service of users in cell edge by handling PF parameter. In order to analyzing performance, we just focus on the PF scheduling algorithm since PF scheduling algorithm is one prominent example of compromise between fairness and high system throughput [7]. However, the proposed scheme is not restricted by PF algorithm. In this paper, we propose the adaptive scheduling based on MCS efficiency of groups and available overall system resources for multicast service. It improves not only cell edge performance but also increase the overall throughput. Finally, we compare the proposed scheme with the conventional scheme in wireless OFDM systems. We also analyze and compare the system performance of PF scheduling based multicast transmission scheme in terms of overall blocking probability, throughput and revenue. The rest of this paper is organized as follows. In Section II, proposed transmission scheme is described, and then In Section III, we develop the system model for analyzing blocking probability. The system performance between proposed scheme and conventional scheme are compared, and the system performance of the proposed scheme is evaluated in Section IV. Finally, conclusions are presented in Section V.

 ISBN: 978-1-61208-133-5

II. PROPOSED TRANSMISSION SCHEME

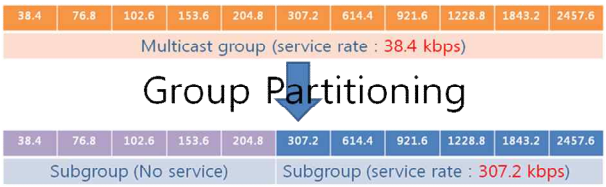


Figure 1. Multicast group partitioning (No service vs. service)

In this section, we describe a problem of the conventional transmission schemes for multicast and propose an adaptive scheduling scheme.

A. *Problem statement*

According to AMC, in unicast service, high spectral efficiency can be achieved by selecting the highest modulation and coding rate with a given acceptable Bit Error Rate (BER) constraint. However, in the multicast case the transmission rate must be the minimum value of a multicast group. This makes system throughput performance degrade extremely since the overall system capacity is limited by the worst channel user. One possible way to improve the system throughput is to split the multicast group into two subgroups and to serve the better channel subgroup only [6-9]. Fig. 1 shows the example how the partitioning method can improve the system throughput.

Although splitting the multicast group can enhance the throughput, the cell-edge users are sacrificed. Therefore, it is important how to increase the throughput with minimizing cell-edge blocking probability. In this respect, it is our contribution to propose the efficient adaptive scheduling scheme for the multicast user group with considering group MCS efficiency and available radio resources in a cell.

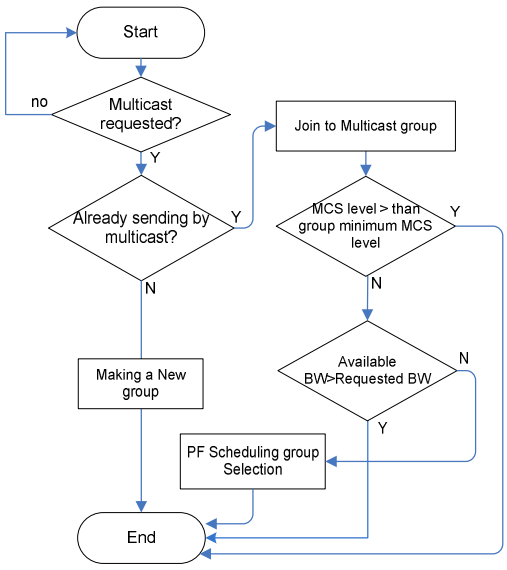


Figure 2. Proposed transmission scheme procedure

B. *Proposed adaptive scheduling scheme*

In this section, we address the proposed adaptive scheduling scheme. The proposed transmission scheme is involved with two cases : sparse phase and dense phase. The sparse phase means that the system has enough bandwidth to support the worst channel users in multicast groups. On the other hands, the dense phase, the system has not enough bandwidth to support the worst channel users because many groups are located in a cell. Fig. 2. expresses the overall procedure for proposed transmission scheme.

1) Select a transmission scheme

if available bandwidth size < requested bandwidth size

- dense mode is executed
- PF Scheduling group selection.

Else if

- Sparse mode is executed

2) Evaluate the group MCS level in dense mode

Measure the current Signal to Noise ratio (SNR) values of multicast users

- Measure the current average rate $R_k\ (t)$ of user k from SNR values at time frame t which updates as follows

$$R_k(t+1) = \frac{(T-1)\cdot R_k(t) + r_k^{\min}(t+1)}{T},\ if,\ k \in U_S$$

$$\frac{(T-1)\cdot R_k(t)}{T},\ elsewhere\ [7] \qquad (1)$$

It is known that a proportionally fair allocation should maximize the sum of logarithmic average user rates [2]. Therefore, PF scheduler maximizes sum of $R_k(t)$ by the property of PF allocation.

- Select and save a group MCS level for multicast group i: [7]

$$\therefore l^* = \arg\max_l \left\{ \prod_{k\in\{i|r_i(t)\geq r^l} \left(1 + \frac{r^l}{(T-1)\cdot R_k(t)} \right) \right\} \qquad (2)$$

From above index l*, we can extract MCS[i] as a group MCS level for multicast group i.

To reduce overall blocking probability in the system, PF scheduling group selection procedure is executed. BN is the number of blocking user in group i, it affects the overall blocking probability in the system. Finally, we consider BN value to choose PF scheduling group by following below algorithm..

for (i=1:the number of group(N))

for (k=1:multicast group users in group I (U_s))

 ISBN: 978-1-61208-133-5

$$if\ \ MCS[i][k] < MCS[i]$$
$$BN_i ++ \qquad (3)$$
end
end

$$G_{PF,i} = \min\{BN_1, BN_2, \ldots, BN_N\} \qquad (4)$$

Since we select PF scheduling group by adopting G_{PF}, we can execute PF scheduling algorithm to group $G_{PF,i}$ which has minimum number of blocking users.

3) Evaluate the group MCS level in sparse mode [6].

- Measure the current MCS values of multicast users
- Select and save the lowest MCS value among U_s multicast users:

$$\therefore C_M = \min\{MCS_1, MCS_2, \ldots, MCS_{Us}\} \qquad (5)$$

Therefore, in proposed adaptive scheduling scheme for multicast service, we concentrate to enhance multicast traffic efficiency.

III. ANALYSIS OF TRANSMISSION SCHEME

To analyze the proposed transmission scheme, we can model our proposal with $M^x/M/C/C$ for OFDM subcarrier allocation system [11][12]. From the viewpoint of analytical purpose, we may obtain statistical average number of used sub-channel in MCS level. Every subcarrier has the same average data rate. The number of sub-channel C generally denotes the system capacity in an NG cell. Because sub-channel is contained 28 subcarriers, the cell has in total $28CR_b$ rate resources, where R_b represents the average data rate per subcarrier. In our model, minimal data requests limited by sub-channel not subcarriers which depend on real service. Therefore, a multicast service (call) can request multiple sub-channels to fulfill its transmission requirement. Hence, this case is considered as a batch (group;bulk) arrival.

TABLE I. NOTATIONS FOR NUMERICAL ANALYSIS

Notations	Explanation
C	System capacity (The maximum number of sub-channel).
X_k	The number of requested sub-channel in PF.
x_k	The number of requested sub-channel in Minimum.
$\pi(k)$	State probability of state k.
Ω_{block}	Call Blocking probability for Adaptive PF.
$\Omega_{pf-block}$	Blocking probability when PF is used and fully blocked
$\Omega_{pf-non-block}$	Blocking probability when PF is used and not fully blocked.
γ	Throughput in multicast service.
μ	Average service rate of multicast stream.
P_{PF}	average blocking probability of PF algorithm

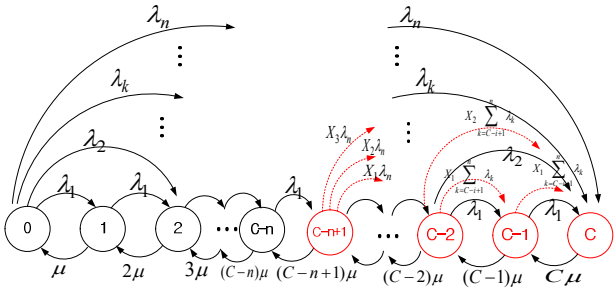


Figure 3. State-transition-rate diagram of $M^x/M/C/C$ for the OFDM sub-channel allocation system in adaptive PF algorithm

Assume the customers arrive in groups following a Poisson process with the mean group-arrival rate λ. The service times (call holding times) are independently exponentially distributed with the parameter μ. The system propability sequence $\{x_k\}$ and $\{X_k\}$ means the probability of requesting k sub-channel based on traditional minimal and PF scheduling based environment. Let λ_k denote the batch arrival rate where $\lambda_k = x_k\lambda$.

$$\sum_{k=1}^{n} x_k = 1, \ \ where, \ \ 1 \le k \le n \le C \qquad (6)$$

The model is equivalent to the standard Erlang loss system [11]. Fig. 3 depicts the state-transition-rate diagram of the model. Red circle means that there is possibility of PF algorithm due to lack of available bandwidth. Finally, red dotted line shows the PF transition rate when requested sub-channel is higher than available sub-channel in multicast environment. The equilibrium (steady-state) equations written below are run to obtain the steady-state probabilities of the model.

i) m=0;

$$\lambda\pi(0) = \mu\pi(1), \ \ where \ \ 1 \le n \le C \ \ and \ \ \lambda = \sum_{k=1}^{n} \lambda_k \qquad (7)$$

ii) $1 \le m \le C - n$

$$(m\mu + \sum_{k=1}^{n} \lambda_k)\pi(m) = \sum_{k=1}^{min(m,n)} \lambda_k \pi(m-k) + (m+1)\mu\pi(m+1) \qquad (8)$$

iii) $C - n + 1 \le m \le C$

$$(m\mu + \sum_{k=1}^{\min(n-1,C-m)} \lambda_k + \sum_{k=1}^{\min(n-1,C-m)} (X_k \sum_{k=C-m+1}^{n} \lambda_k))\pi(m)$$
$$= \sum_{k=1}^{\min(m,n)} \lambda_k \pi(m-k) + \sum_{k=1}^{m-(C-n+1)} \pi(m-k) X_k \sum_{j=C-(m-k)+1}^{n} \lambda_j + (m+1)\mu\pi(m+1) \qquad (9)$$

Reforming (1) and (2) yields

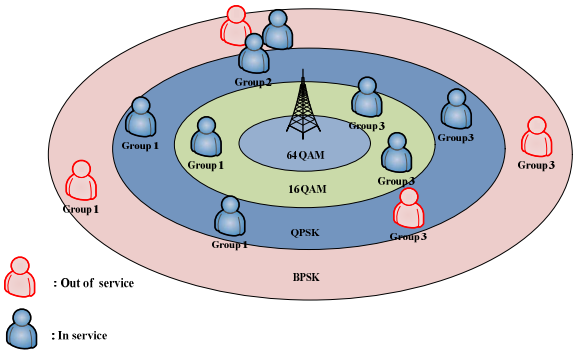


Figure 4. Mobile terminal distribution example

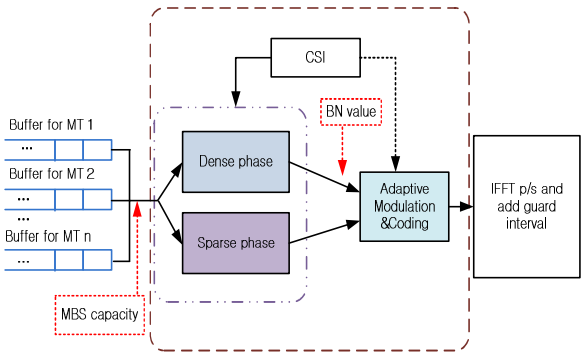


Figure 5. System model for adaptive scheduling scheme

$$\pi(1) = \pi(0)\lambda / \mu \tag{10}$$

$$\pi(m+1) = \frac{(m\mu + \sum_{k=1}^{n}\lambda_k)\pi(m) - \sum_{k=1}^{min(m,n)}\lambda_k \pi(m-k)}{(m+1)\mu} \tag{11}$$

Where, $1 \le m \le C-n$.

$$\pi(m+1) = [(m\mu + \sum_{k=1}^{\min(C-m,n)}\lambda_k + \sum_{k=1}^{\min(n-1,C-m)}(X_k \sum_{k=C-m+1}^{n}\lambda_k))\pi(m)$$

$$- \sum_{k=1}^{\min(m,n)}\lambda_k \pi(m-k) - \sum_{k=1}^{\min(m-(C-n+1),n)}\pi(m-k)X_k \sum_{j=C-(m-k)+1}^{n}\lambda_j]$$

$$/(m+1)\mu \tag{12}$$

Where, $C-n+1 \le m \le C$.

Recursive programs cannot always solve the equations, owing to overabundant recursive levels for large C. Therefore, an iterative procedure is adopted to solve the equilibrium equations. Let initial value P_0^*=1; then other steady state probability value can be extracted by global balance equation. According to the normalizing condition (summation of steady state probability equals one) the equilibrium probabilities of all states are written as follows:

$$\pi(m) = \pi(m)^* / \sum_{i=0}^{C}\pi(i)^*, where \ \ 0 \le m \le C. \tag{13}$$

The Call Blocking Probability (CBP) of the model is explained in the following. Basically, PF algorithm contains static blocking probability that means rates of blocking users who don't satisfy determined MCS level. This probability is expressed as P_{PF}. Sometimes, available bandwidth can't satisfy determined bandwidth which is extracted from PF algorithm, it is fully blocked. Finally, the CBP can contain two blocking cases : Fully block and Static block. Thus, the CBP can be expressed as

$$\Omega_{block} = \Omega_{pf-block} + \Omega_{pf-non-block} \tag{14}$$

$\Omega_{pf-block}$ expresses the blocking probability of fully block case. $\Omega_{pf-non-block}$ expresses the blocking probability of static block case. Equations for both cases contain $P_{min\text{-}out}$, because, when available bandwidth can't satisfy minimal scheme, PF scheduling will be conducted. Finally two cases can be differentiated by $P_{PF\text{-}out}$ and $P_{PF\text{-}nonout}$.

$$\Omega_{pf-block} = \sum_{m=C-n+2}^{C}\pi(m)\cdot P_{\min-out}\cdot P_{PF-out} \tag{15}$$

$$\Omega_{pf-non-block} = \sum_{m=C-n+1}^{C}\pi(m)\cdot P_{\min-out}\cdot P_{PF-nonout}\cdot P_{PF} \tag{16}$$

$$where, \ P_{PF-out} = 1 - \sum_{k=1}^{\min(n-1,C-m)}X_k, \ P_{\min-out} = \sum_{k=C-m+1}^{n}x_k$$

$$P_{PF-nonout} = \sum_{k=1}^{\min(n-1,C-m)}X_k \tag{17}$$

And, Throughput equation follows Erlang's Loss Formula.

Next, we introduce the service provider's reward/penalty cost model to expect service providers' revenue. We assumed that when the base station successfully serves the multicast service without blocking, the service provider receives a reward value of R. On the other hand, if a user is rejected, we assume that the service provider loses a value of L immediately [2].

In prior art under the resource allocation policy, for example, if the system on average services N client per unit time and reject M client per unit time, then the system revenue is

$$\sum N\cdot R - \sum M\cdot L \tag{18}$$

Finally, we define the total system revenue as follow:

$$revenue = \sum_{m=1}^{C} m\mu\pi(m)\times R - \lambda\Omega_{block}\times L \tag{19}$$

 ISBN: 978-1-61208-133-5

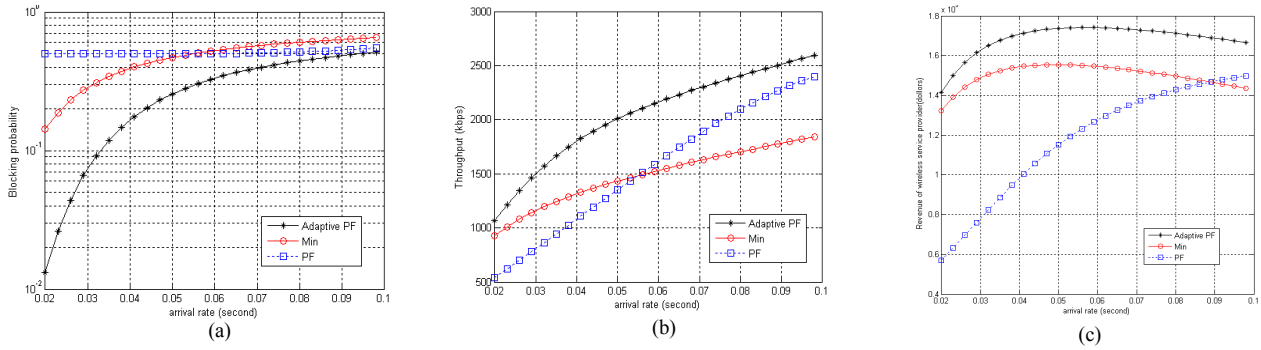


Figure 6. Results with performance comparison for (a) blocking probability (b) throughput (c) revenue between Adaptive PF, Min and PF

IV. PERFORMANCE EVALUATION

Fig 4 shows the mobile terminal (MT) distribution example. MTs are directly mapping the MCS level by considering path loss in large scale fading. In this case, each MT requests the multicast service from BS. After requesting the service, each packet goes to the BS. Fig 5 illustrates the adaptive scheduling scheme for a packet-switched OFDM system. We focus on downlink transmission of multicast data traffic. Therefore, base station (BS) that makes scheduling decisions for packet transmission based on MBS capacity. BS can choose two scheduling method based on MBS capacity adaptively. At Media Access Control (MAC) layer, upon each packet arrival, the BS puts the packet into its corresponding buffer which is assumed to have infinite space. At Physical (PHY) layer, we assume perfect channel state information (CSI). With this CSI, the BS can implement AMC to maximize the throughput on each subcarrier [13]. In performance evaluation, we measure distribution of MCS level in worst channel user in multicast transmission and PF scheduling based multicast transmission. In this case, we consider the arrival rate has a discrete uniform distribution and arrival multicast user has MCS level which is uniformly distributed from 1 to 10. We simulate uniform distributed user in cell to execute min based algorithm and PF based algorithm. Finally, we extract P_{PF} which is static blocking probability when PF algorithm used. Whole simulation procedure follows:

Step1. Each user has their MCS level depend on channel condition. We just consider path loss in large scale fading.

Step2. We randomly group users as multicast group.

Step3. Apply two scheduling algorithm to each multicast group in same environment – PF algorithm and Min algorithm.

Step4. Extract blocking probability and MCS level distribution.

Finally, we assumed that total channel capacity C is 40 and multicast streaming is 300kbps. Service rate is 0.0055 since we just focus on ucc contents environment which have average 3 minutes (180 seconds) running time [12]. We analysis performance by using various parameters in simulation and numerical analysis in terms of blocking probability, throughput and revenue.

A. *Call Blocking Probability by proposed adpative PF algorithm*

Fig. 6 (a) shows the performance comparison among the adaptive PF, min and PF in terms of blocking probability as arrival rate increase from 0.02 to 0.1 [12]. In PF algorithm, although arrival rate is small, it can make blocking situation, because it determines MCS level. As arrival rate increase, PF algorithm is better than min algorithm because min algorithm saturate faster than PF algorithm due to choose worst channel user. In adaptive PF, when arrival rate is low (unused sub-channels are enough to support requested sub-channel), it doesn't use PF algorithm. After arrival rate is high, it uses PF algorithm to enhance multicast sub-channel efficiently. Finally, overall blocking probability patterns show that adaptive scheduling enhances user blocking rates by adaptively choosing algorithm. It also affects cell edge performance since most of blocking user might be cell edge user in large scale fading environment.

B. *Throughput by proposed adaptive PF algorithm*

Fig. 6 (b) shows the performance comparison among the adaptive PF, min and PF in terms of throughput as arrival rate increase from 0.02 to 0.1 [12]. As result of blocking

TABLE II. SYSTEM PARAMETERS IN OFDMA ENVIRONMENT

MCS Level	Modulation	Coding rates	Maximum Data rate (Mbps)	# of used Sub-channel by streaming (300kpbs)
1	QPSK	1/2, 6x	0.75	11
2	QPSK	4x	1.13	8
3	QPSK	2x	2.26	4
4	QPSK	1x	4.51	2
5	QPSK	3/4	6.77	2
6	16QAM	1/2	9.02	1
7	16QAM	3/4	13.54	1
8	64QAM	2/3	18.05	1
9	64QAM	3/4	20.30	1
10	64QAM	5/6	22.56	1

 ISBN: 978-1-61208-133-5

probability, throughput shows the bandwidth utilization of each case. Since throughput is depend on blocking probability, adaptive PF ensures higher bandwidth utilization than min and PF algorithm based multicast. Conventional PF is smaller than min algorithm when arrival rate is low. As arrival rate is increased, conventional PF is better than min algorithm as shown in Fig. 5.

C. Revenue by proposed adaptive PF algorithm

Now we evaluate revenue of each scheme in the aspect of wireless service provider. In this case, we assume that wireless service provider provide IPTV service. Fig. 6 (c) shows the result obtained by service providers' reward/penalty cost model. We assumed that is the base station successfully serves the multicast service without blocking, the system receives a reward value of R(=$10). On the other hand, if a user is rejected, we assume that the service provider loses a value of

L(=$5) immediately. This figure shows that as the arrival rate increase, the revenue of each algorithm is slightly decreasing because of its blocking probability of services. In this case, our proposed algorithm can offer higher revenue than others. And PF can't compensate within most of our observe point since PF algorithm contains static blocking probability.

II. CONCLUSION

Multicast transmission makes efficient utilization of sub-channel in wireless environment. Although conventional PF enhance multicast channel utilization in hot-spot situation, it is not suitable for low arrival rate situation as our analysis. In OFDM environment, since multicast part in OFDM is dedicated, unused multicast sub-channel in certain time slot means inefficient resource allocation. To enhance efficiency of resource allocation, our proposed scheme has been suggested. Our analysis have shown that adaptive scheduling scheme adaptively allocate sub-channel to multicast users depending on MBS channel capacity which is same as available bandwidth. As result of comparison, our proposed scheme guarantees more serviced users in certain time slot and more efficient bandwidth utilization. Since most of blocking users are in the cell edge, we can also enhance cell edge performance. Further work will extend the proposed scheme with considering weight factor of the number of users in multicast group and apply general scheduling algorithm.

ACKNOWLEDGMENT

This work was supported by the IT R&D program of MKE/KEIT. [KI001822, Research on Ubiquitous Mobility Management Methods for Higher Service Availability]

REFERENCES

[1] Joohyung Lee, S.H. Shah Newaz, JongMin Lee, JunKyun Choi, "Context Profile Handling Mechanism for Source Specific Multicast in Proxy Mobile IP", icoin, Jan 2010.

[2] Jong Min Lee, Hyo-Jin Park, "Adaptive Hybrid Transmission Mechanism for On-Demand Mobile IPTV Over WiMAX", *IEEE Trans. on Broadcasting*, vol. 55, no. 2, pp. 468-477, June 2009.

[3] IEEE 802.16e "IEEE Standard for Local and Metropolitan Area Network, Part 16: Air Interface for Fixed Broadband Wireless Access System," Feb. 2006.

[4] Multimedia Broadcast/Multicast Service (MBMS); Architecture and Functional Description, 3GPP Std. TS 23.246, Rel-9 Dec. 2009.

[5] CDMA2000 High Rate Broadcast-Multicast Packet Data Air Interface Specification, 3GPP2 Std. C.S0054-A, v1.0, Mar. 2006.

[6] Seon Yeob Baek, Young-Jun Hong, "Adaptive Transmission Scheme for Mixed Multicast and Unicast Traffic in Cellular Systems", *IEEE Trans. on Vehicular Technology*, vol. 58, no. 6, pp. 2899-2909, July 2009.

[7] Chung Ha Koh, Young Yong Kim, "A proportional Fair Scheduling for Multicast Services in Wireless Cellular Networks", *IEEE Vehicular Technology Conference*, pp. 1-5, fall 2006.

[8] Kyungtae Kang, Jinsung Cho, Heongshik Shin, "Dynamic packet scheduling for cdma 2000 1XEV-DO broadcast and multicast services", *IEEE Wireless Communications and Networking Conference*, vol. 4, pp. 2393-2399, 2005.

[9] Parveen Kumar Gopala, Hesham EI Gamal, "On the Throughput-Delay Tradeoff in Cellular Multicast, *IEEE Wireless Networks, Communications and Mobile Computing*, vol. 2, pp. 4101-1406, 2005.

[10] Xu Ning, Viver Guilaume, Zhou Wen, Qiang Yongquan, "A Dynamic PF Scheduler to Improve the Cell Edge Performance", IEEE *Vehicular Technology Conference*, pp. 1-5, fall 2008.

[11] Jui-Chi Chen, Wen-Shyen E. Chen, "Call Blocking Probability and Bandwidth Utilization of OFDM Subcarrier Allocation in Next-Generation Wireless Networks", *IEEE Communications Letters*, vol. 10, no. 2, pp. 82-84, Feb. 2006 .

[12] Yan Zhang, Yang Xiao, Hsiao-Hwa Chen, "Queueing Analysis for OFDM Subcarrier Allocation in Broadband Wireless Multiservice Networks", *IEEE Trans. on Wireless communications*, vol. 7, no. 10, pp. 3951-3961, Oct 2008.

[13] Jinri Huang, Zhisheng Niu, "A Cross-layer Proportional Fair Scheduling Algorithm with Packet Length Constraint in Multiuser OFDM Networks", *IEEE Global Telecommunications Conference*, pp. 3489-3493, 2007.

 ISBN: 978-1-61208-133-5

Provisioning Service Differentiation for Virtualized Network Devices

Suk Kyu Lee, Hwangnam Kim, Jun-gyu Ahn, Kwang Jae Sung, and Jinwoo Park
School of Electrical Engineering
Korea University, Seoul, Republic of Korea
Email: {sklee25, hnkim, dubhe, kjsung80, jwpark}@korea.ac.kr

***Abstract*— In order to efficiently utilize the network bandwidth and flexibly enable one or more networks to be combined or subdivided into virtual networks, it is essential to virtualize network devices and then to provide service differentiation for the virtualized network devices. In this paper, we propose a virtualizing method for network devices based on the virtual machine and offers a differentiated scheduling scheme to satisfy QoS requirements that are imposed on virtualized devices. We have built the network virtualization framework combining the Virtual Box, time-slot-based time-sharing scheme, and leaky-bucket controller, and then we have conducted a performance evaluation study with real testbed. The empirical study indicates that the service differentiation for virtualized network devices is successfully supported by the proposed framework.**

Keywords **-** Network Virtualization, Scheduling Policy, Virtual Box, Virtual Machine

I. INTRODUCTION

There has been a large improvement in the field of virtualization in the past decade. As noted by Goldberg [7], the idea of the virtual machine emerged around 1970s, but, due to the lack of computing power, the field of virtualization has arisen in the early 2000s. The hardware virtualization allows many users and corporations to reduce the expenditure of buying multiple physical machines to support various applications since it runs those applications with multiple virtual machines in a physical machine. Additionally, the virtualization motivates us to provide an efficient way to run multiple networks, each combined with many networks and/or parts of networks into a virtual network or each isolated with a suite of applications in an independent execution environment with a pseudo network interface. The network virtualization gives the network service providers economic benefits since it decouples network infrastructure installment from network service deployment by running multiple virtual networks over a physical network. Also, the network virtualization benefits consumers with customized programmable network services by encapsulating one or more services into a single virtual machine and activating one or more of them according to customer's demand. One of key components for realizing the network virtualization is to isolate one set of network services from another and to control and manage their access to network resources according to QoS specifications. Therefore, the scheduler among virtualized network devices should be implemented with priority. The works such as the Xen [1] and VMWare [16] have mainly dealt with how to distribute the CPU usage fairly amongst the virtual machines (VMs). Moreover, the work such as the Denali [19] has been focused on scheduling I/O fairly amongst the VMs. MultiNet [3] has devised a framework that virtualizes the IEEE 802.11 wireless LAN card, and has proposed a fair scheduling algorithm among virtualized network interface cards. As we can see with the existing works (that are described in Section II), many of the virtualization techniques have been focused on the fairness among virtual machines' CPU and I/O.

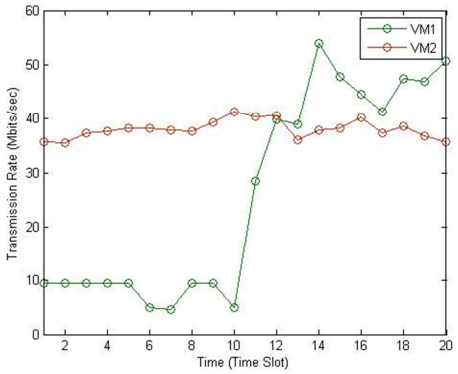


Figure 1. Comparison of network resource usage between two virtual machines without any scheduling scheme.

However, there has not been much research conducting on how to provide service differentiation for the network resources. Specifically, the network service provider or customer may want to allocate a different amount of network resources, e.g. network bandwidth, to each virtualized network device according to QoS specification. Thus, depending on the QoS specification, the network resource should be distributed differently to each VM. On the other hand, with current existing technology, if n VMs exist, then each VM should have $1/n$ rate of the work. However, such the fair allocation cannot be always guaranteed. A unfair resource allocation is presented in Figure 1, where the network bandwidth usage is compared when two virtual machines compete for the network device. We can observe from the figure that the result of current scheduling scheme for virtual machines is not effective in perspective of service differentiation over the virtualized network devices.

Based on this motivation, we propose an internal network virtualization framework to virtualize network devices, which is based on virtual machine, and also, we present a differentiated scheduling scheme to support service differentiation that is imposed on the virtualized devices. We implemented the service differentiation by juxtaposing the leaky-bucket controller [11] and the time slot-based resource allocator [12]. Conclusively, the major contributions are three-fold:

1. To provide service differentiation for internal network virtualization for the first time;
2. To build up the leaky bucket controller and time slot-based resource allocator for virtualized network device;
3. To carry out performance evaluation study in real testbed in terms of (*a*) network performance and (*b*) inter-packet delays to evaluate the service differentiation for

 ISBN: 978-1-61208-133-5

virtualized network devices.

To the best of knowledge, this is the first trial of provisioning the service differentiation for the virtualized network devices.

The rest of the paper is organized as follows. In Section II, we summarize related work in the area of scheduling methods for virtual machines. Then we describe both the specific design and the architecture for implementing the internal network virtualization framework for service differentiation in Section III. With a real testbed, we discuss about the performance and feasibility of the proposed service differentiation framework in Section IV. Finally, we conclude the paper with Section V.

II. PRELIMINARY

In this section, we briefly summarize previous work focused on scheduling methods in virtual machines, and then we explain the specific motivations.

A. Related Work

Many approaches are available to address the scheduling problem within the virtualization. However, there has not been any paper related with provisioning service differentiation for the virtualized network resources. We thus simply explain existing scheduling methods for virtualization according to two categories: I/O based and CPU based.

The CPU based fair scheduling approach focuses entirely on the virtual CPU in order to distribute the host machine's CPU fairly amongst the virtual machines. Govindan *et al.* [8] proposed to use credit-based scheduling algorithm to distribute the CPU resource fairly amongst the VMs by devising a credit scheduler to assign and monitor the credits for each VM. Gulati *et al.* [9] studied on how to proportionally schedule the virtual CPU amongst the VMs in order to improve the CPU fairness by using the Adaptive DRR. Scheduling I/O based fair scheduling approaches extended the mechanism on top of the credit scheduler [8] by adding the BOOST state on the credit scheduler. Instead of only using the Under and Over state, Ongaro *et al.* [14] implemented the BOOST state where it prioritizes the I/O scheduled VM. With this implementation, it provides better chance for I/O-bounded work to control the CPU of the host machine.

Note that all the previous approaches aim at encouraging fair scheduling for CPU or I/O based, extensively relied on the Xen [1] hypervisor. On the contrary, our proposed work aims at how to support service differentiation among multiple VMs without modifying the guest OS. By providing service differentiation method for virtualized network devices, we can dynamically control the network usage for each VM, based on the type of work or a given QoS specification.

B. Motivation: Limitation in Existing Scheduling Schemes

The scheduler in a virtual machine is responsible for assigning computing resources to each virtual machine. It usually exists at the virtual machine monitor (VMM), which is a software layer where it virtualizes many of the resources of the physical host machine. In order for the VMM to handle the task, the resources such as the CPU, network device, I/O devices, and physical memory need to be virtualized. Additionally, there are still many challenges in order to fairly schedule these devices. For example, Virtual Box [18]'s VM scheduling algorithm basically depends on the host machine's thread scheduling mechanism, where it gives the impression of distributing the resource fairly to the VM until the VM is dead. However, with some portion of accuracy, it is not quite true. On the other hand, the scheduling algorithms, such as Xen's Credit Scheduler [8] and Ongaro *et al.* [14], have been taken to distribute the resources fairly to the CPU and the I/O devices. These works focused on how to schedule the resource in order to distribute the resource into *n* number of VMs. The advantage of these algorithms is that it can almost distribute and share the physical resource of the host machine *almost equally* amongst the virtual machines. However, the question of how to provide a service differentiation according to a QoS specification is still unclear.

The main objective of this work is to implement a scheduling algorithm in order to realize service differentiation by coordinating the progress of multiple VMs according to a given QoS specification. For example, if VM1 has been assigned to work for 30% of CPU usage then it is mandatorily use 30% CPU usage as well as the other VM use the rest of the CPU usage, 70%.

III. DESIGN AND IMPLEMENTATION

In this section, we describe the internal network virtualization framework for virtualizing network devices, and the scheduling scheme of providing the service differentiation.

A. Architecture

In order to virtualize network devices, we chose a virtual machine (VM) approach based on the Virtual Box OSE 3.16 SDK[17]. The Virtual Box OSE is a open source software developed by Oracle. Each VM is considered as an EMT, Emulation thread, when the host machine schedules the threads. Unlike the Xen's Credit Scheduler [8], Virtual Box does not have a customary scheduler where it schedules effectively amongst the VMs. However, since the Virtual Box SDK 3.16 [17] provides interfaces to interact with the VMM and virtual device to control the VMs that run concurrently, we implement a scheduling scheme for implementing service differentiation. Figure 2 presents the architecture for the proposed network virtualization framework with the differentiated scheduling scheme.

B. Scheduling Scheme for Service Differentiation

In order to realize the service differentiation in scheduling scheme for the virtualized network devices, we chose two basic building blocks. The first one is the leaky bucket controller, which has been used in packet switched networks and the telecommunications networks in order to regulate the data transmission with the credit-generating rate [1] and the burstiness[11]. In our proposed scheduling scheme, the controller generates the credits according to the QoS specification. The other one is time-slot based resource allocator, with which we can decide the basic time allocation unit instead of using infinitesimal time unit. With these two schemes, we designed a scheduling scheme for providing

[1] We adopted the concept of credit to assign CPU resource to each VM.

 ISBN: 978-1-61208-133-5

service differentiation for virtualized network devices. The scheduler is presented in Figure 3.

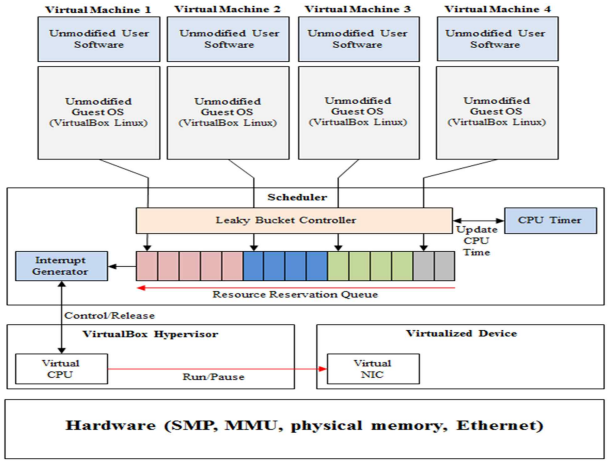


Figure 2.The architecture for the proposed network virtualization.

Once the system starts, the scheduler is periodically executed to produce credits via the leaky-bucket controller and assign the credits to each VM. For example, if there are two VMs and the ratio of CPU usage is given with 65:35 as the QoS specification for two VMs, VM1 should be assigned with use the resource usage of 65% and VM2 should acquire 35% of resource usage. Based on this ratio, the scheduler assigns the credits to each designated VM. Once every VM acquires its own credits, the scheduler assigns the time slot to it. In order to distribute the time slot to the VMs, we used the following equation:

$$VM_i TimeSlot = \frac{VM_i Credit}{\left(\frac{\sum_i^n VM_i Credit}{\#ofTimeSlot}\right)} \quad (1)$$

Each VM is basically inserted into a scheduling queue, and then it is scheduled in a round robin way. If the credits allocated to a VM are used up, then the VM should wait till the scheduler assigns additional credits to it. Otherwise, the VM runs during the time slot, and then, it is reinserted to the queue after the time slot is expired.

```
procedure scheduler()

assign workload (VMi) according to a QoS;
struct cpu_reservation_schedule q;

while(system is running) {
  compute all of the credits for all VMi;
  for i=0 to all VM
       compute the schedule for VMi with (1);
       insert VMi to q(ω);
  update time();
  for ω =0 to sizeof(q)
        if q (ω)=VMi.D
            run VMi for one time slot
            pause rest of the VM within the q(ω)
        else
            decrement credit of the rest of
                  the VM within the q(ω);
   update time();
}
```

Figure 3. The deterministic scheduling algorithm.

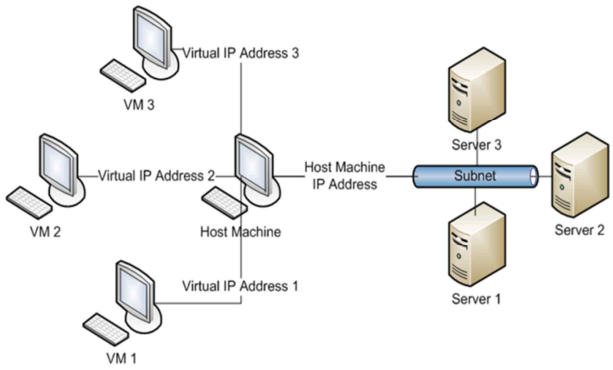


Figure 4. The testbed for performance evaluation study.

IV. PERFORMANCE EVALUATION

In this section, we present the results of performance evaluation, which has been done with some application level benchmarks in the *real testbed*, in order to investigate the performance of the proposed network virtualization framework for providing service differentiation for virtualized network devices

All the empirical experiment have been conducted on a 3.06GHz Intel Core2 Duo with 3MB of L2 cache, 4GB of RAM, and 10/100/1000BASE-T Gigabit Ethernet card. The operating system is Ubuntu 10.04 and the guest OS for the VMs are Ubuntu 9.10. As aforementioned, the Virtual Box OSE 3.16 is employed as the VMM. To generate the UDP and the TCP traffic, we employed the *iperf* [10] utility, which is supposed to constantly generate packets from the VM to the server, and we also used *ping* flood to generate ICMP traffic. The testbed for this study is in Figure 4, where three VMs are resident at one physical machine and each VM communicates with its corresponding real server over the network.

As for the schedulable resource, we used the CPU usage under the assumption that the time amount of using network devices is proportional to that amount of using CPU. As for the performance metrics, we use three metrics, the network bandwidth (transmission rate), inter-packet delay and CPU usage to verify if the proposed scheduling scheme can achieve the service differentiation according to the QoS specification. Note that we selectively present the empirical results in terms of network bandwidth and inter-packet delay due to the space limit. Finally, we have conducted two empirical evaluation study sets: the one is when we activated two virtual machines, and the other is when we used three virtual machines.

A. With Two Virtual Machines

Firstly, we conducted an empirical study with two virtual machines, and we employed UDP, TCP, ICMP traffic to verify if the differentiation is achieved.

In the case of UDP traffic: We have tested in two scenarios: the ratio of CPU usage between VM1 and VM2 is 50:50 (%), and the ratio is 60:40. Figure 5 shows the result of the first case, whereas Figure 6 shows the other case.

As for the results in Figure 5, the average bandwidth of each VM was very similar to each other and it is consistent over time. In specific, the average bandwidth of VM1 is 46.90 Mbits/sec whereas the average of VM2 is 46.95 Mbits/sec. As

 ISBN: 978-1-61208-133-5

for the results in Figure 6, we can observe that the average bandwidth of VM1 is 55.41 Mbits/sec and average bandwidth of VM2 is 38.29 Mbits/sec, which indicates that VM1 used about 60% of total network bandwidth whereas the VM2 used nearly 40% of the bandwidth.

Additionally, we investigate the impact of the differentiated scheduling on the inter-packet delay. Figure 7 presents the fluctuation of inter-packet delays that we observed from the original Virtual Box system (without any change). Specifically, the average delay is 0.303*ms* and its standard deviation is 0.0539*ms* for VM1, and those values for VM2 are 0.327*ms* and 0.0593*ms*, respectively. However, when we used the proposed differentiated scheduling scheme with ratio of 50:50, we could observe *stable and fair dynamics of inter-packet delays* which is presented in Figure 8. In particular, as for VM1, the average delay and standard deviation are 0.132*ms* and 0.0159*ms,* respectively, and, as for VM2, those values are 0.131*ms* and 0.0167*ms*.

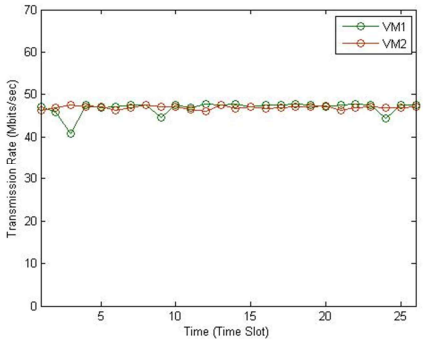


Figure 5. Comparison of network bandwidth when the ratio of using network bandwidth between VM1 and VM2 is 50: 50 and UDP traffic is employed.

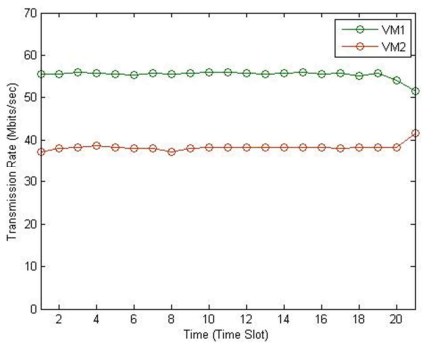


Figure 6. Comparison of network bandwidth when the ratio of using network bandwidth between VM1 and VM2 is 60: 40 and UDP traffic is employed.

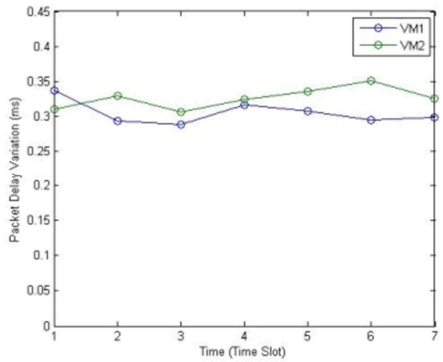


Figure 7. The fluctuation of inter-packet delays that are measured with the original Virtual Box when two VM are activated and UDP traffic is used.

In the case of TCP traffic: When we used TCP traffic, we made a similar observation. Figure 9 compares two network bandwidths when the ratio of network bandwidth usage between two VMs is 60:40. From the figure, we observed that the required differentiation is successfully achieved; specifically, the average of VM1 is 52.85 Mbits/sec and VM2 is 33.72 Mbits/sec.

***In the case of ping (ICMP) traffic*:** We have used the ping flood to verify if the proposed service differentiation is still effective in ICMP traffic. The ratio between VM1 and VM2 for the QoS specification is set to 50:50. Figure 10 compares two network throughputs each of which is used by VM1 and VM2, respectively. The reported average bandwidth of VM1 is 20.05 Mbits/sec, and that of VM2 is 20.26 Mbits/sec.

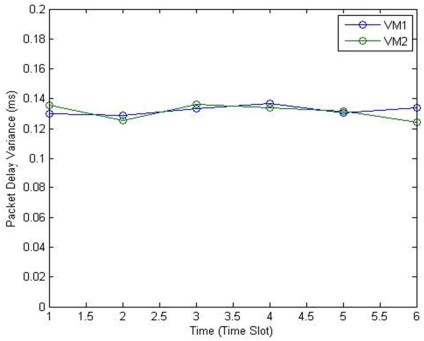


Figure 8. The stable fluctuation of inter-packet delays that are measured under the proposed differentiated scheduling scheme when the ratio of network bandwidth usage between VM1 and VM2 is 50:50 and UDP traffic is used.

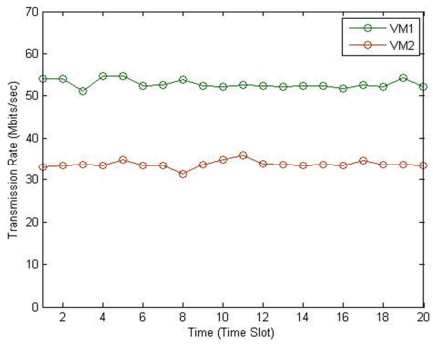


Figure 9. Comparison of network bandwidth when the ratio of using network resources between VM1 and VM2 is 60:40 and TCP traffic is employed.

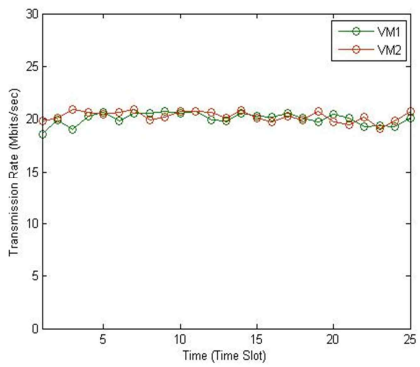


Figure 10. Comparison of network bandwidth when the ratio of using network resources between VM1 and VM2 is 50: 50 and ICMP traffic (generated by ping traffic) is employed.

 ISBN: 978-1-61208-133-5

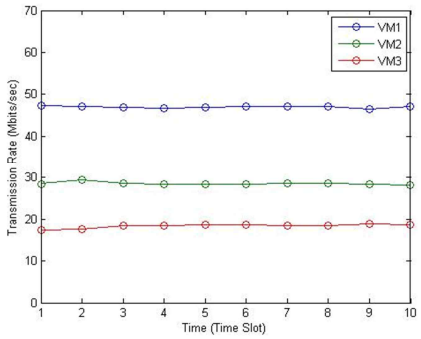


Figure 11. Comparison of three network bandwidths when the ratio among VM1, VM2 and VM3 is 50:30:20 and UDP traffic is employed.

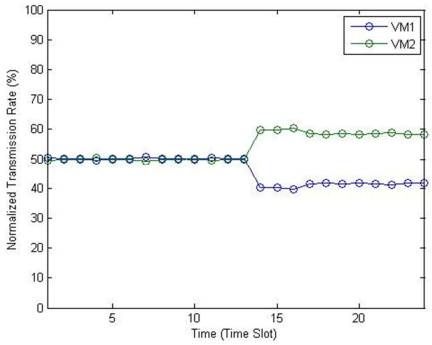


Figure 12. Comparison of two network bandwidth when the ratio between VM1 and VM2 is changed from 50:50 to 60:40 and TCP traffic is employed.

B. With Three Virtual Machines

As the second empirical study, we used three VMs in order to check whether or not the differentiated scheduling scheme is immune to the number of VMs. Figure 11 compares three network bandwidth usages when we use UDP traffic and the ratio for the VMs is set to 50:30:20. We made similar observations to previous empirical studies: the service differentiation is successfully achieved among VMs.

C. Dynamic Service Differentiation

As the last empirical study, we used a dynamic QoS scenario where the ratio betwen two VMs is changed from 50:50 to 40:60. The results are presented in Figure 12. Even though the ratio is changed, the service differentiation that is supported by the proposed scheduling scheme for the virtualized network devices is not affected by the change.

V. CONCLUSION

In this paper, we proposed an internal network virtualization framework based on Virtual Box, and also we built up a scheduling scheme for providing service differentiation among VMs. We specifically presented the proposed architecture for virtualizing network devices and scheduling those devices according to QoS specifications. Then we have demonstrated that the service differentiation can be successfully achieved through both the proposed virtualization framework and the differentiated scheduling scheme, regardless of network traffic, the number of VMs, or dynamic change of QoS specification. Note that the proposed scheduling scheme can cooperate with any framework that supports Virtual Box without the modification.

In the future work, we would like to devise various scheduling resources that can be used elaborately to schedule the virtualized network devices in the proposed framework. We also plan to examine the effect of the proposed scheduling scheme on real multimedia traffic. The study corroborates the effectiveness of the proposed virtualization framework for network devices.

ACKNOWLEDGEMENT

This work was supported in part by the IT R&D program of MKE/KEIT [KI001822, Research on Ubiquitous Mobility Management Methods for Higher Service Availability], and in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2010-0014060)

REFERENCES

[1] P. Barham, B. Dragovic, K. Fraser, S. Hand, T. Harris, A. Ho, R. Neugebauer, I. Pratt, and A. Warfield. Xen and the Art of Virtualization. In *Proc. 19th SOSP,* Lake George, NY, Oct 2003.

[2] E. Bugnion, S. Devine, and M. Rosenblum. Disco: Running Commodity Operating Systems on Scalable Multiprocessors. In *Sixteenth ACM Symposium on Operating System Principles*, October 1997.

[3] R. Chandra, V. Bahl, and P. Bahl. MultiNet: Connecting to multiple IEEE 802.11 networks using a single wireless card. In *INFOCOM*, 2004.

[4] G. W. Dunlap, S. T. King, S. Cinar, M. Basrai, and P. M. Chen. ReVirt: Enabling Intrusion Analysis through Virtual-Machine Logging and Replay. In *Proceedings of the 5th Symposium on Operating Systems Design and Implementation (OSDI 2002)*, ACM Operating Systems Review, Winter 2002 Special Issue, pages 211.224, Boston, MA, USA, Dec. 2002

[5] K. Fraser, S. Hand, R. Neugebauer, I. Pratt, A. Warfield, and M. Williamson. Safe hardware access with the Xen virtual machine monitor. In *Proceedings of the Workshop on Operating System and Architectural Support for the On Demand IT InfraStructure (OASIS)*, Oct. 2004.

[6] T. Garfinkel, B. Pfaff, J. Chow, M. Rosenblum, and D. Boneh, "Terra: A Virtual Machine-Based Platform for Trusted Computing," in *Proc. 9th ACM Symposium on Operating Systems Principles*, 2003, pp. 193–206.

[7] R. Goldberg. *Architectural Principles for Virtual Computer Systems*. PhD thesis, Harvard University, 1972.

[8] S. Govindan, A. R. Nath, A. Das, B. Urgaonkar, and A. Sivasubramaniam. Xen and co.: communication-aware cpu scheduling for consolidated xen-based hosting platforms. In *VEE '07: Proceedings of the 3rd international conference on Virtual execution environments*, pages 126–136, New York, NY, USA, 2007. ACM Press.

[9] A. Gulati, A. Merchant, M. Uysal, and P. Varman, Efficient and adaptive proportional share I/O scheduling, *Technical Report HPL-2007-186, HP Labs*, November, 2007.

[10] Iperf, http://sourceforge.net/projects/iperf

[11] J. Kurose and K. Ross, Computer Networking: A Top-Down Approach 4th Edition, page 675-678, Boston, MA, 2008, Pearson Education International

[12] J. Kurose and K. Ross, Computer Networking: A Top-Down Approach 4th Edition, page 477, Boston, MA, 2008, Pearson Education International

[13] G. Neiger, A. Santoni, F. Leung, D. Rodgers, and R. Uhlig. Intel virtualization technology: Hardware support for efficient processor virtualization. *Intel Technology Journal 10*, 3 (2006).

[14] D. Ongaro, A. Cox, and S. Rixner. Scheduling I/O in virtual machine monitors. In Fourth ACM SIGPLAN/SIGOPS International Conference on Virtual Execution Environments, pages 1-10, ACM, 2008

[15] P. Pancha and M. El Zarki. Leaky bucket-access control for VBR MPEG video. In *Proceedings of the IEEE INFOCOM'95,* Boston, April 1995

[16] J. Sugerman, G. Venkitachalam, and B. Lim. Virtualizing I/O devices on VMware workstation's hosted virtual machine monitor. In *Proc. 2001 Ann. USENIX Tech. Conf.*, Boston, MA, USA, June 2001.

[17] Sun Virtual Box Programming Guide and Reference 3.1.6, http://www.virtualbox.org

[18] Virtual Box, http://www.virtualbox.org

[19] A. Whitaker, M. Shaw, and S. D. Gribble. Denali: Lightweight Virtual Machines for Distributed and Networked Applications. Technical Report 02-02-01, University of Washington, 2002.

 ISBN: 978-1-61208-133-5

The Effects of Cell Size on Total Power Consumption, Handover, User Density of a Base Station, and Outage Probability

Youngmi Lim, Joo Hyung Lee, and Jun Kyun Choi
Department of Electrical Engineering
Korea Advanced Institute of Science and Technology (KAIST)
Daejeon, Republic of Korea
ymlim86@kaist.ac.kr, joohyung08@kaist.ac.kr, jkchoi59@kaist.edu

Abstract—**The green IT (information technology) issue is one of the important issues on the network department recently. Because the wireless access point has an amount of portion of network power consumption, reducing cell size is introduced to save the wireless network power consumption. In this paper, we investigate the effects of cell size in terms of total base station power consumption, handover rate, user density, and outage probability. Finally, as reducing the cell size, total power consumption and outage probability are decreased. However, the handover rate increase and the number of user in a cell decrease. Since, the multicast transmission scheme is good solution to reduce the bandwidth and delivering the same contents to user, we investigate the energy resource performance based on the multicast transmission system. Finally, these analyses can be helpful for energy efficient cell planning.**

Keywords-cell size; base station power consumption; handover; user population density; outage probability

I. INTRODUCTION

Over the last 5 years, the green IT issue is one of the most important issues on the electrical engineering department. During this period, many researches focused on reducing the power consumption. Moreover, it is also important issues to save energy and reduce the power consumption at network equipment. Especially for network department, wireless access point is the most important issue of green network because it has 40% of total power consumption at the access point in wireless networks [1]. Therefore, some technical issues such as network architecture, cell size, routing, etc., are researched to reduce energy consumption of the wireless access point [1].

Since the wireless access technology has been developed quickly and the number of users who requests various multimedia broadcasting and streaming such as IPTV increases, it is important to allocate resource efficiently [2]. To support quality of service (QoS) of users, wireless multicast transmission can be a good solution to reduce the resource waste for delivering the same contents to user. However, there is lack of consideration in energy efficiency of multicast transmission scheme, since it is just focused on some resource such as bandwidth, delay, capacity and so on. Therefore, we investigate the energy resource considering multicast transmission system. Moreover, Cell planning is main problem in the cellular mobile communication and also it is the key of reducing power consumption of base station. Previous cell planning technology for energy saving is focused on small-sized cell which has advantage of reducing base station power consumption [3].

In this paper, we focus on the cell planning technology for reducing the power consumption especially in wireless multicast environments. In detail, we analyze the base station power consumption and handover rates as varying the cell size. Moreover, we also analyze the multicast outage probability versus cell size. It is energy efficient way on behalf of the base station power consumption. However, a small-sized cell topology makes smaller coverage area and more frequent handover. The more handover causes latency and additional unnecessary power consumption. In addition, small-sized cell makes that the number of users per one base station is reduced. Therefore, our analysis shows the effects of cell size on total power consumption, handover, the number of user per base station, outage probability in this paper. This analysis can be efficient tool for cell planning by considering various points of view.

Moreover considering mobility, S.K. Lee et al. [4] already investigated the wireless access network based on WDM-PON for mobility support. They optimize the mobility management process between the corresponding node and the home agent. From their results, bandwidth waste and long end-to-end packet delay are reduced using their proposed scheme. However, power consumption and cell radius are not considered in their proposed scheme.

The remainder of the paper is organized as follow. Section Ⅱ briefly explains the related works. Section Ⅲ introduces the system model and performance. Section Ⅳ concludes the paper. Last, we present the future work which is to find the optimal solution in section Ⅴ.

II. RELATED WORKS

A. Cell size

Previous study, I. Hakki CAVDAR et al. [5] proposed an algorithm for the TDMA-FDMA mobile cellular communication system. They consider traffic and coverage analysis for procedure of cell planning. As the cell radius increases, transmitted power of base station (BS) and path loss are increased, however the capacity has better performance. They also consider three environment, urban

 ISBN: 978-1-61208-133-5

area, suburban area, and rural area. In case of the urban environment, performance is worse than suburban and rural environment. Jayant Baliga et al. [6] present a comparison of energy consumption of access networks which are passive optical networks, fiber to the node, point-to-point optical systems and WiMAX. Their results show that the optical access technologies provide the most energy-efficient solutions than other access technologies.

B. Handover

Daehyung Hong et al. [7] investigate the performance of cellular mobile radio telephone systems with handoff procedures. They consider the cellular structure, frequency reuse, and handoff for mobile radio telephone systems. In their paper, they also analyze the probability distribution of residing time in a cell and derive the handoff probability when mobile node resides in a cell to which its call is handed off. One of their results shows that mean channel holding time in a cell is increased as the cell radius is increase. Hyun-ho Choi et al. [8] propose the new vertical concept, *Takeover*, which enables a neighbor node to process requests of a mobile node. Their proposed handover scheme has better performance in terms of average handover latency, packet loss and power consumption. In their results, energy consumption per a mobile node is increased if the speed of mobile node is lager.

C. The User Population Density

In [5], they present the power consumption per a user and energy per bit versus average access rate for typical access networks. Power consumption per a user and energy per bit are important factor on energy resource point of view. In their results, power consumption per a user is increased if the average access rate is larger. However, they only consider the maximum average access rates that each technology can achieve at user population densities.

D. The Outage Probability

S.Y. Baek et al. [9] investigate the adaptive transmission scheme for mixed multicast and unicast traffic. They proposed a novel hybrid scheduling scheme which is consider some threshold SNR values for multicast transmission. For users have less than threshold SNR, they transmit the data using unicast transmission scheme not the multicast transmission scheme. They evaluate the system performance of multicast and unicast transmission schemes in terms of system capacity, worst average channel user's capacity, and outage probability for varying cell environments. According to their results, the outage probability of the multicast transmission increases as the cell radius increases. Moreover using their novel hybrid scheduling scheme, multicast capacity improves than conventional scheme. However in this paper, the power consumption is not also considered in this paper.

III. System Modeling and Performance

We first describe a system with a typical urban macro-cell model which has cell radius about 1.5km to 3.5km. Each base station consisted of hexagonal cell and one hexagonal cell area has $A_c = \frac{\sqrt{3}}{2}R^2$ value where the cell radius is R. We assume the traffic density has uniformly distribution. Also, we assume the user popular density of total area is fixed and we only consider the base station's transmission power not mobile node's transmission power. Because the multicast transmission scheme is more efficient way of reducing resource waste, we assume that this system provides the multicast transmission scheme. In multicast system, even though the worst channel user should be guaranteed quality of service.

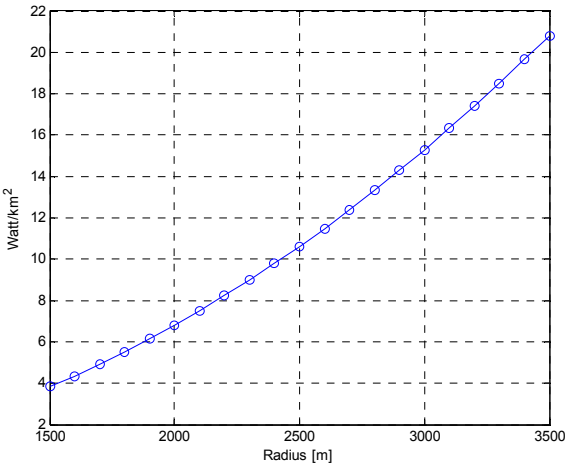


Figure 1. Area power consumption versus cell size

A. Base Station Transmit Power

Consider the propagation model [10] without shadowing, we can define the transmitted power of BS where the signal level is at least P_{min} follows

$$P_{tx} = \frac{P_{min}}{K} R^{\lambda}, \quad (1)$$

where P_{tx}, P_{min}, and λ denote transmit power, minimum transmit power at the cell boundary, and path loss exponent, respectively. In multicast service environment, the power at the cell boundary has minimum requirement power value because multicast service should provide the requirement data rate to worst channel user. Therefore, we fix the minimum power value at the cell boundary to guarantee the data rate. In macro cell environment, the effective propagation parameters are supported as path loss and the factor K, 4.00 and 2.2751, respectively.

Figure 1 shows the area power consumption as radius of BS is increased. The transmit power of BS is proportional to cell radius and also area power consumption, which is power consumption in unit area not a base station's transmit power, is increased when cell radius is increased.

B. Handover Rate

We take into account the handover rate to increase cell size. In more frequent handover environment, user's mobility

 ISBN: 978-1-61208-133-5

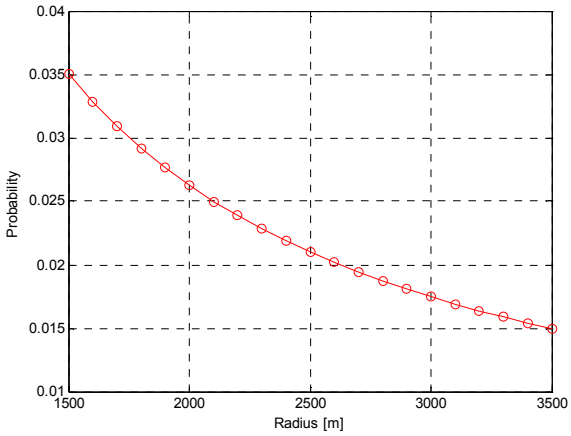


Figure 2. Probability of handover reside time versus radius

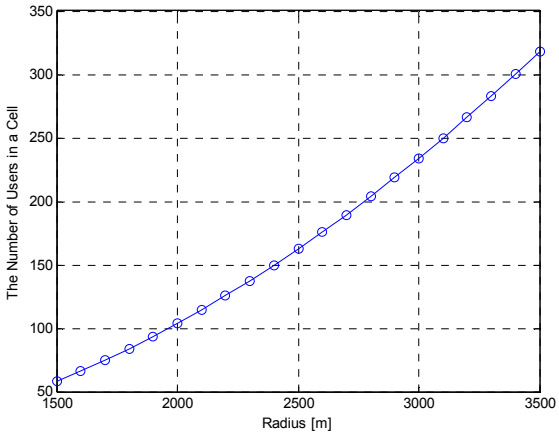


Figure 3. The number of users in a cell versus cell size

could not be guaranteed. Therefore, the handover rate is important factor in wireless system supporting mobility. In [7], the probability of the time, T_h, a mobile resides in a cell to which its call is handed off is defined as

$$P_{T_h} = \begin{cases} \frac{2}{\pi}\arcsin\left(\frac{V_{max}t}{2R_{eq}}\right) - \frac{4}{3\pi}\tan\left[\frac{1}{2}\arcsin\left(\frac{V_{max}t}{2R_{eq}}\right)\right] + \\ \frac{1}{3\pi}\sin\left[2arcsin\left(\frac{V_{max}t}{2R_{eq}}\right)\right], \quad for\ 0 \le t \le \frac{2R_{eq}}{V_{max}} \\ 1 - \frac{8R_{eq}}{3\pi V_{max}}\frac{1}{t}, \quad for\ t \ge \frac{2R_{eq}}{V_{max}} \end{cases}, \qquad (2)$$

where V_{max}, and R_{eq} are maximum speed of mobile terminal and radius of approximation circle, respectively. In this model, we set the speed is 30km/h. Also, we assume the maximum handoff time is fixed and then we can find the probability of handover varying cell radius.

Figure 2 shows the probability of the handover which a mobile node resides in a cell when its call is handed off and maximum handover time is fixed. The probability of that is decreased when cell radius is increased. Following this results, the handover is occurred more frequently when the cell radius goes to smaller. The probability is 0.035 when the cell radius is 1.5km and the probability goes to 0.015 when cell radius goes to 3.5km. However, the effect of cell radius for handover is very small. Therefore, we can ignore the effects of the handover rate by reducing cell size not considering the handover latency in macro cell environment.

C. The Number of Users per One Base Station

The number of users per one BS is lower when the cell size is decreased. When the cell size is decreased, coverage of a BS is also decreased. In this paper, we consider the number of users per one BS and the transmit power of BS per the number of user in a cell.

First, we investigate the number of users in a cell versus cell size. We assume that there are 10 users in $1km^2$ and users are uniformly distributed. Therefore, the number of users in a cell is expressed as

$$N_{BS\_user} = N_{sample\_user} \times \frac{A_{BS}}{A_{sample}}, \qquad (3)$$

where $A_{BS}$is the area of BS and A_{sample} is the sample area. It is proportional to R^2 because the BS coverage is proportional to R^2 and the other terms are constant value. Figure 3 shows its result so that the number of users in a cell is increased when the cell size is increased.

Next, we investigate the power per one user in a cell. It is important factor considering energy resource management. We can say that the energy efficiency is low when the power consumption per user is low even though the total power consumption is large. Therefore, we also consider the transmit power of BS per users in a cell and it is defined as

$$P_{user} = \frac{P_{tx}}{N_{BS\_user}}. \qquad (4)$$

Figure 4 shows the transmit power of BS per the number of users in a cell. Even though the number of users in a cell is increased when the cell radius is increased, the transmit power of BS per the number of users in a cell is increased. Therefore, reducing cell size is more energy efficient way than others in terms of allocated amount power for one user.

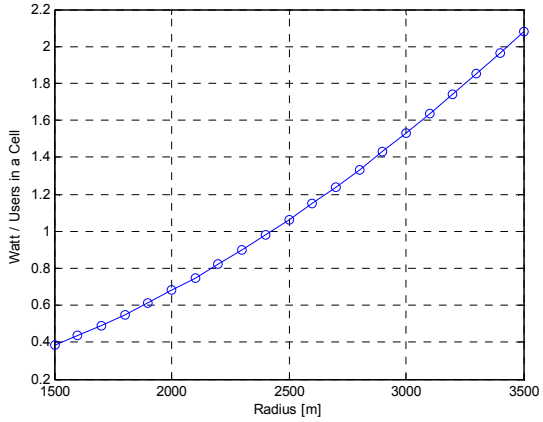


Figure 4. Base Station Transmit Power per Users in a Cell

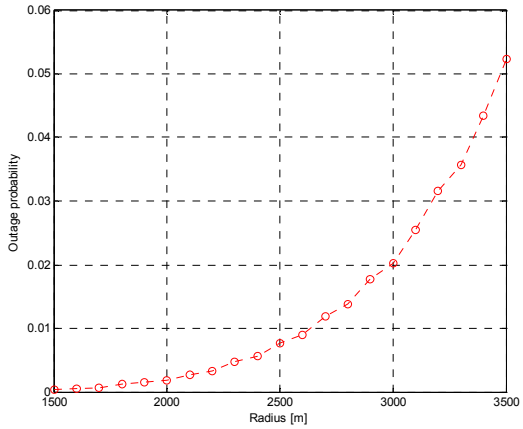


Figure 5. The outage probability versus cell size

D. The outage probability

In this paper, we consider the multicast transmission system. In multicast transmission, the outage probability is increased as the cell radius increases. From [8], the conditional pdf of selected users' SNR value is expressed as

$$f_{z|i}(z|i) = L\left(\frac{2\rho_0^{\frac{2}{n}}}{nR^2}\right)^{L} z^{-\frac{2L}{n}-1}\gamma\left(\frac{2}{n}+1,\frac{R^n z}{\rho_0}\right) \times \left[\gamma\left(\frac{2}{n},\frac{R^n z}{\rho_0}\right)\right]^{L-1}. \quad (5)$$

The parameter L means the number of users in a cell and R is the cell radius. We can calculate the number of users in a cell using ratio of BS coverage and sample area. Then, the outage probability of the multicast transmission scheme is expressed as

$$\Pr\{Z \le \Gamma\} = \int_0^{\Gamma} f_{z|i}(z|i)dz. \quad (6)$$

Figure 5 shows that the outage probability is increased when the cell radius is increased. The outage probability is 4.312×10^{-4} when the cell size is 1km and the outage probability is 0.0522 when the cell size is 3.5km. Therefore, there is advantage of reducing cell size in the aspect of outage probability.

IV. CONCLUSION

In this paper, we focus on the total base station power consumption, handover rate, the allocated power per user in a cell, and outage probability as varying cell size. Our analyses show that the power consumption and outage probability are proportional to the cell size. However, the handover probability and the number of users in a cell are inversely proportional to the cell size. Therefore considering energy efficiency, reducing the cell size is the most energy saving way in the aspect of base station power consumption. However in terms of handover, reducing the cell size does not guarantee the user's QoS or mobility. Finally, these two different aspects have trade-off relationship. Therefore, it is helpful for cell planning because our analyses show that these aspects should be considered.

On the contrary, the total power consumption will be increased from some point when the frequent handover is occurred because of the handover signaling. In addition, more base stations can consume more power consumption because of the initial power consumption of BS.

In future work, we will investigate the some optimal energy efficient point based on WDM-PON for HMIPv6 mobility support and also base station's power consumption.

ACKNOWLEDGMENT

This work was partly supported by the IT R&D program of MKE/KEIT. [KI001822, Research on Ubiquitous Mobility Management Methods for Higher Service Availability]

REFERENCES

[1] "Working Group Meeting Report", Green Touch, June 2010

[2] J.H. Lee, S.H. Shah Newaz, J.M. Lee, and J.K. Choi, "Context Profile Handling Mechanism for Source Specific Multicast in Proxy Mobile IP", ICOIN 2010, pp. 1-4, January 2010.

[3] A.G. Spilling, A.R. Nix, and M.A. Beach, "Adaptive Cell Sizing in Cellular Networks", IEEE Colloquium on capacity and range enhancement techniques for the third generation mobile communications, no. 11, pp.4/1-4/5, Feburuary 2000.

[4] S.K. Lee, Y.C. Jeon, T.H. Lim, K.H. Lee, and J.W. Park, "A Wireless access network based on WDM-PON for HMIPv6 mobility support", Wireless Networks, vol. 16, no. 6, pp. 1707-1722, August 2010.

[5] I.H. Cavdar and O. Akcay, "The Optimization of Cell Sizes and Base Stations Power Level in Cell Planning", VTC 2001, vol. 4, pp. 2344-2348, May 2001.

[6] J. Baliga, R. Ayre, W.V. Sorin, K. Hinton, and R.S. Tucker, "Energy Consumption in Access Networks", OFC/NFOEC 2008, pp. 1-3, Feburuary 2008.

[7] D.H. Hong and S.S. Rappaport, "Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures", IEEE Trans. Vehicular Technology, vol. 35, no. 3, pp. 77-92, August 1986.

[8] H.H. Choi and D.H. Cho, "Takeover: a new vertical handover concept for next-generation heterogeneous networks", VTC 2005 Spring, vol. 4, pp. 2225-2229, May 2005.

[9] S.Y. Baek, Y.J. Hong, and D.K. Sung, "Adaptive Transmission Scheme for Mixed Multicast and Unicast Traffic in Cellular Systems", IEEE Trans. Vehicular Technology, vol. 58, no. 6, pp. 2899-2908, July 2009.

[10] F. Richter, A.J. Fehske, and G.P. Fettweis, "Energy Efficiency Aspects of Base Station Deployment Strategies for Cellular Networks", VTC 2009 Fall, pp. 1-5, September 2009.

 ISBN: 978-1-61208-133-5

Mobile QoS provisioning by Flow Control Management in Proxy Mobile IPv6

Taihyong Yim, Tri M. Nguyen, Youngjun Kim and Jinwoo Park
School of Electrical Engineering
Korea University
Seoul, Rep. of Korea
{autonome, trinm, youngjun2, jwpark}@korea.ac.kr

***Abstract*- In the next-generation mobile wireless networks, mobility support and QoS provisioning are two critical issues. When it becomes much easier to access the internet from mobile devices, the real-time service over mobile network will be on high demand. To satisfy these requirements we must consider finest level of QoS guarantee in the mobile network. In this paper, we propose a QoS Provisioning Method based on flow-level traffic management for guaranteed service in Proxy Mobile IP.**

Keywords-QoS; Mobility; PMIP; Flow-based traffic management; Admission Control

I. INTRODUCTION

Mobility and QoS mechanism is the key issue in future wireless mobile networks. Future wireless mobile networks are expected to provide efficient mobility support with quality-of-service (QoS) guarantees. Services are required to maintain their network connectivity with the same QoS during handoff.

From Mobile IP (MIP)[1] to Proxy Mobile IP (PMIP)[2], many mobility management protocols are proposed to maintain the session continuity for higher layer in the IP-based networks. Especially PMIP [2], the Internet Engineering Task Force (IETF) standard, can serve as the basic network-based mobility management in the IP-based mobile networks. PMIP aims to solve the host-based mobility support scheme such as MIP. PMIP relies on the proxy mobility agents in the network to detect the MN's attachments and detachments and then signal this information, in the form of binding updates without the active participation of the MN itself. However PMIP is not enough to support the service continuity for guaranteed service. If the network suffers from congestion on the specific link, connections to networks may be broken and QoS also may be degraded. It is because PMIP is a legacy of IP, which is based on "best effort service".

To cope with this problem, QoS mechanisms have been largely studied in both wired and wireless environments. For example, Integrated Service (IntServ) [3] and Differentiated Service (DiffServ) [4] can be used in IP based wired and wireless networks. IntServ can provide QoS through admission control, classifier, packet scheduler and resource reservation. In IntServ, a QoS signaling protocol, Resource Reservation Protocol (RSVP) is used for this purpose. RSVP enables end applications requiring certain guaranteed services to signal their end-to-end QoS requirements to obtain service guarantees from the network. In IntServ network resources are reserved for a session according to a specific QoS requirement can support QoS per flow level though the reservation by exchanging explicit signaling messages. However, it has a scalability problem since it requires signaling messages to be exchanged between terminals periodically. Moreover, it results additional delay during handoff. Therefore RSVP is not suitable for mobile networks.

On the other hand, DiffServ is a direct extension to the work done by IntServ. While IntServ provides per-flow guarantees, DiffServ follows the Class of Service (CoS) of mapping multiple flows into a few service levels. DiffServ controls only traffic classes rather than each session within a traffic class. For CoS the SLA (Service Level Agreement) a central component of DiffServ, which is a service contract between a customer and a service provider. The SLA specifies the details of the traffic classifying and the corresponding forwarding service a customer should receive. DiffServ uses code point (DSCP) values in the IP header to deliver the CoS according to the SLA. However, DiffServ lacks controllability such as admission control and it cannot satisfy of per-flow QoS required in the various services.

For these reasons, simple and efficient QoS architecture is needed with a traffic management schemes in flow-level admission control, packet scheduling, policing, which do not use expensive signaling messages. In this manner, Flow-Aware Networking (FAN) is introduced by France Telecom in [5][6] as a new way of providing QoS in the IP networks. The main goal of this proposal is to ensure the proper QoS in packet networks in an implicit way. That is, no signaling is required to control the network. Each node makes locally optimal decision based on local observation. In the congestion state, new flows are blocked to protect existing flows by flow-level admission control of IP packets. On the other hand, Flow-State-Aware (FSA) technologies were

 ISBN: 978-1-61208-133-5

developed for NGN transport technologies [7]. FSA defines the service types based on typical examples of Internet services: maximum rate (MR), guaranteed rate (GR), variable rate (VR), and available rate (AR), and divides the network resource into two portions: fixed rate (FR) and network rate (NR). In FSA, signaling procedure requires every node to exchange requests and responses according to service types. Through this signaling capability in controlling transit nodes FSA can support QoS in flow-level.

In this paper, we proposed a Mobile Flow-Aware access network which can provide a mobile QoS provisioning of flow-level for PMIP. The rest of this paper is organized as follows. In Section II, we discuss the related work to this research. Section III describes our proposed network architecture and scheme. Finally, the conclusion and further work are presented in Section IV.

II. RELATED WORK

A. Proxy Mobile IP

Proxy Mobile IPv6 protocol is intended for providing network-based IP mobility management support to a mobile node, without requiring the participation of the mobile node in any IP mobility related signaling [2]. The mobility entities in the network will track the Mobile Node (MN)'s movements and will initiate the mobility signaling and set up the required routing state. Therefore, an MN is exempt from participation in any mobility-related signaling, and the proxy mobility agent in the serving network performs mobility-related signaling on behalf of the MN. Once an MN enters its PMIPv6 domain and performs access authentication, the serving network ensures that the MN is always on its home network and can obtain its HoA on any access network. That is, the serving network assigns a unique home network prefix to each MN, and conceptually this prefix always follows the MN wherever it moves within a PMIPv6 domain. From the perspective of the MN, the entire PMIPv6 domain appears as its home network. Accordingly, it is needless (or impossible) to configure the CoA at the MN. The new principal functional entities of PMIPv6 are the mobile access gateway (MAG) and local mobility anchor (LMA). The MAG typically runs on the AR. The main role of the MAG is to detect the MN's movements1 and initiate mobility-related signaling with the MN's LMA on behalf of the MN. In addition, the MAG establishes a tunnel with the LMA for enabling the MN to use an address from its home network prefix and emulates the MN's home network on the access network for each MN. On the other hand, the LMA is similar to the HA in MIPv6. However, it has additional capabilities required to support PMIPv6. The main role of the LMA is to maintain reachability to the MN's address while it moves around within a PMIPv6 domain, and the LMA includes a binding cache entry for each currently registered MN. The binding cache entry maintained at the LMA is more extended than that of the HA in MIPv6 with some additional fields such as the MN-Identifier, the MN's home network prefix, a flag indicating a proxy registration, and the interface identifier of the bidirectional tunnel between the LMA and MAG. Such information associates an MN with its serving MAG, and enables the relationship between the MAG and LMA to be maintained.

B. QoS mechanisms based on flow-based traffic management

To cope with limitation of IP based on best effort service, QoS mechanisms have been largely studied in both wired and wireless environments such as IntServ and DiffServ. IntServ enables end applications requiring certain guaranteed services to signal their end-to-end QoS requirements. On the other hand DiffServ controls only traffic classes rather than each session within a traffic class, which enables network to be scalable. However, both InServ and DiffServ have the limitation of scalability and controllability respectively.

For these reasons, simple and efficient QoS architecture is needed with a traffic management schemes in flow-level admission control, packet scheduling, policing. As the network processor and memory technologies developed, routers can recognize packets as a flow which is sequence of packets with the same 3-tuples or 5-tuples information. This enables the network to associate packets dynamically. That is, traffic control can be done at flow-level. The definition of flow is not fixed, but it could be defined in various ways according to the requirements of the user or service provider. A flow could be defined as a traffic flow which shares the 5-tuple IP header fields. Several schemes have proposed in this manner.

FAN is a new way of providing QoS in the IP network [4][5]. It is designed for providing state information to conventional IP router with stateless information for specific classification of the IP packet. In the FAN, packets are treated by the flow level. Through CAC per flow ongoing service can be maintained even in the situation of overload. It can guarantee more specific level of QoS compared with class-based traffic control architecture such as DiffServ. The main goal of FAN is to ensure the proper QoS in packet networks in an implicit way [6]. That is, no signaling is required to control the network. Each node makes locally optimal decision based on local observation. In the congestion state, new flows are blocked to protect existing flows by flow-level admission control of IP packets. If a packet comes into the system, the selected hashing function will generate a hash value. The hash value is used to find the flow state entry for the flow of the packet. If the packet is the first packet of the flow, no flow state entry for the flow exists. Therefore, a new flow state entry must be created for the flow with the appropriate forwarding and QoS information. On the other hand, if there is already a flow state entry for the flow, the packet is just processed

 ISBN: 978-1-61208-133-5

according to the information in the flow state table. Since a flow is uniquely identified by its 5-tuple fields, the lookup for the flow state table should be an exact match instead of longest-prefix match as in the IP forwarding table lookup. A flow state entry is created and maintained when the first packet enters the system [8]. Once flows are identified and maintained in the system, traffic management can be done for each flow.

III. PROPOSED FLOW CONTROL MANAGEMENT

In this section, we describe our proposed QoS provisioning scheme for guaranteed service in Proxy MIP. As mentioned earlier, our ultimate aim is to overcome the limitation of Proxy MIP and benefit from the QoS support capability of flow-based traffic management. Proposed access networks is based on the integration of PMIP [2] and Flow-aware technologies [5][6][7]. That is, mobility management is performed according to PMIP and QoS provision is obtained by Flow-aware technologies. In the following, we present the operation of our proposed network architecture, namely, Mobile Flow-Aware Access Network and mobility management schemes and QoS provisioning method in flow-level.

A typical architecture for Mobile Flow-Aware access network is shown in Fig.1. We assume that a Mobile Flow-Aware access network exist between the Mobile Flow-Aware Local Mobility Anchor (MFA-LMA) and the Mobile Flow-Aware Mobile Access Gateway (MFA-MAG). The architecture is based on a two-level hierarchy. At the higher level is the MFA-LMA that performs the role of the LMA as it of PMIP [2] with flow-based traffic management function. At the second level is the MFA-MAG that is responsible for tracking the MN's movements to and from the access link as conventional MAG in PMIP. MFA-MAG also has a function of flow-level traffic management The MFA is an intermediate node that route packets with function of flow-level traffic management.

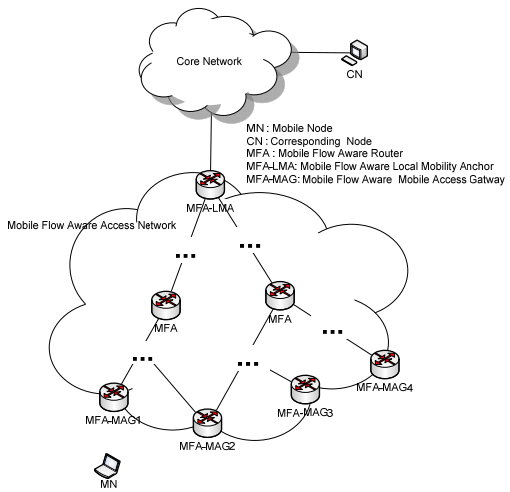


Figure 1. Architecture of a proposed flow- aware access network

A. QoS Provisioning for Guaranteed Service

For QoS provisioning two types of flows are defined: elastic and inelastic. Elastic flow is usually used for data transmission, served with the best effort regime such as web service. On the other hand, inelastic flow is used for delay-sensitive services, served with the specific fixed data rate like VoIP services. The packets of the latter have a higher priority than that of the former. The goal of the proposed QoS provisioning is to guarantee the inelastic flows even though the congestion is occurred at the link. For this purpose, each MFA node should store the list of the ongoing inelastic flow, namely, Flow Cache Entry (FCE) at each interface. Fig. 2 shows the structure of FCE. FCE include the 5-tuple of packets (Source/Destination IP address, Source/Destination port number, higher layer protocol) and interface of the MFA link. Flow entry is maintained by soft state, that is, no explicit signaling is needed.

The main elements of the proposed flow-based traffic management scheme are shown in Fig..3. On a packet arriving at node incoming interface, the packet is classified into elastic or inelastic flow by the classifier. While elastic flows are forwarded to admission control block directly, inelastic flows are checked whether new flows or not. If there are matching entries in the FCE, flows are forwarded to sub queue at the outgoing interface directly. If a packet of flow is determined to new flow, then FCE may be updated according to result of the admission control. The admission control uses traffic measurement of waiting time in sub-queues for inelastic flows. Congestion state can be defined as the state that satisfy the following inequation,

$$\sum_{n=1}^{k} \frac{l_i(n) \times 8}{r_i} \geq \varepsilon_i \qquad (1)$$

where li(n) is the length (byte) of n-th packet of inelastic flow i, ri is the service rate (bits/sec) of flow i and εi is the delay constraint for flow i. If the total sum of the expected service time for each flow which are waiting in the sub-queue is longer than the delay time for QoS of flow i, new elastic flows are blocked to protect existing inelastic flows. According to this admission control Each MFA node makes locally optimal decision based on local observation. The

Flow Cache Entry

interface	Source IP address	Destination IP address	Source port number	Destination Port number	Protocol	time
...	...	...	...	...	...	...

Figure 2. Flow Cache Entry

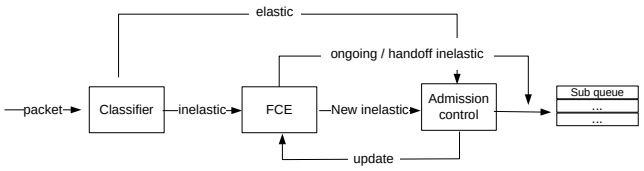


Figure 3. Flow-based traffic management

main advantage of proposed QoS provisioning is simplicity. It requires no signaling for QoS. Only implicit admission control is required upon congestion.

B. QoS Provisioning for handoff flow

When MN moves between MAGs, the flow path may be changed. Mobility can cause some problem in flow-based QoS control such as the failure of flow identification. To protect ongoing inelastic flows in the congestion, MFA node should keep the FCE. However, FCE is maintained locally, therefore some MFA nodes on the newly changed path according to handoff do not have flow list in FCE for the handoff flow. That is, handoff flow can be treated as a new inelastic flow and blocked in the congestion state. To avoid handoff flows treating as new flows, two types of FCE are proposed: Local FCE and Global FCE. Both Local FCE and Global FCE have the same structure as shown in Fig. 2. The only difference between two FCEs is the coverage of the contained flows in the list. That is, the Local FCE is the FCE that is managed by each node respectively while the Global FCE is the FCE that is managed by MFA-LMA. In other words, Local FCE contains the list of inelastic flows that are treated independently by a MFA and Global FCE contains the list of all inelastic flows in the domain. Fig. 4 shows the admission control for handoff flow identification. MFA checks Local FCE first and then checks Global FCE additionally. Through this simple mechanism handoff flow can be detected at the node. Therefore the QoS for handoff flows can be support like ongoing flows.

Basically local MFA do not need to maintain the Global FCE. Although the FCE is maintained in soft state, to maintain the FCE is a burden to the MFA. Therefore the small size of the FCE is good for MFA. For this reason, MFA refer the Global FCE only when MN moves to its

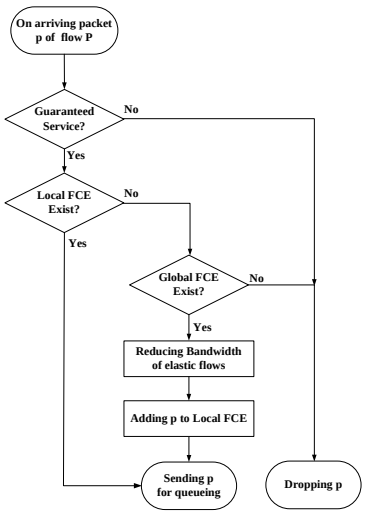


Figure 4. Admission control

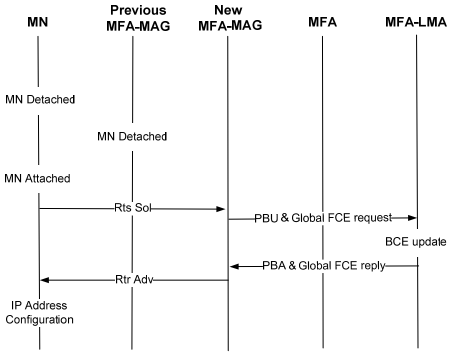


Figure 5. Global FCE request/reply during handoff procedures

local area through the PBU. The Global FCE request/reply procedures are shown in the Fig. 5. Handoff procedures follow the procedures of conventional PMIP [2] basically. The Global FCE is requested and replied with PBU and PBA. Instead of noticing list of all the flows, MFA-LMA just sends the list of flows of MN for reducing the burden of unnecessary work.

IV. PERFORMANCE ANALYSIS

In this section, we analyze the QoS provisioning of flow-based traffic management mechanism that has been proposed in the previous sections. We did not build any concrete numerical methodologies to analyze the queue management schemes; therefore, we provide computer simulation results. The network topology for the simulation is shown in Fig. 6. The links between nodes are set to have a link speed of 1Gbps. Background traffic of 100 elastic flows with CBR 10 Mbps is generated by MN1 and sent to CN1, and MN2 and MN3 each generate 50 inelastic flows with CBR 20 Mbps to send to CN2. Therefore, a total of 2Gbps traffic is trying to be sent between MFA and MFA-LMA, which causes congestion at the link. The packet size was set to 1,000 bytes.

Fig. 7.(a) shows MFA-MAG1's throughput when proposed QoS provisioning is not applied. As we can expect, the rates of the flows are fairly distributed; 100 flows share 1Gbps fairly; therefore, each flow receives about 10 Mbps. This means QoS of inelastic flows is not provided. If we need to guarantee the bandwidth of a certain flow to 20 Mbps, it is not possible in that architecture.

In Fig. 7.(b), our mechanism can guarantee 20 Mbps for a inelastic flow, and the rest of the bandwidth can be fairly shared among the other flows.

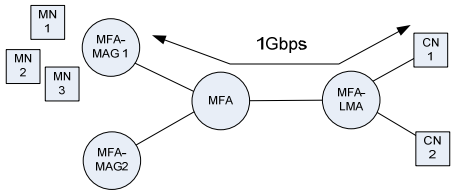
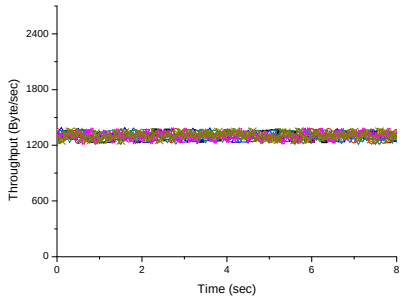
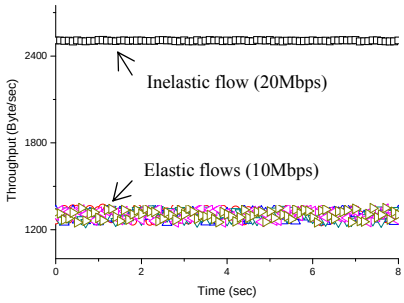


Figure 6. Simulation network topology



(a)



(b)

Figure 7. Throughput : (a) conventional scheme, (b) proposed scheme

V. CONCLUSION

This paper described a new QoS provisioning based-on flow-level traffic management in the PMIP for guarantee the QoS of inelastic flows. The proposed scheme shows how flow-level QoS provisioning can be evaluated in the PMIP domain. Through proposed classification, admission control, buffer managements, QoS for inelastic flows are guaranteed even when the network link is congested. Also through managing the two kinds of FCE, mobile inelastic flows also can be treated with the same priority of ongoing flows and protected on congestion state.

As a further works, we'll analyze the performance of proposed scheme with numerical methodologies to prove the advantage of our proposed scheme.

ACKNOWLEDGMENT

This work was supported by the IT R&D program of MKE/KEIT[KI001822, Research on Ubiquitous Mobility Management Methods for Higher Service Availability] and partly by the BLS project from the Seoul Metropolitan City[WR080951, Seoul R&BD Program]

REFERENCES

[1] C. Perkins, "IP Mobility Support for IPv4," RFC 3220, Jan. 2002.

[2] S.Gundavelli,Ed et al., "Proxy Mobile IPv6," RFC5213, Aug. 2008.

[3] R. Braden et al., Integrated Services in the Internet Architecture: An Overview, IETF RFC 1633, June 1994.

[4] S. Blake et al., An Architecture for Differentiated Services, IETF RFC 2475, Dec. 1998.

[5] Lawrence G. Roberts, "The Next Generation of IP - Flow Routing", Proceedings of SSGRR 2003, Italy, 2003.

[6] S. Oueslati, J. Roberts, A new direction for quality of service: Flow-aware networking, Proceedings of NGI 2005, Rome, June 2005.

[7] ITU-T Recommendation Y.2121, Requirements for the Support of Stateful Flow-Aware Transport Technology in an NGN, Sept. 2007.

[8] Nam-Seok Ko, Sung-Back Hong, Kyung-Ho Lee, Hong-Shik Park, and Nam Kim, "Quality-of-Service Mechanisms for Flow-Based Routers," ETRI Journal, Volume 30, Number 2, April 20

[9] N. Yamagaki, H. Tode, and K. Murakami, "DMFQ Hardware Design of Flow-Based Queue Management Scheme for Improving the Fairness," *IEICE Trans. Comm.*, vol. E88-B, no. 4, Apr. 2005, pp. 1413-142

 ISBN: 978-1-61208-133-5

Resource-Efficient Class-based Flow Mobility Support in PMIPv6 domain

Jiwon Jang[*], Seil Jeon[*], Younghan Kim[*#], and Jinwoo Park[†]
[*]School of Electronic Engineering, Soongsil University, Dongjak-gu, Seoul, 156-743, Korea
{jwjang84, sijeon, yhkim}@dcn.ssu.ac.kr
[†]Department of Electronics and Computer Engineering, Korea University, Seongbuk-ku, Seoul, 136-713, Korea
jwpark@korea.ac.kr

***Abstract*— Flow-based mobility support is becoming increasingly common in multi-interface environments because it provides flexible network selection per application flow and better network experience for mobile users. Recently, several drafts related to flow mobility have been being handled in IETF, but mobility handling per individual flow leads to signaling overhead and power consumption issues because individual flow always wants to have the best connected service with all the available network interfaces in its own demand. Power saving communication is becoming a worldwide issue in the mobile communication field. To make resource-efficient flow mobility, we propose a class-based flow mobility (CFM) mechanism. Through the performance analysis and results, we confirm that a CFM mechanism is superior to an individual flow mobility (IFM) mechanism in terms of signaling overhead and power consumption.**

Keywords - Proxy Mobile IPv6; PMIPv6; flow mobility; class-based flow mobility

I. INTRODUCTION

Multi-interface on mobile devices is becoming increasingly common. In such an environment, flow handover, which controls individual application flows from one interface to another even when the mobile node (MN) does not physically switch its network interface, is becoming the one of the most critical issues in the research field of next generation wireless network. Flow handover can provide a better network experience for end users and can also enable a network operator to balance the load appropriately depending on the availability of network capacity.

For this reason, in the Internet Engineering Task Force (IETF), several proposals [1][2] for flow handover are being handled over Proxy Mobile IPv6 (PMIPv6) [3] to provide network-based mobility management support to an MN without requiring its IP mobility-related signaling. However, it has several drawbacks. First, it can easily bring about signaling overhead that enables all flows to have the best connected network. Second, it can quickly run out of battery power because it preferentially considers flows' performance with all the available network interfaces.

To complement these drawbacks of the individual flow mobility mechanism (IFM), we propose a class-based flow mobility (CFM) mechanism by classifying the application flows into groups and performing group-based flow handling. Through the performance analysis, we confirm that CFM is more resource-efficient than IFM in terms of signaling overhead and power consumption.

The rest of this paper is organized as follows. In Section II, we explain the IFM mechanisms proposed in IETF. In Section III, we propose CFM mechanism. Section IV evaluates a performance of IFM and CFM mechanism based on an analytical model and presents the numerical results. In Section V, we offer a conclusion.

II. FLOW MOBILITY IN PMIPV6

Proxy Mobile IPv6 (PMIPv6) provides network-based mobility management for an MN that is connecting to a PMIPv6 domain. PMIPv6 introduces two new functional entities: the local mobility anchor (LMA) and the mobile access gateway (MAG). The MAG detects the MN's movement and provides IP connectivity. The LMA assigns one or more home network prefixes (HNPs) to the MN and is the topological anchor for all traffic belonging to the MN. The PMIPv6 allows MNs to connect to the network through multiple interfaces for simultaneous access. The MN can send packets simultaneously to the PMIPv6 domain over multiple interfaces. However, for supporting flow handover over PMIPv6, two issues should be resolved.

First, an HNP is assigned to one interface at a time because PMIPv6 employs per-MN prefix model. Therefore, when the flow mobility occurs, some of these flows are moved to a new interface while the other flows remain transmitted via the old interface. For keeping the sessions, the HNP should be assigned to multiple interfaces simultaneously. To solve the issue, a logical interface-based approach is proposed as one option to hide the changes at the physical interfaces from the IP layer [4].

Second, the PMIPv6 does not support flow-based routing because the LMA performs an HNP-based packet routing. To make packets route at the flow-based level, the LMA binding cache is required to be extended [2]. By applying these solutions, the flow handover can provide flexible network selection and better network experience for end users. However, two representative flow mobility solutions proposed in IETF do not consider signaling overhead and battery power that is consumed to control all the flows. Thus, it leads to a waste of resource for an MN and an access network. To complement these drawbacks, a novel flow handover scheme is required.

#Corresponding author

 ISBN: 978-1-61208-133-5

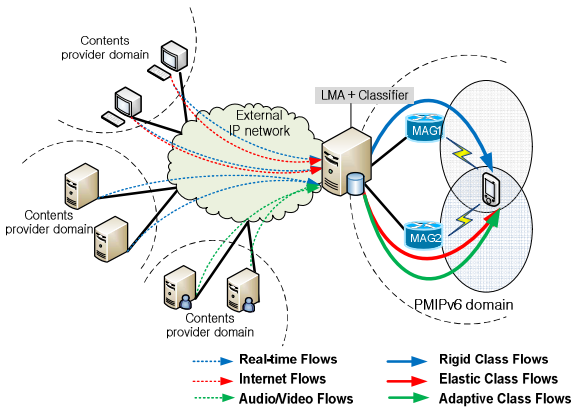


Figure 1. Class-based flow mobility reference network model

III. CLASS-BASED FLOW MOBILITY (CFM) SCHEME

The CFM scheme maximizes the user's performance and minimizes signaling overhead and power consumption. Specifically, it classifies and splits application flows of same class into groups, then it performs CFM scheme targeting a same kinds of class flows.

To classify flows according to application type, a classifier is required within LMA as shown in Figure 1. Several flows are divided into three categories: rigid, elastic, and adaptive. Generally, real-time services such as VoIP are classified as rigid class. Traditional Internet flows, such as FTP and Web are classified as elastic class. And delay-adaptive audio/video streaming or rate-adaptive multimedia flows belong to the adaptive class [5]. Such classification methods are divided into header-based and payload-based methods. Recent services are frequently running on non-standard ports, so the header-based classification method that checks the packet header is difficult to classify correctly. On the contrary, the payload-based classification method that checks the entire protocol payload requires a lot of computational power, and leads to significant overhead [6]. Therefore, we propose a new flow classification algorithm to support the CFM as shown in Figure 2.

To facilitate class-based flow forwarding, several binding entry information is needed on the LMA; therefore, class binding entry (CBE) consisting of three attributes such as class ID, QoS Parameter, and binding ID and data structure are offered, as shown in Figure 3.

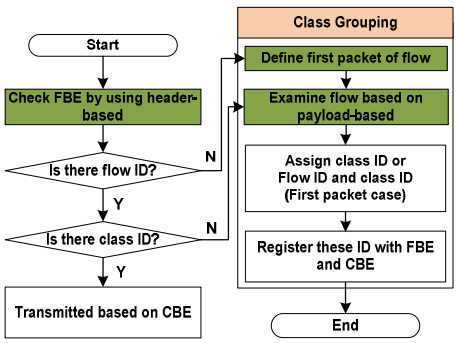


Figure 2. Proposed classification algorithm

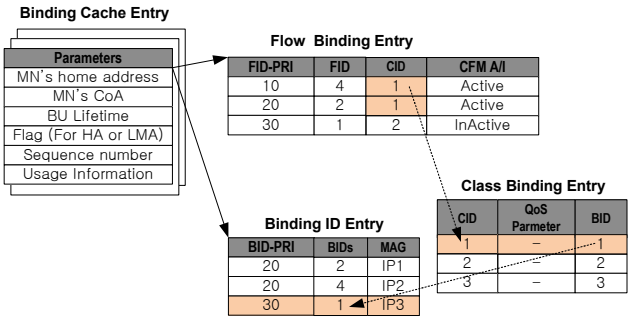


Figure 3. Extended binding cache entry for proposed CFM

The classifier assigns the flow ID and class ID. Assigned IDs are managed within flow binding entry (FBE) and CBE. When flow handover occurs, the same kind of class flows are moved to target MAG through confirming CBE.

In CFM, because each flow within the same class has its own requirement, moving these flows to the single target network with meeting the requirement requires the appropriate algorithm. As one of the solutions for this requirement, we can use the fairness algorithm proposed in [7]. From this solution, we can decide whether to move grouped flows to another interface.

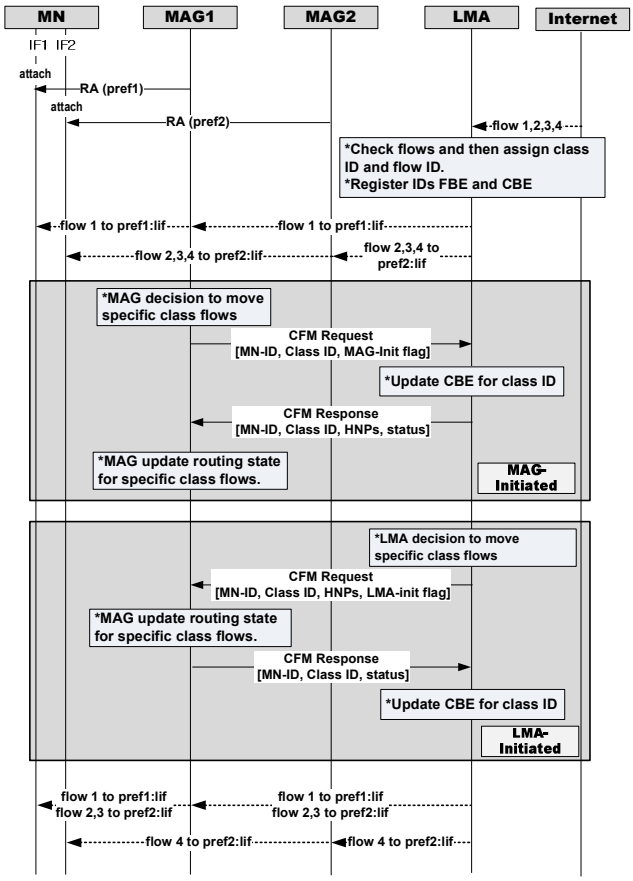


Figure 4. LMA and MAG-Initiated CFM procedure

After performing flow classification and enabling class-based flow forwarding, and deciding fairness values, the flow information such as MN-ID, class-ID, and HNP between MAG and using class flow mobility request/response (CFM Request/Response) is announced. These signaling methods are operated differently for two cases (e.g. MAG-initiated and LMA-initiated), as shown in Figure 4 in details. Flows 2 and 3 belong to the same class.

IV. PERFORMANCE ANALYSIS AND NUMERICAL RESULTS

This Section presents a performance analysis of the CFM and the IFM mechanisms. For ease of analytical modeling, we assume that the network is always possible to admit all flows and the bandwidth required for individual flow within same class is equal. Under these assumptions, we analyze signaling overhead and power consumption of two mechanisms. And we offer the numerical results by comparing their performances.

A. System Model

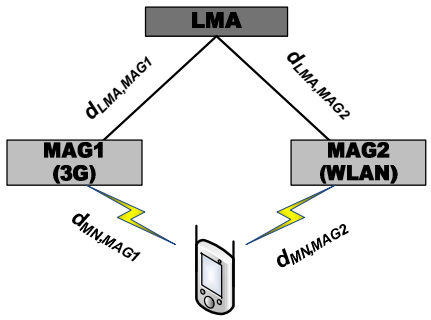


Figure 5. Network topology model for performance analysis

Figure 5 illustrates the network model for performance analysis. $d_{x\text{-}y}$ denotes the hop distance between two network entities x and y. We define μ_s and λ_s as the cell-crossing rate for which the MN still keeps its residence in the same domain and session arrival rate. From them, we obtain the average number of movement, $E(N_s)$ and express it as follows:

$$E(N_s) = \mu_s / \lambda_s. \tag{1}$$

B. Signaling Cost

Cz denotes signaling cost to conduct flow handover operation of z scheme. And, the signaling cost is defined as product of hop distance and signaling message generated between LMA and MAG, considering MN's movement and processing cost issued from MAG, LMA. The signaling cost for CFM and IFM are expressed by

$$\begin{aligned} C_{CFM} &= E(N_s)\cdot(\tau\cdot(d_{LMA,MAG2}\cdot L_{CFM\,Req}) \\ &+ P_R\cdot(d_{LMA,MAG2}-1)+P_{MAG2}+\tau\cdot(d_{LMA,MAG2}\cdot L_{CFM\,Res}) \\ &+ P_R\cdot(d_{LMA,MAG2}-1)+P_{MAG2}), \end{aligned} \tag{2}$$

$$\begin{aligned} C_{IFM} &= E(N_s)\cdot(\tau\cdot(d_{LMA,MAG2}\cdot L_{PBA}) \\ &+ P_R\cdot(d_{LMA,MAG2}-1)+P_{MAG2}+2\cdot P_{LMA}+P_{MAG1} \\ &+ \tau\cdot(d_{LMA,MAG1}\cdot L_{PBA})+P_R\cdot(d_{LMA,MAG1}-1)+P_{LMA}), \end{aligned} \tag{3}$$

where τ and L_m are the unit transmission cost over wired link and the amount of the signaling message, respectively. P_R is the routing processing cost between routers, while P_{LMA} and P_{MAG} denote the processing cost required in LMA and MAG.

C. Power Consumption Cost

The power consumption cost is defined as the amount of power consumed in MN. It is dependent on cell paging, scanning, beacon operation, time of data communication, and the amount of data being received (or sent) by a particular type of application. It is computed using the sum of time of data communication and amount of data. It can be expressed by the following:

$$P = r_d\cdot d + r_t\cdot t + c. \tag{4}$$

Here, r_d and d refer to the power consumption rate for data and the amount of data, respectively. r_t and t are the power consumption per unit time and the transaction time, respectively, and P and c refer to the total power consumption cost to receive d amount of data. Two kinds of application types such as video streaming or VoIP are considered. Corresponding equation is derived from [9].

$$P = t\cdot[r_t + R_{req}\cdot r_d] + c, \tag{5}$$

where R_{req} is the data rate required by the specific session.

D. Numerical Results

We employ some of parameter values used in the literature [8] [9], which are shown in Tables I and II.

Figure 6 shows the signaling cost of IFM and CFM as MN's velocity increases. In the case of CFM, it is assumed that the flows are grouped within the same class. The result shows that IFM increases proportionally to the number of flows, and that the cost of CFM is lower than that of IFM, regardless of the number of flows. $d_{LMA,MAG1}$ and $d_{LMA,MAG2}$ are 4, and $d_{MN,MAG1}$ and $d_{MN,MAG2}$ are 1, respectively.

Figure 7 illustrates power consumption according to four case scenarios. It is assumed that an MN has 1 VoIP and 2 video streaming flows where packet arrival rate for VoIP and video are 80 Kbyte/s and 200~300 Kbyte/s, respectively. In case 3, the MN uses 3 sessions through 3G interface when a single video streaming session is moved to the WLAN interface at 60 seconds.

TABLE I. PARAMETERS USED FOR NUMERICAL RESULTS

Parameters	Values	Parameters	Values
τ	0.5	P_R	0.008s
L_m	100 (byte)	μ_s	0.01
λ_s	10	t	30~150s

 ISBN: 978-1-61208-133-5

TABLE II. POWER CONSUMPTION IN 3G/WLAN INTERFACE

Mode	Parameters	
	r_t (W)	r_d (J/Kbyte)
3G	0.45	0.001
WLAN	0.9	4.12E-04

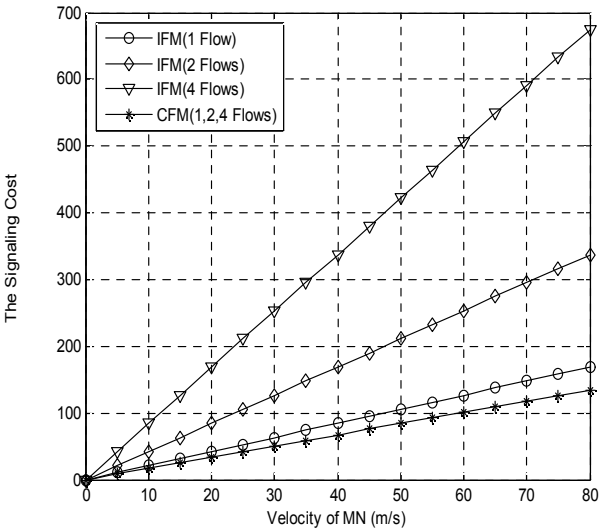


Figure 6. Singnaling ost as MN's velocity increases

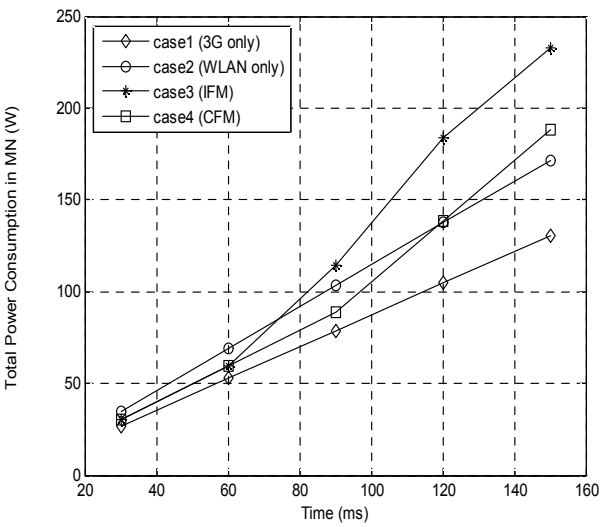


Figure 7. Power consumption cost

At this time, the power consumption cost increases significantly. Then, at 90 seconds, the other video streaming session is also moved to the WLAN interface. At that time, the power consumption cost becomes much higher than in Cases 1 and 2 because WLAN and 3G interfaces are used at the same time. In Case 4, two video sessions of same class are moved to the WLAN interface at 90 seconds using the fairness algorithm, and the power consumption cost rapidly increases. From these results using simple cases, we confirm that the CFM can avoid unnecessary signaling overhead in network side and also reduce power consumption in host side.

V. CONCLUSION

Flow mobility is an effective mobility technique that can provide flexible network selection per application flow and better network experience for mobile users. But individual flow mobility schemes introduced in IETF bring about signaling overhead and power consumption issues due to the pursuit of only the performance of individual flow with high priority.

To solve these issues, we propose a CFM mechanism, which classifies the application flows into groups and performs group-based flow handover.

Through the performance analysis, we confirm that the CFM mechanism is more resource-efficient than the IFM mechanism in terms of signaling overhead and power consumption. For future work, we will evaluate additional performance factor by using simulation.

ACKNOWLEDGMENT

This work was partially supported by the MKE(The Ministry of Knowledge Economy), Korea, under the "program for CITG" support program supervised by the NIPA(National IT Industry Promotion Agency), and the IT R&D program of MKE/KEIT [KI001822, Research on Ubiquitous Mobility Management Methods for Higher Service Availability].

REFERENCES

[1] CJ. Bernardos, M. Jeyatharan, R.Koodli, T. Melia, and F. Xia, "Proxy Mobile IPv6 Extensions to Support Flow Mobility," draft-bernardos-netext-pmipv6-flowmob-00, July, 2010.

[2] T. Tran, Y. Hong, and Y. Han, "Flow Tracking Procedure for PMIPv6," draft-trung-netext-flow-tracking-01, July, 2010.

[3] S. Gundavelli, K. Leung, V. Devarapalli, K. Chowdhury, and B. Patil, "Proxy Mobile IPv6," RFC 5213, August 2008.

[4] T. Melia and S. Gundavelli, "Logical Interface Support for Multi-mode IP Hosts," draft-ietf-netext-logical-interface-support-00, August, 2010.

[5] L. Breslau and S. Shenker "Best Effort versus Reservation: A Simple Comparative Analysis", ACM Computer Communications Review vol. 28, no. 4, pp.131-143, 1998.

[6] A. W. Moore and K. Papagiannaki, "Toward the Accurate Identification of Network Applications," LNCS 3431, pp. 41-54, 2005.

[7] M. Kim and S. Lee, "Load Balancing and Its Performance Evaluation for Layer 3 and IEEE 802.21 Frameworks in PMIPv6-based Wireless Networks," Wireless Communications and Mobile Computing, vol. 10, no. 11, pp.1431-1443, November 2010.

[8] S. Jeon, N. Kang, Y. Kim and W. Yoon, "Enhanced PMIPv6 Route Optimization Handover," IEICE Transactions on Communications vol. E91-B, no.11, pp. 3715-3718, November 2008.

[9] K.Mahmud, M. Inoue, H. Murakammi, M. Hasegawa, and H. Morikawa, "Energy Consumption Measurement of Wireless Interfaces in Multi-Service User Terminals for Heterogeneous Wireless Networks," IEICE Transactions on Communications, vol. E88-B, no.3, pp. 1097-1110, March 2005.

 ISBN: 978-1-61208-133-5

User to User adaptive routing based on QoE

Hai Anh TRAN, Abdelhamid MELLOUK, Said HOCEINI,
University of Paris-Est Creteil Val de Marne (UPEC)
Image, Signal and Intelligent Systems Lab-LiSSi Lab
Transport Infrastructure and Network Control Group - TINC
122 rue Paul Armangot, 94400 Vitry sur Seine, France
{hai-anh.tran, mellouk, hoceini}@u-pec.fr

***Abstract*—Service quality can be defined as "the collective effect of service performances which determine the degree of satisfaction of a user of the service" [1]. In other words, quality is the customer's perception of a delivered service. As larger varieties of services are offered to customers, the impact of network performance on the quality of service will be more complex. It is vital that service engineers identify network-performance issues that impact customer service. They also must quantify revenue lost due to service degradation. The Quality of Experience (QoE) becomes recently the most important tendency to guarantee the quality of network services. QoE represents the subjective perception of end-users using network services with network functions such as admission control, resource management, routing, traffic control, etc. In this paper, our main focus is routing mechanism driven by QoE end-users. With the purpose of avoiding the NP-complete problem and reducing the complexity problem for the future Internet, we propose two protocols based on user QoE measurement in routing paradigm to construct an adaptive and evolutionary system. Our first approach is a routing driven by terminal QoE basing on a least squares reinforcement learning technique called Least Squares Policy Iteration. The second approach, namely QQAR (QoE Q-learning based Adaptive Routing), is a improvement of the first one. QQAR basing on Q-Learning, a Reinforcement Learning algorithm, uses Pseudo Subjective Quality Assessment (PSQA), a real-time QoE assessment tool based on Random Neural Network, to evaluate QoE. Experimental results showed a significant performance against over other traditional routing protocols.**

***Index Terms*—Quality of Service (QoS), Quality of Experience (QoE), Network Services, Routing System, Autonomous System, Pseudo Subjective Quality Assessment (PSQA), Reinforcement Learning.**

I. INTRODUCTION

In order to reach new opportunities and improve market competitiveness, network service providers are offering new value-added services, such as video on demand (VoD), IPTV, voice over IP (VoIP), etc. Consequently, improving the quality of the services as perceived by the users, commonly referred to as the quality of experience (QoE), has a great effect as well as a significant challenge to the service providers with a goal to minimize the customer churn yet maintaining their competitive edge. Based on this kind of quality competition, the new term of QoE has been introduced, combining user perception, experience and expectations without technical parameters such as QoS parameters. In fact, the network provider's aim is to provide a good user experience at minimal network resource usage. It is important from the network operator to be aware of the degree of influence of each network's factor on the user perception. For users, also for operators and Internet service providers, the end-to-end quality is one of the major factors to be achieved. QoE takes into account the needs and the desires of the subscribers when using network services, while the concept of QoS just attempts to objectively measure the service delivered. Furthermore, e2e QoS with more than two non correlated criteria is NP-complete (proved in [2]). With the evolution of the Internet, both technologies and needs continue to develop, so complexity and cost become limiting factors in the future evolution of networks. In order to reduce this complexity problem, one has integrated QoE in network systems. Firstly, as an important measure of the end-to-end performance at the service level from the user's perspective, the QoE is an important metric for the design of systems and engineering processes. Secondly, with QoE paradigm, we can reach a better solution and prevent the NP-complete problem because our goal is just maintaining QoE criteria instead of optimizing multiple QoS criteria.

Routing mechanism is key to the success of large-scale, distributed communication and heterogeneous networks. In this section, a review of some related works reveals that various approaches have been proposed to take account of QoS requirements. However the goal of every traditional algorithm is to maximize many QoS criteria simultaneously. So they meet the NP-complet problem as we mentioned before.

The idea of applying reinforcement learning to routing in networks was firstly introduced by [3]. Authors described the Q-routing algorithm for packet routing. Reinforcement learning module is embedded into each node of a switching network. In [3], each node to keep accurate statistics on which routing decisions lead to minimal delivery times uses only local communication. However, this proposal focus on optimizing only one basis QoS metric (delivery times). So user perception (QoE) is not yet considered in this approach. [4] proposed an application of gradient ascent algorithm for RL to a complex domain of packet routing in network communication. This approach updates the local policies while avoiding the necessity for centralized control or global knowledge of the networks structure. The only global

 ISBN: 978-1-61208-133-5

information required by the learning algorithm is the network utility expressed as a reward signal distributed once in an epoch and dependent on the average routing time. In [5], K-Optimal path Q-Routing Algorithm (KOQRA) is presented as a QoS based routing algorithm based on a multi-path routing approach combined with the Q-routing algorithm. The global learning algorithm finds K best paths in terms of cumulative link cost and optimizes the average delivery time on these paths. The technique used to estimate the end-to-end delay is based on the Q-Learning algorithm to take into account dynamic changes in networks. In [6] AV-BW Delay Q-Routing algorithm uses an inductive approach based on trial/error paradigm combined with swarm adaptive approaches to optimize three QoS different criteria: static cumulative cost path, dynamic residual bandwidth and end-to end delay. Based on KOQRA, the approach presented here adds a new module to this algorithm dealing with a third QoS criterion which takes into account the end-to-end residual bandwidth. In [7], authors use heuristics to determine a source-to-destination path that satisfies two or more additive constraints based on edge weights. [8] presented a polynomial time approximation algorithm for k multi-constrained path using a shortest path algorithm such as Dijkstra algorithm. In [9], authors proposed a randomized heuristic that employs two phases: 1) a shortest path is computed for each of the k QoS constraints as well as for a linear combination of all k constraints; 2) a randomized breadth-first search is performed for a k multi-constrained problem.

We can see that all of these approaches above do not take into account the perception and satisfaction of end-users. In other words, QoE concept is ignored. That poses the problem of choosing the best QoS metric that is often complex. However QoE comes directly from the use and represents the true criteria to optimize. In taking into account this lack, other proposals are presented in [10].

In [11], authors presented an overlay network for end-to-end QoE management. The purpose is QoE optimization by routing around failures in the IP network and optimizing the bandwidth usage on the last mile to the client. Components of overlay network are located both in the core and at the edge of the network. In [12], authors propose an extended version of the Optimized Link State Routing(OLSR) protocol. It uses fuzzy logic to build a fuzzy system that aims to optimize networks resources, solve the problem of using multiple metrics for routing and try to improve the user perception. However, these proposals do not use any adaptive mechanism. Furthermore, they do not consider QoE as a user feedback.

[13] presented a new adaptive mechanism to maximize the overall video quality at the client. Overlay path selection is dynamically done based on available bandwidth estimation, while the QoE is measured using PSQA tool, the same measurement tool we have used. After receiving a client demand, the video server chooses an initial strategy and an initial scheme to start the video streaming. Then, client uses PSQA to evaluate the QoE of the received video in real time and sends this feedback to server. After examining this feedback, the video server will decide to keep or to change its strategies. This approach has well considered end-users perception. However the adaptive mechanism is quite simple because it is not based on the learning method. Furthermore, the problem of this approach is the fact that authors use source routing.

Instead of trying to optimizing many QoS criteria like approaches above, our algorithm just based on user perception (QoE).

In the recent years, there are many researches, proposals that are made in order to measure, evaluate, and improve QoE in networks. Our work aimed to construct an adaptive routing method that can retrieve environment information and adapt to the environment changes. This adaptive routing mechanism maintaining the required QoE of end-users is very necessary for network systems that have great dynamics (i.e. unreliable communication) and multiple user profiles where the required QoE levels are different. For better user's perception, it is preferable that a routing protocol adapts itself to these QoE levels.

Our two proposals are routing systems driven by terminal QoE based on Reinforcement Learning (RL) concept [14] [15]. They aimed to maintain user satisfaction using QoE feedback of end-users. The first algorithm is based on Least Squares Policy Iteration (LSPI) [16], a RL technique that combines least squares function approximation with policy iteration. The second algorithm, an improvement of the first one, is based on Q-Learning [15] which is one of the RL algorithms. However, native Q-Learning taking into account rewards at all nodes in the system is inadequate to our target problem because the QoE reward is available only at the last node (QoE is evaluated at end-users). In order to improve the first algorithm, we evaluate the QoE at any node in the whole system. The QoE measurement is realized by using a hybrid between subjective and objective evaluation method called Pseudo Subjective Quality Assessment (PSQA) tool [17].

The paper is structured as follows: Section II describes the QoE measurement method we have used. We present our approaches in section III. The experimental results are shown in section IV. Paper is ended with conclusion and some future works in section V.

II. QoE MEASUREMENT METHOD

It is not easy to evaluate the perceived quality of a multimedia stream. The best way to assess it is to have real people do the assessment because quality is a very subjective concept. There are standard methods for organizing subjective quality assessments, such as the ITU-P.800 [18] recommendation for telephony, or the ITU-R BT.500-10 [19] for video. However, subjective evaluations are very expensive and cannot be a part of an automatic process. As subjective assessment is useless for real time evaluations, a significant research effort has been done to obtain similar evaluations by objective methods. The most commonly used objective measures for speech / audio are Signal-to-Noise Ratio (SNR), Segmental SNR (SNRseg),

 ISBN: 978-1-61208-133-5

Perceptual Speech Quality Measure (PSQM) [20], Measuring Normalizing Blocks (MNB) [21], ITU E–model [22], Enhanced Modified Bark Spectral Distortion (EMBSD) [23] and Perceptual Analysis Measurement System (PAMS) [24]. For video, some examples are the ITS' Video Quality Metric (VQM) [25], Color Moving Picture Quality Metric (CMPQM) [26], and Normalization Video Fidelity Metric (NVFM) [27]. These quality metrics often provide assessments that do not correlate well with human perception, and thus their use as a replacement of subjective tests is limited.

So we decide to use PSQA tool, a hybrid between subjective and objective evaluation. PSQA tool measures QoE in an automatic and efficient way, such that it can be done in real time. It consists of training a Random Neural Network (RNN) to behave as a human observer and to deliver a numerical evaluation of quality, which must be close to the average value that a set of real human observers would give to the received streams. PSQA method includes the following steps: a) Identifying a set of parameters having an important impact on the perceived quality, b) Building a platform allowing to send a video sequence through an IP connection, c) Performing a subjective testing experiment, d) Training the RNN.

PSQA used RNN for the learning phase. Sequences are input data of RNN that will give a real function as output. So for any configuration, the function returns a number close to the associated MOS (Mean Opinion Score)[1] value [28]. Our testbed using PSQA tool is described in detail in section IV. Using PSQA method gives us a function f expressed as:

$$f : R^3 \rightarrow R \quad (1)$$

This function f takes a combination of three parameter values as mentioned above (*delay time*, *loss rate* and *conditional loss rate*) to obtain a single output equivalent to the appropriate MOS score. So from now onwards the expression of "applying PSQA tool" means using function f in Equation 1 with three parameters as input to obtain the MOS score.

III. User Perception Based Routing System

Our idea to take into account end-to-end QoE consists to develop adaptive mechanisms that can retrieve the information from their environment (QoE) and adapt to initiate actions. The action choice should be executed in response to end-users feedback, in other words the QoE feedback. Concretely, the system integrates the QoE measurement in an evolutionary routing system in order to improve the user perception based on Q-Learning algorithm to choose the "best optimal QoE paths" (Fig. 1). So in that way, the routing process is build according to maintaining the best user perception.

In this section, we present our two approaches: our routing system driven by terminal QoE and his improvement, QQAR algorithm.

[1]Mean Opinion Score (MOS) gives a numerical indication of the perceived quality of the media received after being transmitted. MOS is expressed in a number, from 1 to 5, 1 being the worst and 5 the best. The MOS is generated by averaging the results of a set of standard, subjective tests where a number of users rate the service quality

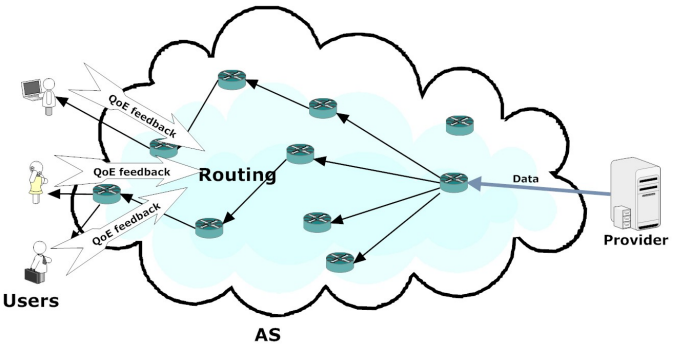


Fig. 1: Integration of QoE measurement in routing system

A. Routing system driven by terminal QoE

In order to integrate RL notion into our routing system, we have mapped RL model to our routing model in the context of learning routing strategy (Fig. 2). We consider each router in the system as a state. The states are arranged along the routing path. Furthermore, we consider each link emerging from a router as an action to choose. The system routing mechanism corresponds to the policy π. After data reach end-

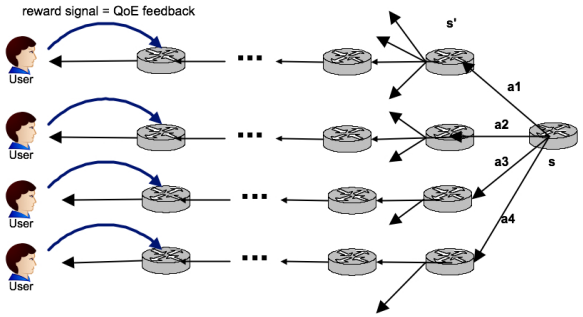


Fig. 2: Routing system based on reinforcement learning

users, QoE evaluation is realized to give a QoE feedback to the system. We consider this feedback as environment reward and our purpose is to improve the policy π using this QoE feedback. Concretely, the policy π is chosen so that it is equal to $argmax$ of action value function Q in policy π:

$$\pi_{t+1}(s_t) = argmax Q^{\pi_t}(s_t, a) \quad (2)$$

Least-Squares Policy Iteration (LSPI) [16] is a recently introduced reinforcement learning method. Our choice is based on the fact that this technique learns the weights of the linear functions, thus can update the Q-values based on the most updated information regarding the features. It does not need carefully tuning initial parameters (e.g., learning rate). Furthermore, LSPI converges faster with less samples than basic Q-learning.

In this technique, instead to calculate directly action-value function Q, this latter is approximated with a parametric function approximation. In other words, the value function is

 ISBN: 978-1-61208-133-5

approximated as a linear weighted combination:

$$\hat{Q}^{\pi}(s,a,\omega) = \sum_{i=1}^{k} \phi_i(s,a)\omega_i = \phi(s,a)^T\omega \quad (3)$$

where $\phi(s,a)$ is the basis features vector and ω is weight vector in the linear equation. The k basis functions represent characteristics of each state-action pair.
We have to update the weight vector ω to improve system policy.
Bellman equation and Eq 3 can be transformed to $\Phi\omega \approx R + \gamma P^{\pi}\Phi\omega$, where Φ represent the basis features for all state-action pairs. This equation is reformulated as:

$$\Phi^T(\Phi - \gamma P^{\pi}\Phi)\omega^{\pi} = \Phi^T R \quad (4)$$

Basing on equation 4, the weight ω of the linear functions in equation 3 is extracted:

$$\omega = (\Phi^T(\Phi - \gamma P^{\pi}\Phi))^{-1} \times \Phi^T R \quad (5)$$

For a router s and forwarding action a, s' is the corresponding neighbor router with $P(s'|s,a) = 1$. Our learning procedure is realized as follows: when a packet is forwarded from node s to s' by action a, which has been chosen by current Q-values $\phi(s,a)^T\omega$, a record $< s,a,s',\phi(s,a) >$ is inserted to the packet. The gathering process (Fig. 3) is realized until the packet arrives at the destination (end-users). Thus with

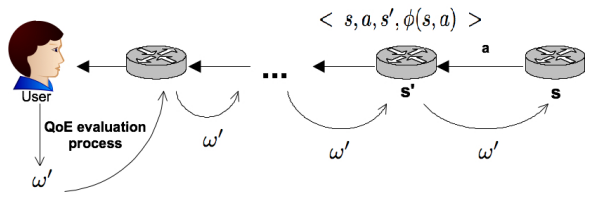


Fig. 3: Learning procedure

this way, one can trace the information of the whole routing path. At the end-users, a QoE evaluation process is realized to give a QoE score that is the value of *R* vector in equation 5. Furthermore, with the gathered information, the new value of ω is determined using equation 5. Then this new weight value ω' is sent back to the system along the routing path in order to improve policy procedure in each router on the routing path. With the new weights ω', policy improvement is realized in each router on the routing path by selecting the action a with the highest Q-value:

$$\pi(s|\omega') = argmax_a \phi(s,a)^T\omega' \quad (6)$$

The next subsection presents the improvement of this algorithm in using QoE measurement tool at all nodes of the routing system.

B. QQAR algorithm

In order to evaluate QoE in entire system, we have applied PSQA tool into all nodes (routers) including the last one representing end-user station (Fig. 4). In fact, measuring QoE anywhere in the system facilitates applying Q-Learning into our model with the reward at any node. That is the improvement factor of our first approach.
The proposed routing mechanism can be formulated as follows:

- *First step - Data packet flow*: the provider sends data packet to end-user. After receiving this data packet, each node in the routing path forwards the packet to the next node with a selection process presented in detail in subsection 2. It simultaneously evaluates QoE by using PSQA tool and saves this result.
- *Second step - At end-user side*: After data reach end-user, QoE evaluation is realized by using PSQA tool to give a QoE feedback as ACK message to the routing path that this data flow just passes through.
- *Third step - ACK message flow*: Each time a node receives a ACK message, it updates the Q-value of the link that this ACK message just passes through. The update process is introduced ci-below (subsection 1). It then attaches the QoE measurement result that it saved above into the ACK message and forwards it to the previous neighbor.

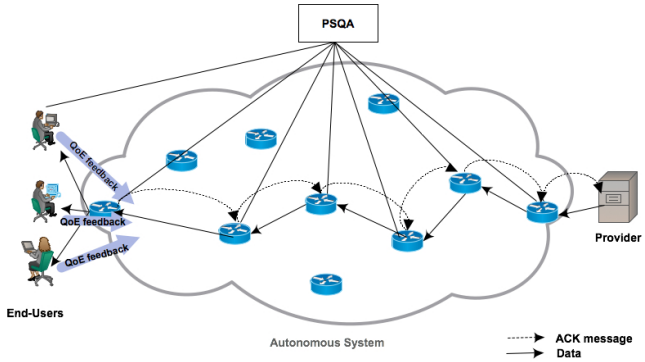


Fig. 4: QQAR routing system

With regard to a selection process in each node after receiving a data packet, we have to consider the tradeoff between *exploration* and *exploitation*. This tradeoff is one of the challenges that arises in RL and not in other kinds of learning. To obtain a lot of reward, a RL agent (router) must prefer actions that it has tried in the past and found to be effective in producing reward. But to discover such actions, it has to try actions that it has not selected before. The agent has to *exploit* what it already knows, but it also has to *explore* in order to make better action selection in the future. There are some mathematical issues to balance exploration and exploitation. In our approach, we choose softmax method as selection process that will be presented in the subsection 2.

1) Learning process: In our model, each router has a routing table that indicates the Q-values of links emerging from this router. For example in Fig. 5, node y has a routing table containing Q values: $Q_{yz_1}, Q_{yz_2}, Q_{yz_3}...Q_{yz_n}$ corresponding to n links from y to z_i with $i = 1..n$. Based on this routing table, the optimal routing path can be trivially constructed by a sequence of table look-up operations.

 ISBN: 978-1-61208-133-5

As mentioned above, after receiving a data packet, the last

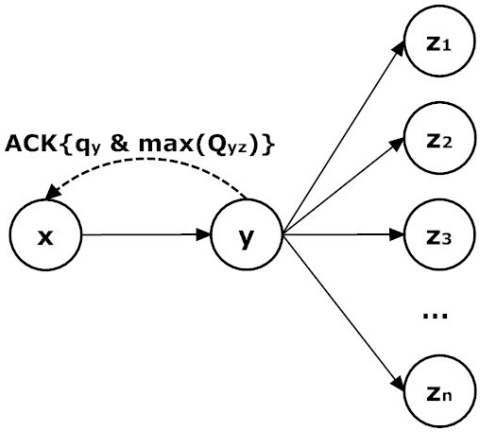


Fig. 5: Learning process

node representing end-user evaluates the QoE, it then sends back feedback as ACK message to the routing path. Each router x in this routing path receives this message containing information: the PSQA result of the previous router (q_y) and the maximum value of Q-values ($\max Q_{yz}$) in the routing table of node y (Fig. 5). Router x then updates the Q-value of the link connecting to y. Our update function based on the native Q-Learning (Eq **??**) is defined in Equation 7:

$$\underbrace{Q_{xy}}_{\text{new value}} = \underbrace{Q_{xy}}_{\text{old value}} + \alpha \Big[\underbrace{\beta\,(q_y - q_x) + \gamma \max_i Q_{yz_i}}_{\text{new estimation}} - \underbrace{Q_{xy}}_{\text{old value}} \Big] \tag{7}$$

Where Q_{xy} and Q_{yz_i} are Q-values of links xy and yz_i. q_x and q_y are results obtained (MOS score) by using PSQA tool at node x and y. α is the learning rate, which models the rate updating Q-value. The two discount factors β and γ balance the value between future reward and immediate reward.

2) Selection process: As mentioned above, our selection process must balance between the *exploration* and *exploitation* phase. It cannot always *exploit* the link that has the maximum Q-value because each link must be tried many times to reliably estimate its expected reward. Therefore, we have chosen softmax method using Boltzmann distribution [15]. So with this softmax action selection rules, after receiving a packet, node x chooses its neighbor y_k among its n neighbors y_i ($i = 1..n$) with probability presented in Equation 8:

$$p_{xy_k} = \frac{e^{\frac{Q_{xy_k}}{\tau}}}{\sum_{i=1}^{n} e^{\frac{Q_{xy_i}}{\tau}}} \quad (0 \le k \le 1) \tag{8}$$

Where Q_{xy_i} represents Q-value of link xy_i and τ represents a temperature parameter of Boltzmann distribution. High temperature causes the link selection to be all equi-probable. Low temperatures generates a greater difference in selection probability for links that differ in their Q-values. In other words, more we reduce the temperature τ, the more we exploit the system. In that way, we reduce the temperature τ after each time of forwarding packet as shown in Equation 9:

$$\tau = \phi \times \tau \quad (0 < \phi < 1) \tag{9}$$

where ϕ is the weight parameter.
So in that way, we initially balance *exploration* and *exploitation*. After that system is quite converged, we then increasingly exploit the system.

IV. Experiment

In order to validate our proposed approach, this section presents firstly our testbed for collecting dataset to PSQA tool. We then describe our simulation results using OPNET simulator.

A. Testbed for PSQA

Training RNN for PSQA tool needs a real dataset of the impact of the network on the perceived video quality. To construct this dataset, we conducted an experiment consist of selecting a number of human and asked him to score the perceived quality of video using the MOS score. The testbed (Fig. 6) is composed by client-server architecture and a network emulator. The client is VLC video client [29] and the server is VLC video streaming server [29]. The traffic between client and server is forwarded by the network emulator NetEm [30]. NetEm provides the way to reproduce a real network in a lab environment.

The current version of NetEm can emulate variable delay,

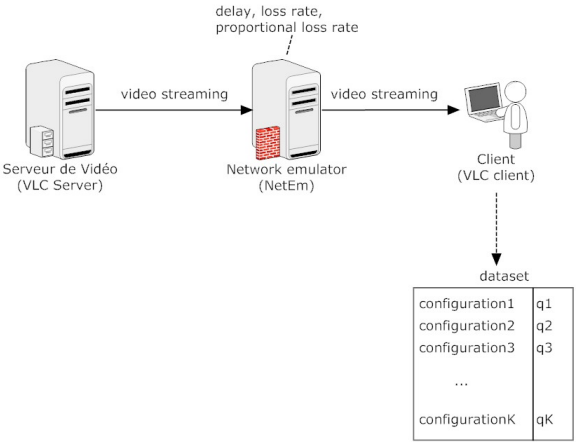


Fig. 6: Testbed for PSQA tool

loss, duplication and re-ordering.
The experimental setup consists on forwarding video traffic between server and client. Then, we introduce artificial fixed delay, variable delay and loss on the link to disturb the video signal.
According to ITU-R [31], the length of the video should be at least 5sec. We choose the sintel video trailer [32]. This video is of 52 seconds duration, 1280 x 720 dimensions and 24 frames per second cadence and uses H.264 codec. This video was chosen because it alternates high and slow movements.
Experiments were conducted with fixed delay values of [25, 50, 75, and 100ms], variable delays of [0, 2, 4, 6, 8, 16,

 ISBN: 978-1-61208-133-5

32ms], loss rate values of [0, 2, 4, 6, 10, 15, 20, 25, 30%] and successive loss probability of [0, 30, 60, 90%]. These values were chosen to cover the maximum of QoE range.
To collect data, we chooses viewers with a strong cinematic experience.
Nowadays, as a major part of monitors are LCD, we used the same ones. The particular screen used is a 19" size screen "LG flatron L194wt–SF" which has 1440x900 resolution.
The obtained dataset of this testbes is used for learning process of PSQA tool. We then obtain the function f in Equation 1 and apply the latter to our system as QoE measurement tool (PSQA).

B. Routing system simulation results

We have used Opnet simulator version 14.0 to implement our approach in an autonomous system. Regarding network topology, we have implemented an irregular network with 3 separated areas including 38 routers, each area is connected to each other by one link, all links are similar (Fig. 7). The network system is dynamically changed with an average period of 200 seconds.

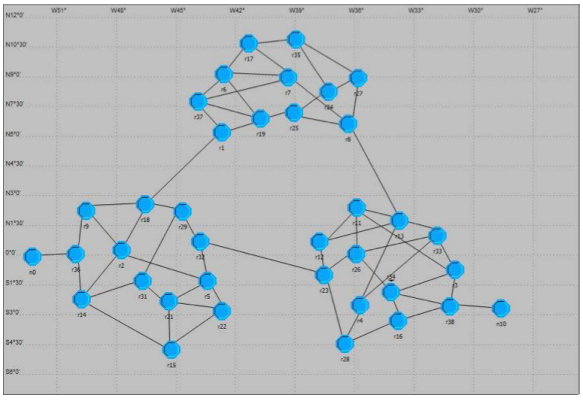
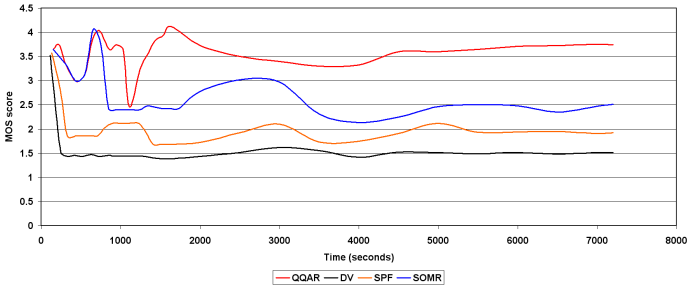


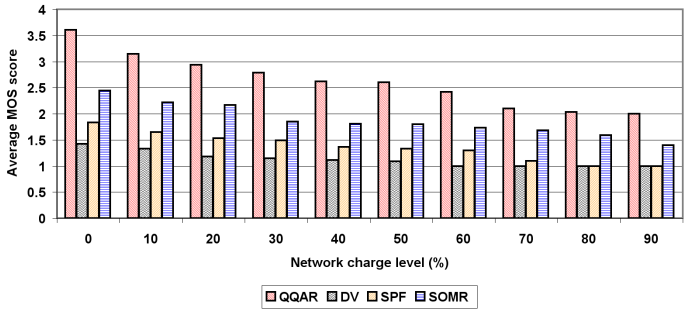
Fig. 7: Simulation network topology

To validate our results, we compare our approach with three kinds of algorithm:

- *Those based on Distance-Vector (DV) algorithm*. In this algorithm, routers can maintain the optimal route by storing the address of the next router in the routing table so that the number of hops to reach the destination is minimal. The roads are updated every 30 seconds.
- *Those based on Link-State algorithm: SPF (Short Path First)*. In this algorithm, each router establishes relations with its neighbors by sending hello messages. Each router then forwards the list of networks it is directly connected by messages (LSA-Link State Advertisements) to spread gradually to all routers in the network. The set of LSAs forms the database links Link-State Database (LSDB), which is identical for all participating routers. Each router then uses Dijkstra's algorithm to determine the shortest route to each network known in the LSDB.
- *Those based on Standard Optimal QoS Multi-Path Routing (SOMR) algorithm* where routing is based on finding K Best Optimal Paths and used a composite function to optimize delay and link cost criteria simultaneously.



(a) User perception in low load traffic network



(b) User perception in different load traffic levels of network

Fig. 8: User perception

Fig. 8 illustrates the result of average MOS score of four algorithms (non-charged network in 8a and different charge levels in 8b). In the charged system case, we have generated a traffic that stress the network. The charge level represents the rate of number of charged link and number of total link:

$$\text{level} = \frac{n_s}{N} \tag{10}$$

where n_s is number of charged links and N is number of total links in the system.
In observing Fig. 8a, we can see that results of all four algorithms fluctuate very much in the first 1500 seconds. That is explained by the execution of initialization process. In other words, these four algorithms do not have any information about the system in this first period, they try to explore it. In these first 1500 seconds, the MOS score of QQAR varies between 2.5 and 4, SOMR between 2.4 and 4.1, SPF between 1.7 and 3.6, DV between 1.4 and 3.5. After the first 1500 seconds, protocols gradually become stable. QQAR varies between 3.4 and 3.7. DV and SPF are quite stable with average results respectively 1.4 and 1.9. SOMR varies much more but the maximum value (obtained in period from 2100th to 2300th second) is still lower than QQAR.
Fig. 8b gives us the average of these four algorithms in different charge levels formulated in (10): from 10% to 90%. We can see that more the system is charged, the higher the average score decreases. However, at any charge level, the average MOS score of QQAR is better than three other algorithms. With a charge level lower than 10%, the MOS

 ISBN: 978-1-61208-133-5

score of QQAR is higher than 3 (in MOS score range, 3 represents a fair quality). Regarding the three other protocols, the maximum value obtained by SOMR is just 2.4 with charge level 10%.

QQAR gives a better e2e QoE perception in both cases than three other algorithms in the same delay. So with our approach, despite network environment changes, we can maintain a better QoE without any other e2e delay or any other QoS metric. Thus, QQAR is able to adapt its decisions rapidly in response to changes in the network dynamics.

Our experiment works consist also the survey of overheads met by these protocols. The type of overhead we observe is control overhead that is determined by the proportion of control packet number to the number of all packets emitted. To monitor this overhead value, we have varied node number in adding routers to network system. So the observed node numbers are [38, 50, 60, 70, 80]. The obtained result is showed in Fig. 9. We can see that the control overhead of

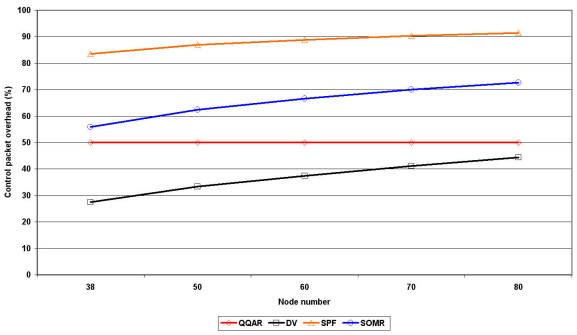


Fig. 9: Control overhead

our approach is constant (50%). That is explained by the equal of control packet number and data packet number in QQAR. Each generated data packet leads to an acknowledge packet generated by destination node. The control packet rates of DV, SPF and SOMR are respectively 0.03, 0.4 and 0.1 (packet/second). This order explains the control overhead order in Fig. 9. Whereas the SPF algorithm has the highest value because of the highest control packet rate (0.4 packet/second) with multiple type of packet such as Hello packet, Link State (LS) Acknowledgement packet, LS Update packet, LS State request packet, etc. , DV algorithm has the smallest overhead value with a control packet rate value of 0.03. We can see also that the higher the number of node, the higher the overhead is. So with a stable overhead, our approach is more scalable than these three others.

V. Conclusion

We present in this paper two approaches for routing systems driven by terminal QoE. We have integrated QoE measurement to routing paradigm for an adaptive and evolutionary system. Our second approach based on Q-Learning algorithm uses PSQA tool, a hybrid of subjective and objective method, for QoE evaluation. The simulations obtained demonstrates that our proposed approach yields significant QoE evaluation improvements over traditional approaches.

Finally, extensions to the framework for using these techniques across hybrid networks to achieve end-to-end QoE needs to be further investigated. Also, some future works includes applying our protocol to large scalable real testbed to verify its feasibility.

References

[1] "ITU-T recommendation E.800. Quality of telecommunication services: concepts, models, objectives and dependability planning. Terms and definitions related to the quality of telecommunication services," September 2008.

[2] Z. Wang and J. Crowcroft, "Quality of service routing for supporting multimedia applications," *IEEE Journal on selected areas in communications*, vol. 14, no. 7, pp. 1228–1234, 1996.

[3] J. A. Boyan and M. L. Littman, "Packet routing in dynamically changing networks: A reinforcement learning approach," *Advances in Neural Information Processing Systems*, p. 671, 1994.

[4] L. Peshkin and V. Savova, "Reinforcement learning for adaptive routing," in *Neural Networks, 2002. IJCNN'02. Proceedings of the 2002 International Joint Conference on Neural Networks*, vol. 2. IEEE, 2002, pp. 1825–1830.

[5] A. Mellouk, S. Hoceini, and S. Zeadally, "Design and performance analysis of an inductive qos routing algorithm," *Computer Communications*, vol. 32, no. 1371-1376, 2009.

[6] S. Hoceini, A. Mellouk, and B. Smail, "Average-Bandwidth Delay Q-Routing Adaptive Algorithm," in *ICC'08. IEEE International Conference on Communications*. IEEE, 2008, pp. 1840–1844.

[7] J. Jaffe, "Algorithms for finding paths with multiple constraints," *Networks*, vol. 14, no. 1, pp. 95–116, 1984.

[8] S. Sahni, *Data Structures, Algorithms and applications in C++*. Universities Press, 2005.

[9] B. Quoitin and S. Uhlig, "Modeling the routing of an autonomous system with C-BGP," *Network, IEEE*, vol. 19, no. 6, pp. 12–19, 2005.

[10] H. A. Tran and A. Mellouk, "Qoe model driven for network services," *Wired/Wireless Internet Communications*, pp. 264–277, 2010.

[11] B. D. Vleeschauwer, F. D. Turck, B. Dhoedt, P. Demeester, M. Wijnants, and W. Lamotte, "End-to-end qoe optimization through overlay network deployment," *International Conference on Information Networking*, 2008.

[12] R. Gomes, W. Junior, E. Cerqueira, and A. Abelem, "A QoE Fuzzy Routing Protocol for Wireless Mesh Networks," *Future Multimedia Networking*, pp. 1–12, 2010.

[13] G. Majd, V. Cesar, and K. Adlen, "An adaptive mechanism for multipath video streaming over video distribution network (vdn)," *First International Conference on Advances in Multimedia*, 2009.

[14] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *Journal of AI Research*, vol. 4, pp. 237–285, 1996.

[15] R. S. Sutton and A. G. Barto, "Reinforcement learning: An introduction," *IEEE transactions on neural networks*, vol. 9, 1998.

[16] M. G. Lagoudakis and R. Parr, "Least-squares policy iteration," *Journal of Machine Learning Research*, vol. 4, p. 1149, 2003.

[17] G. Rubino, "Quantifying the quality of audio and video transmissions over the internet: The psqa approach," *Design and operations of communication networks: a review of wired and wireless modeling and management challenges. Imperial College Press, London*, 2005.

[18] "ITU-T P.800. Methods for subjective determination of transmission quality - Series P: telephone transmission quality; methods for objective and subjective assessment of quality," August 1996.

[19] *Methodology for the Subjective Assessment of the Quality of Television Pictures, Recommendation ITU-R BT.500-10*, ITU Telecom. Standardization Sector of ITU, August 2000.

[20] J. Beerends and J. Stemerdink, "A perceptual audio quality measure based on a psychoacoustic sound representation," *JOURNAL-AUDIO ENGINEERING SOCIETY*, vol. 40, pp. 963–963, 1992.

[21] S. Voran, "Estimation of perceived speech quality using measuring normalizing blocks," in *Speech Coding For Telecommunications Proceeding, 1997, 1997 IEEE Workshop on*. IEEE, 2002, pp. 83–84.

[22] *ITU-T Recommendation G.107: The E-model, a computation model for use in transmission planning*, International Telecommunication Union, August 2008.

 ISBN: 978-1-61208-133-5

[23] W. Yang, "Enhanced modified bark spectral distortion (EMBSD): An objective speech quality measure based on audible distortion and cognition model," Ph.D. dissertation, Temple University, 1999.
[24] A. Rix, "Advances in objective quality assessment of speech over analogue and packet based networks," in *Data Compression: Methods and Implementations (Ref. No. 1999/150), IEE Colloquium on*. IET, 2002, p. 10.
[25] S. Voran, "The development of objective video quality measures that emulate human perception," in *Global Telecommunications Conference, 1991. GLOBECOM'91.'Countdown to the New Millennium. Featuring a Mini-Theme on: Personal Communications Services*. IEEE, 2002, pp. 1776–1781.
[26] C. van den Branden Lambrecht, "Color moving pictures quality metric," in *Image Processing, 1996. Proceedings., International Conference on*, vol. 1. IEEE, 2002, pp. 885–888.
[27] C. van den Lambrecht, "Perceptual models and architectures for video coding applications," Ph.D. dissertation, Ph. D. dissertation, EPFL, Switzerland, 1996.
[28] "Recommendation p.801: Mean opinion score (mos) terminology," ITU-T Rec P.801, 2006.
[29] Videolan. [Online]. Available: http://www.videolan.org/
[30] S. Hemminger, "Network emulation with netem," in *Linux Conf Au*, April 2005.
[31] "Recommendation 500-10: Methodology for the subjective assessment of the quality of television pictures," ITU-R Rec. BT.500, 2000.
[32] Sintel video trailer. [Online]. Available: http://www.sintel.org/

 ISBN: 978-1-61208-133-5

A Hierarchical Wireless Network Architecture for Building Automation and Control Systems

Mohammad Mostafizur Rahman Mozumdar*, Alberto Puggelli*, Alessandro Pinto†
Luciano Lavagno‡ and Alberto L. Sangiovanni-Vincentelli*
*EECS, University of California Berkeley, CA
{mozumdar,puggelli,alberto}@eecs.berkeley.edu
†United Technologies Research Center Inc. Berkeley, CA
alessandro.pinto@utrc.utc.com
‡Politecnico di Torino, Italy
lavagno@polito.it

***Abstract*—The building automation industry is experiencing a sudden increase in the complexity of control systems, mainly due to the push towards energy efficient buildings. These systems are necessarily distributed and rely on a communication network to gather data from sensors, produce intermediate results, and send commands to actuators. These networks should be cost effective, and should be flexible enough to be easily reconfigured if the building usage changes over the years. Wireless sensor and actuator networks are, therefore, key enablers. In this paper, we propose a hierarchical wireless network architecture for building automation and control systems and a protocol to manage it. We implement gradient based routing (for collecting data) and label switching table (for disseminating configuration commands), thereby supporting upstream and downstream data flows across the network.**

***Keywords*-Sensor Networks, Building Automation Systems, Implementation**

I. Introduction

Residential and commercial buildings in the United States are responsible for up to 73% of the total energy consumption [14]. To reduce the energy consumed by the building stock, both new constructions and existing buildings must be equipped with energy efficient solutions. These solutions are not only architectural, but rely on the use of advanced control algorithms that, based on measurements collected by sensors, compute an optimal control policy and send commands to actuators. Thus, the sensor-actuator network is a key element of building automation systems for energy efficiency. The selection of an optimal network is driven by several concerns including cost. Installation cost is one of the major concerns in building retrofits mainly due to wiring. For new constructions, although wiring is still a concern, the major problem is in making sure that the network is flexible enough to accommodate changes in the building usage that are common over its life-cycle.

For these reasons, wireless networks have been always considered as an interesting technology. Wireless sensor networks (WSN) find applications in factory and building automation, environmental monitoring, security systems and in a wide variety of commercial and military systems. However, wireless networks operation cost (which is mainly due to battery replacement) and low reliability have long been the major roadblocks to their adoption. Moreover, the whole-building control policy is often hierarchically structured: room-level controllers communicate to zone-level controllers, which in turn communicate to floor-level controller and so forth up to the building management system. This structure follows the geometry of the building. At each level, some computational power is required to execute control algorithms which is a challenge in the case of wireless sensor networks.

Based on these observations, we argue that the right networked system for building automation is in between a wired and a wireless solution. The first contribution of this work is the selection a design point that seems to be a good compromise between these two extremes. We distinguish between the power network and the data network. The proposed solution is entirely wireless for what concerns the transmission of data. However, we propose a hybrid approach for the power network where sensors are battery powered and are used only for measurement of physical quantities; actuators can be battery powered or not (a power source might be available close to them); and the rest of the components are powered by a wired network. These components provide also some computational power to execute control algorithms. This architecture responds to the needs of being able to reconfigure the sensing platform without major efforts, and to support the structure of a building control system, namely hierarchical and based on building geometry. The second contribution of this paper is a protocol to manage the network so that devices can be easily plugged into the systems, and the network is robust and reliable. The proposed protocol is based on a minimal set of APIs assumed to be provided by the lower layer of the stack. Thus, the proposed protocol can be implemented on other standard protocols such as Zigbee [11]. The third contribution of this paper is to illustrate the implementation details of

 ISBN: 978-1-61208-133-5

a hierarchical sensor network taking Building Automation System (BAS) as a context, so that the application engineers for BAS can use the proposed protocol implementation directly or can customize it according to their specifications.

The remainder of the paper is organized as follows: in Section 2, we review related works including existing standards, and we present some existing routing protocols that could be suitable candidates for building automation. In Section 3, we introduce our proposed architecture and in Section 4, we describe the network formation and maintenance protocol in details. We present the modeling and evaluation of the architecture in Section 5. In Section 6, we provide a set of conclusion and our future research directions.

II. Related Work

Building automation systems went through different phases of sophistication, from point-to-point centralized implementations to more complex networked systems. The implementation of these systems rely today mainly on wired networks. EIB or its successor KNX [10], LonWorks [13] and BACNET [12] are among the most used standard protocols.

Recent advancements in wireless and sensor technologies have pushed forward new emerging protocols and standards such as ZigBee, Z-Wave[15], WirelessHART [16], and others. In general, the deployments of WSN based systems is *convergecast* [9][2], in which a cloud of sensor nodes transmit data to one sink. Sensors are connected to the sink over a mesh or tree network. Since mesh networking is in general resource hungry, tree networks are typically used. To send configuration commands to sensor nodes, one of the most widely used mechanism is controlled *flooding* [4], [1] which is the simplest solution to implement but it has several drawbacks in sparse networks. A more complex solution is source routing [5], in which the sink defines the route of the command packet based on its routing table.

Little attention has been paid to defining wireless protocols and architectures that are suitable for building automation systems, which is one the main motivations of this work. We put particular emphasis on the network organization which could become the underlying layer of any wireless standard. Recently, the IETF work group ROLL [1] has selected gradient based routing for collecting data from sensor networks [6] and we envision that a customized variant of this technique could be a good fit for BAS.

In gradient based routing [3], the backbone network is initially formed by using controlled flooding. The first node that starts the network formation broadcasts a message by setting the gradient height to 1. Nodes that receive gradient messages set their gradient to the received value, and broadcast a gradient value which is equal to their gradient plus 1.

[1]The goal of the IETF ROLL is to standardize a routing protocol for wireless sensor networks.

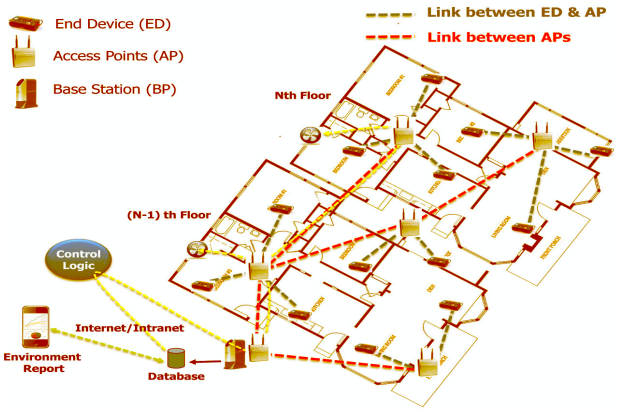


Figure 1. A snapshot of the BAS

This constructs a network organized into layers where the higher gradient nodes report to the lower gradient ones. After network formation, one of the main concerns is to maintain the gradient levels at each node. In fact, wireless links may appear or disappear at any time which may cause problems [8]. If a link disappears, then a node might have to select a new parent among a set of reachable ones. In this cases, a node always joins the network so to reduce its gradient level [7].

III. An Overview of the Architecture

An example instance of the proposed architecture is shown in figure 1. The network architecture is organized into a hierarchy of components that include end devices, access points, base station, and a server connected to a database. The role of each component in the network is the following:

- An **end device** is a sensor or an actuator. Both components can be battery powered. Having battery powered sensors give the flexibility of quickly relocating them. Each device has a *floor* and a *room* ID and can join in the network through any access point on the same floor (in a star configuration), which also provides for in-network load balancing. A sensor node could be configured for periodic or threshold crossing reporting depending on the quantity to be measured (e.g temperature, light and occupancy). To conserve battery power, sensor nodes should sleep most of the time and should wake up periodically to acquire data from sensors and transmit them to the associated access point. On the other hand, actuator nodes (which drive actuators), should periodically interrogate the access point for any pending command from the base station and then tune the actuator based on the received configuration.
- An **access point** is part of the backbone network that is used for data collection and command routing and is connected to direct power supply. These devices can be always on and capable of low power listening

 ISBN: 978-1-61208-133-5

to minimize energy consumption. Each device has a *floor* ID and a network-wide unique ID (for routing, as described later in this paper). These devices permit end devices to join the network and send data to collection points, construct aggregated packets and route them to the base station. The base station uses these nodes to configure sensors and actuators.

- A **Base station** has wireless connections with access points and an Ethernet connection for LAN or web access. A base station works as a master and initiates the formation of the backbone network. It collects the sensor data and logs them into the database which can be analyzed for trending and optimization. There could be one or more base stations for the whole system depending on the network size.

The architecture also includes server, storage database, control logic and web applications to control and analyze the behavior and data of the deployed networks, because of the space limitation we are excluding the details of these components from this paper.

IV. Network Formation and Maintenance

The network is formed and activated by following a series of phases. First, the backbone network is formed. The base station and the access points participate to this phase. Then, end devices join with access points. Each access point sends a report of the connected end devices to the base station. The control logic sends configuration commands to activate/deactivate sensors and actuators. After activation, end device sensors periodically report to the associated access point node which aggregates sensor data to construct a single packet that is routed to the base station. The following sections describe these phases in detail, including the issue of node failure.

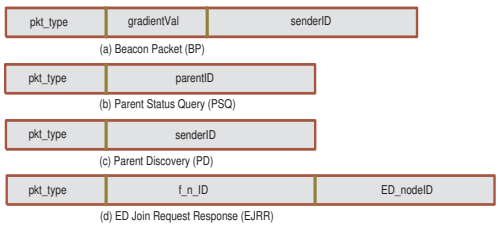


Figure 2. Packet format for network formation and maintenance

A. Backbone network between base station and access points

The main purpose of the backbone network is to support routing of messages in the network. We choose gradient-based routing to form the backbone. The formation of the backbone network is initiated by the base station which constructs and broadcasts the Beacon Packet (BP) shown in figure 2(a), with *gradientVal=1* and *senderID = BaseStationID*. Initially, all active access points set their parent to *NIL*, and the gradient to a very high value. Access points that are in the radio range of base station receive the BP and set their gradient to 1, and their parent to the *BaseStationID*. After receiving the BP, an access point updates its gradient if any of the following holds:

- It has no parent.
- It has a parent but BP is received from an access point node that has lower gradient.
- It receives the BP from its existing parent but the parent's gradient has been changed (this scenario might arise when an access point is added or deleted in the network, as described in detail later).

An access point updates the values of its *gradient* and *parent* taken data from *gradientVal* and *senderID* of the BP respectively. An access point always broadcasts the BP after updating its gradient and/or parent value. When broadcasting, an access node modifies the BP by incrementing the gradient by one, and by setting *senderID* to its own ID. While broadcasting the BP, the node waits for a random time and uses a simple CSMA/CA protocol at the MAC layer to reduce collisions. This process of controlled flooding continues until the backbone network is formed.

B. Tributary network between access points and end devices

After power-up, each end device constructs and broadcasts the EJRR (ED Join Request Response) packet shown in figure 2(d). While constructing the EJRR packet, end devices set *f_n_ID = floorID*, *ED_nodeID = nodeID*, and the value of *pkt_type*. The *pkt_type* field can take three different values (used for the end device joining process) that are listed below:

- pkt_type = requestVal, when the end device broadcasts a join request
- pkt_type = responseVal, when the access point sends a response to the end device
- pkt_type = confirmVal, when the end device sends confirmation of joining to the access point

Any access point node after receiving the EJRR packet checks the *f_n_ID*. If the end device joining request is coming from other floors it ignores the request, otherwise it modifies the EJRR packet by changing the value of packet type (responseVal) and by setting *f_n_ID = its nodeID* and then rebroadcasts it. An end device might get multiple responses from different access points, but proceeds with the first response and sends a confirmation EJRR packet by simply changing the value of packet type (confirmVal). The specified access point (by checking the *f_n_ID*) adds the end device into its children list and routes all data from it to the base station. It also sends the configuration packets from the base station to the end device.

C. Handling access points node failure

Since access points create the backbone network, and there is only one path from any access point / sensor

 ISBN: 978-1-61208-133-5

node to the base station, if a link becomes unavailable between two access points the network can be disconnected. That's why each access point checks its parent status at a periodic interval. After choosing a parent, each access point sends a PSQ (Parent Status Query) packet shown in figure 2(b) at random *nQT* interval. In every *nQT* interval, an access point constructs a PSQ packet with *pkt_type* = *psq_request*, *parentID* = *its parentID* and broadcasts it. After receiving the PSQ packet, the parent of the requesting access point sends a response by simply updating *pkt_type* = *psq_ response*. This response might be received by all or some children (access points) including the requesting one. The requesting node schedules an event at current time plus its *nQT* to check its parent status again, and all other child nodes, who have heard the response postpone their immediate PSQ and schedule another future event to check the status of their parent. If a child does not receive the PSQ response from its parent for three consecutive times, it sets its gradient/parentvalue to NIL and searches for a new parent.

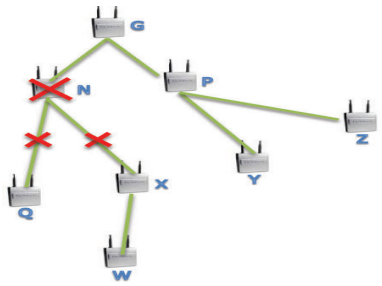


Figure 3. Example network setup for searching new parent

D. Searching for a new parent

Searching for a new parent is required when a new access point is added to the network and when the children of a dead access point search for a new parent, as described by the example shown in figure 3. Let us consider the case where access point node N dies and node X has failed to get a PSQ response for three consecutive times (same for node Q, but let us assume that the *nQT* timer of node X expires first). Node X constructs the PD (Parent Discovery) packet shown in figure 2 (c) with *senderID*=*X*. Suppose that this PD packet is received by P, Q, Y and W and they proceed as follows:

- A node after receiving a PD packet checks its *senderID* field first.
- If *senderID* = *Node's parentID* then the node does nothing (e.g. node W). This prevents loop formation.
- If *senderID != Node's parentID* then the node checks whether its parent is alive or not (by sending PSQ packet). If its parent is not alive, then the node does not do any further processing (e.g. node Q). This step prevents following a dead route.
- If the node's parent is alive, the node simply broadcasts a BP (e.g. nodes P and Y).

Node X might receive two BPs, either from P first, and then from Y, or vice versa.

CASE 1: First P then Y

- **BP from Node P:** Node X sets P as its parent and changes its gradient value. Since the beacon's gradient value is modified, node X broadcasts it. The beacon packet from X might be received by nodes Q and W. Since node Q is also looking for a parent, it sets X as its parent and broadcasts the beacon packet. Node W, after receiving the beacon packet form X, finds that the beacon is from its parent. If the gradient has been changed, node W modifies its gradient and broadcasts it again.
- **BP from Node Y:** Node X finds that the beacon packet contains a higher gradient value, so it does not modify its gradient value.

CASE 2: First Y then P

- **BP from Node Y :** Node X sets Y as its parent and changes its gradient value. Since the beacon's gradient value is modified, node X broadcasts it. The BP from X might be received by nodes Q and W. Nodes Q and W follows the same procedure as described before.
- **BP from Node P:** Node X finds that the beacon packet contains a lower gradient value, so it changes its parent to P and broadcasts the beacon packet. Nodes Q and W update their gradient value but their parent remains the same (Node X).

Both cases lead to the same network setup. Let us consider the scenario of node Q that is also looking for a new parent approximately at the same time as X.

- Node Q broadcasts the PD packet which might be received by node X and W.
- Suppose that node X is still in the process of finding its parent (currently it's parent is not alive, so it will not respond).
- Since Q is not W's parent and the parent of W (node X) is alive, W broadcasts the beacon packet which might be received by both parentless nodes Q and X.
- Both Q and X set their parent as W, which leads to formation of an isolated graph (it could also be a loop if X sets Q as a parent)
- When node X receives the BP either from node P or Y, the loop is broken and graph becomes connected again. A node needs to send the PD packet after knowing that its parent is not alive, even though it receives a BP earlier than sending the PD packet.

E. Upstream and downstream

Each access point collects sensor data from sensors and *upstream* flows route these data to the base station. Since

 ISBN: 978-1-61208-133-5

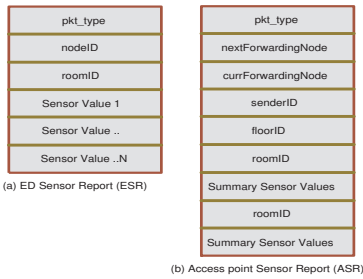


Figure 4. Packet formats for sensor data reporting and routing

each access point has a unique parent in the backbone tree and the base station is the root of that tree, data from each access point follows a unique path to the base station. Whenever a new end device (sensor/actuator) joins an access point, it immediately sends a notification report to the base station. If needed, the base station sends configuration commands to the end devices using the *downstream* flow. After configuration, each end device constructs ESR packet (shown in figure 4(a)) periodically or after crossing a specified threshold, and sends it to its access point. The ESR packet contains room identification and also collected sensor values. The access point node constructs an ASR packet (shown in figure 4(b)) at periodic intervals and sends it to the base station. While constructing the ASR packet, the access point node puts the following routing information in the header:

- nextForwardingNode = Its parent ID,
- currForwardingNode = Its NodeID
- sender = Its NodeID

Whenever an ASR packet flows along the upstream path to the base station from the originating access point, the intermediate access points change the value of *nextForwardingNode* and *currForwardingNode* accordingly and record the following information:

- Sender NodeID (sender field of the ASR packet)
- Immediate downstream router ID (currForwardingNode field of the ASR packet)

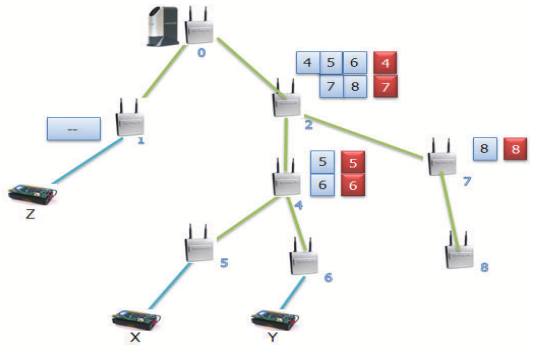


Figure 5. Upstream and downstream routing

These information are needed for downstream command/configuration packet routing. Let us consider the simple network setup shown in figure 5. In the case of upstream flow, node 5 sends its data to the base station through nodes 4 and 2. Nodes 2 and 4 update their downstream routing table for node 5. Whenever node 2 receives a command/configuration packet for node 5, it simply routes it to node 4, and node 4 routes it to node 5. The same process applies to all other intermediate routing nodes. In the up/down stream routing, each access point uses its parent node for upstream routing and the one-hop label switching routing table for downstream routing.

V. Modeling and Evaluation

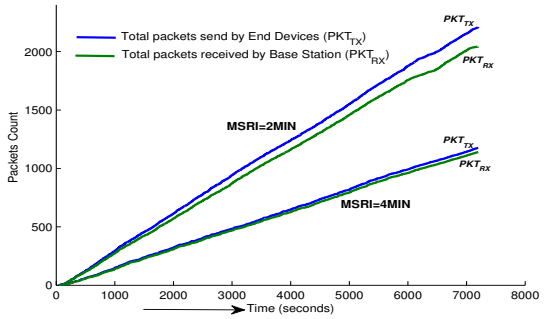


Figure 6. Packet loss count in the whole network

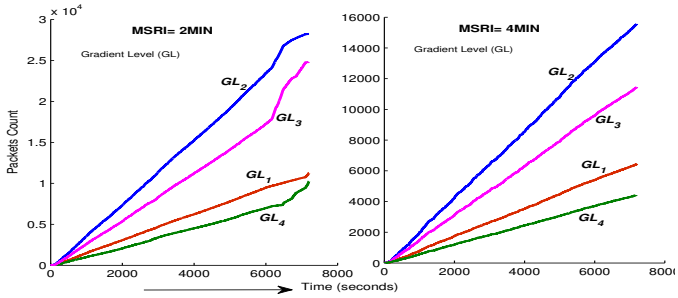


Figure 7. Traffic load at different gradient level

We developed an OPNET model for the proposed architecture. OPNET has three hierarchical component levels: the network level creates the topology of the network, the node level defines the behavior of the node and controls the flow of data between different functional elements inside the node, and the process level describes the underlying protocols by using finite state machines (FSMs). We developed three kinds of nodes, as described in our architecture, with the routing logic described in Sections III and IV. For the evaluation, we created the first setup with 1 Base station, 18 Access Points and 21 End device nodes deployed in a seven-storey building. We configured two scenarios for collecting

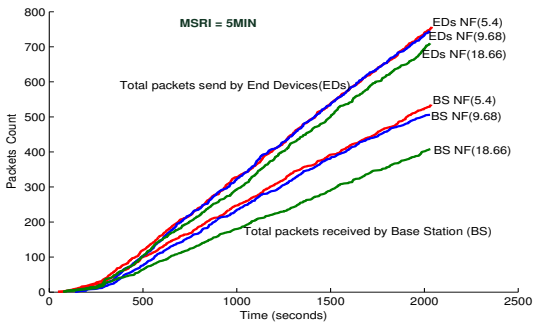


Figure 8. Packet transmissions and receptions at different Node Failure (NF) rate

data from end devices by setting the MSRI (Max Sensor Reporting Interval) at 2min and 4min respectively (The base station sends these configuration commands when the backbone network has been formed and end devices joined into the network). Thus, sensor reporting will be scheduled at every $uniform_{dist}(1.00) * MSRI$ interval. Figures 6 and 7 depict the packet loss and traffic load characteristics of the example network setup. The results show that at the higher interval (MSRI=4min) the network has less packet loss. To minimize the height of the backbone tree, we put the base station at the 4th floor (near the middle of the building), so the the backbone tree expands both up and down. This is reflected in figure 7 which depicts traffic loads (both sending and receiving packets) at different gradient levels. Since the base station (gradient value 1) most of the time receives packets but sends only a few configuration packets, its load is less than that of the 2nd and 3rd level nodes. To understand the effect of random node failures on the proposed architecture, we created another network setup with 1 Base station, 75 Access Points and 100 End device nodes deployed in a fifteen storey building. Figure 8 depicts the loss of packets at different node failure rates (NF (nodes/sec) = 5.4, 9.68 and 18.66). Here a node fails at any random time, then revives and joins back in the network. Figure 8 depicts that the packet delivery ratio remains relatively the same for NF rates up to 13% (NF=9.68) and afterwards it drops gradually. This means that the proposed architecture can tolerate 13% of access points failure in the most extreme case. Since the access points are directly connected to the power line, such high node failure rate is unlikely (we expect software/hardware malfunction to be rare).

VI. Summary and Conclusions

We presented a hierarchical wireless network architecture for building automation and control systems, and a protocol that efficiently solves the problems of forming and maintaining the network. The proposed architecture is hybrid with respect to power: some nodes are battery powered and others are plugged into a power supply. This hybrid approach provides a good trade-off between reliability and flexibility. We modeled the network components and protocol using OPNET. Our future work includes tools to generate application code for each node in the network, which entails generating drivers for the underlying low level protocol to implement the protocol primitives presented in Section IV. We will target standard platforms such as TinyOS and ZigBee.

References

[1] D. Braginsky and D. Estrin, Rumor routing algorthim for sensor networks, In *WSNA '02*, pages 22–31.

[2] T. Chen, H. Tsai, and C. Chu, Adjustable convergecast tree protocol for wireless sensor networks, *Computer Communications*, 33(5):559–570, 2010.

[3] J. Faruque, K. Psounis, and A. Helmy, Analysis of gradient-based routing protocols in sensor networks, In *DCOSS*, pages 258–275, 2005.

[4] C. Intanagonwiwat, R. Govindan, and D. Estrin, Directed diffusion: a scalable and robust communication paradigm for sensor networks, In *MobiCom '00*, pages 56–67.

[5] D. B. Johnson and D. A. Maltz, Dynamic source routing in ad hoc wireless networks, In *Mobile Computing*, pages 153–181, Kluwer Academic, 1996.

[6] R. D. Team, Rpl: Routing protocol for low power and lossy networks, In *IETF ROLL WG, IETF Internet-Draft*, 2009.

[7] T. Watteyne, I. Augé-Blum, M. Dohler, S. Ubéda, and D. Barthel, Centroid virtual coordinates - a novel near-shortest path routing paradigm, *Journal of Computer and Telecommunications Networking*, 53(10):1697–1711, 2009.

[8] T. Watteyne, K. Pister, D. Barthel, M. Dohler, and I. Auge-Blum, Implementation of gradient routing in wireless sensor networks, In *GLOBECOM*, pages 1–6, 2009.

[9] H. Zhang, A. Arora, Y. Choi, and M. G. Gouda, Reliable bursty convergecast in wireless sensor networks, In *MobiHoc '05*, pages 266–276.

[10] KNX Association, 2003, `http://www.knx.org/`

[11] Zigbee alliance, 2004 `www.zigbee.org/`

[12] BACnet, A data communication protocol for building automation and control networks, 2003, `www.bacnet.org/`

[13] Echelon, Lonworks technology overview, 2002, `http://www.echelon.com/`

[14] EIA, Annual energy outlook 2009, technical report, doe/eia-0383 (2009), In *US Department of Energy, 2009*.

[15] , Z-wave alliance, 2007, `www.z-wavealliance.org/`

[16] WirelessHART, Hart communication foundation, 2007, `http://www.hartcomm.org/`

 ISBN: 978-1-61208-133-5

An Efficient Scheduling Algorithm for Multiple MSSs in IEEE 802.16e Network

Wen-Hwa Liao and Chen Liu
Department of Information Management
Tatung University
Taipei, Taiwan
whliao@ttu.edu.tw

Sital Prasad Kedia
Samsung India Software Operations Pvt. Ltd.
Bangalore, India
sital.kedia@samsung.com

***Abstract*—The IEEE 802.16e standard introduced the concept of mobile subscriber stations (MSS) to provide mobility support. Five quality of service (QoS) classes have been defined to support QoS requirement different connections between the base station and the subscriber station. Out of these QoS classes, UGS has been designed to support real-time service flows that periodically generate fixed-size data packets. Most of the existing scheduling schemes for UGS class consider scheduling of a single MSS or even if they consider scheduling of multiple MSSs, the QoS requirement of the MSSs is not satisfied properly after scheduling. In this paper we have proposed a scheduling scheme, which schedules multiple MSSs with UGS connections so that QoS requirement of each MSS can be satisfied after scheduling.**

Keywords- IEEE 802.16e; scheduling; WiMAX; wireless.

I. INTRODUCTION

Wireless network is often considered as a cheaper and time saving alternative to its wired counterpart. In addition, some developing countries and uncivilized regions lack in infrastructures for deployment of wired network. To overcome the above-mentioned limitations of the wired network and to satisfy the huge demand for wireless services, worldwide interoperability for microwave access (WiMAX) was advocated as IEEE 802.16 wireless technology with high throughput over long distance (up to 30 miles). Later, the IEEE 802.16e standard, also known as Mobile WiMAX was proposed, which introduced the concpet of Mobile Subscriber Station (MSS) [1]. Basic WiMAX network includes a Base Station (BS) and several Subscriber Stations (SS) that are served by the BS. There are five QoS classes defined in 802.16e standard: UGS (Unsolicited Grant Service), rtPS (real-time Polling Service), ertPS (Extended Real-time Polling Service), nrtPS (non-real-time Polling Service), and BE (Best Effort). UGS is designed for services that periodically generate fixed-size data, such as T1/E1 and Voice over IP (VoIP). The BS assigns fixed grant to UGS connections. Hence, the MSSs need not send bandwidth request to BS every time they need to transmit data and thus saving the bandwidth used to send the bandwidth request to the BS.

There are several research works which have focused on optimizing the power consumption in IEEE 802.16e networks. The works in [3][4] apply the Chinese remainder theorem to decide the start time of each connections of an MSS. Due to different start time combination, the wake up time of the MSS is reduced. However, they didn't take into account the bandwidth used by each connection. So an MSS may need to handle more number of connections than it can in a certain time, which debase the feasibility of their approach. The goal of works in [5][9] is to minimize the wake up time of an MSS having multiple connections of different service classes. They gather the bursts of all the connections of different service classes and transmit these bursts together so that the wakeup time of the MSS can be reduced and thus saving significant amount of energy. Their approach is efficient but they didn't consider the multiple MSSs environment. Three energy efficiency scheduling algorithms for multiple MSSs were proposed in [6][7][8]. The work in [6] classifies MSSs into two catagories i.e., primary and secondary, based on their QoS requirement. A primary MSS is allowed to use the bandwidth in burst mode, whereas a secondary MSS is given the necessary bandwidth only to meet the requirement of its delay constraint. This approach can save significant amount of energy and also can avoid interference between the MSSs. However, in the real world environment, when the traffic load of all the MSSs is high, it is hard to classify the MSSs as primary and secondary. The work in [7] proposes a scheduling algorithm for mulitple MSSs environment. The algorithm gathers the bursts of all the connections in an MSS and transmit them together in order to minimize the wake up time of the MSSs. Minimum wake up time and multiple MSSs environment were considered for the first time in their work. In [8], authors have applied the Ford-Fulkerson algorithm to decide the time slots used by the MSSs. However, after applying the algorithm, the QoS requirement of the MSSs is not guranteed to be satisfied.

Most of the related works consider single MSS for scheduling, or even if they consider multiple MSSs, the QoS requirement of the MSSs is not satisfied after scheduling. Therefore, we have proposed a scheduling scheme to assign time slots used by multiple MSSs with UGS connection so that QoS requirement of each connection is satisfied after scheduling.

This paper is organized as follows. Section II discusses the scheduling scheme for multiple MSSs. Simulation results are presented in section III and Section IV concludes this paper.

 ISBN: 978-1-61208-133-5

II. SCHEDULING ALGORITHM FOR MULTIPLE MSSS

In the IEEE 802.16e networks, one BS may serve several MSSs simultaneously. However, at a particular timeslot, the BS can serve only one MSS [2]. If two or more MSSs are scheduled for data transmission in the same time slot, it will cause interference at the BS side. Therefore, an efficient scheduling algorithm is required, which can avoid the potential interference between the MSSs and also can satisfy the QoS requirements of all the MSSs. Next, we list some notations used in the rest of the paper in Table I and then we present our scheduling scheme.

TABLE I. NOTATIONS

U	Set of connections waiting for transmission.
T	The repeat cycle length.
G_i	Grant transmit interval of connection i.
I_i	Idle interval of connection i.
ST_i	Start time of connection i.
C_i	Cycle of connection i, equals to G_i+I_i.
W_i	Weight of connection i.
ω_i	Waiting time of connection i.
ω_{max}	Maximum waiting time of all the connections.
d_i	Delay constraint of connection i.
t	Target connection.

The scheduling algorithm for multiple MSSs has been given in Fig. 1. Initially, set U contains all the connections that are waiting to be scheduled. The first step of our scheduling algorithm is to select a connection from the set U having maximum weight as the target connection (t) for scheduling. Then we will check if there are sufficient empty slots to satisfy the bandwidth requirement of the target connection. If not, we will remove the target connection from U because its bandwidth requirement cannot be fulfilled in this round of scheduling, but it will be given higher priority for selection in the next round. After that, we will check if the target connection needs to be modified. If the target connection is modified then we need to verify whether the requirement for delay constraint is satisfied after modification. If not, then the target connection cannot be scheduled in this round. Connection modification may produce some surplus connections, which are scheduled only after all the connections in set U have been scheduled. However, we need to reserve the timeslots used to schedule these additional connections in advance after connection modification. We then determine the start time of the target connection and perform the connection separation if required. At last, we will schedule the target connection based on its start time, grant transmit interval, and the target connection is removed from the set U.

Algorithm: Scheduling algorithm for multiple MSSs

```
1:  for all the connections in set U
2:      Select a connection with maximum weight
        from U as target connection t.
3:      if the remaining slots are not enough
4:          goto step 16.
5:      endif
6:      if the target connection needs to be modified
7:          Perform connection modification.
8:          if the delay constraint is not satisfied
9:              goto step 16.
10:         endif
11:         Reserve timeslots for scheduling
            additional connections.
12:     endif
13:     Determine the start time of the target connection.
14:     Perform connection separation if needed.
15:     Schedule the target connection.
16:     Remove target connection from set U.
17: endfor
18: Schedule the additional connections resulted from
    connection modification.
```

Figure 1. Scheduling algorithm for multiple MSSs.

A. Weight of a Connection

$$W_i = \frac{\frac{\omega_i}{\omega_{max}} + \frac{I_i}{d_i}}{C_i} \quad (1)$$

Weight formula is given in (1). The weight of connection i (W_i) is associated with waiting ratio (ω_i/ω_{max}), delay constraint ratio (I_i/d_i), and length of the cycle (C_i). The weight of the connection increases with increase in the waiting ratio because a connection with longer waiting time should be served earlier. Similarly, when the delay constraint ratio is large, it means that the connection has stringent delay constraint, hence it should be given higher priority for scheduling. The length of cycle (C_i) is inversely proportional to the weight because we found out that smaller is the length of the cycle, more is the chance that the selected connection can be scheduled without any modification. The connection modification will be described later in this section.

B. Connection Modification

In this step of scheduling, we will check whether the target connection is having interference with any of the already scheduled connection. If it is, then the target connection needs to be modified. We transform the target connection into another UGS connection having different grant transmit interval and idle interval to avoid interference with the already scheduled connections.

 ISBN: 978-1-61208-133-5

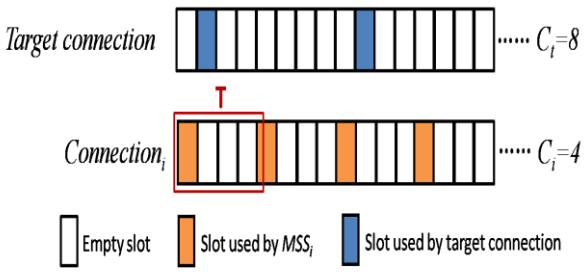


Figure 2. Example of the relationship between connections.

In Fig. 2, we assume that a connection i has already been scheduled, so the repeat cycle length T will be equal to the cycle (C_i) of connection i which is four in this case. Now we want to schedule the target connection t. We observe that if the cycle of target connection (C_t) is a multiple of T then t will have no interference with the already scheduled connections and hence it can be scheduled without any modification. Moreover, the new repeated cycle will be equal to C_t, and the grant transmit interval and idle interval of target connection will remain unchanged.

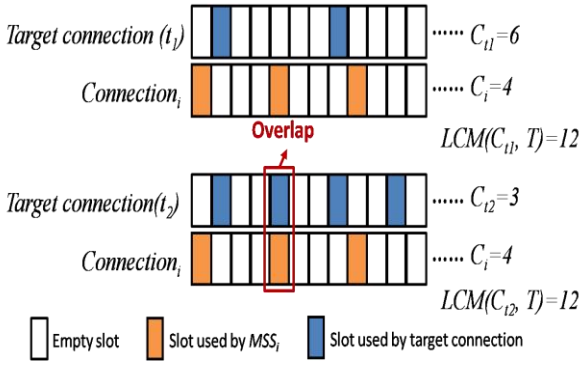


Figure 3. Interference between multiple connections.

However, if C_t is not a multiple of C_i, then the target connection may or may not overlap with the already scheduled connections. As shown in Fig. 3, C_t is not a multiple of C_i. In the first case, t_1 has no overlap with the already scheduled connection i, which means that t_1 can be scheduled without modification and the new repeated cycle T will be equal $LCM(T, C_{t1})$. However, in the second case, t_2 has an overlap with the already scheduled connection, so we need to modify the connection t_2. In general, let ST_t be the start time of the target connection t. If the $(ST_t + n*C_t)$th slot is not free, then the connection t has overlap with the already scheduled connections and it needs to be modified. Here n varies from 0 to $(LCM(T, C_{t1})/C_t) - 1$.

When a target connection needs to be modified, it is converted into another UGS connection having cycle (C_m) of length T. We use (2) to decide the new grant transmit interval (G_m) and idle interval (I_m) of the target connection.

$$G_m = G_t \times \left\lfloor \frac{T}{C_t} \right\rfloor$$
$$I_m = T - G_t \qquad (2)$$

Fig. 4 shows the modified connection for the target connection t_2 of Fig. 3. From (2), we can get G_m=1 and I_m=3 for the modified connection. We can see that t_2 has been converted to a new UGS connection having cycle of length 4 and it can be scheduled without any overlap. However, a surplus slot is left as shown in Fig. 4, which has to be put back as an additional connection to ensure the QoS requirement of the target connection. We see that the surplus slots are needed in each $LCM(T, C_t)$ slots. Thus we can calculate the length of the cycle(C_a), grant transmit interval (G_a) and idle interval (I_a) of the additional connection as given in (3). When the grant transmit interval of the additional connection is more than 1, we will further split it into G_a connections with grant transmit interval 1.

Figure 4. Modified connection for target connection t_2.

$$C_a = LCM(C_t, T)$$
$$G_a = G_t \times (C_a/C_t) - G_t \times \left\lfloor \frac{T}{C_t} \right\rfloor \times C_a/T \qquad (3)$$
$$I_a = C_a - G_a$$

C. Scheduling Additional Connections

Additional connections, which resulted from connection modification, are scheduled only after all the connections in set U are scheduled. However, we need to reserve the timeslots for these additional connections in advance. Let's consider the example shown in Fig. 5. The scheduling pattern has a cycle (T) of length 8 and the additional connection has a cycle (C_a) of length 12. Now suppose we select the third time slot as the start time of the additional connection, then the next timeslot occupied by the additional connection will be the 15th one. Here we can observe that the timeslots used by the additional connection is the third slot of each four timeslots which is equal to $GCD(8, 12)$. For example, the first timeslot used by the additional connection is the third slot of the first four slots and the second timeslot used is the third timeslot of fourth four slots. Thus we need to reserve one timeslot in each $GCD(T, C_a)$ cycle for scheduling of additional connections and the total number of timeslots that are reserved for scheduling of additional connection will be equal to $T/GCD(T, C_a)$.

 ISBN: 978-1-61208-133-5

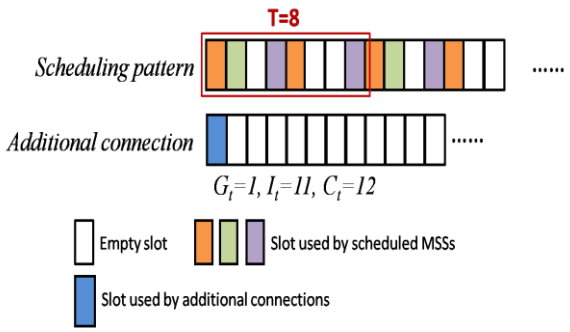


Figure 5. Scheduling additional connections.

Moreover, when more than one additional connections have the cycle of same length, they can use the same *GCD*(T, C_a) cycle. For example in the previous case, the additional connection used the first and fourth *GCD*(T, C_a) cycle, which means the second and third *GCD*(T, C_a) cycle is empty and can be used to schedule another additional connection having cycle of length C_a. The number of connections that can be scheduled together is given by C_a/*GCD*(T, C_a). For this reason, we split an additional connections into G_a connections with grant transmit interval 1 so that they can use the same *GCD*(T, C_a) cycle. It should be noted that in case the number of additional connection having cycle of length C_a is less then C_a/*GCD*(T, C_a), then some of the timeslots reserved for additional connections will remain empty, which results in wastage of bandwidth. We have considered this bandwidth waste as a parameter for performance evaluation in our simulation results.

D. Delay Constraint

For UGS connections, the idle interval between packet generation and packet transmission must be less than the predefined delay constraint. In our work, this idle period is the idle interval between two grant transmit intervals. Due to connection modification, the idle interval of the target connection may change. Hence, before scheduling the target connection, we have to check that the idle interval of the target connection is less than its predefined delay constraint. If not, it is not possible to satisfy the QoS requirements of the target connection, so it will not be scheduled in this round.

E. Assigning Start Time for Connections

For connections that are not modified, we will select the first empty slot of the cycle as the start time (ST_i) of those connections. The first empty slot should be larger than the summation of the grant transmit interval of all the connections scheduled so far. On the other hand, when assigning start time for the modified connections, we will find the last empty slot in the repeat cycle to meet requirements of the delay constraint. For example in Fig. 4, the start time of target connection will be the fourth timeslot. When we assign the last empty slot of repeat cycle as the start time, we can simply use the idle interval of target connection to check that the delay constraint is satisfied and ensure that the data needed to be sent is generated before the slot assigned to transmit it.

F. Connection Separation

In case the grant transmit interval of the target connection is more than the number of consecutive empty slots starting from ST_t, then timeslots assigned to the target connection will overlap with the already scheduled connections. For example in Fig. 6, G_t of the target connection is 4, while there are only 2 consecutive empty slots after timeslot 2 in the scheduling pattern. To solve this problem, we will split the target connection into two separate connections, connection 1 and connection 2, as shown in Fig. 6. The grant transmit interval (G_1) of connection 1 is the number of consecutive empty slots starting from ST_t and the idle interval (I_1) will be C_t–G_1. Connection 1 will be immediately scheduled and its start time (ST_1) will be the original start time (ST_t) of the target connection. The connection 2 will be treated as the new target connection whose grant transmit interval (G_2) will be equal to $G_t - G_1$ and idle interval(I_2) will be $C_t - G_2$.

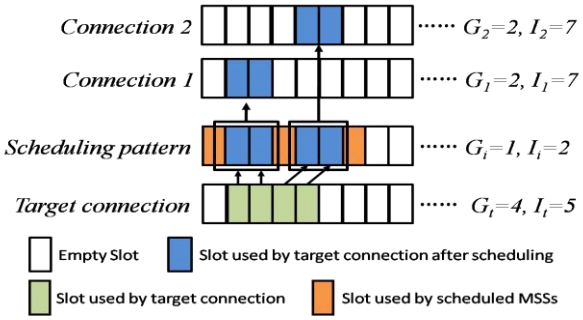


Figure 6. Connection separation.

III. Simulation Results

In this section, we have presented the simulation results to evaluate the performance of our proposed scheduling algorithm. As already discussed in section I, most of the related works consider single MSS for scheduling. Hence, it is not possible to compare the performance of our proposed algorithm with any of the related works. We have used bandwidth utilization, connections selection rate, and bandwidth waste for arranging additional connections as the parameters for performance evaluation. Bandwidth utilization is the percentage of used slots to the total number of slots. Connection selection rate is the rate at which connections are being selected for scheduling. Bandwidth waste, as discussed previously, is the wastage of the bandwidth incurred by reserving time slots for scheduling of additional connections. The bandwidth waste is given by reserved slots*number of additional connections that can be scheduled/number of additional connections actually scheduled.

A C-coded custom simulator is used to evaluate the performance of our scheduling algorithm. All the simulation results were obtained by running the scheduling algorithm for 50 times and then taking the average. The simulation parameters are listed in Table II.

 ISBN: 978-1-61208-133-5

TABLE II. SIMULATION PARAMETERS

Parameters	Value
Grant transmit interval	1-5
Idle interval	1-25
Number of connections	1-15
Ratio of grant transmit	1:1 – 1:10
Delay constraint	2 * Idle interval

Fig. 7 and Fig. 8 present the bandwidth utilization and connection selection rate for varying number of connections and varying ratio of grant transmit interval to idle interval. The number of connection varies from 1 to 10 and the ratio of grant transmit interval to idle interval varies from 1:1 and 1:10.

In Fig. 7, the bandwidth utilization increases with increase in the number of connections and it reaches up to 90% when the number of connection is 10. This is because of the fact that, with increase in number of connections, more data is available for transmission and the bandwidth utilization increases. Similarly, when ratio of grant transmit interval to idle interval increases, the bandwidth utilization is increased due to the same reason.

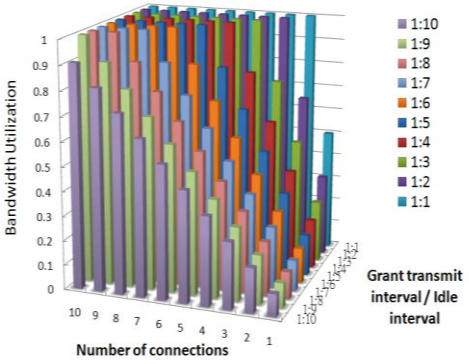


Figure 7. Bandwidth utilization.

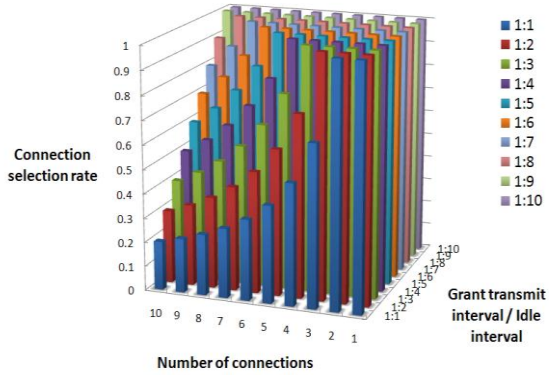


Figure 8. Connection selection rate.

In Fig. 8, the connection selection rate decreases as the number of connections increases because more is the number of connections, lesser is the chance for a connection to be selected. The connection selection rate also decreases with increase in the ratio of grant transmit interval to idle interval due to increase in amount of data for transmission.

We have also considered the parameters of UGS connection for simulation. The UGS parameters are listed in Table III. We classified UGS connections into three classes i.e., class A, class B, and class C having packet size 32 bytes, 64 bytes, and 128 bytes respectively.

TABLE III. UGS SIMULATION PARAMETERS

Parameters	Value
Packet Size	32, 64, 128 bytes
Frame duration	5 ms
Slot duration	0.1 ms
Data rate	30 kbps
Idle interval	10 – 50 ms
Delay constraint	150 ms – 400 ms
Number of connections	1-15

The simulation results of bandwidth utilization and connection selection rate for varying UGS parameters have been presented in Fig. 9 and Fig. 10 respectively. In Fig. 9, the bandwidth utilization for class A and class B is less than that for others because of smaller packet size of class A and class B. On the other hand, the bandwidth utilization for class B and class C is highest because of their large packet size. As shown in Fig. 10, the connection selection rate is highest for class A and class B. We can observe that the connection selection rate decreases with increase in the packet size and is lowest for class B and class C.

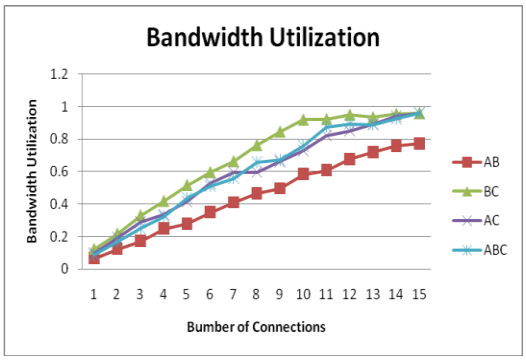


Figure 9. Bandwidth utilization with UGS parameters.

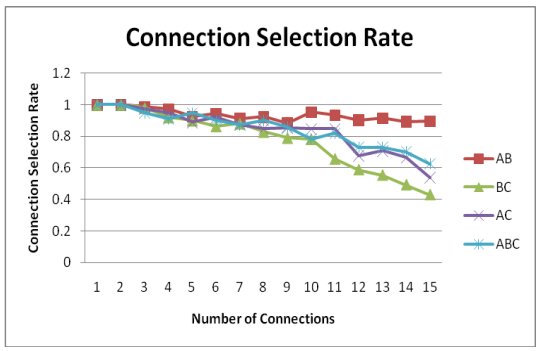


Figure 10. Connection selection rate with UGS parameters.

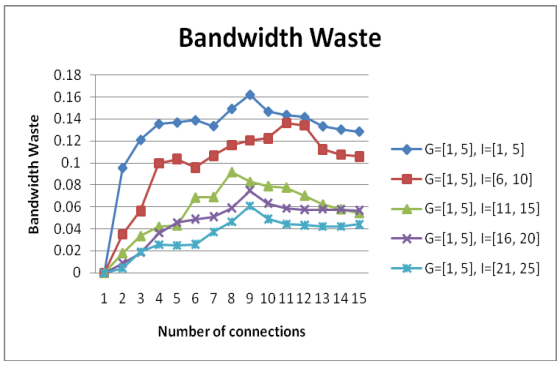


Figure 11. Bandwidth waste.

Fig. 11 shows the bandwidth waste for different values of ideal interval and number of connections. The grant transmit interval value is set to a random number between 1 to 5. Initially, the bandwidth waste increases with increase in the number of connections but it tends to decrease as the number of connections increases further. This is because of the fact that initially with increase in the number of connections, more number of additional connections are produced and we have to reserve more timeslots to schedule these additional connections. Hence, the bandwidth waste increases. However, when the number of connections increases further, the chances that additional connections can be scheduled together increases, which result in decrease in the bandwidth waste. Thus, our approach will not have large bandwidth waste when number of connections is large.

IV. CONCLUSION

In this paper, we proposed a scheduling algorithm for multiple MSSs with UGS connection in IEEE 802.16e networks. Our approach can avoid the potential interference between the MSSs and satisfy the QoS requirement of the MSSs after scheduling. The simulation results show that our approach can achieve more than 90% bandwidth utilization and it will not have large bandwidth waste when number of connections is large. Thus, our approach can achieve good performance and scalability in the real world environment.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Council, Taiwan, under grant NSC 99-2221-E-036-036-MY2, and Tatung University, under grant B99-N03-065.

REFERENCES

[1] IEEE Standard 802.16e-2005, "IEEE standard for local and metropolitan area networks part 16: air interface for fixed and mobile broadband wireless access system," Proceedings of IEEE Std., Feb. 2006.

[2] S.-I. Chakchai, R. Jain, and A.-K. Tamimi, "Scheduling in IEEE 802.16e mobile WiMAX networks: key issues and a survey," IEEE Journal on Selected Areas in Communications, Vol. 27, No. 2, pp. 156-171, February 2009.

[3] T.-C. Chen, Y.-Y. Chen, and J.-C. Chen, "An efficient energy saving mechanism for IEEE 802.16e wireless MANs," IEEE Transactions on Wireless Communications, Vol. 7, No. 10, pp. 3708-3712, October 2008.

[4] T.-C. Chen, J.-C. Chen, and Y.-Y. Chen, "Maximizing unavailability interval for energy saving in IEEE 802.16e wireless MANs," IEEE Transactions on Mobile Computing, Vol. 8, No. 4, pp. 475-487, April 2009.

[5] C.-Y. Lin, H.-L. Chao, Y.-J. Liao, and T.-J. Tsai, "Least-awake-slot scheduling with delay guarantee for IEEE 802.16e broadband wireless access networks," IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), 2008.

[6] G. Fang, E. Dutkiewicz, Y. Sun, J. Zhou, J. Shi, and Z. Li, "Improving mobile station energy efficiency in IEEE 802.16e WMAN by burst scheduling," IEEE GLOBECOM, 2006.

[7] S. C. Huang, R. H. Jan, and C. Chen, "Energy efficient scheduling with QoS guarantee for IEEE 802.16e broadband wireless access networks," International Wireless Communications and Mobile Computing Conference (IWCMC), 2007.

[8] S.-C. Huang, C. Chen, R.-H. Jan, and C.-C. Hsieh, "An energy-efficient scheduling for multiple MSSs in IEEE 802.16e broadband wireless," IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), 2008.

[9] C.-Y. Lin and H.-L. Chao, "Energy-saving scheduling in IEEE 802.16e networks," IEEE Conference on Local Computer Networks (LCN), 2008.

 ISBN: 978-1-61208-133-5

Data Gathering System for Watering and Gas Pipelines Using Wireless Sensor Networks

Radosveta Sokullu
Department of Electrical & Electronic Engineering
Ege University
Bornova IZMIR 35100 TURKEY
radosveta.sokullu@ege.edu.tr

Mustafa Alper Akkaş
Department of Electrical & Electronic Engineering
Ege University
Bornova IZMIR 35100 TURKEY
alper.akkas@ege.edu.tr

Fahrettin Demirel
2nd Regional Directorate of State Hydraulic Works
Bornova IZMIR 35100 TURKEY
fahrettind@dsi.gov.tr

***Abstract*—In this paper we discuss a small size experimentally implemented pipeline watering system, in which the pressure sensors are placed at carefully selected points. In our prototype, the TelosB motes, integrated with the pressure sensors, communicate in real time with each other and also with the base station according to the specifically designed wireless network protocol. The reason to choose a WSN as an infrastructure in our study is to let the system work without the need of extra cabling. Such a system incorporates both efficiency and flexibility, and provides the users with an automatic controlled water/gas system based on the sensor data. In our work, data was continuously measured using a pressure sensor and transferred to a central monitoring station via IEEE 802.15.4 wireless sensor network for storage and display. TelosB wireless motes were programmed with nesC and a graphical user interface was used to capture and display incoming measurements for all selected points being monitored. Evaluation of the system was done based on the WSN performance criteria (packet loss and network lifetime) and also based on the accuracy of the collected pressure data. In both respects the system performed very satisfactory and has been successfully implemented.**

Keywords: - wireless sensor networks; remote monitoring;nesC; TinyOS; water pipeline; TelosB; pressure sensor.

I. INTRODUCTION

The increase in the processing and integration capacity of electronic devices, as well as the advances of low power wireless communications have enabled the development of unwired intelligent sensors for a wide set of applications. The advent of small, low-cost and power efficient wireless sensor hardware is driving the development of applications in different industrial sectors, remote process control and also agriculture. Of particular interest here is the ability to remotely monitor water and gas pipeline pressure data. Efficient and accurate use of water resources has become very significant especially in recent time due to the global warming issues. Least but not last, the watering pipelines used in agriculture fields are to be controlled according to their water use. If pressure sensors are used in the existing pipelines of watering systems data can be gathered about the operation of the system.

In this work we present the experimental system design for remote monitoring the pressure of a watering and gas pipeline system. The idea is to investigate the possibilities to support farmers and organizations involved in continuous monitoring and operation of water and gas pipelines. The novelty of this work is in two aspects: first it uses pressure information to both control the performance in the systems and discover leaks and failures; a new, simple but very efficient static cluster tree routing protocol is defined and experimentally tests that can be used with such systems.

From here on the paper is organized as follows: in the next section we outline the characteristics of similar projects done before. In Section III, the developed system is described. In Section IV, we concentrate on the tested and the experimental results, and in Section V, we conclude the paper.

II. RELATED WORK

Min Lin et al. [1], suggested an interesting application for wireless sensor networks, specifically a water distribution network monitoring system. They propose a possible communication model for the water distribution monitoring network, and describe the channel measurement approach for the determination of an appropriate path-loss model. The accuracy of the proposed measurement approach has been confirmed using the flat earth two-ray model [1].

Yiming Zhou et al. [2], proposed a wireless solution for intelligent field irrigation system dedicated to Jew's-ear planting in Lishui, Zhejiang, China, based on ZigBee technology. Instead of conventional wired connection, the wireless design made the system easy to install and maintain.

 ISBN: 978-1-61208-133-5

The hardware architecture and software algorithm of wireless sensor/actuator node and portable controller, acting as the end device and coordinator in ZigBee wireless sensor network respectively, were elaborated in detail.

Yunseop (James) Kim et al. [3], describe details of the design and instrumentation of a wireless sensor network for variable rate irrigation and software for real-time in-field sensing and control of a site-specific precision linear-move irrigation system. Field conditions were site-specifically monitored by six in-field sensor stations distributed across the field based on a soil property map, and periodically sampled and wirelessly transmitted to a base station. An irrigation machine was converted to be electronically controlled by a programming logic controller that updates geo referenced location of sprinklers from a differential Global Positioning System (GPS) and wirelessly communicates with a computer at the base station. Communication signals from the sensor network and irrigation controller to the base station were interfaced using low-cost Bluetooth wireless radio communication. Graphic user interface-based software was developed.

Ivan Stoianov et al. [4], discuss how wireless sensor networks can increase the spatial and temporal resolution of operational data from pipeline infrastructures and thus address the challenge of near real-time monitoring and control. They focus on the use of WSNs for monitoring large diameter bulk-water transmission pipelines. They outline PipeNet, a system they have been developing for collecting hydraulic and acoustic/vibration data at high sampling rates as well as algorithms for analyzing this data to detect and locate leaks. Challenges include sampling at high data rates, maintaining aggressive duty cycles, and ensuring tightly time synchronized data collection, all under a strict power budget. They have carried out an extensive field trial with Boston Water and Sewer Commission in order to evaluate some of the critical components of PipeNet.

Liting Cao et al. [5], a remote real time AMR (Automatic Meter Reading) system based on wireless sensor networks is presented. The useful remote AMR sensors are analyzed and an efficient wireless network structure is suggested. The remote measurement system for water supply is taken as a typical example in their experiments. The structure of system employs distributed wireless sensors and consists of measure meters, sensor nodes, data collectors, server and a wireless communication network.

In their work Baoding Zhang et al. [6], introduce a design schema of wireless smart water meter based on the study of existing water meters. The main communication is based on Zigbee. The design is appropriate for modern water management and the efficiency can be improved.

Most of the studies mentioned above imply limited flexibility and reconfigurability. They are expensive to develop since they do not use off-the-shelf solutions to implement wireless sensor networking in a power-efficient and user-friendly way. With the proliferation and standardization of wireless sensor devices the trend is towards simpler and much cheaper solutions based on standardized nodes and networks.

In this study, pressure sensors and IEEE 802.15.4 compliant wireless modules are used to implement a mesh network and water distributing pressure data are stored and displayed on a PC connected to local gateway or base station. The proposed system has the advantage of simplicity, scalability and modularity over other alternatives. It is fully based on off-the-shelf components and freely available hardware and allows easy and reasonably priced setup and tailoring [7, 8]. Even though very similar in principle to others described in literature, our system is more robust, more cost effective and the provided user interface is more elaborate and flexible.

III. SYSTEM DESCRIPTION

A. *System architecture*

A conceptual view of the system is shown in Fig. 1. Each TelosB mote is connected to a remote monitoring system, which allows the observer to track the pressure. The readings are transmitted wirelessly from the specific points on the pipelines through an infrastructure of routing nodes to a central monitoring system (the base station). Depending on the pipeline's distance from the base station, messages can pass through multiple nodes to reach the base station. The base station is connected to a host computer running Moteiv Trawler to interpret, store and display the collected data. There are three main subsystems involved: wireless network structure, data measurement subsystem and the base station with its graphical interface.

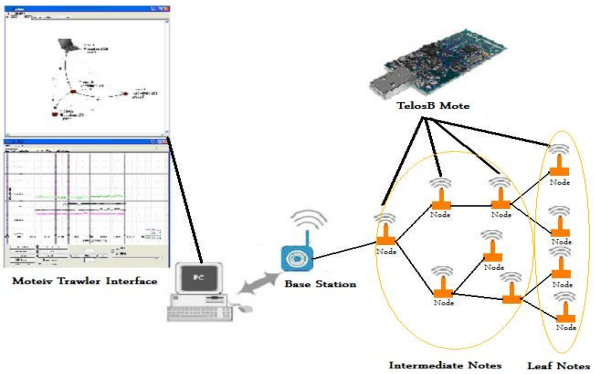


Figure 1. General System Architecture

B. *Freescale Mp3v5050gp Pressure Sensor*

The MPxx5050 [9] is a family of pressure sensors integrating on-chip, bipolar op amp circuitry and thin film resistor networks to provide a high output signal and temperature compensation. (Fig. 2)

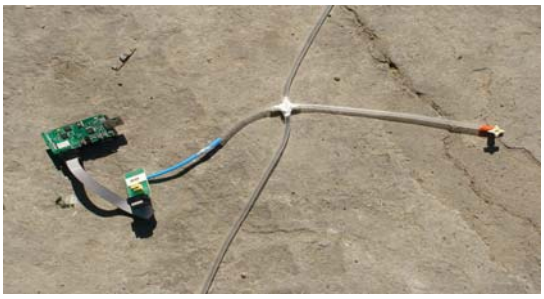


Figure 2. Mp3v5050gp Pressure Sensor Connected to TelosB Mote

C. *Wireless Module and Software*

This wireless communication platform uses TelosB wireless motes programmed in nesC and the Freescale Mp3v5050gp sensors are externally connected to the motes for collecting pressure data from the selected points on the pipelines that is to be transferred to a central database over 802.15.4 based wireless network. The major reason for choosing TelosB, off-the-shelf wireless module, is the ease of programming as well as the low power consumption. The software platform is based on TinyOS [10], an open-source operating system designed for wireless embedded sensor networks and the motes are programmed using nesC, an extension of C. It features a component-based architecture, which enables rapid implementation while minimizing code size [11].

IV. TESTBED SETUP AND EXPERIMENTAL RESULTS

A. *Prototype System Description*

The architecture described above has been implemented and tested in real environment. The network structure is a static hierarchical tree where the motes are mounted on predefined places on the pipes. Such deployment allows for collecting data from specifically selected points as well as precise location of problems. There are two types of nodes defined based on their functionality: leaf nodes and intermediate (collector) nodes. (Fig. 1) Leaf nodes transmit only readings from their own sensors while intermediate nodes transmit both their own and other nodes information doing aggregation when necessary. The base node collects all the data and transfers it to the PC via USB.

The network topology is static hierarchical tree and is given on Fig. 3. Dark colored nodes are leaf nodes, red colored nodes are intermediate nodes doing aggregation and the blue one is the base node. Both leaf and intermediate nodes do sampling and threshold comparison but while leaf nodes transmit only their own data, intermediate nodes also

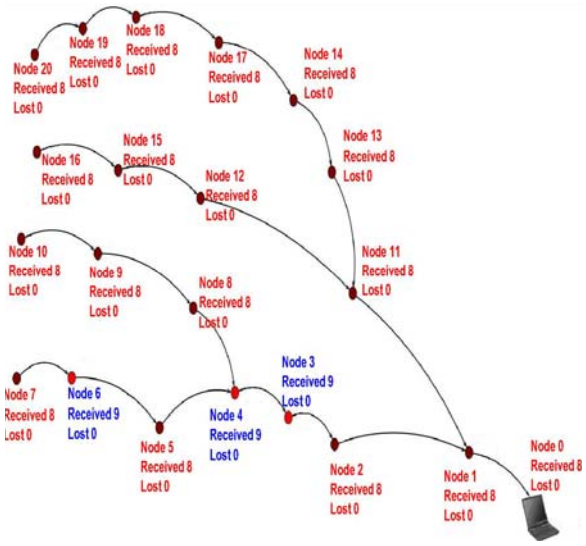


Figure 3. Schematic Network Tree Topology

transmit data from other nodes and if necessary do aggregation. The base node does not do any sampling but only transmits collected data.

As it is well known a major source of energy depletion for a wireless node is receiving and transmitting. In order to prevent this and provide longer lifetime two thresholds were introduced. As long as the sensor readings stay within these thresholds the node would be sending data at longer intervals (1 – 2 min). However, as soon as the reading falls outside these thresholds the data is transmitted in an order of seconds. A flow chart of the protocol operation is provided in Fig. 4.

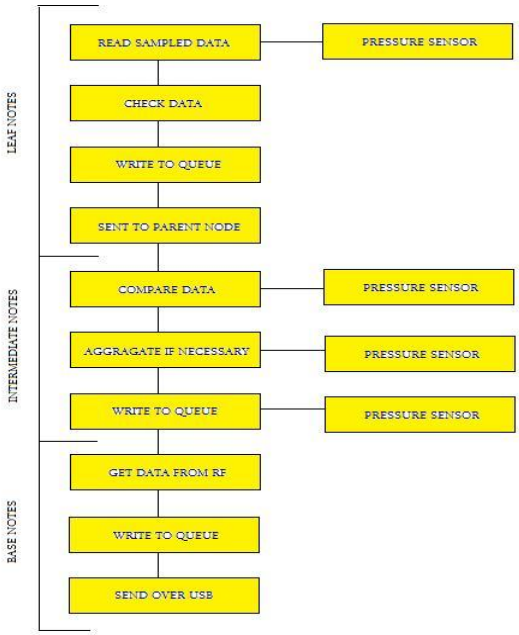


Figure 4. Flow Chart of the Network Protocol

 ISBN: 978-1-61208-133-5

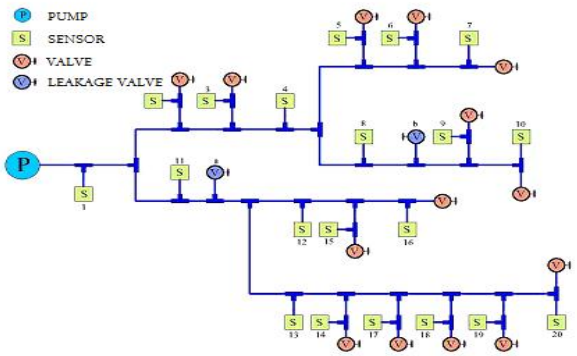


Figure 5. Schematic Deployment Plan

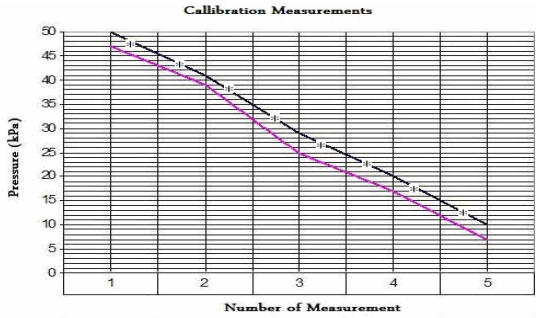
For safety reasons test were done using pressurized air instead of gas or water. The test setup included 20 sensors and 20 nodes plus the base node, equipment for providing pressurized air and plastic pipes as well as equipment to simulate user connection points (marked in red) and controlled pressure leaks (marked in blue). Schematic deployment plan is provided on Fig. 5. Additional filter elements (including a low pass filter) were used in the connection of the sensors to the TelsoB motes.

The test setup covers an area of approximately 200 m^2 with motes placed at distances from 1.5 to 5 m apart. In reality this will represent a very small farmer's field.

B. Testbed Setup and Operation

Using this experimental setup three groups of tests were carried out: calibration tests, user oriented tests and network performance tests. Calibration tests were carried out with the aim to prove the correctness of the pressure samples taken. User oriented tests were carried out to verify the correctness of the collected data as well as the system's proper operation related to the two threshold and its ability to diagnose leakage and other problems.

1) System Calibration: Before testing the performance of the system in real time the pressure sensor measurements were compared with analog manometer measurements. Results are given in Fig. 6. and show a constant minimal difference of 3 kPa which is due to the analog nature of the manometer. As the difference proved linear and quite small it was assumed that it does not influence the performance of the system.



Manometre	50	41	29	20	10
Freescale mp3v5050	47	39	25	17	7

Figure 6. Callibration Measurements

2) User Oriented Tests: These set of measurements aimed to test the general user perceived performance of the network. Using the additionally mounted equipment, different users' behavior (different water usage) and unwanted leakage were simulated.

3) Network Performance Tests: In this group of tests the basic performance of the network protocol was tested. Measurements provided information for the packet loss as a function of the duty cycle, packet queue length, distance and load (number of packets per unit time).

Furthermore, the lifetime of the network was evaluated as a function of the duty cycle, the distance and the load distribution. Measurements of the node's energy consumption were taken under different circumstances and lifetime calculations were provided. This allowed us to determine the nodes which are critical from power consumption point of view. Such information is very important and can be utilized at deployment time to overcome possible operational problems or to provide for maximizing the lifetime of the network as a whole. On the other hand, the energy required for the transmission of a single packet was also calculated.

As a result, the performance of the designed protocol can be optimized further to provide better performance once the system is realized at large scale.

C. Evaluation of the Results and Discussion

Evaluation of the results is done form the point of view of the user (pressure information) and the point of view of the network performance (network performance parameters).

1) Correctness of Pressure Data: As for the correctness of the collected information the measurements both in between the two thresholds and outside the thresholds proved that the node data algorithms are working properly. Data related to the time for reporting a possible pressure leak is given as an example in Fig. 7. and Fig. 8. The first one presents the information from all the nodes, while the second one (a zoomed version) shows more clearly the difference in the pressure and the time delay between the two nodes – node number 10 and node number 20.

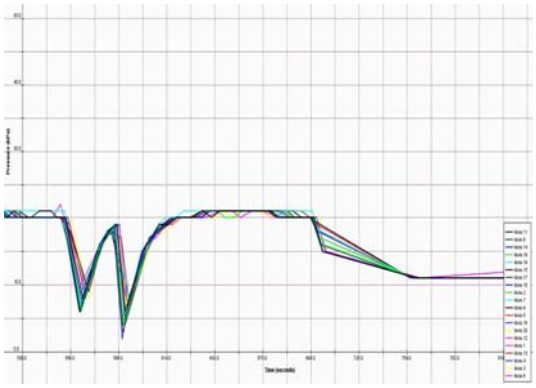


Figure 7. Graph of Changes Pressure Changes for Different Nodes

In Fig. 8. each different color represents and different node. Because the size of the experimental setup is quite small there is a very small time and pressure difference

between the nodes (Δt and Δp in Fig. 8). In a larger system these changes will be more easily perceived but still within the limits required for proper reaction.

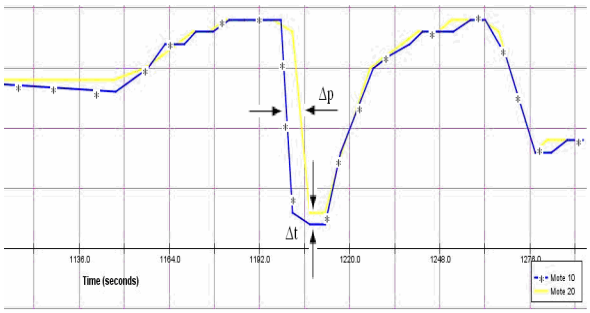


Figure 8. Zoomed Version Showing Two Separate Nodes

These tests and results prove that the suggested system can be used for monitoring a watering system and collecting information about possible leaks in due time. It can be easily attached to an automatic actuator system that will react to such failures and provide automatic closing of certain valves in order to prevent water waste.

2) Network Performance Evaluation: As mentioned above a number of different tests were carried out, however due to space limitations only some are presented here: the packet loss as a function of the sampling period, the length of the queue and the distance; the power consumption as a function of the sampling period as well as the graph of the power consumed by different nodes for a fixed sampling period.

The sampling period is an important parameter which influences both the correctness of the results, the network load and the overall lifetime of the system. Increasing the sampling period reflects in changing the network load – if we assume a 28 bytes packet size, for a sampling period of 1 second the load will be 560 bytes/sec while for a sampling period of 60 seconds it will be 9.3 bytes/sec. Thus, as it can be seen from Fig. 9. and Fig. 10. regulating the sampling period is an important tool in reducing the packet loss and also determining the power consumption. Voltage reduction is the difference between the battery level at the beging and at the end of the experiment. It is important to note that the largest change in power consumption is observed for very small intervals and a decision of interval duration of 2 or 10 sec can drive the required voltage from 0.157 V to 0.0182 V, or nearly 8.7 times less while at the same time packet loss is increased 2.6 times.

Furthermore, measurements in the battery loss for a sampling period of 60 sec show very distinctly (Fig. 11) that due to the hierarchical tree structure some nodes deplete their energy at a much faster rate then others. As these are usually key intermediate nodes, a possible solution is providing alternative energy sources or recharging probabilities, like for example solar rechargeable batteries. Such solutions are still expensive and our study gives an opportunity to evaluate more realistically the need for such implementations. According to our calculations, for the worst case, if a sampling period of 60 sec is applied the TelosB nodes will be able to work for 190 days.

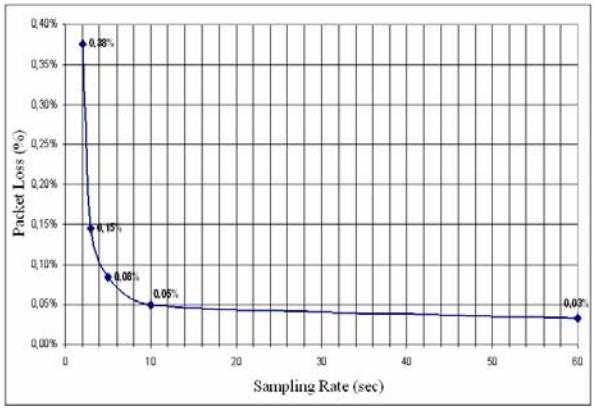


Figure 9. Packet Loss as a Function of Sampling Rate

Another factor which can be regulated to tune the network performance is the length of the queue of packets to be sent (or in other words the sending buffer size). For example, for a queue length of packets per node 7 the packet loss is 0.017% compared to 0.195% for a length of 2 packets. We have also observed that 7 is an optimal value since an increase to 8 o9 has also leaded to increase in the packet loss to 0.043 (Fig. 12).

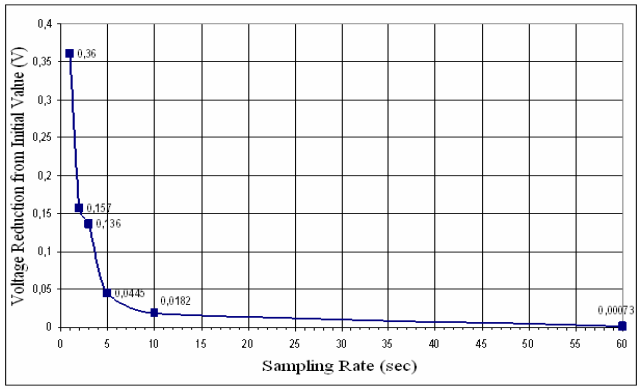


Figure 10. Voltage Reduction as a Function of Sampling Rate

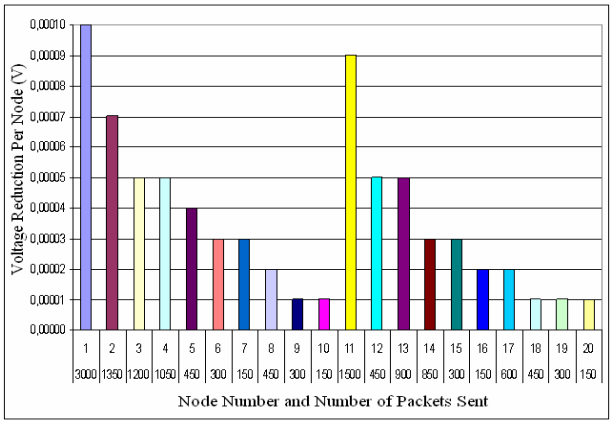


Figure 11. Voltage Reduction Per Node for 60 sec. Sampling Rate

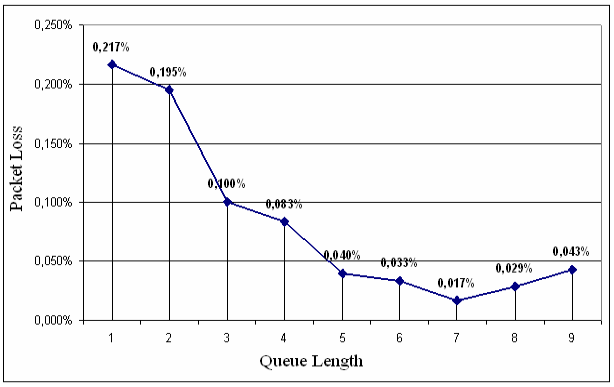


Figure 12. Packet Loss as a Function of Queue Length

Finally, the packet loss as a function of the distance between nodes is presented in Fig. 13. According to the datasheets TelosB modes have a range of up to 125 m. However our tests, done in the open field have shown that at a distance of 120 m the packet losses are unacceptably high (nearly 60%). Accordingly, packet losses less than 2% can be achieved if the nodes are deployed at up to 70 m apart.

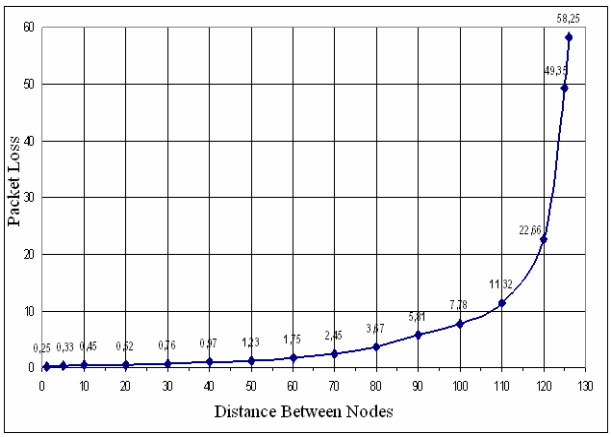


Figure 13. Packet Loss as a Function of the Distance Between Nodes

V. CONCLUSION

In this study, we have presented the design considerations and results from the experimental prototype of a wireless sensor based network for collecting pressure data that can be used for watering systems and gas pipelines. The system is based on TelosB motes organized in a WSN with a suitable designed network protocol which provides also data control and aggregation. The field-tests aim to provide insights into the possibilities for optimizing the network performance and increasing the network lifetime as a whole. The prototype is based on requirements provided by the 2nd DSI (2nd Regional Directorate of State Hydraulic and Water Works) for a water project in Usak, Turkey.

This theoretical and experimental work will provide an example of the possibilities to apply WSN for better and more effective control and implementation of farming watering systems. Such systems are static and the selected hierarchical tree network architecture and the pertinently designed protocol provide both very effective use of network resources and possibilities for fast determination of leaks and problems. The performance of the network and its power consumption were examined in detail and the parameters and methods to increase both its performance and lifetime were outlined. To reduce traffic load data aggregation and sampling rate optimization was suggested. Due to the adopted addressing scheme the coordinates of the problems can be efficiently determined without the need of GPS data.

In the future the system can be further enhanced by connecting to an actuator network that will provide timely reactions to prevent water waste.

As a result, the work presented in this paper provides the background for further research and implementation of wireless sensor networks is agriculture related fields as automatic watering systems. With small adjustments and additions the suggested system can be used for gasoline pipeline monitoring and control as well.

REFERENCES

[1] Min Lin, Yan Wu and Ian Wassell, “Wireless Sensor Network: Water Distribution Monitoring System”, Computer Laboratory, Cambridge University, UK, 2008.

[2] Yiming Zhou, Xianglong Yang, Liren Wang and Yibin Ying, "A Wireless Design of Low-Cost Irrigation System Using ZigBee Technology," nswctc, vol. 1, pp. 572-575, 2009 International Conference on Networks Security, Wireless Communications and Trusted Computing, 2009

[3] Yunseop Kim; Evans, R.G. and Iversen, W.M.; “Remote Sensing and Control of an Irrigation System Using a Distributed Wireless Sensor Network” IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT, VOL. 57, NO. 7, JULY 2008, pp. 1379-1387

[4] Stoianov, I., Nachman, L., Madden, S., Tokmouline, T. and Csail, M., “PIPENET: A Wireless Sensor Network for Pipeline Monitoring”, IPSN 2007, pp. 264 – 273, 2007.

[5] Liting Cao, Jingwen Tian, and Yanxia Liu, “Remote Real Time Automatic Meter Reading System Based on Wireless Sensor Networks” Beijing Union University, China, C07-02, ICICIC-2008-2304

[6] Baoding Zhang and Jialei Liu; “A Kind of Design Schema of Wireless Smart Water Meter Reading System Based on Zigbee Technology”, ICEEE , 7-9 Nov. 2010 , pp. 1 – 4

[7] Crossbow Technology Inc., "MPR/MIB User’s Manual", Rev. B, Document 7430-0021-06, 2005, http://www.xbow.com/Support /Support_pdf_files/MPRMIB_ Series_Users_Manual.pdf.

[8] http://www.xbow.com/Support/Support_pdf_files/MoteWorks_Getting_Started_Guide.pdf

[9] http://www.freescale.com/webapp/sps/site/prod_summary.jsp?code=MPXx5050

[10] http://www.xbow.com/Support/Support_pdf_files/MoteWorks_Getting_Started_Guide.pdf

[11] Sokullu, R., Akkas and M.A., Çetin, H.E., “Wireless Patient Monitoring System” Sensor Technologies and Applications (SENSORCOMM), 2010 Fourth International Conference on, 18-25 July 2010, pp. 179-184.

[12] B. Orhan, E. Karatepe and R.Sokullu “Modelling Energy Consumption for RF Control modules”, UBICOMM 2010, 4th International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies, October 25-30, 2010 - Florence, Italy

 ISBN: 978-1-61208-133-5

Design of a Control Algorithm for a 2x3 Optical Switch

Fakher Eldin M. Suliman
Department of Electronics Engineering
Sudan University of Science and Technology
Khartoum, Sudan
e-mail: fakhereldinmohamed@sustech.edu

Samia K. Hassan
Department of Computer Engineering
University of Gezira
Khartoum, Sudan
e-mail: samia.khaleel@geziracollege.edu.sd

Abstract--In this paper, a control algorithm is proposed for a 2x3 nonblocking photonic switch. The switch is a space division multistage network using 2x2 optical switching elements which can serve as basic element for larger size networks. The idea behind the proposed algorithm is presented. The wide-sense nonblocking property of the switch under this control algorithm is tested and discussed. The results indicate that the algorithm is capable to maintain the wide-sense nonblocking property for all possible switch configurations.

Keywords- photonic switch; multistage network; wide-sense nonblocking; control algorithm

I. INTRODUCTION

Photonic switching architectures based on 2 x 2 optical switching elements (SEs) are attractive, since they can be constructed from directional couplers. The directional coupler switch is a device with two inputs and two outputs, both of which are optical signals [1]. The state of the device, as shown in Fig. 1, is controlled electrically by applying different levels of voltage on the electrodes.

Although other materials can be used as a substrate, lithium niobate is the most mature technology for directional coupler-based optical switch fabrication. A feature of these switches is their ability to route optical information regardless of its bit rate or coding format [1]. Several directional coupler-based architectures had been proposed in the literature [2][4][6][8]. This hybrid device will be the SE of the 2x3 network proposed in this paper.

For a good switching architecture from system considerations, the number of SEs for a given switch size, *N*, should be as small as possible [2]. When the number is large, implementation is expensive and the optical path is subject to large power loss and crosstalk. When designed to reduce the SE number in total and in each path, a switch can have a large internal blocking probability. The internal blocking should be avoided or reduced. It can be reduced to zero by using a good switching control or by rearranging the current switching configuration. These cases are called wide-sense nonblocking and rearrangeably nonblocking, respectively [3]. If a blocking condition never arises in a switch, it is said to be strictly nonblocking [3][4].

In this paper, a control algorithm for a 2x3 nonblocking photonic switch which is derived based on 2x2 SEs is proposed. The idea behind the proposed algorithm is presented. The wide-sense nonblocking property of the switch under this control algorithm is tested and discussed.

The paper is organized as follows; Section II provides an overview of the 2x3 architecture. We explain how to design it using planar switches. In Section III, the development of the control algorithm is explained. The wide-sense nonblocking property of the switch under this control algorithm is tested and discussed in Section IV. Section V concludes the discussion.

II. THE 2x3 SWITCH

The *N*-stage planar switch has *N*/2 odd stages and *N*/2 even stages. The odd stages are of *N*/2 SEs each, while the even stages are of *N*/2 – 1 SEs each [5]. In general an *N* x *N* network requires *N* stages, where *N* may be even or odd. The total number of SEs is:

$$SE_S = N/2(N/2 + N/2 - 1) = N/2(N - 1) \quad (1)$$

The maximum number of SEs in a connection path is obtained when the optical signal crosses a SE in every stage of the switching system, that is, when it crosses *N* SEs. Fig. 2 illustrates a 3x3 *N*-stage planar switch.

An algorithm for deciding whether a given network is nonblocking or not is described in [7]. Using this algorithm the 3x3 switch of Fig. 2 was proved to be blocking unless rearranged [5][7].

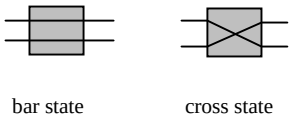


Figure 1. The states of a 2 x 2 switch element

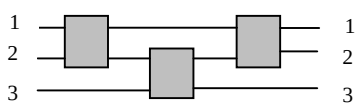


Figure 2. A 3 x 3 planar switch

Now, if only inputs 1 and 3 of Fig. 2 are used, instead of using all its three inputs, this will give the 2x3 planar switch

 ISBN: 978-1-61208-133-5

shown in Fig. 3. Again, the same algorithm described in [7] can be used to decide if this switch network is nonblocking.

Because the switch network is simple and small, all its possible states can be manually studied on paper. However, both methods lead to the same outcome. That is, the network is nonblocking in the wide sense if all the states in which SE A is cross (x) and SE B is bar (=) are avoided. In other words, if SE A is in the cross state, SE B should not be allowed to get into a bar state, and vice versa. Such a state, which can cause blocking for a network, is said to be a forbidden state. The set of states of a network, that allow any required switching without bringing the network into a forbidden state, was called preservable by Benes [3].

If the 2x3 switch is flipped horizontally, as shown in Fig. 4, the network will be a 3x2 switch with the same nonblocking rule still applicable.

The preservable state of the 2x3 network is given in Fig. 5a. The state of the last SE does not affect the state of the network and this is the reason why it is left blank. The preservable state of the 3x2 switch is shown in Fig. 5b from which it is clear that neither input 1 nor input 2 should be connected to output 1 through SE B. The elements of Fig. 3 and Fig. 4 will be called 2W3 and 3W2, respectively. If these elements follow the algorithms given in Fig. 5, any future connection can always be made without blocking or additional rearrangement of the existing paths.

The 2W3 and 3W2 elements can be used to build a 4x4 wide-sense nonblocking network. The 4x4 network will consist of two 2W3 switches, three 2x2 switches, and two 3W2 switches as shown in Fig. 6.

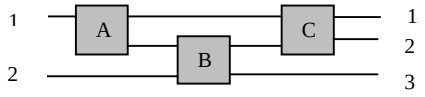


Figure 3. A 2 x 3 planar switch

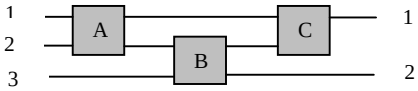
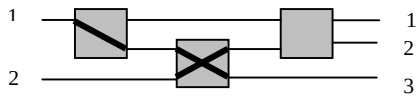
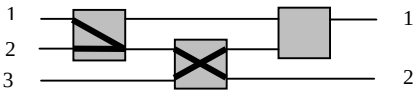


Figure 4. A 3 x 2 planar switch



(a)



(b)

Figure 5. The preservable states for: (a) the 2x3 switch and (b) the 3x2 switch

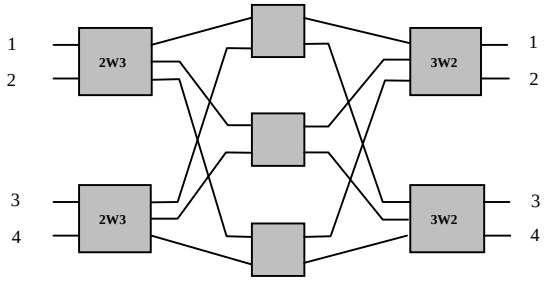


Figure 6. A 4 x 4 wide-sense nonblocking network

The 2W3 and 3W2 elements can also be used, recursively, to build larger wide-sense nonblocking networks with the basic 2x2 SE, always, representing the smallest possible subnetwork.

III. THE CONTROL ALGORITHM DESIGN

The 2W3 switch, and all the switches designed, recursively, based on it, will need a control algorithm to maintain their wide-sense nonblocking property explained in the previous section. In this section, a control algorithm is developed to control the 2W3 switch keeping in mind that algorithms for controlling larger sizes of switches based on the 2W3 and 3W2 elements should be addressed separately, and that the recursive approach is not applicable here.

The number of possible configurations of a switch, of size *N*, is given by *N*!. Thus, the 2x3 switch has 3! = 6 different configurations as shown in Fig. 7 denoted by S_1, S_2, S_3, S_4, S_5, and S_6. The state of each SE for obtaining these six configurations is shown in table 1 with "0" used to represent the bar state and "1" used to represent the cross state.

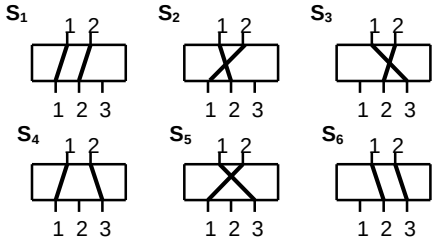


Figure 7. Possible configurations of the 3x2 switch

Observing Table 1, it can be noted that the highlighted states, i.e., state number 5 and state number 6, are the forbidden ones. These states can be compensated for by state number 2 and state number 1, respectively, since they give the same switch configuration for each case.

 ISBN: 978-1-61208-133-5

TABLE 1. THE DETAILED SE STATES OF THE POSSIBLE 6 CONFIGURATIONS

NO	SE A	SE B	SE C	Configuration
1	0	0	0	S_4
2	0	0	1	S_6
3	0	1	0	S_1
4	0	1	1	S_2
5	1	0	0	S_6'
6	1	0	1	S_4'
7	1	1	0	S_3
8	1	1	1	S_5

An algorithm code was written using Visual Basic tool to control the switch and maintain its nonblocking property. This tool is event-driven and is governed by an event processor. When an event is detected, the event procedure will then be executed [9]. The flow chart of the algorithm code is shown in Fig. 8. A Graphical User Interface (GUI) was also developed using Visual Basic.

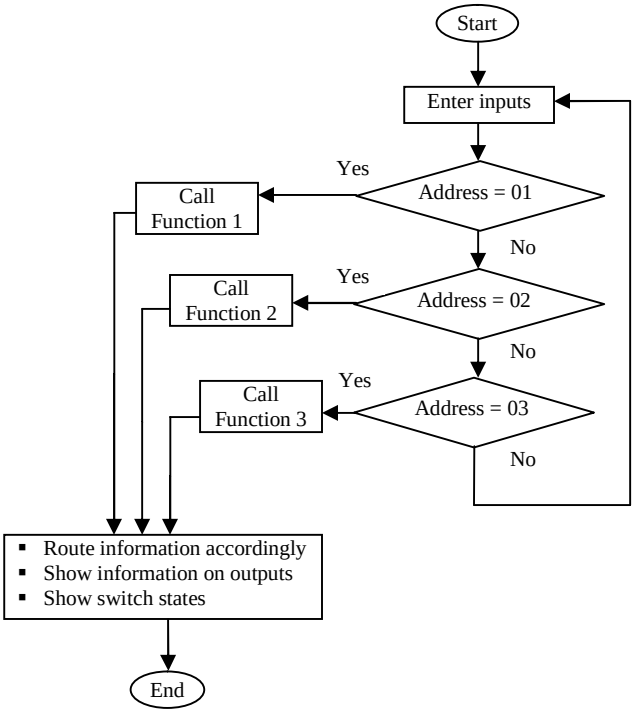


Figure 8. The flow chart of the algorithm code

As illustrated in the flowchart, one of the routing functions 1, 2, or 3 is called depending on the destination targeted by each input. Function 1, function 2, and function 3 are algorithm codes designed to control the state of SE A, SE B, and SE C, respectively. Each function contains tow subfunctions, one for the bar state, and the other for the cross state, of the corresponding switch. In each subfunction, all possible relations between the inputs and outputs of the SE are examined and, then, the route established if, and only if, it maintains the preservable states of the 2x3 switch as shown in Fig. 5a.

IV. PERFORMANCE ANALYSIS AND DISCUSSIONS

In this section, some results obtained after running the code will be presented and discussed. The following simulate window which appears when the start button of the start window is pressed is shown in Fig. 9. Here, the switches A, B, and C, are shown as Switch1, Switch2, and Switch3, respectively. These switches are shown without their current state which will change accordingly when the simulate button is pressed after a user inputs the destination output followed by the data as illustrated in Fig. 10, Fig. 11, Fig. 12, Fig. 13, Fig. 14, and Fig. 15 for different switching configurations.

In Fig. 10, the data at input data_1 is routed to output out_3 through switch 1 and switch 2 which both must be in the cross state to establish the configuration. If the data from the same input is to be routed to output out_1 or output out_2, the switch configuration will look as illustrated in Fig. 11 and Fig. 12, respectively. Note how the forbidden state, in which switch 2 should be brought into a bar state, was avoided in the configuration set to establish the new connection.

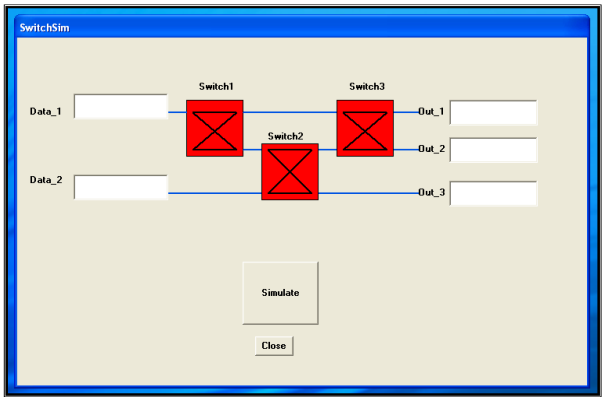


Figure 9. The simulate window of the designed GUI

Figures 13, 14, and 15 present simulation results obtained when the algorithm was tested for two inputs applied at parallel to the 2x3 switch. From these figures, of special consideration is the transition of the switch configuration from Fig. 14 to Fig. 15. From a control point of view, it was easier just to change switch 2 into a bar state to reconfigure Fig. 14 to establish the configuration shown in Fig. 15. If switch 2 was changed into a bar state, input data_2, would

have been blocked from reaching output out_1, as long as input data_1 is connected to output out_2 through switch 2. That is the reason why the control algorithm instead changed the states of all the switches to avoid the easy, but yet, forbidden reconfiguration.

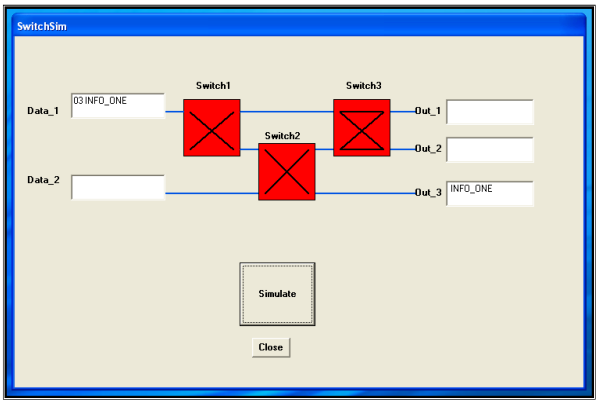


Figure 10. Input data_1 to output out_3 switch configuration

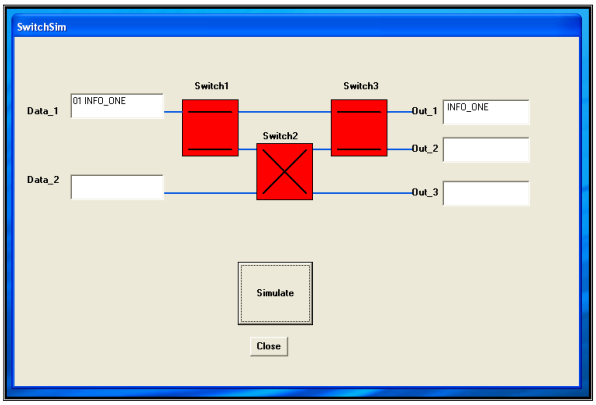


Figure 11. Input data_1 to output out_1 switch configuration

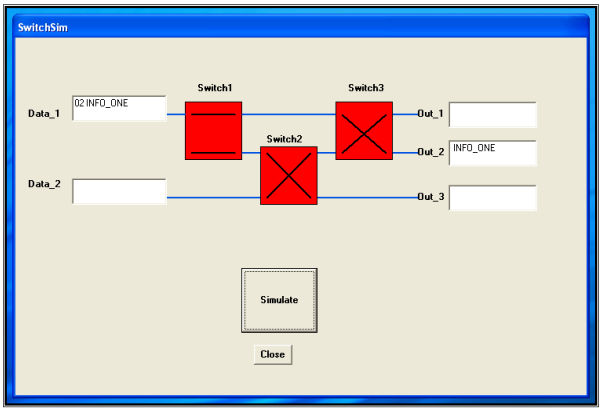


Figure 12. Input data_1 to output out_2 switch configuration

V. CONCLUSION AND FUTURE WORK

A control algorithm for a 2x3 wide-sense nonblocking photonic switching network has been proposed, designed, and simulated. Some simulation results of the proposed control algorithm are presented and discussed.

The results indicate that the designed algorithm is capable to maintain the wide-sense nonblocking property for all possible switch configurations. Algorithms for controlling larger sizes of switches based on the 2W3 and 3W2 elements should be addressed separately since the recursive approach can not be applied. Authors are now working on designing a control algorithm for the 4x4 optical switch mentioned in section II.

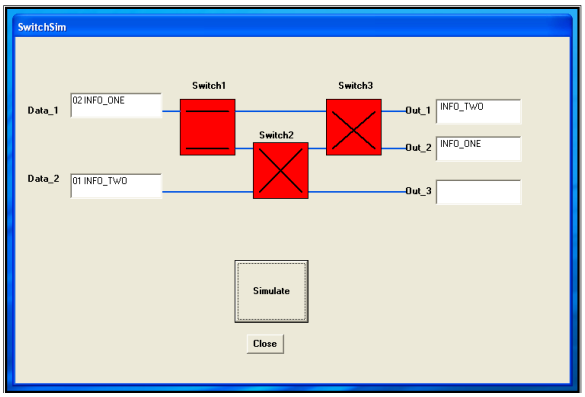


Figure 13. Input data_1 to output out_2 and input data_2 to output out_1 switch configuration

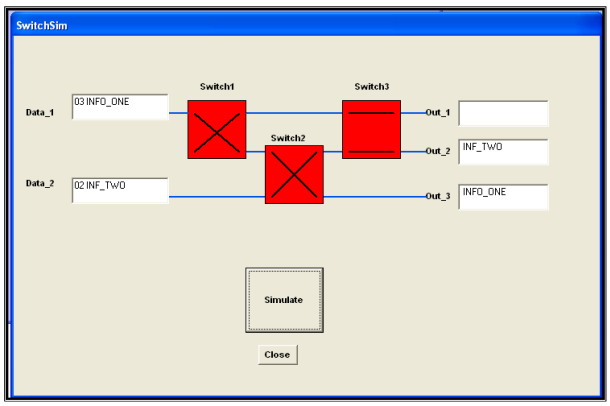


Figure 14. Input data_1 to output out_3 and input data_2 to output out_2 switch configuration

 ISBN: 978-1-61208-133-5

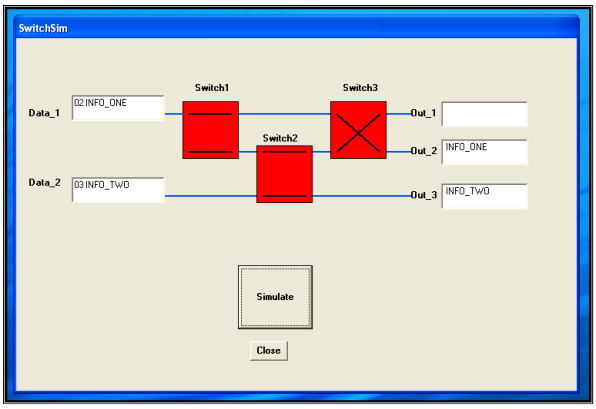


Figure 15. Input data_1 to output out_2 and input data_2 to output out_3 switch configuration

ACKNOWLEDGMENT

We would like to thank the Department of Computer Engineering, Faculty of Engineering and Technology, University of Gezira where the work has been carried.

REFERENCES

[1] D. K. Hunter, “Optical switching in ultrafast communications networks”, PhD Dissertation, University of Strathclyde, Glasgow G1 1XW, 1991.

[2] F. M. Suliman, “Design of nonblocking high-density photonic switches” PhD Dissertation, Universiti Teknologi Malaysia, October 2003

[3] V. E. Benes, Mathematical Theory of Connecting Networks and Telephone Traffic, Academic Press, 1965.

[4] R. A. Spanke, “Architectures of guided-wave optical space switching systems,” IEEE Comm. Magazine, vol. 25, no. 5, pp. 42-48, 1987.

[5] R. A. Spanke and V. E. Benes, “An N-stage planar optical permutation network,” Applied Optics, vol. 26, no. 7, pp. 1226-26, 1987.

[6] F. M. Suliman, A.B. Mohammad, and K. Seman “A new nonblocking photonic switching network” Global Telecommunications Conference, 2001. GLOBECOM '01. IEEE, USA, Volume: 4, 2001, pp. 2071 –2076, November 2001

[7] C. J. Smith, ‘Nonblocking photonic switch networks,’ IEEE Journal on Selected Areas in Comm., vol. 6 no. 7, pp. 1052-62, August 1988.

[8] A. Jajzczyk, “A class of directional-coupler-based photonic switching networks,” IEEE Trans. Commun. vol. 41, no 4, pp. 599-603, April 1993.

[9] Visual dBASE Programmer’s Guide, 1995
Borland International, Inc., 100 Borland Way
P.O. Box 660001, Scotts Valley, CA 95067-0001

 ISBN: 978-1-61208-133-5

Energy-efficient Optimizations of the Authentication and Anti-replay Security Protocol for Wireless Sensor Networks

Laura Gheorghe, Răzvan Rughiniş, Nicolae Ţăpuş
University POLITEHNICA of Bucharest
Bucharest, Romania
{laura.gheorghe, razvan.rughinis, ntapus}@cs.pub.ro

Abstract **- Wireless Sensor Networks are an emerging technology used for environmental monitoring. Security is a major concern when deploying a network for critical applications, such as military or medical surveillance. We have previously developed a security protocol that provides authentication, anti-replay, integrity and reliability. This paper presents further optimizations in order to minimize energy consumption. We have implemented the Energy-efficient Authentication and Anti-replay Security Protocol in TinyOS and we have tested its functionality and performance using the TOSSIM simulator. We have developed a mathematical model to evaluate energy consumption, determining the control overhead of the security protocol and the energy saved through the optimizations and thus proving the efficiency of the optimizations. The optimized EAASP represents a security protocol that provides strong authentication and anti-replay, integrity and reliability, and energy-efficiency.**

Keywords - wireless sensor networks, security, authentication, anti-replay, integrity, reliability, energy-efficiency.

I. INTRODUCTION

Wireless Sensor Networks (WSNs) consist of a large number of small embedded devices with limited capabilities and low power consumption that have the abilities to self-organize into a network and to perform sensing, communicating and processing tasks [1].

WSNs are used to monitor their environment. Standard applications that use WSNs are environmental, medical, and military surveillance and emergency detection [2]. Such applications require high levels of security.

Securing sensor networks is a challenging task because of their specific constrains, such as the limited capabilities of sensor node hardware, and the deployment context [3].

We have previously developed the Authentication and Anti-replay Security Protocol (AASP), which provides authentication, integrity, anti-replay and reliability. Our contributions in this paper consist in several optimizations to the AASP in order to reduce the energy consumption, such as minimizing the control overhead, reducing the number of handshake packets, using negative acknowledgements instead of positive ones, and aggregating sensed data.

We have also developed a mathematical model that evaluates the energy consumption introduced by the security protocol, and we used it in order to perform a formal evaluation of the control overhead and energy savings.

The rest of the paper is structured as follows: Section II presents related work, Section III discusses the protocol design, Section IV describes the implementation of the protocol, Section V presents the mathematical model and the evaluation results, and Section VI discusses the values of the protocol and concludes the paper.

II. RELATED WORK

Several significant security solutions for WSNs include the ZigBee Security Architecture, SPINS, and TinySec.

ZigBee Security Architecture consists in a coordinator that acts as "Trust Center", which allows other devices to join the network and provides them keys [4]. ZigBee works with three roles: the trust manager that authenticates devices that want to join the network, the configuration manager that manages and distributes keys, and the configuration manager that provides end-to-end security. The infrastructure operates in two modes: the Residential Mode that is used for low security residential applications and the Commercial Mode that is used by high-security commercial applications.

SPINS is a suite of security protocols optimized for WSNs, consisting of two building blocks: the Secure Network Encryption Protocol (SNEP) and the "micro" version of the Timed, Efficient, Streaming, Loss-tolerant Authentication Protocol (μTESLA) [5]. SNEP provides confidentiality using encryption, authentication and integrity by Message Authentication Codes (MAC). μTESLA provides authenticated broadcast by emulating asymmetry through a delayed disclosure of symmetric keys. SPINS has been implemented on top of TinyOS [6].

TinySec has been designed as the replacement of SNEP and provides confidentiality, authentication, integrity and anti-replay protection [7]. It implements the Cipher Block Chaining (CBC) mode with cipher text stealing for encryption and the Cipher Block Chaining Message Authentication Code (CBC-MAC) for authentication. TinySec uses the TinySec-Auth format for authenticated packets and the TinySec-AE for authenticated and encrypted packets.

We aim to develop a security protocol that provides strong authentication through the establishment of an authentication connection, strong anti-replay through binding the packet to its context, integrity and reliability, while also being energy-efficient.

III. AN ENERGY-EFFICIENT SECURITY PROTOCOL

The purpose of this work is to develop a lightweight, energy-efficient security protocol that provides authentication, freshness and integrity for Wireless Sensor Networks. EAASP has been designed by minimizing the

 ISBN: 978-1-61208-133-5

energy consumption of the Authentication and Anti-replay Security Protocol (AASP) [8] [9].

In order to improve the energy efficiency of a protocol one has to reduce the number of packets communicated by nodes, and the average packet size. We aim to optimize the security protocol by reducing the control overhead.

A. Authentication and Anti-replay Security Protocol

We have previously developed a security protocol that provides two-way authentication, anti-replay protection and integrity [8].

Authentication is ensured by the use of the Message Authentication Code (MAC), which is computed from the payload of the message and a secret key, by applying a collision resistant hash function. We have used the Hash Message Authentication Code (HMAC) implementation.

The anti-replay protection derives from binding the packet to its context, specifically to the previous packet between the same source and destination, and its sequence number. The mechanism consists in including the MAC of the previous packet in the current packet for all messages sent between the same source and destination.

Integrity is ensured by including the MAC of the current message in the packet.

For the first packet, authentication is performed by checking the MAC of the current message. Still, this measure does not protect against replay attacks. If the exact sequence of packets between the same source and destination is captured, it can be easily replayed later.

In order to further strengthen the authentication and anti-replay protection, an authenticated connection has to be established before sending any data packets. The authenticated connection is established in AASP after a four-step handshake.

In order to prevent the de-synchronization of the anti-replay mechanism through loss of packets, we have implemented an acknowledgement mechanism [9]. The next packet is not sent until the previous packet is acknowledged by the destination. The source node waits for the acknowledgement for a specified period of time and, if it times out, the packet is re-sent.

The authenticated connection has to be re-initiated if one of the communicating nodes loses its connection data or if the anti-replay mechanism is de-synchronized. The connection times out after a specified period of time in which no data or ACK message is received from the other node. In that moment, connection data is erased and an authenticated connection can be re-initiated.

The AASP is effective against malicious injection and replay attacks. In order to increase its energy efficiency we aim to reduce the number of control packets and the control overhead from the data packets.

B. Energy-efficient Authentication and Anti-replay Security Protocol

In this paper, we present a lightweight security protocol that uses the basic mechanisms of AASP and has higher energy efficiency. In order to reduce energy consumption, we reduce the number of packets and the control overhead.

1) Reducing the control overhead

The AASP protocol has a header with the following fields: Previous Hash (P_Hash) – 2 bytes, Current Hash (C_Hash) – 2 bytes, Authentication (Auth) – 1 byte, Acknowledge (ACK) – 1 byte and Sequence (Seq) – 1 byte.

EAASP is designed with a protocol header as presented in Table 1.

TABLE I. HEADER STRUCTURE

	EAASP Header Fields		
	Hash	***Type***	***Seq***
Number of bytes	2	1	1

The Hash field contains a MAC computed from the current payload, the previous payload, the secret key and the sequence number, as shown in Formula (1). The Type field encodes the type of packet, as presented in section C.

$$Hash_i = MAC(Payload_i, Payload_{i-1}, Seq, K) \quad (1)$$

The sequence number is taken into account when computing the hash in order to avoid packet altering by intermediate nodes.

2) Reducing the number of control packets

AASP has two types of control packets: handshake packets and acknowledgement packets.

We aim at reducing the number of handshake packets, while still providing powerful authentication. The authentication method can be strengthened by sending random challenges to each other.

We have reduced the number of handshake packets to three, as presented in Figure 1.

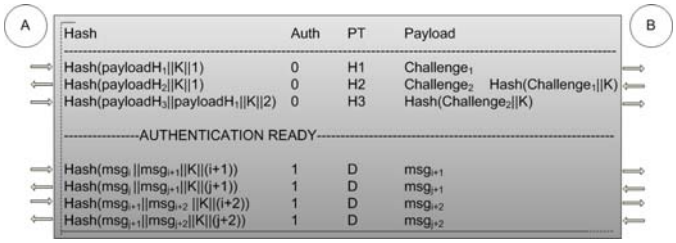


Figure 1. Energy-efficient Three Step Handshake

The first packet, designated H1, contains a challenge number randomly generated by the initiating node A. The second packet, designated H2, is the answer from node B, and it contains a hash based on the first challenge and the secret key, and the challenge randomly generated by node B. The third packet, designated H3, is the response of node A and contains a hash based on the second challenge and the secret key. After the three-step handshake is completed, data packets can be exchanged.

We have used negative acknowledgements to reduce the number of acknowledgement packets while still maintaining a certain level of reliability.

The destination is able to detect packet loss when receiving out-of-sequence packets. The destination stores the sequence number of the last received and valid packet, and it

 ISBN: 978-1-61208-133-5

expects the packet with the next sequence number. When it receives a packet with a sequence number greater than the expected one, it generates and then sends a Negative Acknowledgement (NACK) packet back to the source node.

The out-of-order packet is not dropped at the destination node, and it is stored until all previous packets are received. The NACK packet contains the sequence number of the first lost packet – L_Seq, and the sequence number of the received out-of-order packet – R_Seq, in order to prevent the source node to deliver the out-of-order packet again. The source node resends all packets with a sequence number greater or equal to L_Seq and less than R_Seq, when receiving a NACK packet. The destination performs an integrity and an anti-replay check on the re-sent and out-of-order packets, and it delivers them to the application.

C. Protocol message structure

The EAASP header contains the following fields: Hash (2 bytes), Type (1 byte) and Seq (1 byte), as presented in Table 1. The Hash field contains the MAC computed from the payloads of the previous and current packets, the secret key, and the sequence number of the current packet. The sequence number is used by the destination node to detect lost packets. The Type field encodes the type of a certain packet and contains the following fields: Auth (1 bit), NACK (1 bit), PT (3 bits) and QoS (3 bits), as represented in Table 2.

The Auth flag is set to 0 during the three-step authentication handshake, and it is set to 1 after the authenticated connection has been established and data packets have been exchanged. The NACK flag is set to 1 in a NACK packet and to 0 otherwise.

TABLE II. TYPE FIELD STRUCTURE

	Type Field			
	Auth	***NACK***	***PT***	***QoS***
Number of bits	1	1	3	3

PT represents the Packet Type and is 001 for H1 packets, 010 for H2 packets, 011 for H3 packets, 100 for Data packets.

The QoS field is used for assigning a priority value to packets. Because it is implemented on 3 bits, it provides for 8 priority levels. Certain packets can be assigned a higher priority than others, such as the re-send or control packets.

IV. EAASP IMPLEMENTATION

The EAASP was implemented in TinyOS, an event-driven, component-based operating system for Wireless Sensor Networks [6].

1) Implementing the security protocol

EAASP consists of two layers, which were introduced in the communication stack of the operating system: the MAC layer and the Authentication layer.

The MACLayerSender component is placed between the AMSenderC component and the ActiveMessageC component, and it has access to all packets sent by the application layer. In the AASP version this layer placed the MAC of the previous and current payload in the analyzed packet, and it stored the MAC of the current payload in the component, for further use. In the current, optimized version, this process has been moved into the Authentication layer, because it stores a predefined number of sent packets for further retrieval. If lost, they can be requested by a NACK packet. This measure avoids duplicating the packet list. Therefore, the MACLayerSender contains only the routing protocol implementation.

The MACLayerReceiver is placed between the ActiveMessageC component and the AMReceiverC component, and it has access to all packets received by the node before reaching the application layer.

If the component receives an out-of-order packet, it stores it, and it delivers it to the upper layers only after all packets with a lower sequence number have also been delivered.

The MAC is computed from the payload of the current packet, the payload of the previous packet, the secret key and the sequence number of the current packet. If the MAC does not match the one found in the Hash field of the packet, the Altered flag is set. If the MAC is correct, the packet is stored for further use when verifying the MAC of the next packet. In either case, the packet is delivered to the upper layers.

The Application layer is placed on top of the operating system, and it uses the components AMSender and AMReceiver to send to and respectively to receive packets from the medium. We cannot send control packets, such as handshake and ACK/NACK packets, from the MACLayer components; therefore, we must divide the Application layer into two sublayers: the Authentication layer and the actual application layer.

Whenever a packet is received by the Authentication layer from the actual application layer, without having an authenticated connection with the destination node, the Authentication layer initiates the three-step handshake by sending an Authentication Request (H1) packet. The Authentication layer performs the handshake in order to establish the authenticated connection.

The Authentication layer keeps track of the sequence number and stores a list of sent packets that can be used when packets are lost and a NACK is received from the destination. The layer computes and writes the protocol header for each sent packet: sequence number, type and hash. If it receives an out-of-order or altered packet, it generates and sends a NACK to the source node. When sending a NACK, a timer is configured to be fired after a predefined period of time. If the lost data packets are not retrieved in that interval, the NACK is considered to be lost and the NACK is resent.

If the Authentication layer receives a NACK, it re-sends all lost packets. If multiple packets are lost, the first packet is re-sent from the Receive.receive() event, and the subsequent packets, except for the out-of-order packet, are sent from the AMSend.sendDone() event. If it receives a correct data packet, it delivers it to the actual application layer and stores the sequence number of this packet.

2) Implementing the routing protocol

TinyOS provides single-hop communication via the Active Messages stack. In order to ensure multi-hop

 ISBN: 978-1-61208-133-5

communication, we introduce layer 3 information in packets and we implement a routing protocol. We use the AM addresses as layer 2 addresses and we introduce a layer 3 source ID and destination ID in the packets.

The MAC layer contains the implementation of a simple routing protocol. This layer stores a routing table that contains the next hop associated with a certain destination. When the MACLayerSender component receives a packet from the application layer, it checks the routing table in order to determine the next hop towards the destination. After that, it sends the packet to the next hop node by setting it as the destination in the AM packet.

When the MACLayerReceiver receives a packet with the layer 3 destination different from the node ID, it checks the routing table to find the next hop and sends the packet to that node. A discussion of routing procedures in the EAASP lies beyond the scope of this article.

V. EVALUATION RESULTS

We evaluate EAASP by determining its energy efficiency and its scalability.

A. Simulation results

A first evaluation relies on several test scenarios implemented with TOSSIM, a simulation tool for TinyOS applications [10].

1) Single-hop scenario

The first test scenario has the purpose of demonstrating basic single-hop functionality. We determine the proportion of lost packets by computing an average value across 20 instances of scenario execution.

Figure 2 presents the TOSSIM output for a single-hop authentication initialization, connection establishment, and data packets exchange.

Each line has the following format: The ID of the node, the component that generates the output, the type of packet sent or received, the fields, the source and destination of the packet.

```
(3): AuthenticationLayer: H1 packet sent [payload=234 hash=56843 type=8 seq=1 (3->1)]
(1): AuthenticationLayer: H2 packet sent [payload=57195 hash=42756 type=16 seq=1 (1->3)]
(3): AuthenticationLayer: H3 packet sent [payload=56185 hash=47406 type=24 seq=2 (3->1)]
(3): AuthenticationLayer: Managed to authenticate myself to node 1
(1): AuthenticationLayer: Managed to authenticate myself to node 3
(3): ApplicationC: Data packet sent [payload=1235 hash=41396 type=160 seq=3 (3->1)]
(1): ApplicationC: Data packet received [payload=1235 hash=41396 type=160 seq=3 (3->1)]
```

Figure 2. Handshake and data packets

We can observe from Figure 2 that the Authentication layer is responsible for performing the handshake and for establishing the connection, and the Application layer has the role of sending and receiving data packets.

In Figure 3, we can observe that a packet with sequence 11 is lost and the subsequent packet is received by the destination. It is detected as an out-of-sequence number and a NACK packet is sent to the source node, which resends the packet. The out-of-order packet is stored in the MACLayerReceiver and it is delivered to the application layer after the lost packet is received. After that, the next packet is sent and received correctly at the destination.

```
(3): ApplicationC: Data packet sent [payload=1243 hash=43279 type=160 seq=11 (3->1)]
(3): ApplicationC: Data packet sent [payload=1244 hash=43260 type=160 seq=12 (3->1)]
(1): AuthenticationLayer: Out-of-order packet received [payload=1244 hash=43260 type=161 seq=12 (3->1)]
(1): AuthenticationLayer: NACK packet sent [payload=11 hash=42686 type=64 seq=2 (1->3)]
(3): AuthenticationLayer: NACK packet received [payload=11 hash=42686 type=64 seq=2 0 (1->3)]
(3): AuthenticationLayer: Data packet re-sent [payload=1243 hash=43279 type=162 seq=11 (3->1)]
(1): ApplicationC: Data packet received [payload=1243 hash=43279 type=162 seq=11 (3->1)]
(1): ApplicationC: Data packet received [payload=1244 hash=43260 type=160 seq=12 (3->1)]
(3): ApplicationC: Data packet sent [payload=1245 hash=43243 type=160 seq=13 (3->1)]
(1): ApplicationC: Data packet received [payload=1245 hash=43243 type=160 seq=13 (3->1)]
```

Figure 3. Lost and recovered data packets

In order to determine the proportion of lost packets we have used a scenario in which we have generated 100 packets, we have counted the number of lost packets and we have computed the percent of lost packets. The scenario has been run for 20 times in order to compute an average value. The resulting average value for the single-hop case is 1.55% lost packets.

2) The multi-hop scenario

As a simple multi-hop scenario we choose a 3 node chain topology and we send packets from one end to another, as presented in Figure 4.

```
(0): RadioCountToLedsC: Data packet sent [payload=1257 hash=43030 type=160 seq=24 (0->2)]
(1): RoutingLayer: Routing packet through 2 [payload=1257 hash=43030 type=160 seq=24 (0->2)]
(2): RadioCountToLedsC: Data packet received [payload=1257 hash=43030 type=160 seq=24 (0->2)]
```

Figure 4. Multi-hop packet routing

To determine the proportion of lost packets for a multi-hop scenario, we have used 10 nodes placed in a chain topology. We have generated 100 packets, we have run the scenario for 20 times, and we have obtained an average of 9.55% lost packets for the 10 node chain topology. The average distance (in hops) from the source node where the packets are lost is 4.45.

B. Energy consumption

We have developed a mathematical model designated as the Sent/Received Bytes Evaluation Model that allows us to determine a measurement of energy consumption in order to

 ISBN: 978-1-61208-133-5

evaluate the control overhead and to compare EAASP with AASP.

Similar mathematical models such as [11] include 2nd layer information, which is not useful when analyzing a security protocol.

In order to evaluate energy consumption, our mathematical model takes in consideration only the number of bytes sent and received by the nodes. We did not include the relatively insignificant levels of energy consumed when executing code on sensor nodes, given that 1 bit transmitted in a sensor network consumes as much power as 800 -1000 instructions [12].

We consider 2 scenarios with the maximum number of hops between two nodes of 10, and respectively 100 hops. For each of these scenarios we transfer 10, 20, 50 and 100 packets between two nodes with the maximum number of hops between them. We use data payloads of 4 and 8 bytes.

Formula (2) evaluates power consumption when no packet is lost. Nb_{H1}, Nb_{H2} and Nb_{H3} represent the size (in bytes) of the handshake packets; Nb_D represents the size of a data packet; NP is the number of packets and NH is the number of hops between source and destination.

$$EC = (Nb_{H1}+Nb_{H2}+Nb_{H3}+ NP*Nb_D)*2*(NH+1) \quad (2)$$

Formula (3) evaluates power consumption when packets are lost and NACKs are used. PPL represents the percentage of lost packets; ADLP represents the average distance where packets are lost, Nb_{NACK} represents the size (in bytes) of the NACK packet. This formula takes into consideration that packets go through a number of nodes before being lost and that NACK packets are used to retrieve those packets.

$$EC = (Nb_{H1}+Nb_{H2}+Nb_{H3}+ NP*Nb_D)*2*(NH + 1) + NP*PPL*Nb_D*(1 + 2*ADLP*NH) + NP*PPL*Nb_{NACK}*2*(NH + 1) \quad (3)$$

1) 10 hops scenario

The results for sending 10, 20, 50 and 100 packets on a 10 hop chain are presented in Table 3 and Figure 5. All values are computed in bytes. Energy consumption for a byte depends on the hardware platform.

We assume that the average distance where the packets are lost is 50% from the total number of hops, a similar situation to the one determined experimentally.

TABLE III. 10 HOPS SCENARIO, 4 BYTE PAYLOADS

Packet loss rate	Number of packets			
	10	*20*	*50*	*100*
No packet loss	2200	3960	9240	18040
10% packet loss	2420	4400	10340	20240
20% packet loss	2640	4840	11440	22440
30% packet loss	2860	5280	12540	24640
40% packet loss	3080	5720	13640	26840

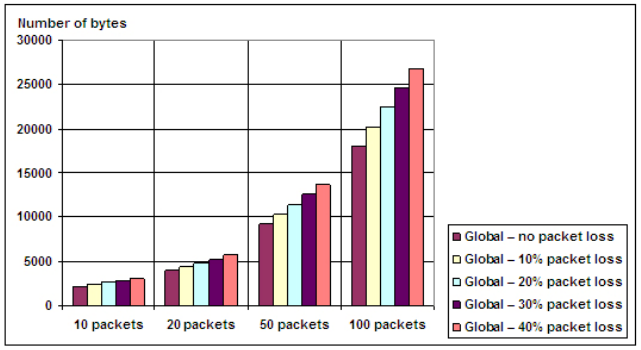


Figure 5. 10 hops scenario, 4 byte payloads

2) 100 hops scenario

The results for the 100 hops scenario are presented in Table 4 and in Figure 6.

TABLE IV. 100 HOPS SCENARIO, 4 BYTE PAYLOADS

Packet loss rate	Number of packets			
	10	*20*	*50*	*100*
No packet loss	20200	36360	84840	165640
10% packet loss	22220	40400	94940	185840
20% packet loss	24240	44440	105040	206040
30% packet loss	26260	48480	115140	226240
40% packet loss	28280	52520	125240	246440

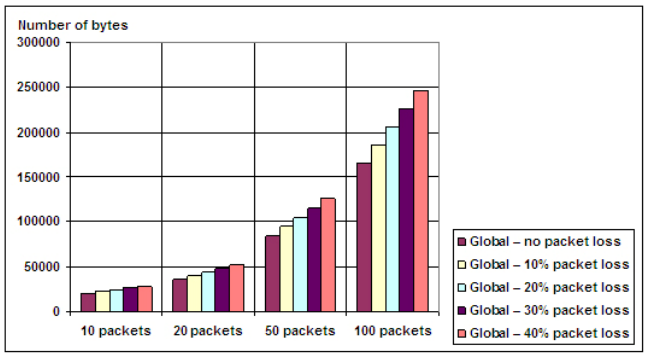


Figure 6. 100 hops scenario, 4 byte payloads

3) Control overhead

We aim to determine the control overhead (CO) of the security protocol for 4 and 8 byte payloads. We use Formula (4) to determine the control overhead. Nb_{HD} is the dimension in bytes of the EAASP header.

$$CO = (Nb_{H1}+Nb_{H2}+Nb_{H3}+ NP*Nb_D)*2*(NH + 1) + NP*PPL*Nb_{HD}*(1 + 2*ADLP*NH) + NP*PPL*Nb_{NACK}*2*(NH + 1) \quad (4)$$

Table 5 and Figure 7 present the control overhead for the 10 hops scenario, for 100 transferred packets.

TABLE V. CO FOR 4 AND 8 BYTE PAYLOAD PACKETS

Packet loss rate	CO bytes	CO for 4 bytes	CO for 8 bytes
No packet loss	9240	51%	34%
10% packet loss	11000	54%	37%

Packet loss rate	CO bytes	CO for 4 bytes	CO for 8 bytes
20% packet loss	12760	57%	40%
30% packet loss	14520	59%	42%
40% packet loss	16280	61%	44%

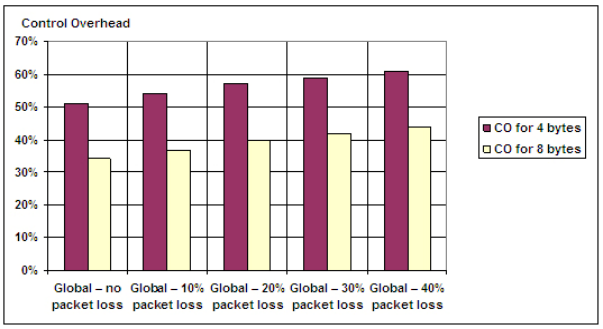


Figure 7. Control overhead for 4 and 8 byte payload packets

The percentage of control overhead decreases as the payload dimension is increased. For 4 byte payloads, CO goes from 51% for no packet loss, to 54% for 10% packet loss, to 61% for 40% packet loss. For 8 bytes, the CO goes from 34% for no packet loss, to 37% for 10% packet loss, to 44% for 40% packet loss.

4) EAASP vs. AASP

We compare the two versions of the protocol in order to determine the extent to which EAASP is more energy-efficient.

The difference in energy consumption between the two versions of the protocol is determined in Formula (5). Nb_H represents the dimension of the handshake packets in AASP; Nb_{ACK} is the dimension of the ACK packet; Nb_{De} is the dimension of EAASP data packet and Nb_{Da} is the dimension of the AASP data packet.

$$AASP - EAASP = [4*Nb_H + NP*(Nb_{Da}+Nb_{ACK} - Nb_{De} - PPL*Nb_{NACK}) - Nb_{H1} - Nb_{H2} - Nb_{H3}]*2*(NH + 1) + NP*PPL*(Nb_{Da} - Nb_{De})*(1 + 2*ADLP*NH) \quad (5)$$

We present results for 10 hops scenario and 4 byte payloads in Table 6 and Figure 8. All values are represented in number of bytes, because the energy depends directly on the number of sent and received bytes.

TABLE VI. AASP VS. EAASP – 10 HOPS, 4 BYTE PAYLOADS

Packet loss rate	AASP	EAASP	Saved energy
No packet loss	42592	18040	24552
10% packet loss	43802	20240	23562
20% packet loss	45012	22440	22572
30% packet loss	46222	24640	21582
40% packet loss	47432	26840	20592

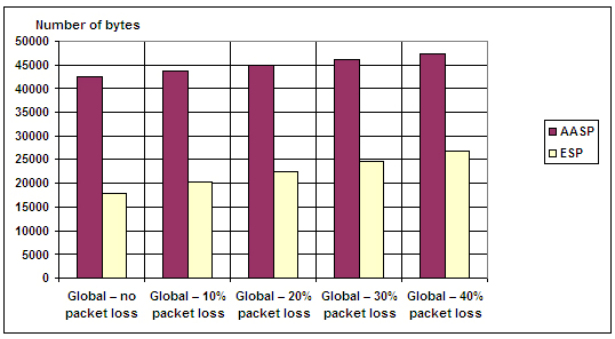


Figure 8. AASP vs. EAASP

Results indicate that EAASP is considerably more energy efficient than AASP: the saved energy amounts to 24KB when no packet is lost, to 23KB for 10% packet loss, and to 20KB when 40% packets are lost. The saved energy decreases slightly as the percentage of lost packets increases.

5) Data aggregation

We consider sending two 4 byte values into a single payload in order to reduce energy consumption.

In Table 7 and Figure 11 we compare energy consumption when sending 50 packets with an 8 bytes payload and when sending 100 packets with a 4 bytes payload, for the 10 hops scenario.

TABLE VII. 8 VS 4 BYTE PAYLOADS

Packet loss rate	50 packets 8 bytes data	100 packets 4 bytes data	Saved energy
No packet loss	13640	18040	4400
10% packet loss	14960	20240	5280
20% packet loss	16280	22440	6160
30% packet loss	17600	24640	7040
40% packet loss	18920	26840	7920

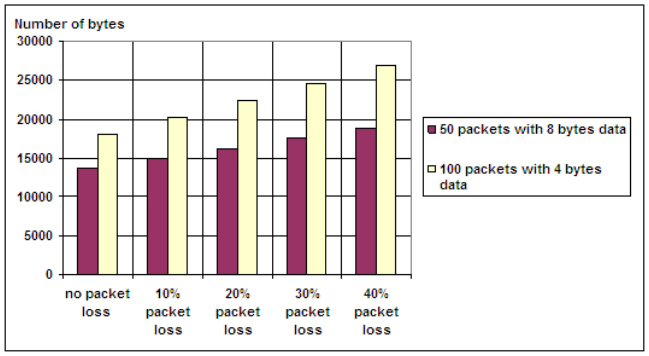


Figure 9. 8 vs. 4 byte payloads

We observe that, in each case, sending 50 packets with 8 byte payloads consumes less energy than sending 100 packets with 4 byte payloads. The saved energy is estimated to 4.4 MB when no packet is lost, 5.2MB for 10% packet loss, and 7.9MB when 40% of the packets are lost. As the percentage of lost packets increases, the saved energy also increases.

 ISBN: 978-1-61208-133-5

VI. CONCLUSIONS

We have previously developed a security protocol (AASP) that provides authentication, integrity, anti-replay and reliability. In this paper we present protocol optimizations that reduce energy consumption: the control overhead has been minimized, the number of handshake packets has been reduced, NACKs have been used for selective data retransmission, and benefits of data aggregation have been evaluated.

We implemented the improved security protocol (EAASP) in TinyOS as two layers in the communication stack, and we have used TOSSIM to run several test scenarios in order to demonstrate its functionality and to evaluate its performance.

We have developed a mathematical model in order to determine energy consumption. In several test scenarios we have estimated the energy consumption, we have evaluated the control overhead, and we have determined the energy saved by optimizations and also by aggregating data.

The formal evaluation proves that EAASP provides substantial energy savings in relation to AASP. Data aggregation can be further used, when possible, to increase energy efficiency.

The Energy-efficient Security Protocol is an appropriate choice when authentication, integrity and anti-replay are required for low-power devices. We have used simulation and mathematical results to prove the energy-efficiency of all introduced optimizations.

In further work we will use the QoS bits to prioritize certain packets, such as negative acknowledgements and re-sent packets, in order to speed up the retrieval of lost packets. The QoS bits may also be used in order to reduce energy consumption caused by traffic congestion.

In a future evaluation, we will compare our security protocol with other state of the art security protocols, such as TinySec and SPINS, regarding energy consumption, and the strength of authentication, freshness and reliability mechanisms.

ACKNOWLEDGEMENTS

The work has been partially funded from the Sectoral Operational Programme Human Resources Development 2007-2013 of the Romanian Ministry of Labour, Family and Social Protection through the Financial Agreement POSDRU/88/1.5/S/60203, partially from the national project "Excellence in research through postdoctoral programs in priority areas of knowledge-based society(EXCEL)", Project POSDRU/89/1.5/S/62557 and partially from POSCCE project GEEA 226 - SMIS code 2471, which is co-founded through the European Found for Regional Development inside the Operational Sectoral Program "Economical competivity improvement" under contract 51/11.05.200.

REFERENCES

[1] J. Zheng and A. Jamalipour, *Wireless Sensor Networks: A Networking Perspective*, Wiley-IEEE Press, 2009.

[2] M. Winkler, K.D. Tuchs, K. Hughes, and G. Barclay, "Theoretical and practical aspects of military wireless sensor networks," *Journal of Telecommunications and Information Technology*, vol. 2, 2008, pp. 37-45.

[3] T. Kavitha and D. Sridharan, "Security Vulnerabilities In Wireless Sensor Networks: A Survey," *Journal of Information Assurance and Security*, vol. 5, 2010, pp. 31-44.

[4] D. Boyle and T. Newe, "Security protocols for use with wireless sensor networks: A survey of security architectures," *Third International Conference on Wireless and Mobile Communications*, 2007, pp. 54-59.

[5] A. Perrig, R. Szewczyk, J. Tygar, V. Wen, and D.E. Culler, "SPINS: Security protocols for sensor networks," *Wireless networks*, vol. 8, 2002, pp. 521-534.

[6] P. Levis, S. Madden, J. Polastre, R. Szewczyk, A. Woo, D. Gay, J. Hill, M. Welsh, E. Brewer, and D. Culler, "TinyOS: An operating system for sensor networks," *Ambient Intelligence*, 2004, pp. 115-148.

[7] C. Karlof, N. Sastry, and D. Wagner, "TinySec: a link layer security architecture for wireless sensor networks," *Proceedings of the 2nd international conference on Embedded networked sensor systems*, ACM, 2004, pp. 162-175.

[8] L. Gheorghe, R. Rughiniş, R. Deaconescu, and N. Ţăpuş, "Authentication and Anti-replay Security Protocol for Wireless Sensor Networks," *The Fifth International Conference on Systems and Networks Communications, ICSNC 2010*, 2010, pp. 7-13.

[9] L. Gheorghe, R. Rughiniş, R. Deaconescu, and N. Ţăpuş, "Reliable Authentication and Anti-replay Security Protocol for Wireless Sensor Networks," *The Second International Conferences on Advanced Service Computing, SERVICE COMPUTATION 2010*, 2010, pp. 208-214.

[10] P. Levis, N. Lee, M. Welsh, and D. Culler, "TOSSIM: Accurate and scalable simulation of entire TinyOS applications," *Proceedings of the 1st international conference on Embedded networked sensor systems*, ACM, 2003, pp. 126-137.

[11] M. Amiri, "Evaluation of Lifetime Bounds of Wireless Sensor Networks," *Arxiv preprint arXiv:1011.2103*, vol. abs/1011.2, 2010.

[12] J. Hill, R. Szewczyk, A. Woo, S. Hollar, D. Culler, and K. Pister, "System architecture directions for networked sensors," *ASPLOS-IX: Proceedings of the ninth international conference on Architectural support for programming languages and operating systems*, vol. 35, Dec. 2000, pp. 93-104.

 ISBN: 978-1-61208-133-5

Stateful or Stateless Flooding Attack Detection?

Martine Bellaïche
Génie Informatique et Génie Logiciel
École Polytechnique de Montréal
Montréal, QC, CANADA
email: martine.bellaiche@polymtl.ca

Jean-Charles Grégoire
INRS–EMT
Montréal, QC, CANADA
email: gregoire@emt.inrs.ca

***Abstract*—SYN flooding attacks exploit the 3-way handshake characteristic of TCP connection setup to cause denials of service. Many techniques have been proposed for the detection of flooding attacks; most are stateless while a few are stateful. A stateful method keeps specific information on flows of packets while stateless methods will only keep counters on specific packet features. The low performance impact of stateless methods has led to their predominance in practical deployments. We introduce a methodology to support a comparison between methods, which allows to quantify all key factors which can be used to assess and compare performance and see how they can be built into a metric. In this article, we evaluate and compare the performance of two key DoS detection techniques, one stateless and one stateful, and investigate their relative merits.**

***Keywords*—Denial of Service; SYN Flooding; TCP Handshake; Network Security.**

I. Introduction

Most internet based services rely on the TCP protocol. Establishment of TCP connections is based on a handshake, more specifically a 3–way handshake (exchange of 3 packets), to reserve and announce suitable resources at both ends before data exchange can proceed. This mechanism has however proven to be quite vulnerable to denial of service (DoS) attacks on servers, which have for objective to stop legitimate users from using a service by overloading it with connection establishment requests. A distributed DoS attack (DDoS) occurs when a large number of nodes wage such an attack simultaneously.

SYN flooding attacks represent 90 % of most DDoS attacks [1]. The goal of the attack is to tie the memory of server machines with half–open connections. A large number of attack machines send an important number of connection set-up requests to a single server and, consequently, legitimate clients cannot connect any more to the server, whose resources have been depleted.

Many techniques have been proposed for the detection of flooding attacks; most are stateless while a few are stateful. A stateful method keeps specific information on packets crossing the router while stateless methods will only keep counters. The low overhead of stateless methods has led to their predominance in practical deployments; yet we would want to know if a stateful method performs significantly better than a stateless one: As we use more memory and, to a lesser degree, more CPU in stateful techniques, we want significant improvements in detection time, detection rate and rate of false alarms to justify their use. We also want to know if they are more robust towards evasion of detection.

In this article, we describe several stateless and stateful techniques for flooding attack detection and compare the performance of two key techniques, representative of each kind. The originality of this work lies in: (1) reviewing of the state of the art in stateless and stateful attack detection, (2) presenting a method for evaluating performance detection system, and (3) comparing stateless and stateful methods to establish their relative merit.

This paper is organized as follows. Sections II looks over previous work in detection and isolates two key methods for our comparison. In Section III, we introduce elements of comparison between methods and proceed to an evaluation of one stateful technique vs. a stateless one in Section IV. We discuss our results and conclude in Section V.

II. State of the Art

Because of space constraint, we do not cover extensively the full range of stateless and stateful techniques, restricting our study to two, representative techniques and highlight differences which are specific to each kind. The reader will be referred to the bibliography for further reading.

A. State of the Art for Stateless Techniques

Stateless techniques use packet counting and statistical analysis (e.g. CUSUM) to detect an attack. Packet information is tallied in a random variable X_n over an observation period (and not continuously). X_n has taken many forms, based directly on protocols (Blazek and al. [2]), traffic correlation (Jin and al. [3]) or behaviour (Ohsita and al. [4]), the presence of specific packets or packet sequences (Siris and al. [5], Shin and al. [6]).

Wang and al. [7] propose a simple mechanism to detect SYN flooding attacks by monitoring the normal behaviour of TCP. Their stateless Flooding Detection System (FDS) has low computation overhead. For the normal behaviour of TCP, there must be a match between the number of TCP FIN (or RST) packets and TCP SYN packets over all TCP connections. Using the CUSUM method, they

 ISBN: 978-1-61208-133-5

record the number of SYN and FIN (or RST) packets and detect discrepancies. The difference between the number of SYN and FIN (RST) is normalized by an estimated average number of FIN (RST), to ensure that the FDS is independent of sites and access patterns. As most TCP connections last from 12 to 19 s, they set the duration of observation periods to 20 s.

The weakness of stateless detection methods is that the attacker can send a mix of packets to thwart detection. Also, such counting strategies are vulnerable because Internet traffic is bursty and the detection may therefore raise false alarms (see e.g. [8]). They also lead to rather long detection periods.

B. State of the Art for Stateful Technique

Stateful techniques rely on a memory of past events such as occurrence of source addresses (Schuba and al. [9]), analysis of current condition changes in traffic patterns due to congestion (Xiao and al. [10]) or other factors (Gil and Poletto [11], Cheng and al. [12]).

In [13] we have proposed the Unusual Handshake Detection (UHD) method. TCP handshakes whose sequence does not follow the 3 steps standard are recognized as unusual handshakes. Those are typically the result of network congestion and—sometimes—router errors or unreachable ports; but during DDoS, they can also be the result of the attack. This work concentrates on detection from the server side, at the last mile router, and looks at handshakes from that perspective only. A dedicated data structure stores all information on the TCP handshakes. For each flow, the IP source and destination addresses, the source and destination ports, the arrival time of the last SYN packet of a new TCP flow and the flag of the TCP packet are stored. The data structure keeps track of an estimate of TCP connection latency (RTT, plus delay for memory allocation for the TCP data structure) per source network, to set the detection delay for the unusual handshakes.

Stateful techniques require memory to support monitoring. How memory is managed is critical as the available space may be exhausted with increasing traffic [10] [11] [12] [13]. Such detection techniques must therefore be able to detect attacks very quickly to be resource–effective.

III. Framework for Comparaison

As we are interested in comparing stateful and stateless detection, we use and adapt the methodology we have presented in [14] for detection. Our purpose is to quantify all factors which can be used to assess and compare performance. It is also possible to construct an aggregate metric from the different factors used to evaluate the quality of detection to end up with a unique performance number.

In order to protect the victim efficiently, the essential objectives are to detect attacks quickly, with accuracy and with minimal deployment costs. Deployment costs will reflect the complexity of the detection method, measured according to the changes it requires compared to a defenceless service architecture. These overall objectives translate into the following criteria: accuracy, latency—or detection time, deployment cost and robustness, which can be related to specific measures.

A. Performance Measures

a) Detection Rate: The detection rate is the percentage of attacks that are detected as compared to the total number of attacks [5]. This metric—associated with the detection time—validates the detection mechanism for each attack. Similarly, the non-detection rate—or false negative rate—is a way of determining the errors made by defences for not identifying the attacks. It corresponds to the percentage of non-detected attacks compared to the total number of attacks. It is the complement of the detection rate.

b) Rate of False Positives: The rate of false positives or the rate of false detection alarms [15] is another way of assessing detection errors made by identifying an attack when none occurred. This rate is the ratio between the number of erroneously-reported attacks and the total number of attacks. This metric verifies that the detection mechanism does not make (significant) mistakes. For example, we want to know if an increase in traffic or a flash crowd can cause false alarms.

c) Latency: The detection time—or latency—metric reflects the delay in the detection of attacks. The detection time of the attack is the duration of the time interval between the beginning of an attack and its detection. The detection time is important because an attack should be detected before any severe damage is done.

The latency depends on a number of elements: there are architectural constraints, for example a polling cycle to acquire data, and algorithmic constraints, such as the existence of a time window to average the information over several acquisition cycles.

B. Deployment Costs

The deployment costs of the defence system depend on the computation time, the memory overhead, the bandwidth overhead and the system complexity as explained below. In fact, we want to evaluate the increase of these costs due to the deployment of the detection system. To evaluate the different elements, we need to perform two experiments: a first one to find the baseline value of deployment costs, in the absence of attacks and a second one to evaluate the increase from the baseline value.

1) *Computation overhead* in ms: The time required to process the measured data.
2) *Memory* in Kbytes: The storage space necessary for the implementation of the detection mechanism.
3) *Bandwidth overhead* in %: Should the detection method imply the transmission of some form of control messages (e.g. throttle), then this in turn would yield a reduction of the available bandwidth.
4) *Deployment complexity*: The deployment complexity, measured from 1 (low) to 4 (high cost). This measure

 ISBN: 978-1-61208-133-5

Deployment	Cost fct.	Priority
Complexity y	$f_y(y) = y \times Y$	p_y
Bandwidth overhead b	$f_b(b) = B$ (%)	p_b
Computation overhead c	$f_c(c) = C$ (ms)	p_c
Memory m	$f_m(m) = M$ (KB)	p_m

TABLE I: Cost Functions

depends on whether the detection strategy involves one or several nodes and whether it involves numerous and substantial modifications to the network.

Finally, the installation of the detection system should not increase the deployment costs, i.e. it should be integrated with another network device.

C. A Composite Metric

We have shown in [14] how these different measures can be combined into a composite metric, through a weighted sum to emphasise certain metrics. Such a composite gives a global evaluation of the objectives and quality of the method. Here, we show how to effectively compute such a metric. To that end, we need (1) a list of the relative priorities of these elements, (2) a cost function for each element to have uniform comparison units, (3) values for the weighting coefficient determined by the priority list.

The priority values are attributed according to the cost functions. A low priority value represents the cost of an easy deployment. For example, the composite metric for the deployment cost DC is expressed by:

$$DC = \alpha_c \times f_c(c) + \alpha_m \times f_m(m) + \alpha_b \times f_b(b) + \alpha_y \times f_y(y) \quad (1)$$

with computation overhead c, memory m, bandwidth overhead b, and deployment complexity y, and the matching weighting factors $\alpha\_$, the cost functions $f\_$ and the priority p between $[1, 4]$ (see Table I).

Similarly, for the performance measure, we build D as follows.

$$D = \alpha_l \times l + \alpha_n \times n + \alpha_p \times p \quad (2)$$

with latency l in s, rate of false negatives n in %, and rate of false positives p in %. Each performance measure has a priority p between $[1, 3]$. As the performance measure is not a cost, we do not use cost functions. An ideal detection technique must have a short latency l, as well as, a rate of false negatives n and positives p as low as possible.

We develop further in Section IV how the weighting factors can be chosen to build a meaningful composite.

D. Robustness

Robustness is a critical evaluation of a defence and, unlike previous metrics, it is difficult to define it in terms of a specific cost.

What we require, in our evaluation of defences, is the assessment of the effectiveness of the detection as an attack proceeds. In this case we cannot simply reduce such assessment to a unique metric, as we expect performance not to be constant, but to depend on the legitimate traffic load and characteristics.

For different quality factors, e.g. false positives, false negatives, latency, and sensitivity (low threshold) we want to identify the detection weaknesses:

1) **False Positives (or False Alarms)**: which traffic conditions increase the rate of false positives?
2) **False Negatives**: at which rate is it possible to avoid detection?
3) **Latency**: how significantly does detection increase latency?
4) **Sensitivity**: how do we establish a detection threshold?

On this last point, we note that whereas the value of the threshold is not too significant when a server is unloaded, it must be kept as low as possible when we have a high usage, to preserve useful traffic while providing detection. Sensitivity is thus denoted by the ability to set a threshold to allow detection while keeping rates of false positives and false negatives low.

We therefore propose that robustness be examined as a standardized test, at a specific usage level. We must note however that this picture is not complete as stateless measures can be fooled by specific forms of attacks which supply the relative number of TCP messages they expect, hence creating a large number of false negatives. Such an assessment is required to complement sensitivity/quality tests. In the following comparison, we will look at the behaviour of the specific parameters of robustness assessment without attempting to build a composite picture.

IV. Comparison between Stateful and Stateless Detection Methods

We choose the following techniques for a comparison according to the methodology presented above: the FDS of Wang and al. [7] for the stateless case and our own UHD [13] for the stateful one. These techniques use for the detection the behaviour observation of the TCP protocol and the CUSUM algorithm to confirm the attack.

The $\alpha\_$ values are set as follows.

- As memory is cheap, we assign 1 to the priority because it is not a significant contribution to deployment costs. We calculate $\alpha_m = 0.1$.
- As the processors are more and more powerful, for the computation overhead, we assign 2 to the priority and obtain $\alpha_c = 0.2$
- As we want save network resources, for the bandwidth overhead, we assign 3 to the priority and set $\alpha_b = 0.3$.
- To encourage minimal deployment complexity, we assign 4 to the priority $\alpha_y = 0.4$.

These values are chosen because some resources are easy to obtain and are not too significant, while some elements are very important and play a fundamental role in the computation of the deployment cost (DC).

A. Evaluation of FDS

a) Performance Measures: From [7], the detection rate is within the range $[70\%, 100\%]$ and the rate of false positives is null. The detection time is within $[20\,s, 487\,s]$.

 ISBN: 978-1-61208-133-5

We have set an interval for the overall evaluation of detection of $D = [6.6, 160.7]$. But in practice (see Section IV-A0c) the method can trigger false positives during flash crowds.

b) Evaluation of Deployment Costs: The computation overhead for the FDS system represents the time required for packet classification and addition, and is not evaluated in the article. As the stateful technique also needs packet classification, it is not very important to determine its computation overhead. Also, the addition operation time is rather insignificant. The computer overhead is therefore fixed to 0. For storage, FDS uses only two integers for compiling the number of SYN and FIN packets, so the memory is only 8 bytes. There is no bandwidth overhead. We fix the level of the deployment complexity to 1, because it is very easy. With the weighting coefficient and the cost function, we evaluate the composite metric to $DC = 0.1 \times 8 \times 10^{-3} \times f_m(m) + 0.4 \times 1 \times f_y(y)$.

c) Robustness: The count of SYN and FIN packets of the FDS technique is not very reliable for the following reasons:

False Positives or False Alarms. The FDS does not consider whether a SYN packet is retransmitted, which does not follow proper TCP behaviour that associates one FIN packet for one SYN packet. This discrepancy can lead to false alarms. In one observation period we could observe a large number of TCP connections with a duration significantly longer than average, or alternatively a flash crowd, either of which could trigger a false alarm because the FIN packet will be counted in a later observation period and, as a consequence, would lead to an imbalance between the count of SYN and FIN packets.

False Negatives. The weakness of counting SYN–FIN pairs is that the attacker can flood a mixture of SYNs and FINs in equal numbers. As the consequence, the clever attacker can evade the FDS detection technique.

Latency. As the observation period corresponds to the duration of TCP connections ($20\,s$), the detection time is the number of the observation periods. In most cases, the detection is therefore triggered after the TCP connection ends.

Sensitivity. FDS fixes the threshold and varies the attack rate to evaluate the detection rate.

For all these reasons, the robustness of FDS is low.

B. Evaluation of UHD

a) Performance Measures: As indicated in the paper, the detection rate is 100 % and the false alarm rate 0 %. The detection time is therefore between $30\,s$ and $70\,s$. With a weighting coefficient of 0.33, we evaluate the overall evaluation in between $[9.9, 23.1]$

b) Evaluation of Deployment Costs: We have used a form of XOR-folding, also called bit-folding or bit-extraction as a hashing function. It is a practical manipulation of bits combining shifts, masking and logical combinations. With XOR-folding, it is easy to construct a function which will be robust to the permutation of information; the only challenge is to find the combination which generates the most dispersion and this can depend on the nature (i.e. regional character) of the server. Such issues are however beyond the scope of this paper.

On the Intel Duo processor at 3.00 GHz used for our experiments, the insertion time and the scanning time are $3.95\,ms$ and $20.62\,ms$ respectively. These values are very small. As the networks are identified by 24 bits of an IP address, the hashing function uses XOR-folding of the two halves of the 24 bits address into 12 bits—for a basic table size of 16384 B. In this case, we have observed on the same data an average length of the chains of 1.65, and an occupancy rate of 31.7 %. Under normal conditions, we require extra memory which amounts to at most 21 KB for the handshake information.

The technique does not use any bandwidth for detection. The deployment is very easy and this technique can be built into a last mile router. With the weighting coefficients and the cost functions, DC evaluates as: $0.1 \times (21 + 16.384) \times f_m(m) + 0.2 \times (3.95 + 20.62) \times f_c(c) + 0.4 \times 1 \times f_y(y)$.

c) Robustness: As the principle of the detection is stateful, the attack can exhaust the memory with enough variety of unusual handshakes. But the detection is fast, and flow information is reset every period, so the attack is detected quickly, before using up all the router memory.

False Positives or False Alarms. As the technique does not count the packets, UHD is not vulnerable to flash crowds and consequently, it does not produce false alarms.

False Negatives. Of course, if the attacker knows the principle of the detection, he can try to send a flood of SYN packets followed by ACK packets to keep a reasonable balance of SYN vs. non–SYN packets. This will be undetected unless we also keep track of TCP sequence numbers in the data structure. The server, however, would quite likely reset the connection because of wrong sequence numbers, which would lead again to another form of unusual handshake.

Latency. The detection attack can be caught right from the beginning as the technique observes the TCP handshake. However, as the detection time is linked with the observation period, such a short observation period involves a quick detection time.

Sensitivity. From real traces and with merging fictitious SYN flooding attacks, we have run tests with an attack rate fixed at 25 % of normal traffic. In Figure 1, we show the importance of the value the entropy threshold. A "wrong"—or too tight—value can lead to numerous false alarms. As UHD detects attacks when the entropy value is below the threshold, to evaluate the sensitivity of this value to the detection of false alarms, we measure this rate while increasing the threshold. Moreover, the start value of the threshold represents the detection value of a trace. In Figure 1 we see that, as the threshold increases linearly beyond a threshold of .44, the number of false alarms

 ISBN: 978-1-61208-133-5

increases exponentially. In practical terms, this shows that the number of false alarms is highly sensitive to the level of the threshold and decreases exponentially as the threshold is set lower. This, in turn, means that 1) the threshold does not need to be unduly low to be effective and 2) it can be set to resist to some degree of fluctuations in traffic characteristics.

For all these reasons, the robustness of UHD is high.

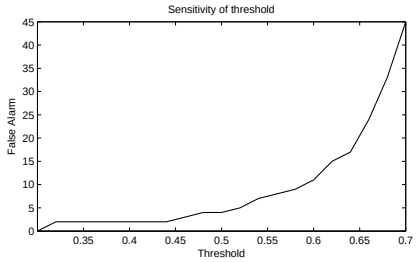


Fig. 1: Number of false alarms as function of threshold

C. Summary

In Sections IV-A and IV-B, we have applied a method for evaluating the performance of the detection mechanisms. For a comparison of the deployment cost of the two techniques, once we know the real value of the cost functions, we can evaluate DC.

The greater value of the DC metric reflects a better evaluation performance for the stateful technique. We can observe a shorter detection time interval for the UHD technique than for the FDS technique and we can conclude that, as the technique is stateful, detection is faster and more accurate. Also, the UHD stateful technique is robust and the FDS mechanism is not as strong as an attacker can evade the mechanism and it can produce false alarms during flash crowds. While stateless methods have varying degree of sensitivity to these issues, they are nevertheless exposed to them.

V. Conclusion and Discussion

From our comparison of two detection techniques and for other techniques, and following our assumption that each technique is representative of its genre, we have observed that stateless detection is slower and less reliable than a stateful detection technique. Also stateless techniques cannot respond to the detection as they do not store information and, as a consequence, cannot as effectively selectively stop the attack packets whereas stateful techniques store data, which can be used to react to the attack once it has been detected such as throttling or blocking attack traffic [16].

Stateful techniques demonstrate significant improvements in (1) the robustness of detection in the presence of detection evasion techniques (false negatives) or errors (false positives), (2) detection time, detection rate and the false alarm rate, and (3) the possibility of using the information collected by the technique to stop or control the attack.

Stateful techniques however use more memory and, to a lesser degree, more CPU cycles and have higher deployment cost: they would tend to require dedicated pieces of equipment whereas stateless techniques would more easily be embedded in routers.

Finally, we should remind the reader that these techniques can be viewed as complementary. While we have established the superiority of stateful techniques close to the edge of the server's network, stateless methods can be useful closer to the core of the network, where resources are scarce, but detection efficiency useful, albeit less critical.

Further research underway aims at better refining and defining this complementarity to extend our framework to hybrid models.

References

[1] D. Moore, G. Voelker, and S. Savage, "Inferring Internet Denial-of-Service Activity," in *USENIX Security Symposium*, August 2001.

[2] R. B. Blazek, H. Kim, B. Rozovskii, and A. Tartakovsky, "A Novel Approach to Detection of Denial of Service Attacks via Adaptive Sequential and Batch Sequential Change Point Detection Methods," in *Workshop on Information Assurance and Security, IEEE*, June 2001.

[3] S. Jin and D.S.Yeung, "A Covariance Analysis Model for DDoS Attack Detection," in *IEEE International Conference on Communications*, vol. 4, June 2004, pp. 1882 – 1886 Vol.4.

[4] Y. Ohsita, M. Murata, and S. Ata, "Detecting Distributed Denial-of-Service Attacks by Analyzing TCP SYN Packets Statistically," in *IEEE, GLOBECOM*, December 2004.

[5] V. A. Siris and F. Papagalou, "Application of Anomaly Detection Algorithms for Detecting SYN Flooding Attacks," in *IEEE, GLOBECOM*, December 2004.

[6] S.-W. Shin, K.-Y. Kim, and J.-S. Jang, "D-SAT: Detecting SYN Flooding Attack by Two-Stage Statistical Approach," in *The Symp. on App. and the Internet, IEEE*, Janurary 2005, pp. 430–436.

[7] H. Wang, D. Zhang, and K. G. Shin, "Detecting SYN Flooding Attacks," in *Proceedings of IEEE INFOCOM*, vol. 3, June 2002, pp. 1530 – 1539.

[8] R. Mahajan, S. Bellovin, S. Floyd, J. Vern, and P. Scott, "Controlling High Bandwidth Aggregates in the Network," *SIGCOMM Comput. Commun. Rev.*, vol. 32, pp. 62–73, July 2002.

[9] C. L. Schuba, I. V. Krsul, M. G. Kuhn, E. H. Spafford, A. Sundaram, and D. Zamboni, "Analysis of a Denial of Service Attack on TCP," in *Proceedings of the IEEE Symposium on Security and Privacy*, 1997, pp. 208–.

[10] B. Xiao, W. Chen, Y. He, and S. E.H.-M., "An Active Detecting Method Against SYN Flooding Attack," in *Proceedings. 11th International Conference on Parallel and Distributed Systems*, July 2005, pp. 709 – 715 Vol. 1.

[11] T. M. Gil and M. Poletto, "MULTOPS: A Data Structure for Bandwidth Attack Detection," in *Usenix Security Symposium*, August 2001.

[12] J.Cheng, J. Yin, Y. Liu, Z. Cai, and M. Li, "DDoS Attack Detection Algorithm Using IP Address Features," in *Lecture Notes in Comp. Sc., Springer Verlag*, vol. 5598/2009, 2009.

[13] M. Bellaïche and J.-C. Grégoire, "SYN Flooding Attack Detection Based on Entropy Computing," in *IEEE GLOBECOM*, November 2009.

[14] ——, "Measuring Defense Systems Against Flooding Attacks," in *IWCMC 2008, International Wireless Communications and Mobile Computing Conference*, August 2008, pp. 600 –605.

[15] R. Chang, "Defending Against Flooding-based Distributed Denial-of-Service Attacks: a Tutorial," *Communications Magazine, IEEE*, vol. 40, no. 10, pp. 42–51, October 2002.

[16] T. Peng, C. Leckie, and R. Kotagiri, "Protection from Distributed Denial of Service Attack Using History-based IP Filtering," in *International Conference on Communications*. IEEE, June 2003, pp. 482 – 486 vol.1.

 ISBN: 978-1-61208-133-5

Investigation of Visible Light Communication Transceiver Performance for Short-Range Wireless Data Interfaces

Hongseok Shin, Sungbum Park, Kyungwoo Lee,
Daekwang Jung, Youngmin Lee
DMC R&D Center,
Samsung Electronics Corp.,
Maetan 3 Dong 416, Suwon, S. Korea
hongseok.shin@samsung.com

Seoksu Song, Jinwoo Park
School of Electrical Engineering,
Korea University
Anam-dong, Seungbuk-gu, Seoul, S. Korea
jwpark@korea.ac.kr

***Abstract* — We investigate the performance of a visible light communication (VLC) transceiver for bi-directional high-speed and short-range wireless data interfaces. The proposed VLC transceivers with optical antenna structure are implemented with edge-emitting laser diode and silicon photo diode, which is primarily designated to operate in a full duplex mode at 120 Mbit/s. The shielding method that is employed to reduce the light cross coupling effect inside the VLC transceiver is proposed and experimentally investigated. The influence of the tilt degree between two transceivers without optical antenna and with optical antenna is investigated. Their bit error rate performance is examined experimentally with respect to the transmission distance, the coverage range and the tilt degree.**

Keywords - Visible light Communication; Free-space optical communications; optical wireless

I. INTRODUCTION

The various communication technologies have been advanced to process the immense amount of data information at a very high speed. Among them, recently, visible light communication (VLC) technology is attracting much attention as short range communication means of high speed. This technology uses light-emitting diodes (LEDs) emitting light with the wavelength interval of 380-700 nm to carry information. The VLC technology is a novel kind of optical wireless communication technology with high potentially which can play a supplementary role of wireless communication which is available at any time and any place. As supplementary system, VLC has many advantages compared to RF-based wireless communications [1]; (1) It can potentially use existing local power line infrastructure for wireless communication as a backbone, (2) The bandwidth which can use is virtually unlimited, (3) The security is very outstanding, that is, it is difficult for an intruder to pick up the signal from outside due to characteristic of light, (4) Transmitters and receivers using LEDs are cheap and there is no need for expensive radio frequency units, (5) Visible light radiations are free of any health concerns, (6) Furthermore, no interference with RF based systems exists, so that the use in airplanes or hospitals is uncritical. Currently, the VLC based on these advantages has been mainly investigated into various applications, such as ubiquitous communication system, intelligent transport communication system and illuminating light communication system [2], [3].

We have paid attention to use VLC technology in the high speed and short range communication as a peripheral interface of hand-held devices such as mobile phones, notebook computers, digital cameras and so on. Since users can actively align communication links by observing the visible beam spot, VLC transceiver does not necessarily demand a wide coverage, implying that its power consumption can be potentially lower than other invisible options. Another key issue to investigate feasibility of wireless optical connectivity technology is channel efficiency. Full duplex operation can significantly increase the efficiency of communication channel, but the self reflection of transceiver degrades the transmission performance. It is also difficult for users to align VLC transceivers to be faced each other for a long time due to the characteristic of portable equipment. Based on those observations, we investigated a wireless optical transceiver especially focusing on the high speed and short range visible communications. In this letter, we demonstrate a practicability of VLC transceiver experimentally, which was designed to operate at 120 Mbit/s in the full duplex mode with expectation to be used as a peripheral interface of hand-held devices.

II. THE PROPOSED VLC TRANSCEIVER

We tried to develop a VLC transceiver in a small package by gathering optical and electronic technologies and components currently available in a commercial market. The optical part of VLC transmitter consisted of three devices; a light source, collimation lens and a diffuser. For the compatibility with physical line speed of Ultra Fast Infrared and for competing with the matured RF-based connectivity technologies such as WiFi or UWB, the VLC bandwidth was set to 120 Mbit/s. It is known unfortunately that the LEDs for ambient artificial illumination or message signboard typically have the modulation limit of about 10 Mbit/s [4]. The resonant-cavity LEDs (RCLEDs) for plastic fiber communications have wide modulation bandwidth but do not have enough power to provide sufficient visibility. Edge emitted LDs in visible wavelength, on the other hand, have

 ISBN: 978-1-61208-133-5

higher optical output power and show the better visibility compared to RCLEDs. In these regards, we adopted edge emitting LDs for the high speed short range visible light source. A diffuser was placed in front of LD to avoid the eye safety regulation strictly applied to laser sources [5]. The center wavelength of edge emitted LD was 635 nm and the full width half maximum spectral width was about 5 nm. The beam divergence of the red edge emitted LD was also engineered to be less than 10° by placing a diffuser and collimation lens in front of the metal can package. The beam spot was visible at the distance of up to 1.2 m in a typical office environment. The proposed VLC system uses on-off keying modulation. The measured optical power after a diffuser was about 1.5 mW.

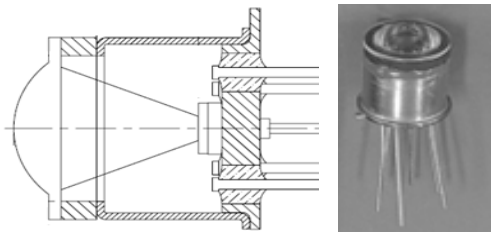


Figure 1. Drawing and picture of the designed optical antenna.

We investigated two types of VLC receiver, which utilize two methods of concentrating incoming light at the photo diode (PD). In the first method, a convex lens is to concentrate light which is available in the market. The size of the convex lens was chosen based on the beam divergence and the transmission distance, noting that a convex lens with bigger diameter has in general the smaller angle of field of view (FOV). In the designed VLC receiver, a convex lens with 7-mm diameter was used in front of PD. The second method uses an optical antenna as shown in Fig. 1, which was particularly designed for VLC receiver applications. Since optical antenna generally requires a more exact alignment, the optical antenna needs to be designed for a wider FOV. The measured FOV of our optical antenna is about ±10° . In addition, the surface of inside of the VLC receiver is coated for full reflection to help concentrating the incoming light. The ambient light is a noise source to the VLC optical receiver. The major source of ambient light inside building is indoor lightings. The power spectrum of the photo detector output in the presence of fluorescent light extends up to 100 kHz [6]. An electrical high-pass filter with 300 kHz cutoff-frequency was equipped right after a photo diode to reduce the influence of ambient light. No additional optical filter was used in the proposed VLC receiver.

III. THE PERFORMANCE EXAMINATIONS

Fig. 2 is a schematic of the experimental setup to examine the designed VLC transceivers. VLC transceiver 1 was connected to a pulse pattern generator (PPG), PPG1 which generated a 2^7-1 pseudorandom binary sequence (PRBS) at 120 Mbit/s. Another transceiver, VLC transceiver 2, was connected to an error detector, so that two VLC transceivers face each other as shown in Fig. 2. Transceiver 2 moved along X axis and Y axis while measuring the performance of the VLC transceiver. The bit error rates (BERs) were measured varying the transmission distance of X cm and the coverage range of Y cm. The distance and coverage are varied from 5 cm to 130 cm and from -10 cm to +10 cm, respectively.

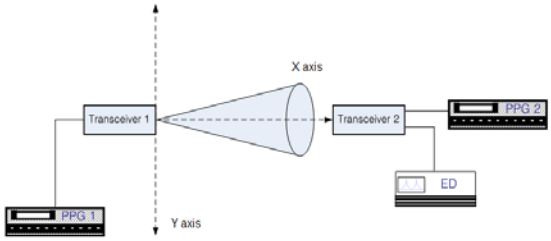


Figure 2. Experimental setup for performance measurements.

We designed the VLC transmitters to have the beam profile which can provide the uniform optical power distribution over the entire shining circle, so that the BERs at a distance were almost same within a coverage range. It should be also noted that a clear boundary of the shining circle is observed, which is beneficial to clear visibility of the spot. One of the major reasons that limit the usage of full duplex mode in optical link is the cross coupling of light. There may be some tricky cases for the optical link applications where detrimental light scattering is serious and a receiver can be blinded by the light of its own transmitter. We carried out the measurements with and without the presence of the cross coupling light. In Fig. 2, Transceiver 2 is arranged to transmit a PRBS signal by turning on PPG2, which generates a cross coupling light interference to Transceiver 1 under investigation.

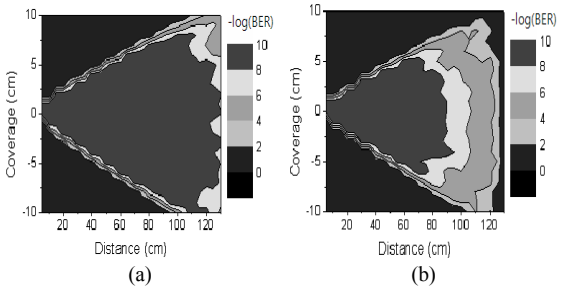


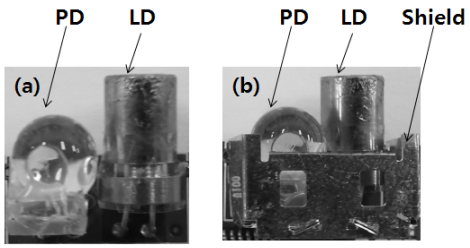
Figure 3. The BER performance of a transceiver at 120 Mbit/s over the visible link as a function of transmission distance and coverage (a) without cross coupling light (b) with cross coupling light.

Fig. 3 is the comparison of the BER performance of the proposed VLC transceivers without and with cross coupling light. Fig. 3 (a) shows that VLC system without cross coupling light can provide BERs lower than 10^{-8} at the distance of about 110 cm and within the coverage of about 17.5 cm. But, Fig. 3(b) reveals that with cross coupling light the distance and the coverage for successful communications were reduced to about 70 cm and about 11 cm respectively. Note that the divergence angles for successful

 ISBN: 978-1-61208-133-5

communications were not affected by the cross coupling light.

A metal shield between LD and PD was utilized to block the scattered or reflected light as shown in Fig. 4(b).



(a) (b)

Figure 4. Pictures of VLC transceivers
(a) without a shield (b) with a shield.

Fig. 5 shows the BER performance improvement achieved by reducing the influence of cross coupling light, extending the transmission distance and coverage range up to about 100 cm and about 15 cm respectively.

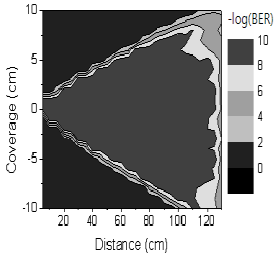
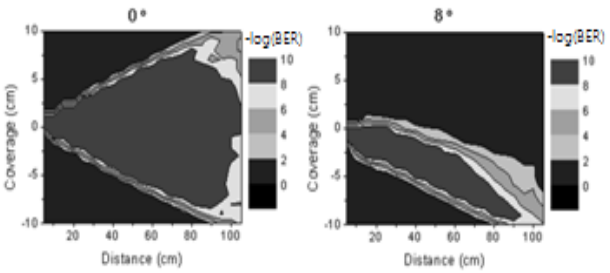
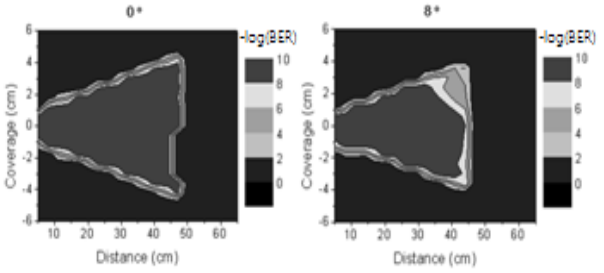


Figure 5. The BER performance of a transceiver with a shield in the presence of cross coupling light.

The tilting effect of the VLC transceiver was investigated because it is generally difficult for users to locate or to maintain the VLC transceivers facing each other exactly in many practical applications. While VLC transceiver 2 was tilted by 8° intentionally in Fig. 2, the BER performance was measured to obtain the Fig. 6. It was found out that the VLC transceiver equipped with an optical antenna experiences greater reduction in the transmission distance for successful communications with the tilting angle, but reveals better coverage range characteristics, in comparison with the VLC transceiver with a convex lens.



(a)



(b)

Figure 6. The BER performance of transceiver at 0 and 8 degree tilt
(a) with a convex lens (b) with an optical antenna.

IV. CONCLUSION

We demonstrate the feasibility of the visible light transceiver for the high speed and short-range peripheral wireless interface applications of hand held devices. The proposed VLC design approach features What-You-See-Is-What-You-Send security by employing visible lights as communication media. This paper proved the practicability of a 120 Mbit/s VLC transceiver using an edge emitted LD and a silicon PD, by presenting the BER performance measurements with respect to the transmission distance and the coverage range. VLC system without cross coupling light can provide BERs lower than 10^{-8} at the distance of about 110 cm and within the coverage of about 17.5 cm. But, the distance and the coverage for successful communications were reduced with cross coupling light to about 70 cm and about 11 cm respectively. A metal shield between LD and PD reduced the influence of cross coupling light and extended the transmission distance and coverage range up to about 100 cm and about 15 cm respectively. It was found out that the VLC transceiver equipped with an optical antenna experiences greater reduction in the transmission distance to about 50 cm for successful communications, but reveals better coverage range characteristics with the tilting angle of 8 degree, in comparison with the VLC transceiver with a convex lens. It may need to be further investigated for improvement in the transmission distance and the coverage range.

ACKNOWLEDGMENTS

This work was supported by the IT R&D program of MKE/KEIT[KI001822, Research on Ubiquitous Mobility Management Methods for Higher Service Availability] and partly by the BLS project from the Seoul Metropolitan City[WR080951, Seoul R&BD Program].

REFERENCES

[1] W. C. Kim, C. S. Bae, S. Y. Jeon, S. Y. Pyun, and D. H. Cho, "Efficient resource allocation for rapid link recovery and visibility in visible light local area networks," IEEE Transactions on Consumer Electronics, 2010, pp. 5240-531.

[2] I. E. Lee, M. L. Sim, F. W. L. Kung, "Performance enhancement of outdoor visible light communication system using selective combining receiver," IET. Optoelectronics, 2009, pp. 30-39.

[3] A. M. Street, P. N. Stavrinou, D. C. Obrien, and D. J. Edwards, "Indoor optical wireless systems—A review," Opt. Quantum Electron., vol. 29, pp. 349–378, (1997).

[4] S.-B. Park, D.K. Jung, H.S. Shin, D.J. Shin, Y.-J. Hyun, K. Lee, and Y.J. Oh, "Information Broadcasting System based on Visible Light Signboard," Proc. Wireless and Optical Communications 2007, pp. 311-313, Montreal, Canada, (2007).

[5] D. J. T. Heatley, D. R. Wisely, I. Neild, and P. Cochrane, "Optical wireless: The story so far," IEEE Communications Magazine, vol. 36, pp. 72-82, (1998).

[6] J.R. Barry, J.M. Kahn, E.A. Lee and D.G. Messerchmitt, "High-speed nondirective optical communication for wireless networks," IEEE Network Mag. 5 44-54 (1991).

 ISBN: 978-1-61208-133-5

Design and Implementation of a BitTorrent Tracker Overlay for Swarm Unification

Călin-Andrei Burloiu, Răzvan Deaconescu, Nicolae Țăpuș
Automatic Control and Computers Faculty
Politehnica University of Bucharest
Emails: calin.burloiu@cti.pub.ro,{razvan.deaconescu,nicolae.tapus}@cs.pub.ro

***Abstract*—The deployment of the BitTorrent protocol in the early 2000s has meant a significant shift in Peer-to-Peer technologies. Currently the most heavily used protocol in the Internet core, the BitTorrent protocol has sparked numerous implementations, commercial entities and research interest. In this paper, we present a mechanism that allows integration of disparate swarms that share the same content. We've designed and implemented a novel inter-tracker protocol, dubbed TSUP, that allows trackers to share peer information, distribute it to clients and enable swarm unification. The protocol forms the basis for putting together small swarms into large, healthy ones with reduced performance overhead. Our work achieves its goals to increase the number of peers in a swarm and proves that the TSUP incurred overhead is minimal.**

***Index Terms*—Peer-to-Peer, BitTorrent, tracker, unification, TSUP, swarm**

I. Introduction

The continuous development of the Internet and the increase of bandwidth capacity to end-users have ensured the context for domination of content-distribution protocols in the Internet. Currently, most Internet traffic is content Peer-to-Peer traffic, mostly BitTorrent. Peer-to-Peer protocols have positioned themselves as the main class of protocols with respect to bandwidth usage.

The arrival of the BitTorrent protocol in the early 2000s has marked a burst of interest and usage in P2P protocols, with the BitTorrent protocol currently being accounted for the biggest chunk in Internet traffic [3]. Modern implementations, various features, focused research and commercial entities have been added to the protocol's environment.

In this paper, we address the issue of unifying separate swarms that take part in a session sharing the same file. We propose a tracker unification protocol that enables disparate swarms, using different .torrent files, to converge. We define swarm unification as enabling clients from different swarms to communicate with each other. The basis for the unification is a "tracker-centric convergence protocol" through which trackers form an overlay network send peer information to each other.

A. BitTorrent Keywords

A *peer* is the basic unit of action in any P2P system, and, as such, a BitTorrent system. It is typically a computer system or program that is actively participating in sharing a given file in a P2P network. A peer is generally characterized by its implementation, download/upload bandwidth capacity (or limitation), download/upload speed, number of download/upload slots, geolocation and general behavior (aggressiveness, entry–exit time interval, churn rate).

The context in which a BitTorrent content distribution session takes place is defined by a BitTorrent **swarm** which is the peer ensemble that participate in sharing a given file. It is characterized by the number of peers, number of seeders, file size, peers' upload/download speed. swarm, one that allows rapid content distribution to peers, is generally defined by a good proportion of seeders and stable peers. We define **stable peers** as peers that are part of the swarm for prolonged periods of time.

Communication of peers in a swarm is typically mediated by a BitTorrent **tracker** or several trackers which are defined in the .torrent file. It is periodically contacted by the peers to provide information regarding piece distribution within the swarm. A peer would receive a list of peers from the tracker and then connect to these peers in a decentralized manner. As it uses a centralized tracker, the swarm may suffer if the tracker is inaccessible or crashes. Several solutions have been devised, such as PEX (Peer EXchange) [1] or DHT (Distributed Hash Table) [11]. The tracker swarm unification protocol presented in this article enables redundancy by integrating multiple trackers in a single swarm.

B. Tracker Swarm Unification Protocol

The goal of the tracker swarm unification protocol is the integration of peers taking part in different swarms that share the same file. These swarms, named **common swarms**, use the same content but different trackers.

We have designed and implemented a tracker network overlay that enables trackers to share information and integrate peers in their respective swarms. The overlay is based on the Tracker Swarm Unification Protocol (TSUP) that allows update notification and delivery to trackers from the overlay. The protocol design is inspired by routing protocols in computer networks.

At this point, as proof of concept, the tracker overlay is defined statically. All trackers know beforehand the host/IP addresses and port of the neighboring trackers and contact them to receive required information. The integration of dynamic tracker discovery is set as future work. Each tracker may act as a "router", sending updates to neighboring trackers that may themselves send them to other trackers.

 ISBN: 978-1-61208-133-5

II. Tracker Unification

A. Motivation

It is also possible that different users create different .torrent files, but with the same files for sharing. If the .torrent files don't refer the same tracker, each one will represent another swarm. The peers from different swarms do not know each other and integration is not accomplished.

By unifying swarms the communication between peers is possible. Every client will have the opportunity of increased connections to other peers, increasing the download speed and decreasing the download time. By having more peers in the swarm, the number of stable seeders also increases and the client can approach its maximum potential.

B. Solution

The protocol proposed in this article, named **Tracker Swarm Unification Protocol (TSUP)**, renders possible the unification of swarms which share the same files, by employing a tracker network overlay. A tracker which implements this protocol will be referred here as an **unified tracker**.

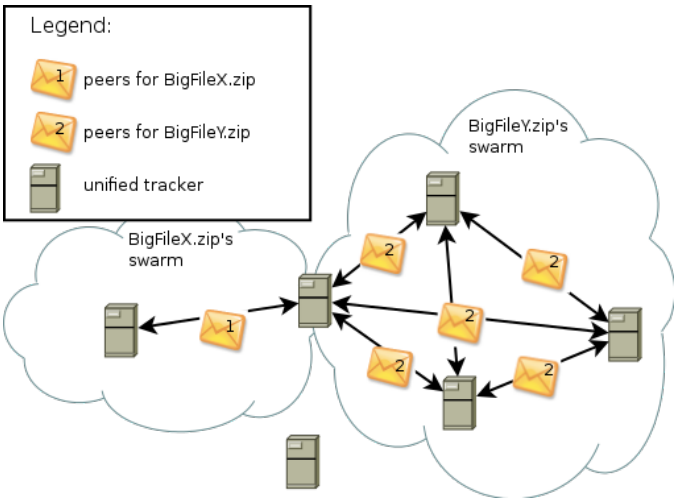


Figure 1. Unified trackers

Torrent files created for the same content have the same *info_hash*. So in swarms that share the same file(s) (common swarms), peers will announce to the tracker the same *info_hash*. Therefore, TSUP capable trackers can "unify" them by communicating with each other in order to change information about peers which contribute to shared files with the same *info_hash*. In order to accomplish this, unified trackers send periodic updates to each other, containing information about the peers from the network.

III. TSUP

As mentioned above, *TSUP* is the acronym for *Tracker Swarm Unification Protocol*. TSUP is responsible with the communication between trackers for peer exchange information in common swarms.

A. Protocol Overview

For transport layer communication, the protocol uses UDP (User Datagram Protocol) to reduce resource consumption. A tracker already possesses a lot of TCP (Transport Control Protocol) connections with each other peers. Adding more TCP connections to each neighboring tracker would increase TSUP overhead too much for a big tracker overlay. The messages passed from one tracker to another do not need TCP's flow control and need a lower level of reliability than TCP as it will be explained below. The simplicity of the UDP protocol gives the advantage of a smaller communication overhead.

In TSUP's operation the following processes may be identified:

- **Virtual connections establishment process:** A three-way-handshake responsible with establishing a UDP "connection" between two linked trackers at the application layer. The process is started by a SYN packet (synchronization packet).
- **Unification process:** The trackers exchange unification packets (named *SUMMARY* packets) during a three-way-handshake in order to find out which are the common swarms.
- **Updating process:** The trackers exchange peers from common swarms, encapsulated in *UPDATE* packets.
- **Election process:** The establishment of a *swarm leader* which is responsible with receiving all updates from the neighboring linked trackers, aggregating them and sending the results back.

The *unification process* includes an updating process in its three-way-handshake, such that the two operations, unification and update, are run in pipeline. Whenever a new torrent file is registered to the tracker, a new swarm is created. The unification process is triggered and a SUMMARY packet is immediately sent to each neighboring tracker, informing the others of the new swarm.

The *updating process* is started periodically, such that UPDATE packets are sent at a configurable interval of time to each tracker in a common swarm. A typical update interval is 30 seconds.

In order to maintain the virtual connections between trackers, HELLO packets are sent periodically, acting as a keep-alive mechanism. A typical HELLO interval is 10 seconds, but its value may be changed from protocol configuration. If no HELLO packet is received during a configurable interval, called disconnect interval, the virtual connection is dropped and the virtual connection establishment process is restarted for that link by sending a SYN packet.

Some packets, such as UPDATE packets, must be acknowledged. If no answer or acknowledgement is received, the packet is retransmitted. For example, UPDATE packets are resent at each hello interval until an acknowledgement is received.

It is not a problem if some UPDATE packets are lost and arrive later to destination. However they need to be

 ISBN: 978-1-61208-133-5

acknowledged and they are retransmitted in order to increase the probability of their arrival. TCP, by offering reliability, provides a faster delivery of updates in case of a network failure which is not needed in the case of TSUP. Lower overhead is considered here more important that fast retransmission. Thus TSUP implements a timer-driver retransmission, as opposed to data-driven used by TCP [10].

Periodically sent packets, the keep-alive mechanism, acknowledgements and retransmissions contribute to the low reliability needed in TSUP. They help exceed the drawbacks of the UDP transport protocol, and also give a more efficient communication than a TCP one.

B. Tracker Awareness

Tracker communication is conditioned by awareness. For this purpose, in the current version of the protocol, each tracker is configured statically with a list of other communicating trackers. Each element of the list represents a *link* which is identified by the tracker host name (URL or IP address) and port. Other parameters for the link may be configured; some of them may be specific to the implementation. If the virtual connection establishment process is successful, the link becomes a virtual connection, which is conserved with keep-alive packets (HELLO packets).

A future version of the protocol will incorporate the design of a tracker discovery mechanism capable of generating the list of communicating trackers for each tracker dynamically, with the benefit of scalability and reduction of the administrative overhead.

C. Tracker Networks

To improve TSUP's scalability, trackers may be grouped together in networks named **tracker networks**. Connections in all tracker networks are full mesh. Two networks are connected with the aid of **border trackers** (see Figure 2).

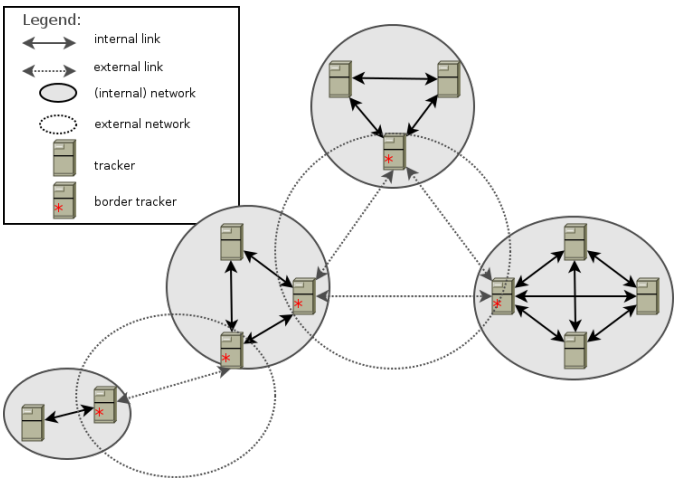


Figure 2. Tracker Networks

To configure a topology which contains multiple networks, each link of each tracker must be set as an **internal link** or an **external link** (see Figure 2). Trackers connected with internal links are part of the same *tracker network*; trackers connected with external links are part of other networks. However, the ones from the first category may also be classified as belonging to an **internal network** and the ones from the second category as being from an **external network**. A complete graph, using internal links as edges is an *internal network*, and a complete graph with external links as edges is an *external network* (see Figure 2). A tracker which has both internal and external links is a border tracker. Peer information received from an internal network originates from **internal peers** while information received from an external network originates from **external peers**. Peers connected to a tracker with TCP, via the BitTorrent protocol, are called **own peers**. The links between trackers in a network, whether internal or external, must be full mesh.

In order to use a scalable and low resource consuming communication within a network, trackers are organized in groups depending on the unified swarm. Therefore a tracker may belong to more than one group at the same time, the number of groups it belongs being equal with the number of swarms present on that tracker. Each group contains a **swarm leader** responsible for sending peer updates to peers in the group. The other group members, instead of sending updates to other members on a full mesh graph, it sends updates to the swarm leader on a tree graph, reducing updating overhead. These updates propagate to other peers in the group.

In accordance to graph theory, the number of updates sent in a swarm without the swarm leader mechanism (full mesh) is computed by using the formula below:

$$UPDATES_{fullmesh} = 2 \cdot \frac{n(n-1)}{2} = n(n-1) \quad (1)$$

The number of updates sent within a swarm using the swarm leader mechanism (tree) are:

$$UPDATES_{swarmleader} = 2(n-1) \quad (2)$$

As may be observed from the formulas above the complexity decreases from $O(n^2)$ in a full mesh update scheme to $O(n)$ with the swarm leader scheme.

Each swarm contains two swarm leaders, one for the internal network (which sends updates through internal links), called *internal swarm leader* and one for the external network (which communicates updates through external links), named *external swarm leader*.

As connections are full mesh in an internal network, the internal peers (received from other trackers from the internal network) are distributed to other internal trackers only by the internal swarm leader and in no other circumstance by another tracker. Through analogy, in an external network, peers (received from other trackers from the external network) are distributed to other external trackers only through the external swarm leader. On the other hand, internal peers are distributed to external trackers and external peers to the internal trackers.

 ISBN: 978-1-61208-133-5

Own tracker peers are distributed both to the internal and the external network.

Swarm leaders are automatically chosen by trackers during the election process which is started periodically. There are metrics used in order to choose the most appropriate leader. The first and most important one prefers as swarm leader a tracker which possesses the smallest *number of swarm leader mandates*. The *number of mandates* is the number of swarms where a tracker is swarm leader. This balances the load of the trackers – as the number of mandates of a tracker increases, its load also increases. In the current version of the protocol the grouping of trackers into networks and the selection of border trackers is done manually (statically) by the system administrator.

When two network trackers A_1 and A_2 are connected indirectly through other network trackers B_j, if A_1 and A_2 use a common swarm and B_j doesn't use this swarm, then the A_i trackers cannot unify unless the border trackers are specified in the configuration. This happens because the configured border trackers must unify with any swarm, although they do not have peers connected from that swarm.

Grouping trackers in networks increases system scalability, but also network convergence time. The update timers can be set to a lower value for border trackers to limit convergence overhead. The system administrator should opt between scalability and convergence and adapt the protocol parameters to the specific topology.

IV. Implementation Details

TSUP is currently implemented in the popular *XBT Tracker* [9], implemented in C++. The extended TSUP capable tracker was dubbed *XBT Unified Tracker*.

The original tracker implements an experimental UDP BitTorrent protocol known as *UDP Tracker*. Because TSUP also uses UDP and communication takes place using the same port, TSUP-specific packets use the same header structure as the UDP Tracker, enabling compatibility.

XBT Tracker uses a MySQL database [8] for configuration parameters [7] and for communication with a potential front end. XBT Unified Tracker adds parameters for configurations that are specific for TSUP and uses a new table in order to remember links with other trackers and their parameters. *Tracker awareness*, as described in III-B, is implemented in the database.

Besides the HTTP *announce* and *scrape* URLs, the original tracker uses other web pages for information and debugging purpose. The unified tracker adds two extra information web pages for *monitoring*. The *trackers web page* shows details about every link and the state of the connection for that link. For every swarm, the *swarms web page* shows the list of peers and the list of trackers connected for that swarm.

V. Experimental Setup

TSUP testing activities used a virtualized infrastructure and a Peer-to-Peer testing framework running on top of it. We were able to deploy scenarios employing various swarms, ranging from a 4-peer and 1-tracker swarm and a 48-peer and 12-tracker swarm. Apart from testing and evalution, the infrastructure has been used to compare the proposed tracker overlay network with classical swarms using a single tracker and the same number of peers. We will show that a unified swarm has similar performance when compared to a single tracker (classical) swarm.

In order to deploy a large number of peers we have used a thin virtualization layer employing OpenVZ [5]. OpenVZ is a lightweight solution that allows rapid creation of virtual machines (also called containers). All systems are identical with respect to hardware and software components. The deployed experiments used a single OpenVZ container either for each tracker or peer taking part in a swarm. A virtualized network has been build allowing direct link layer access between systems – all systems are part of the same network; this allows easy configuration and interraction.

The experiments made use of an updated version of hrktorrent [2], a lightweight application built on top of libtorrent-rasterbar [4]. Previous experiments [13] have shown libtorrent-rasterbar outperforming other BitTorrent experiments leading to its usage in the current experiments. The experiments we conducted used a limitation typical to ADSL2+ connections (24 Mbit download speed limitation, 3 Mbit upload speed limitation).

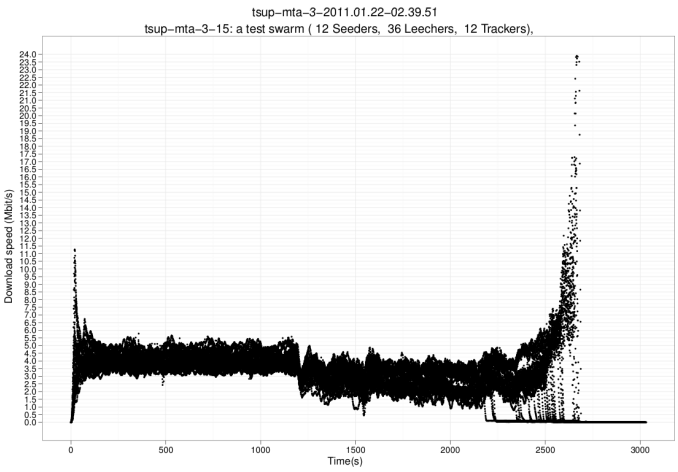


Figure 3. Sample Run Graphic

An automatically-generated sample output graphic, describing a 48 peer session (12 seeders, 36 leechers, 12 trackers) sharing a 1024 MB file is shown in Figure 3. The image presents download speed evolution for all swarm peers. All of them are limited to 24 Mbit download speed and 3 Mbit upload speed.

All peers use download speed between 2 Mbit and 5 Mbit on the first 2000 seconds of the swarm's lifetime. As the leechers become seeders, the swarm download speed increases rapidly as seen in the last part of the swarm's lifetime, with the last leechers reaching the top speed of 24 Mbit.

 ISBN: 978-1-61208-133-5

VI. Scenarios and Results

In order to test the overhead added by TSUP to BitTorrent protocol, we have made a set of test scenarios which compare the average download speed for a swarm with unified trackers and for another swarm with just one non-unified tracker, but the same number of leechers and seeders. Each test scenario is characterized by the shared file sizes, the number of peers and, in the case of tests with unified trackers, by the number of trackers. We shared 3 files of sizes 64MB, 256MB and 1024MB. In the test scenarios with unified trackers for each file we tested the swarm with 1, 2, 4, 8 and 12 trackers. On each tracker there were connected 4 peers, from which 1 is a seeder and 3 are leechers. So, for example, in a scenario with 8 trackers there are 8 seeders and 24 leechers, totalizing 32 peers. Having 3 files and 12 trackers in the biggest scenario we needed to create 36 .torrent files, because for each shared file we made a .torrent file for each tracker. In the corresponding test scenarios with non-unified trackers, there is just one .torrent file for each shared file. We varied the numbers of seeders and leechers connected to the tracker so that they have the same cardinality with the corresponding unified trackers scenarios.

Each test scenario has been repeated *20 times* in order to allow statistical relevance. The average download speed was calculated as an average value from the 20 sessions.

Results may be seen in the table from Figure 4, which depicts the results for each file size, in the top (64MB), middle (256MB) and bottom part of it (1024MB), respectively. For each of this two situations the mean download speed ("mean dspeed") and relative standard deviation ("rel.st.dev.") is depicted. In the right part, titled "perf. Decrease" (performance decrease), shows the percent of download speed decrease induced by the overhead of the TSUP. In the left side of the table the number of seeders and leechers for each scenario is shown. The percentage value for download speed decrease is computed using the standard formula:

$$d = 100\% \cdot \frac{ds_{SingleTracker} - ds_{UnifiedTrackers}}{ds_{SingleTracker}}, \quad (3)$$

where *ds* is an abbreviation from mean download speed.

All positive percentage values from the "perf. decrease" header mark a decrease of performance caused by TSUP overhead. A negative download speed decrease percentage shows that there is an increase instead of a decrease.

The unification which takes place in XBT Unified Tracker introduces an overhead in the BitTorrent protocol in comparison to XBT Tracker. In theory the performance decrease must always be positive. But there are situations where the percentages are negative, which could suggest that TSUP increases the speed, thus reducing the download time. But this performance increase is not due to TSUP, but is caused by another fact. In all scenarios, in the Single Tracker experiments, at the beginning all peers are started almost simultaneously, creating a flash crowd. So in this situation the communication between peers will start immediately, but when multiple trackers are unified, the TSUP imposes a delay before each peer finds out of all the others, because of the convergence time. It is known that sometimes it is better when peers enter the swarm later [12], explaining the presence of negative values for the performance decrease.

		Single Tracker			**Unified Trackers**			
seeders	*leechers*	*mean dspeed (KB/s)*	*rel.st.dev. (%)*	*trackers*	*mean dspeed (KB/s)*	*rel.st.dev. (%)*	*perf. decrease (%)*
$size = 64MB$							
1	3	337.73	1.28	1	335.11	2.87	0.78
2	6	472.27	0.87	6	396.55	16.64	16.03
4	12	475.13	1.87	4	463.42	12.60	2.46
8	24	476.10	6.06	8	497.35	11.10	-4.46
12	36	470.53	9.61	12	496.48	11.48	-5.52
$size = 256MB$							
1	3	358.72	0.45	1	356.55	1.07	0.6
2	6	476.59	0.26	6	407.55	14.67	14.49
4	12	477.56	0.45	4	477.45	10.73	0.02
8	24	492.40	6.64	8	500.56	8.49	-1.66
12	36	486.04	8.33	12	494.93	12.38	-1.83
$size = 1024MB$							
1	3	365.27	0.19	1	365.57	0.31	-0.08
2	6	477.47	0.15	6	437.09	9.12	8.46
4	12	477.54	0.18	4	466.03	5.93	2.41
8	24	423.71	6.19	8	418.69	8.73	1.18
12	36	407.71	4.47	12	418.76	6.27	-2.71

Figure 4. Tracker Networks

From results in Figure 4 several conclusions are drawn. The TSUP overhead becomes more insignificant, on the first hand, when the number of peers increases (and proportionally the number of seeders) and, on the other hand, when the file size increases. When the overhead is insignificant, the percentages have lower values. TSUP convergence time causes

 ISBN: 978-1-61208-133-5

the avoidance of a flash crown at the beginning of each scenario, thus inducing a small performance increase. Starting from 8 seeders and more TSUP performance decrease becomes smaller that the performance increase caused by avoiding the flash crowd. That is why some performance decrease values are negative. The relative standard deviation is generally increasing with the number of peers, but is decreasing when the file size increases. The values can be considered normal, taking into account the number of peers that are part of a swarm.

Due to the small values of performance decrease and relative standard deviation, we concluded that TSUP overhead is insignificant for small to medium-sized swarms (less than 50 peers) which share big files (1GB). BitTorrent is generally used for sharing large files and TSUP allows the increase of swarms size; these two factors come as an advantage for this technology.

Swarm unification increases the number of peers for a shared file, but this fact does not always grant a bigger download speed, as it can be seen in Figure 4. However, having a swarm with a bigger number of peers has three advantages. First of all increases the chance that more seeders will later be available and a big proportion of stable seeders increases download speed. The second reason is that bigger swarms increase shared file's availability by making redundancy. The third reason is that a bigger swarm is more attractive for new users, giving the possibility of creating a big social network, which is an important thing these days.

VII. Conclusion and Further Work

A novel overlay network protocol on top of BitTorrent, aiming at integrating peers in different swarms, has been presented. Dubbed TSUP (Tracker Swarm Unification Protocol), the protocol is used for creating and maintaining a tracker network enabling peers in swarms to converge in a single swarm. Each initial swarm is controlled by a different tracker; trackers use the overlay protocol to communicate with each other and, thus, take part in a greater swarm.

We have used an OpenVZ-based Peer-to-Peer testing infrastructure to create a variety of scenarios employing an updated version of the XBT Tracker, dubbed XBT Unified Tracker. The protocol incurs low overhead and overall performance. The unified swarm is close to the performance of single-tracker swarm consisting of the same number of seeders and leechers with the benefit of increased number of peers, which boosts download speed. The increased number of peers provides the basis for improved information for various overlays (such as social networks) and allows a healthier swarm – given enough peers, if some of them decide to leave the swarm, some peers will still take part in the transfer session.

Experiments have involved different swarms, with respect to the number of peers and trackers, and different file sizes using simulated asymmetric links. The table at the end of the article shows a summary of results, which prove the fact that the protocol has a low overhead.

At this point each tracker uses a statically defined preconfigured list of neighboring trackers. One of the main goals for the near future is to enable dynamic detection of neighboring trackers and ensure extended scalability. We are currently considering two approaches: the use of a tracker index where trackers' IP/host addresses and ports are stored or the use of a completely decentralized tracker discovery overlay similar to DHT's discovery methods.

As proof of concept, our test scenarios have focused on homogeneous swarms. All peers in swarms are using the same implementation and the same bandwidth limitation. All peers enter the swarm at about the same time, with some delay until swarm convergence, in case of the unified tracker protocol. We plan to create heterogeneous swarms that use different clients with different characteristics. The number of seeders and leechers in initial swarms are also going to be altered and observe the changes incurred by using the unification protocol.

Acknowledgment

This paper is suported from POSCCE project GEEA 226 - SMIS code 2471, which is co-founded through the European Found for Regional Development inside the Operational Sectoral Program "Economical competivity improvement" under contract 51/11.05.2009, and from the Sectoral Operational Programme Human Resources Development 2007-2013 of the Romanian Ministry of Labour, Family and Social Protection through the Financial Agreement POSDRU/6/1.5/S/19.

Special thanks go the the P2P-Next [6] team who is working enthusiastically to deliver the next generation peer-to-peer content delivery platform. Their dedication, professionalism and vision are a constant factor of motivation and focus for our work.

References

[1] DHT & PEX. http://lifehacker.com/5411311/bittorrents-future-dht-pex-and-magnet-links-explained. [Online, accessed 31-March-2011].

[2] hrktorrent. http://50hz.ws/hrktorrent/. [Online, accessed 31-March-2011].

[3] ipoque Internet Studies. http://www.ipoque.com/resources/internet-studies/internet-study-2008_2009. [Online, accessed 31-March-2011].

[4] libtorrent (Rasterbar). http://www.rasterbar.com/products/libtorrent/. [Online, accessed 31-March-2011].

[5] OpenVZ. http://wiki.openvz.org/. [Online, accessed 31-March-2011].

[6] P2P-Next. http://www.p2p-next.org/. [Online, accessed 31-March-2011].

[7] XBT Configuration Options. http://www.visigod.com/xbt-tracker/configuration. [Online, accessed 31-March-2011].

[8] XBT Table Documentation. http://www.visigod.com/xbt-tracker/table-documentation. [Online, accessed 31-March-2011].

[9] XBT Tracker by Olaf van der Spek. http://xbtt.sourceforge.net/tracker/. [Online, accessed 31-March-2011].

[10] H. Balakrishnan. Lecture 3; Coping with Best-Effort: Reliable Transport. 2005.

[11] H. Balakrishnan, M. F. Kaashoek, D. Karger, R. Morris, and I. Stoica. Looking up Data in P2P Systems. *Commun. ACM*, 46:43–48, February 2003.

[12] A. R. Bharambe, C. Herley, and V. N. Padmanabhan. Understanding and Deconstructing BitTorrent Performance. *Technical Report MSR-TR-2005-03, Microsoft Research*, 2005.

[13] R. Deaconescu, G. Milescu, B. Aurelian, R. Rughiniș, and N. Țăpuș. A Virtualized Infrastructure for Automated BitTorrent Performance Testing and Evaluation. *International Journal on Advances in Systems and Measurements*, 2(2&3):236–247, 2009.

 ISBN: 978-1-61208-133-5

Dependable Routing Protocol Considering the k-Coverage Problem for Wireless Sensor Networks

Hamza Drid, Samer Lahoud
University of Rennes I–IRISA
Rennes, France
Email: {hdrid, slahoud}@irisa.fr

László Gönczy, Gábor Bergmann
Budapest Univ. of Technology and Economics
Budapest, Hungary
Email: {gonczy, bergmann}@mit.bme.hu

Miklós Molnár
University of Montpellier II–LIRMM
Montpellier, France
Email: miklos.molnar@lirmm.fr

***Abstract*—Fault tolerance and periodical changes of the network topology are two important attributes that should be carefully designed in Wireless Sensor Networks (WSN). In this paper we propose a new routing protocol for improving fault tolerance of WSN which takes into consideration the measurement accuracy requirements (expressed as a lower limit on k-coverage) and network performance. The main idea of the k-coverage problem is to schedule the sleeping time of sensors in order to preserve their energy and maximize network lifetime. The proposed routing protocol offers a protection against link and node failures by computing k disjoint paths. It also takes into account the changes of the network topology caused by the scheduling of sensor sleeping.**

***Keywords*-Fault tolerance; Sensor Networks; k-coverage; Routing; Protection;**

I. Introduction

Wireless sensor networks (WSNs) are generally deployed to monitor areas and provide measurements for surveillance applications. WSNs have several application domains: to monitor the environment/habitat, to collect information, to register and process environmental parameters for optimization or prediction, and/or to insure security ([2] describes many applications and challenges). Often, the measurement or surveillance task of a WSN requires the complete coverage of a target area or a set of target objects. In a general WSN architecture, several sensor nodes send the observation data to Base Stations (BSs) or sinks. Then data can be processed by the sinks and later send to the potential clients. The sensors are performing sensing and communication tasks and the main problems and challenges of this kind of networks are associated to these two activities [1]. The sensor network should be capable to measure in the target area and to process the measured values and transmit them to sink nodes. As several critical applications can depend on the measurement results, reliability of the overall network is a key issue, including both measurement reliability (which often requires multiple nodes to measure the same area) and communication (supported by routing).

Routing in WSNs is an important issue that addresses the delivery of the sensed data from source sensor nodes to the sink. Due to the inherent characteristics of WSNs. The routing protocol should be simple and able to deal with a very large number of nodes (scalable), the limited energy, the limited processing and storage capacities of nodes and also to be self-configurable regarding node failures and changes of the network topology.

Many routing protocols have been proposed in the literature. These routing protocols can be classified in three classes: Data centric routing, hierarchical routing and location-based routing. In data centric routing [8], [12], [4], all nodes are typically assigned equal roles; the base station sends queries to certain regions and waits for data from the sensors. In hierarchical routing [6], [7], [3], the network is divided into clusters, in each cluster the node with high energy is used to process and send the information while the other nodes are used to perform the sensing. In location-based routing [11], [5] there is no addressing scheme available, and nodes are addressed by location. The location is obtained by distance estimation, neighbor discovery or GPS.

Previous works have considered routing constraints such as node mobility, energy limitations, high network density and limited computation capabilities. However, only a few works have actually considered the node failures and the periodical changes of the network topology.

In our previous works [9], we developed the Controlled Greedy Sleep (CGS) algorithm which performs well in prolonging network lifetime and offering reliable measurement results (by ensuring k-coverage) at the same time. It has been proven that this results in a quasi-optimal solution. Here we investigate how an efficient routing algorithm can be implemented in this context, using efficiently ”cross-layer” information, i.e., the scheduling of nodes determined by CGS.

Our main contribution in this paper is to propose a new routing protocol for fault tolerance in WSN. The overall paper is organized as follows. Section 2 presents K-Coverage Problem and the previously developed CGS algorithm. Next, Section 3 describes our assumptions and the proposed routing algorithm, illustrating the operation on an example. Conclusions and directions for future work are presented in Section 4.

 ISBN: 978-1-61208-133-5

II. K-COVERAGE PROBLEM AND CGS

Generally, sensors have limited power capacity but WSNs has to meet relatively strong lifetime requirements, thus the energy conservation is a critical issue in WSN. Mostly, high density of sensors and a *scheduled sleeping* algorithm are employed to preserve energy of the sensors and provide the network services. Sensors periodically alternate between *sleep* and *awake* states using a given period length T, thus saving energy. On the other hand, the notion of *k-coverage* refers to the requirement that each measurement area should be covered by at least *k* sensors. Our aim is to find a balance between these contradicting requirements.

A. Sensor Network Model

Generally the communication range of a sensor is greater than the twice of the sensing one; so, one can suppose that the sensors sharing the measurement/observation task anywhere can also communicate one with the other. Consequently, if the target area is covered with awake sensors, then the connectivity of the WSN is also ensured [10].

We consider a WSN consisting of *n* homogeneous sensors $s_1, s_2, ..., s_n$. A WSN is considered homogeneous if its sensors have the same sensing and communication range. We also assume that sensors are static and each sensor knows its own location (x_i, y_i) and where the sink is located. The position of the sink must be broadcasted throughout the network at the start (or when the sink moves). The information about the actives neighbours such as the location and energy are provided by CGS. Our routing protocol can be executed at the beginning of each CGS period.

B. The Controlled Greedy Sleep Algorithm

In this scheduling nodes have local information on their neighborhood only. Each sensor node q will use a locally known sub-graph $G_q(S_q \bigcup R_q, E_q)$. This sub-graph contains geographical regions R_q covered by q, the set S_q of sensors which participate the coverage of at least one region of R_q. E_q contains the edges between the regions and sensors. The scheduling is based on a particular factor.

The coverage ratio is positive if the region is over-covered, and negative otherwise: in this latter case the operation of all sensors covering r is essential. Moreover, the smaller the energy of q, the larger its *drowsiness*. CGS enforces the sensors in critical positions to go to sleep whenever it is possible, to conserve their energy for times when their participation will become inevitable. A sensor q can go to sleep when its neighbors with larger drowsiness factor decided their state for the next period and q has no critical (not over-covered) region to monitor. Consequently, each sensor should know the drowsiness factor of its neighbors and the decision of neighbors with larger factor. To organize the local communication, a communication delay (DTD) is associated with each sensor. This delay is inversely proportional with the drowsiness factor. So the sensors with large factor broadcast their decision earlier. Only the awake state decision should be broadcasted explicitly, in this way the communication overhead can be minimal.

The main steps of the Controlled Greedy Sleep (CGS) Algorithm are the followings:

1) At the beginning of the period, wake up all sensors whose remaining energy is enough for spending at least one period awake.
2) Alive sensors broadcast local *Hello* messages containing their locations. Based on received *Hello* messages each sensor q builds up its local set of alive neighbors S_q and generates the local bi-partite graph $G_q(S_q \bigcup R_q, E_q)$, and then it calculates its drowsiness factor D_q.
3) Based on D_q each node q selects a Decision Time Delay (DTD_q). Small drowsiness means large DTD, large drowsiness means small DTD. These delays provide priorities when nodes announce their Awake Message (AM). Each sensor q broadcasts its DTD_q and starts collecting DTD and AM messages from the neighborhood. From the received DTD and AM messages it builds a Delay List (DL_q) and a List of Awake Neighbors (LAN_q) respectively.
4) When DTD_q time elapsed the node q makes a decision based upon LAN_q and DL_q:
   - if all regions in R_q can be K-covered using only nodes present in LAN_q and/or nodes u present in DL_q for which $DTD_u > DTD_q$ then go to *sleep*
   - otherwise stay *awake* and broadcast an AM to inform the neigbors of this decision.

Obviously, the communication overhead of the algorithm depends on the length of periods. At the begining of the period there are three time intervals: to exchange *Hello*, DTD and AM messages respectively. A sensor broadcasts at most three messages during T_e (two if it the node will go to sleep, three otherwise) and must stay awake in order to complete the election process. During this extra T_e time sensors consume energy. The scheduling communication and awake-time overhead can be low if the length T of a period is significantly longer than T_e. But, one can state that this period can not be too long either.

The determination of the optimal length T is a hard computation problem. In real cases only empirical and estimated solutions can be formulated. The study of this period length with simulation offers significant elements to choose the period length.

For a detailed presentation of CGS, the reader is referred to [9].

III. ROUTING PROPOSAL

The following requirements were posed for this routing:

1) It should take into account the information on sensor status, determined by the scheduling algorithm.

 ISBN: 978-1-61208-133-5

2) It should be efficient in terms of forwarding messages to nodes which are close to optimal (so that the expected hop count remains low).
3) It should be fault tolerant in an efficient manner (c.f. req.2.), i.e., two disjoint paths should exist for all source-sink pairs. By disjoint paths, we mean to paths which have no common nodes.

Relevant Information obtained from CGS: In order to efficiently use the energy of the sensor network and have as few lost messages as possible, the routing algorithm needs the following information from CGS:

1) The position and remaining energy of neighbors (contained in *Hello* messages). Note that here only the neighbors which measure the same are will be known, not all neighbors within the communication radius. This would lead to situations where not all possible routes are used. The enhancement of this would need a modification of CGS.
2) The sensor status in the next CGS scheduling period should be considered; here the routing will count only on sensors which remain awake. This will be contained in *AM* messages. Note that the periods of CGS do not limit the routing in a sense that message trasnfer is not blocked during the decision mechanism, in contrary, as all (alive) sensors will be awake, all possible routes will be available (although not necessarily utilized).

A. Assumptions

We made the following assumptions during our work:

- The source node is coded in the header of each message.
- The algorithm itself does not ensure k-connectivity, rather it tries to utilize existing connectivity with efficient routing.
- The information of CGS (as described in previous section) is available for the routing algorithm.
- The communication area is contained by the measurement area. If this does not hold, then the number of neighbours wrt. measurement (revealed by CGS messages) does not tell relevant information on the number of nodes wrt. communication. Note that this is true for most practical cases, a counterexample can be a relatively big circle which contains sensors mostly at the perimeter while the sink is in the middle.

B. Proposed Algorithm

The algorithm that we describe here is intended to compute two disjoint paths. In our algorithm, only sensors currently active to k-cover can participate in data forwarding.

When useful data is sensed by the source node si, si sends the sensed data to two neighbours based on its local information. The selection of the two neighbours can be based on different criteria. Our selection is basically based on the remaining energy and the position of the neighbour node relative to the sink node and the sender node. The sensor nodes which have maximum value of F(j) will be selected as relay nodes.

$$F(j) = E(j) \times \frac{d(s,j)}{d(j), Sink} \times cos(a_j) \times awk(j) \qquad (1)$$

where *E(j)* is energy available of sensor node j in the set of candidate set Y, a_j is the critical angle created by the coordination of node j, the sender node s, and the sink. $d_{(s,j)}$, $d_{(j,sink)}$ are distance from the sender node s to node j and distance from node j to the sink. $awk(j)$ is binary variable, it takes the value of 1 if the sensor node$j$ is awake. Otherwise, it is 1.

Each node received the data packet saves the packet head into own cache, and send the data packet to the best relay node which has maximum value of $F(j)$. The number of neighbours Nb in the slice of the sender node is also sent to the relay node as shown later in the example.

If a node j receives the same packet twice(node j checks its cache to verify if it has already received the same packet), the following control packets are exchanged between concerned nodes in order to obtain two disjoint paths.

For all j nodes where $j! = sink$

1) j sends a control packet to the sender k (from which it has received the data packet and that has a bigger number of neighbors).
2) k selects a new relay node in the set of candidate nodes $Y - j$ using (1). (cf. figure 1, k selects z).
3) j still a relay node but for the sender which has the minimum number of neighbors node. (cf. figure 1, j the relay node of A).

The process continues until the data reach the sink. In order to enhance more reliability, we can also impose that, when a sensor forwards a data packet towards the sink, it would also send an ACK to the sender from which it has received the packet.

C. Example

Figure 1 illustrates an example of disjoint paths construction. In order to construct two disjoint paths, the source sends the sensed data to two neighbours A, K based on its local information.

A and *K* nodes received the data packet from the source, *A* and *K* send back the data packet to the best relay node , which is *J* in our case (cf. Fig.1 (a)). *J* receives the same packet twice. In this case *J* informs *K* to change the relay node (cf. Fig 1 (b)). *K* was chosen because it has several neighbours, while A has only *j* as neighbour. *K* chooses another neighbor and the process continues until the data reach the sink as shown in (cf. Fig 1 (d)).

 ISBN: 978-1-61208-133-5

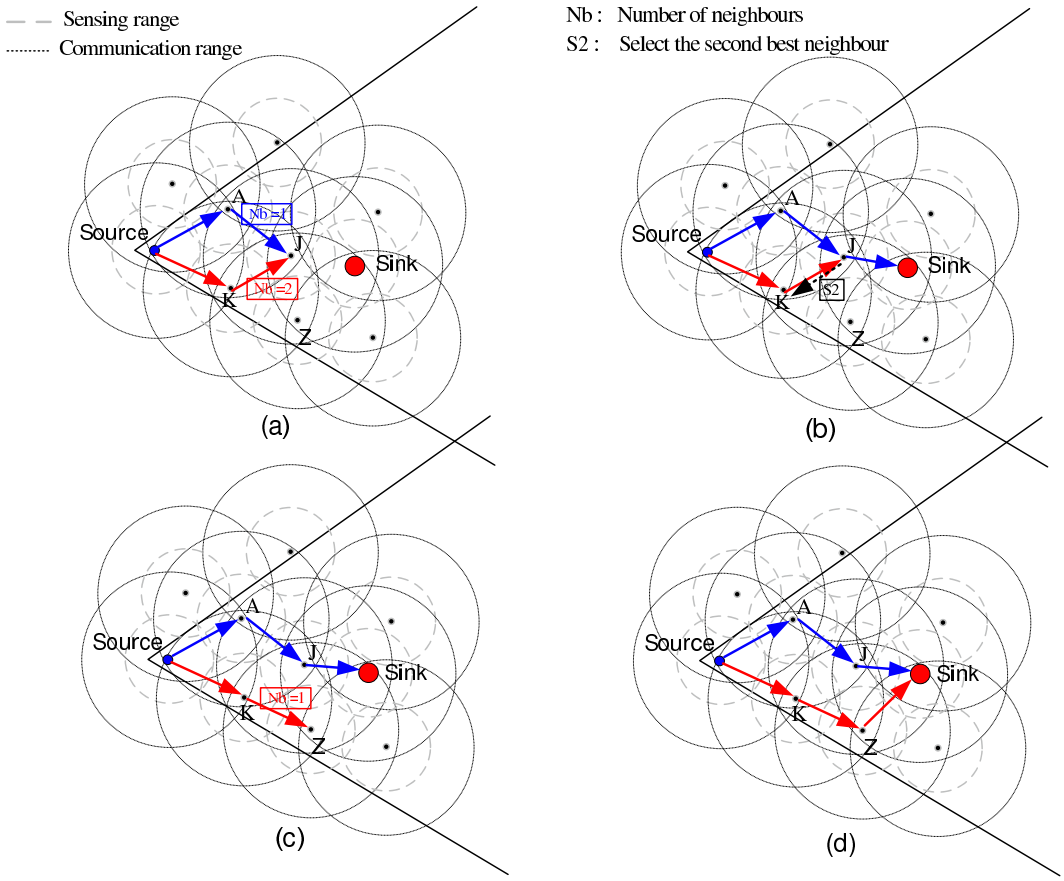


Figure 1. Protection using p-Cycle

IV. Conclusion and Future Steps

Benefits of this routing include a message transfer to the sink considering both energy of sensors and direction, fault tolerance by ensuring two disjoint paths and efficient use of awake sensors based on information obtaind from CGS messages.

The algorithm should be validated in practice by executing simulations on different topologies. Also the effect of period length and communication intensivity (frequency of messages) would be interesting to investigate. Our expectation is that small period lengths will be efficient, especially if there is intense message communication which may cause a faster deprecation than expected (so that the drowsiness factor of some sensors will decrease fast).

A possible enhancement to this algorithm would be to incorporate knowledge on previous decisions. This could be done e.g., by storing a list in each active sensor containing source nodes for which there was a collision (so messages from that particular sensor should be forwarded to the ”second best” neighbor in order to avoid unnecessary control messages). Note that this list should be refreshed for each period, as the routes may change as the scheduling of CGS changes node status.

An important research step would be to investigate the connection between *k-coverage* and *k-connectivity* (where the guaranteed k values can differ for the two properties!). Also the presence of articulation nodes can obviously affect routing (even if k-coverage may be ensured for major parts of the network).

We have investigated how CGS can work together with different sensor types where multiple measurement objectives exist and these can be covered by different types of sensors (e.g., some sensors can measure humidity and temperautre while others only humidity, etc.). It would be interesting to see how the routing can work on top of this (considering that communication itself is not restricted to equivalent sensors).

Also one can consider an extension to information sent in the decision phase of CGS to include all sensors within the communication radius and therefore enable a more efficient routing.

An additional future step is to investigate the effect of aggregation, where messages from different source nodes/ different messages from the same source node are waited for and sent together. This might result in less communication (longer lifetime) with an increased expected message

 ISBN: 978-1-61208-133-5

transfer time.

Finally, the effect of realistic sensing/communication radius wrt. obstacles among sensors remains a future research question.

REFERENCES

[1] N. Ahmed, S. S. Kanhere, and S. Jha. The holes problem in wireless sensor networks: a survey. *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 9(2):pp. 4–18, 2005.

[2] C.-Y. Chong and S. P. Kumar. Sensor networks: evolution, opportunities, and challenges. *Proceedings of the IEEE*, vol. 91(8):pp. 1247–1256, 2003. URL http://dx.doi.org/10.1109/JPROC.2003.814918.

[3] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan. Energy-efficient communication protocol for wireless sensor networks. In *Proceeding of the Hawaii International Conference System Sciences*. 2000.

[4] W. Heinzelman and H. B. J. Kulik. Adaptive protocols for information dissemination in wireless sensor networks. In *Proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking*. 1999.

[5] L. Li and J. Y. Halpern. Minimum energy mobile wireless networks revisited. In *IEEE International Conference on Communications*. 2001.

[6] S. Lindsey and C. Raghavendra. Pegasis: power efficient gathering in sensor information systems. In *IEEE Aerospace Conference*. 2002.

[7] S. Lindsey, C. Raghavendra, and K. Sivalingam. Data gathering in sensor networks using the energy delay metric. In *IPDPS Workshop on Issues in Wireless Networks and Mobile Computing*. 2001.

[8] A. Manjeshwar and D. Agrawal. Teen: a routing protocol for enhanced efficiency in wireless sensor networks. In *Parallel and Distributed Processing Symposium*. 2001.

[9] G. Simon, M. Molnár, L. Gönczy, and B. Cousin. Robust k-coverage algorithms for sensor networks. *IEEE Transactions on Instrumentation and Measurement*, vol. 57, 2008.

[10] G. Xing, X. Wang, Y. Zhang, C. Lu, R. Pless, and C. Gill. Integrated coverage and connectivity configuration for energy conservation in sensor networks. *ACM Trans. Sen. Netw.*, vol. 1(1):pp. 36–72, 2005.

[11] Y. Xu, J. Heidemann, and D. Estrin. Geography-informed energy conservation for ad hoc routing. In *7th Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom 01)*. 2001.

[12] Y. Yao and J. Gehrke. The cougar approach to in-network query processing in sensor networks. In *SIGMOD Record*. 2002.

 ISBN: 978-1-61208-133-5

Practising Problem Solving Using Mobile Technologies

Richard Seaton
The Open University
Walton Hall, Milton Keynes, United Kingdom
r..k.seaton@open.ac.uk

***Abstract*- Mobile phones have exceptional computing abilities and yet are used largely for talking, texting and telling the time. This paper describes a project using the mobile phone to provide a learning platform for students on the United Kingdom Open University's (UKOU) distance learning module T216, Cisco Networking (CCNA). Within the project two different applications were offered and a survey conducted to gather students' opinions of the applications in particular and the general use, by the UKOU, of mobile technologies for student learning and support. The first application allows practice of IP address subnetting, an area of the module that students find problematic. The second gives students the opportunity to practise the types of questions used in the Cisco Academy's online tests. It also allows them the opportunity to monitor their progress. The results of a student survey are promising with students finding that both applications aided their learning, knowledge and understanding. They liked the ability to practise whenever they had a few spare minutes and wherever they found themselves. In general, UKOU students would like to see greater use of mobile technologies within their studies both for learning and for organisation. Reminders about assignments, tutorials and day schools were considered to be a useful service should this be possible. Overall the project has been a success. The use of mobile phones to enhance student learning has been proved. Once the problems regarding use with all mobile platforms are resolved, opportunities for use by the University and other learning organizations look realistic. This is particularly true in areas of the world where the availability of computers and broadband are limited and mobile phones are the only method of distance communication.**

Keywords – Mobile technologies; learning; problem solving.

I. APPROACH

The work of Kukulska-Hulme and Traxier [1] established that mobile technologies can be used successfully to support students' learning through a range of different activities, particularly around data collection and recording. Further, The Open University has undertaken work such as the Mobile Assisted Language Learning [2] project, which involved a wide range of mobile devices including mobile phones. Other UK higher education institutions have reported use of mobile phones to support students with timetabling and assessment dates. Research in this area is being undertaken within the University. A comprehensive Literature Review in Mobile Technologies and Learning (Futurelab 2006) [3] demonstrated many opportunities for use of mobile technologies. It saw the future as offering genuinely learner-centred learning experience that is specific, personal, collaborative and long term. This may be offered by exploiting the ubiquitous qualities of today's mobile phone technology, demonstrated particularly by the smartphone represented by devices such as Apple's iPhone and the RIM Blackberry. These devices offer facilities over and above the specific capabilities of mobile phones: high resolution cameras, PDA features, multimedia players, etc., integrated into devices that fit in a shirt pocket and may be used almost anywhere. Such devices were merely a vision to the authors of the Futurelab report in 2006.

The Communications Market Report (Ofcom, 2008) [4] showed 86% of the United Kingdom population owning a mobile phone, and so it would seem natural to exploit this device for use among distance learning students. The mobile phone has considerable computational abilities and yet anecdotal evidence, confirmed by simple surveys undertaken during presentations of this project, is that most people only use their mobile phones for talking, texting and telling the time. Few of the other applications within even relatively basic mobile phones, let alone those found in smartphones, are used.

Learning involving mobile technologies can introduce some specific issues:

- Context – the ability to personalize the learner's environment may lead to ethical issues requiring secure storage of data;
- Mobility – promotes the need to manage the learning environment outside the more traditional setting;
- Informality – the advantages of using mobile technologies in a formal way may deter students;
- Ownership – the necessity to avoid exclusion;
- Learning over time – the need for organisation and tracking of learning.

These issues were taken into account during the progress of the project.

 ISBN: 978-1-61208-133-5

II. APPLICATIONS

A. *Subnet Exerciser*

The Subnet Exerciser is an application that allows students to practise the technique of subnetting, something this is known to be challenging to students. The application was released to a cohort of 300 Cisco Networking students towards the end of 2007 to fit in with their study of subnetting. This was repeated to increased numbers in April 2008. Informal forum-based questionnaires were undertaken with both cohorts and feedback was received through forum comment and personal email. A similar trial was conducted in April 2009 but followed up with an additional application discussed later and a more formal questionnaire to complete the project.

The Subnet Exerciser, shown in Fig. 1, allows students to practice the skills necessary to manipulate IP addresses and includes:

- Denary to Binary conversion
- Binary to Denary conversion
- Logical AND operation
- IP address classification
- Practice questions on subnetting

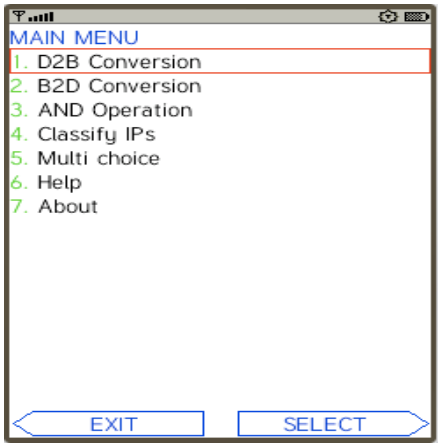


Figure 1. Main menu screen of the Subnet Exerciser

The application is downloaded to students' phones and run stand-alone, in much the same way that a games programme would be used. From the outset it was realised that not all mobile phones were suitable for this application. It was written in a dialect of the Java® programming language designed for mobile phones. Interestingly, the smartphone caused more issues than the basic mobile phone, something which seemed somewhat perverse. This was because the former does not conform to the mobile phone Connection Limited Device Configuration (CLDC) standard [5]. Students were able to find out whether or not their phone was suitable, but it was obvious from students' messages in the forums that this facility was not used extensively. For students unable to run the application on their mobile phone a simulator could be downloaded from Sun Microsystems [6] which allowed the Subnet Exerciser to run on their computers. Care was taken throughout the project to ensure no student was disenfranchised through lack of a suitable phone.

Initial findings suggested that students valued particularly the learning opportunities offered by the multi-choice practice questions. This was because much of the Cisco assessment is based around this methodology. Further, there was a need to address the issues raised earlier regarding mobile learning, specifically those raised by mobility and learning over time.

B. *Multiple choices*

Luzia Research [7] is a company that develops mobile learning applications. Their *uHavePassed.com* application [8], which allows those learning to drive to practise the theoretical part of the United Kingdom driving test using mobile phone and computer platforms, looked particularly useful. Following registration users download a simple client onto their mobile phone, see Fig. 2.

Figure 2. The initial client screen of *uHavePassed* on a mobile phone

This client allows access to different ranges of questions and progression statistics. A range of multi-choice question types, including use of graphics and multimedia, are available for download during the online synchronization process. The selection of questions is dependent upon statistical indication of the student's progress, with failed questions being re-sent along with new ones to replace those which were answered correctly. Importantly, this application does not require the student to be online except when synchronizing with the server. Data transfer to mobile phones can be costly so it is important to keep the need to be online to a minimum.

Luzia Research provided the University with a *uHavePassed* platform (*OuHavePassed*) and server access to banks of questions covering the Cisco Academy CCNA Exploration curriculum. This involved developing over 400 new questions to add to those written for the original Subnet Exerciser. Many questions needed diagrams, some of which

 ISBN: 978-1-61208-133-5

were drawn originally for viewing on a computer screen, others were drawn specifically to suit the reduced size of the mobile phone screen. Blocks of questions were sub-divided into chapters to match the Cisco Exploration curriculum. This ensured that students knew where to seek additional support, over and above the supplied feedback when required. An example of a question and answer is shown in Fig. 3. The smart buttons immediately below the phone screen move pages and either the cursor or the number pad is used to select an answer.

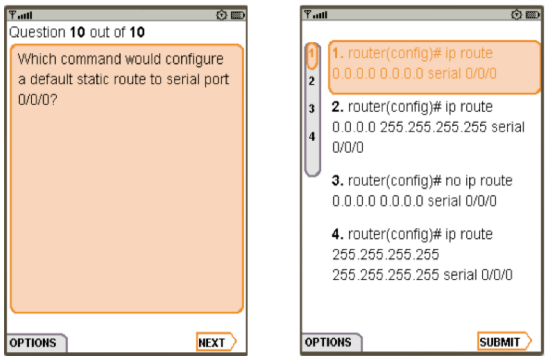


Figure 3. *uHavePassed* screens showing a question and optional answers

C. *Data and evidence gathering*

An online survey by questionnaire was the most appropriate way of reaching all students, for many were based overseas. Evidence was also gathered from student forums and face to face meetings at day schools which form part of the module.

IV. OUTCOMES

A. *Main outcomes and supporting evidence*

1. Students are willing to accept using mobile technologies when they can see a real purpose;
2. Mobile technology can support students' learning but has to be targeted for maximum effect;
3. Students are concerned about costs of using mobile technologies;
4. All mobile platforms must be supported, since students use a wide range of technologies;
5. Students would like administrative support with their studies using mobile technologies.

B. *Student statistics*

Table 1 shows the student involvement with the project taken from the *uHavePassed* server. 32% of the total active cohort registered on the website and of these 56% downloaded the client onto their phones. 612 mobile tests and 182 web-based tests were attempted, indicating that the ability to use this resource was more valued by those using their mobile phones. Making an assumption that those using the client on their phones would not use the website for tests, then nearly seven tests were attempted by students on their phones against just over two tests via the web. This could be because access via the web was probably seen as no more beneficial than accessing the formative tests on the Cisco website, but there is no evidence to support this

TABLE I. STUDENT INVOLVEMENT WITH THE PROJECT

Students:	**Number**	**Percentage of total**
Active on module	500	n/a
Registering on website	160	32
Downloading client	89	56
Taking mobile phone tests	612	n/a
Taking web-based tests	182	n/a

C. *Response to the survey*

A personalized email request was sent to over 400 students taking Cisco modules, and Table 2 shows the student response to the survey. A 10% response was typical for this form of survey and was just acceptable from a statistical view. It was pleasing to see that a cross-section of students responded, not just those who took part in the project.

TABLE II. STUDENT RESPONSE TO THE SURVEY

Students	Number	Percentage
Received details of survey	405	n/a
Responded to survey	40	10
Accessed the website	24	60
Used mobile phone	12	30

D. *Support of outcomes*

D.i. *Students are willing to accept using mobile technologies when they can see a real purpose*

The overall response to the project demonstrates students are willing to embrace mobile technologies or use existing ones in new and novel ways. The biggest concern raised by students was the screen size of their phones with regard to the way in which we expected the students to interact with it. The responses shown in Fig. 4 were from a question asking students about using their phone compared with a similar experience on the website viewed on a conventional PC screen. Approximately 40% stated they found reading text somewhat difficult, and overall, they found the experience of using the phone's screen less acceptable than that of a PC screen. Results showed few problems with text size but

 ISBN: 978-1-61208-133-5

diagrams were more challenging, although this is phone dependent. The problem with diagrams was in part due to not having time to rework those drawn originally for PC screens. Where new diagrams were drawn with the phone screen in mind fewer problems were experienced. There was the ability to exploit 'landscape' by turning the phone on its side, as shown in Fig. 5.

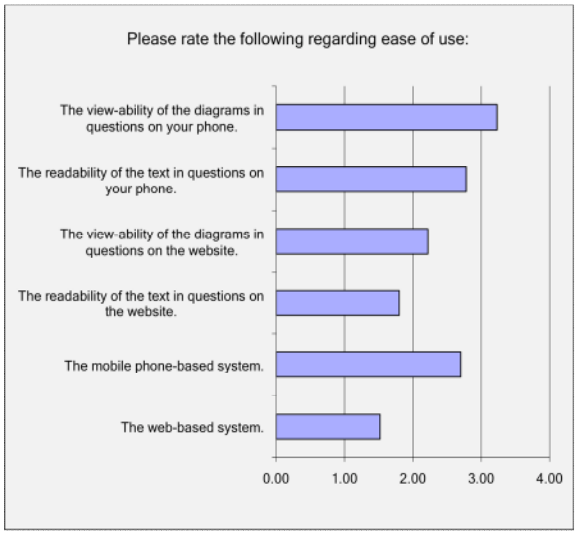


Figure 4. Ease of use, the lower the value the better

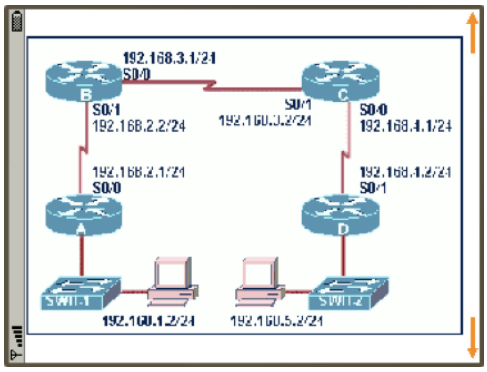


Figure 5. A diagram imported from a PC version viewed in landscape.

Currently, the text size is restricted by the Java CLDC specification. This will become less of a problem when the application becomes available on smartphones as these tend to have both larger screens and ways to resize text.

D.ii. Mobile technology can support students' learning but has to be targeted for maximum effect

The response to a question which asked how helpful students had found the application in supporting their learning is shown in Fig. 6. Over two thirds responded positively to having the questions available on their phone, to the choice of questions and to being able to monitor their progress. In particular they found the ability to retake questions where they been unsuccessful and the provision of context related feedback very helpful. Learning by reinforcement is a key to the behaviourist theory which this project adopts.

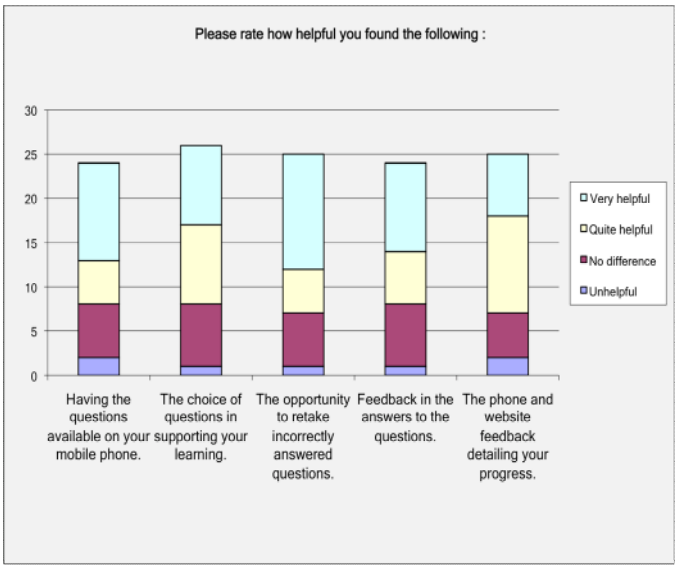


Figure 6. Students' opinions of the application

D.iii. Students are happy to use mobile technologies but are concerned about costs of data

Fewer than 10% of students expressed concern regarding costs, but they must be in one's mind when developing applications for new technologies. Unfortunately, data communication on mobile phones can seem expensive unless combined within a bundled contract. The advantage of the original Subnet Exerciser application was that it involved a single download of around 50 kilobytes. The ability to change the question banks and the availability of many questions does increase data transfer costs. A typical download when synchronizing is in the order of 300 kilobytes. Many smartphones offer network connections by Wi-Fi as well as the mobile network connections which can significantly reduce costs.

D.iv. All mobile platforms should be supported, because students use a wide range of technologies

Unfortunately, known problems in supporting smartphones were not completely solved before the end of the project. In particular, the Apple iPhone App took longer to develop than was expected. Blackberry access was unavailable to corporate users due to problems with firewalls and this affected several students. Applications must work on all major platforms if students are to be well served and the ability to access the project through a website was an important feature of the project even if, in this case, it received little use.

 ISBN: 978-1-61208-133-5

D.v They would like support with their work through mobile technologies

Opinions were sought and responses shown in Fig. 7 as to how students might like to see mobile phones used by the University in supporting their learning. Reminders about dates of particular calendar events were thought to be very worthwhile. Over 70% thought texts (SMS) reminding students when particular events, such as due dates for assignments and results available, a good idea. 60% thought texts when assignments were not received would also be helpful. Interestingly, support for module specific applications were less popular. This may reflect the fact that the current ways of dealing with these applications are sufficient. A suggestion of using text messages to support group work seems an excellent idea bearing in mind the University's aim to include this activity on every module.

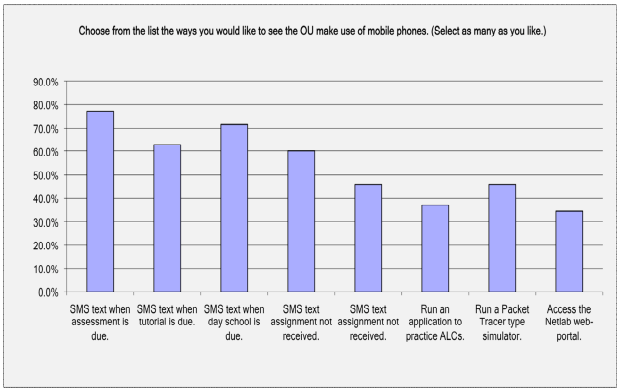


Figure 7. Ways in which the University could use mobile phones

V. IMPACT OF THE PROJECT

V.i. The ways the project impacted on student learning

The student response, shown in Fig. 8, to how they used the facilities offered by the project demonstrated that they had accessed all four blocks, and whilst percentages appear low it should be remembered that only 46% of the respondents took part in the project. Taking this into account 83% of the students who took part accessed block 2 with the other values not far below. The students were moving to block 2 as the project was made available which may well be why this was the most popular block

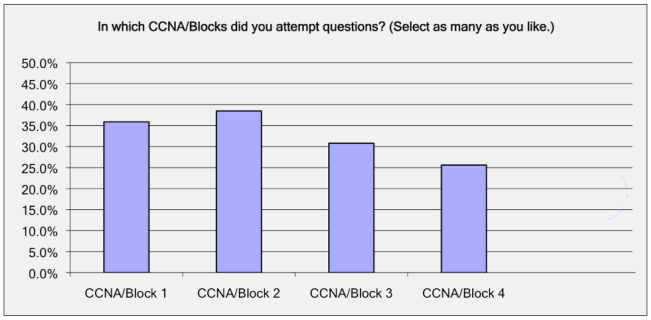


Figure 8 Access to T216 module blocks

The responses to a question regarding student learning, shown in Fig. 9, further demonstrates the value of this project. Over 75% of replies stating that the application had assisted in their learning and understanding of the Cisco materials. Two thirds of students agreed that the application had assisted with the completion of Cisco tests. The lesser impact on University assessments was not unexpected as the project was not focused in these areas.

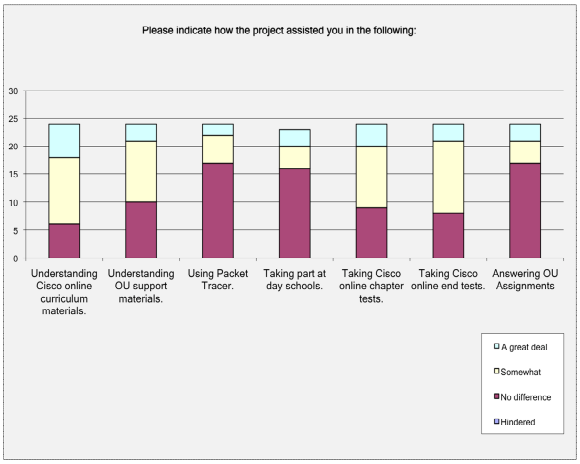


Figure 9. Students' opinions on how the application supported their learning

V.ii. How is your project contributing to increasing student performance?

Any improvement in student performance must lead to improved retention. The opportunity for students to gain more convenient access to module materials and support will support their learning and the evidence from my project demonstrates this to be the case. That students report a benefit to their knowledge and understanding of the areas targeted by this project establish this as fact. Success in the Cisco Tests, for which students receive a Cisco Certificate, can directly affect their employment opportunities. There is anecdotal evidence, backed up by the large number of students who take such modules, that proof of competence in the demonstration of knowledge and understanding of Cisco materials is highly regarded by employers.

 ISBN: 978-1-61208-133-5

VI. CONCLUSION

The use of mobile technologies to further enhance students' learning is evident from the results presented here. Whether sole use of such technologies is possible must rest on the appropriateness of the technology to support the applications and vice versa. Certainly screen size and the resultant legibility of both text and diagrams is a particular problem which may be resolved in future products. The iPhone and others have provided some improvements by incorporating expansion techniques but applications need to be able to exploit these techniques to be of use. An alternative way is that exploited by applications such as the Subnet Exerciser which proved that carefully designed applications may be of considerable help to students.

However, such applications must be available across all mobile platforms. Working to the lowest common denominator is probably the most satisfactory way for the time being.

A possible area to exploit is that of information dissemination. Reminding students of impending assessments and events or availability of results is simple to achieve, costs very little to implement and can be of huge benefit to students.

Disenfranchisement of students by technology is not acceptable in any circumstance. However, the mobile phone often the most advanced technology available to students, particularly in developing countries, so opportunities to provide quality learning through this medium must be taken.

VII. REFERENCES

1. Kukulska-Hulme, A and Traxier, J, (2005) Mobile learning, Routledge, Oxford, UK.
2. Kukulska-Hulme, Agnes and Shield, Lesley (2008). An overview of Mobile Assisted Language Learning: From content delivery to supported collaboration and interaction. ReCALL, 20(3), pp. 271–289.
3. Naismith L, Lonsdale P, Vavoula G, and Sharples M, (2006) Literature Review in Mobile Technologies and Learning, Futurelab, Bristol, UK.
4. http://www.ofcom.org.uk/research/cm/cmr08/ (accessed 18/08/09).
5. http://java.sun.com/products/cldc/ (accessed 07/01/11)
6. http://java.sun.com/products/cldc/ (accessed 07/01/11).
7. http://luziaresearch.com/ (accessed 13/01/11)
8. uhavepassed.com/ (accessed 13/01/11).

 ISBN: 978-1-61208-133-5

An Integrated TDMA-Based MAC and Routing Solution for Airborne Backbone Networks Using Directional Antennas

Yamin Al-Mousa, William Huba and Nirmala Shenoy
Laboratory of wireless networking and security
Rochester Institute of Technology
Rochester, NY, USA
{ysavks, wgh4235, nxsvks}@rit.edu

***Abstract*—Airborne backbone networks are useful in tactical applications to interconnect tactical sub-networks. Major challenge in such networks which normally comprise large numbers of highly mobile nodes lie in the design of medium access control (MAC) and routing protocols that can accommodate scalability and the highly dynamic topology. In this work, we propose an integrated solution of time division multiple access (TDMA) based MAC and routing protocols, both of which use the attributes of a clustering scheme, where clustering was adopted to address scalability. While the clustering scheme also establishes proactive routes between cluster clients and cluster head, the reactive routing protocol uses both the cluster attributes and the proactive routes within the cluster to address the challenges of high dynamics in the airborne network. The MAC is equipped with TDMA scheduler for operation with directional antennas used in the airborne nodes to provide spatial reuse. We assess the performance of the proposed integrated solution in terms of success rate and latency in packet delivery.**

Keywords-Airborne Backbone Network, Cluster formation, TDMA, Routing, MAC, Directional Antennas

I. INTRODUCTION

Backbone networks formed by airborne nodes such as *unmanned aerial vehicles* (UAVs) are of significant importance in tactical applications as they can be used to connect several tactical sub-networks which are at distance from one another. Such airborne nodes have significant processing and computation capabilities and can be equipped with directional antennas Though the targeted application is that of a backbone network, we limit our contribution to demonstrating the capability of the airborne backbone network to forward data reliably between two distant nodes in the backbone network, which could serve as the gateway points to two distant sub-networks.

Airborne backbone networks comprise large number of mobile nodes at speeds ranging from 200 to 300 km/h making scalability and highly dynamic topology two of the main challenges for designing an efficient solution. The contribution of this work is to provide an integrated solution of TDMA-Based MAC and routing protocols that can overcome previous challenges through following features:

- Clustering: was adopted to address scalability. We assume that cluster heads (CHs) are pre-assigned to last for entire mission duration. A cluster contains one CH and several cluster clients (CCs). Proactive routes are formed within a cluster, while routes across clusters are maintained reactively.
- Multiple redundant proactive routes: are formed between a CC and its CH so that if one route is lost another is ready to use, thus supporting dynamic topology. These routes are formed using the *Meshed Tree algorithm* [1] which simplifies proactive route formation and maintenance thanks to its unique naming scheme.
- Reactive routes: are maintained as a sequence of clusters, hence, reactive route discovery and maintenance is done at cluster level. This adds resiliency against mobility since reactive routes are not dependent on specific nodes, in addition to reducing control message flooding. Since reactive routes are concatenation of proactive routes which are continually updated with node mobility, the probability of stale reactive routes is low.
- Hybrid time division multiple access (TDMA) scheduler: is adopted by the MAC and uses the attributes of a multi hop clustering scheme to schedule time slots for CCs in the cluster. The scheduler uses directional antennas and is aware of the routing mechanism and proactive routes naming scheme within the cluster; hence schedules slots for data routing from CH to CCs and CCs to CH in an efficient manner providing spatial reuse.

The paper is organized as follows. In Section II, we highlight related work in the area of cluster based routing and TDMA scheduling. In Section III, we provide details of the physical layer. An abbreviated description of the meshed tree clustering scheme is provided in Section IV, followed by the proposed solution description including the proactive, reactive routing protocol and the TDMA scheduler. In Section V, simulation details with results and performance analysis are presented. Section VI provides concluding remarks and planned future work.

 ISBN: 978-1-61208-133-5

II. RELATED WORK

In this section, we present some work related to cluster based routing and TDMA scheduling. Though our solution combines both schemes effectively, a similar approach is not available in the literature to the best of our knowledge. Hence the first part of the related work deals with cluster based routing. This topic has been researched extensively; we present only those closely related to our approach. The second part of related work deals with TDMA schedulers, especially the ones that use directional antennas and leverage spatial reuse, as they are closely related to our approach.

A. Cluster Based Routing

Several reviews are available that compare across different types of routing protocols [2-6], namely proactive, reactive and hybrid routing protocols. Reactive and hybrid routing protocols are desirable when communications between distant nodes in a MANET are required and the MANET is not very dense. In reactive routing protocols a source node discovers and maintains several cached routes to a destination node. As mobility increases, route caching becomes ineffective as pre-discovered routes break down, requiring repeated route discoveries [10].

Partitioning the MANET through clustering and zoning is useful to limit control messages and also to address scalability. Hybrid routing protocols normally adopt zoning or clustering, and then use proactive routing protocols within the zone and reactive routing protocols to communicate with nodes outside of a zone. The *Zone Routing Protocol* (ZRP) [11] is one such hybrid routing protocol, where each node has a pre-defined zone centered at itself. ZRP proposes a framework, whereby 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 (MDVZRP) [12] is an implementation of ZRP that uses multi path *Destination Sequence Distance Vector* (DSDV) [9] for proactive routing and *Ad-hoc On-demand Distance Vector* (AODV) [7] for reactive routing. *LANMAR* [13] routing protocol defines logical groups to address scalability where by landmark nodes keep track of the groups. A local scope routing scheme based on *Fisheye State Routing* is used in the group. To forward outside the scope, packets are routed towards the landmark in the destination's logical group. In *Hybrid Cluster Routing* [14] multi-hop clusters are established. However in this case intra-cluster uses a reactive routing scheme similar to AODV and *Dynamic Source Routing*,[8] while inter cluster maintenance is done proactively.

Our Approach: is a hybrid cluster based routing protocol. Using the *Meshed Tree algorithm* proactive routes are formed between CCs and CHs. For routing across clusters a reactive routing approach at cluster level, which uses concatenations of proactive routes within the clusters, is adopted. We argue that our approach is different in adopting the *Meshed Tree algorithm* for cluster formation which uses a unique proactive route naming scheme. In more details each route is given a virtual ID (VID) which reveals current route topology information simplifying the task of forming and maintaining routes and helps in calculating efficient TDMA schedules.

B. Schedulers: Time Division Multiple Access

TDMA scheduling requires strict time synchronization among participating nodes [15]. In addition, if the nodes are mobile, periodic changes in the network topology require updated TDM schedules, to be computed, preferably with low complexity and propagated to all concerned nodes in a timely and an efficient manner.

A major challenge in the design of a TDMA scheduler is the generation of schedules. Several algorithms directed towards scheduling can be noted in the literature [16]. Scheduling algorithms fall under two main categories distributed or centralized. In the centralized approach, scheduling is performed by a scheduler that gathers information about all nodes and their links to compute the schedule. This is a difficult task to achieve in a timely and resource efficient manner, especially with large numbers of mobile nodes. On the other hand, distributed scheduling requires complex algorithms with intelligence to enable each node to decide their schedules with minimal conflicts.

Our Approach uses hybrid scheduling, which is possible due to the cluster based approach. Within a cluster the CH is the scheduler that decides the transmission reception schedules for its CCs. However each cluster's schedule is determined independently by its CH giving conflict consideration only to those CCs that are bordering two or more clusters thus making it distributed across clusters. Given that *link assignment strategies* are efficient if employed with directional antennas [15] and as the proposed clustering scheme has such link information available in a cluster we decided to adopt this type of assignment strategy. However, our approach is different in using topology information contained in proactive route naming scheme to calculate efficient TDMA schedule and maximize spatial reuse with the aid of directional antennas.

III. THE PHYSICAL LAYER

In this section, we describe the operational features of the directional antenna system used at the physical layer. All nodes in the airborne network can be equipped with four phased array antennas capable of forming two beam widths. One is focused with a beam angle of 10^o and the other is defocused with a beam angle of 90^o. In the focused beam mode the data rate is 50 Mbps and in the defocused mode the data rate is 1.5 Mbps. Each antenna array covers a quadrant and is independently steerable to focus in a particular direction within that quadrant in the focused beam mode.

We assume each node is equipped with Global Positioning System (GPS) to provide node position. Every node appends its GPS location in the packets it transmits. Receiving (neighbor) nodes log and continuously update a "location" cache with the transmitting node's location. The cache stores a maximum of the last three positions of any node. Cached location information is used to track and estimate the current location of neighboring nodes during packet transmission. The estimated location of a receiver node is used, by a transmitting node, to control the transmit

 ISBN: 978-1-61208-133-5

power and form a directed beam to the receiver node by using the most appropriate of its phased array antennas.

GPS is also used for time synchronization, and we assume that all nodes are synchronized to time slot boundaries and the beginning of new frames. However, a guard time is included in each time slot to offset synchronization errors as well as to allow for beam switching.

IV. MESHED TREE CLUSTERS

It is important to understand the meshed tree cluster formation and proactive routing within the cluster as they are used both by the scheduling algorithm and reactive routing protocol. The clustering scheme adopted in this work, forms multi hop clusters using the *Meshed Tree algorithm* [1], where the root of the meshed tree is at the CH. A single 'meshed tree' cluster formation is described with the aid of Figure 1.

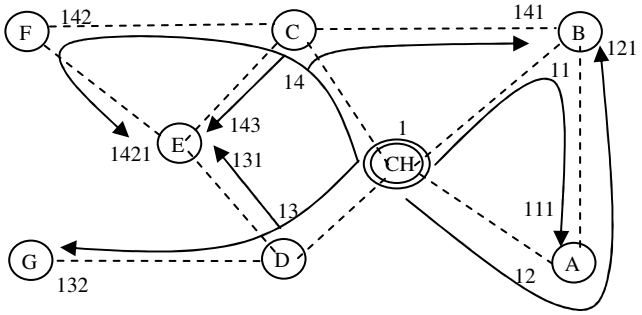


Figure 1. Cluster Formation Based on Meshed Trees

The dotted lines connect nodes that are in communication range with one another at the physical layer. The cluster head is labeled 'CH'. Nodes A to G are the CCs. For simplicity in explanation, the meshed tree formation is 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 CC defining branches, proactive routes, linking it with its CH. Assuming that the CH has a VID '1', all its CCs have '1' as a prefix in their VIDs. Any CC that attaches to a branch is assigned a VID, which will inherit the prefix from its parent node, followed by an integer, which indicates the child number under that parent. In this work we limit the number of children to nine and use a single digit without loss of generality.

A. Proactive Routes in the Cluster

In Figure 1, each branch is a sequence of VIDs that is assigned to CCs connecting at different points of the branch. The branches of the meshed tree thus 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 then 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 routes through the multiple VIDs, allows for robust and ***dynamic route adaptability*** to topology changes in the cluster, as the nodes can request for new VIDs and join different branches as their neighbors change.

B. Inter-cluster Reactive Routing

Nodes bordering two or more clusters are allowed to join the branches originating from different CHs, and will accordingly inform their respective CHs about their multiple VIDs under the different clusters. This information will enable the CHs to avoid conflicts when scheduling timeslots for such border nodes. Also by allowing nodes to belong to multiple clusters, the single meshed tree cluster can be extended to *multiple* overlapping *meshed tree* (MMT) clusters to cover a wider area and address *scalability.*

A node that has to discover a route to a distant node sends a 'route request' message to its CH(s). The CH then identifies the neighboring clusters based on updates from border nodes and forwards a copy of the 'route request' message to the border node, so that they can forward to the CH in the next cluster. The 'route request' message however has an entry for all the clusters that will be receiving the message, to avoid looping of the message. Thus the route request is not forwarded by all nodes, but only by all clusters and follows a path CH-border node- CH and so on.

When the CH of the destination node receives the route request, it will forward the route request directly to the destination node. The clusters forwarding the route request record the original sending node and the last cluster that the route request came from; this information is useful in forwarding the route response message when it returns. The destination node generates the route response and sends to its CH, which then forwards it back to the CH in the originating cluster and the source node along the same *cluster path* the route request took. Along the path back, all forwarding CHs will record the previous cluster and original sender of the route reply. The route between the sender and the destination node is thus initially set up as a sequence of CHs, but maintained as next cluster information. Mobility of nodes does not impact the reactively discovered route, as long as the CHs exist. Note that movement of CHs also does not impact the reactive routes.

C. Scheduling in the Cluster

The meshed tree cluster is formed in a distributed manner, where a node listens to its neighbor nodes advertising their VIDs, and decides to join any or all of the branches. Once a node decides to join a branch, it informs the CH, who registers the node as its CC and confirms its admittance to the cluster and accordingly updates a VID

 ISBN: 978-1-61208-133-5

table of its CCs as shown in Table 1 for the cluster in Figure 1. Thus the 'meshed tree' cluster formation allows a CH to control the nodes it accepts; i.e., a CH can restrict admittance of nodes who are within a certain number of hops and not dmit new nodes to keep the number of CCs in the cluster under a certain value. This is useful to contain the scheduling zone of the CH.

TABLE I. CLUSTER CLIENT'S VIDS LIST AT CH

Node	Multiple VIDs
A	12, 111
B	11, 121, 141
C	14
D	13
E	131, 143, 1421
F	142
G	132

From the *Cluster Client's* VID table, implicit topology information is available to the CH; for example node B has a VID 1421, indicates that it has a link to the node with VID 142. The CH will use this information and its capabilities of controlling and communicating with the CCs to establish recurring time frames with a time slot scheduled for transmission and reception on the links between CCs and between CCs and CH in the cluster. As nodes, join and leave a cluster, the CH updates this table and announces the new schedule. Thus the scheduling operations are closely integrated with the cluster formation process.

D. Slots and Functions

A frame comprises of control and data slots. In this work four slots are preselected for control purposes. These slots are used by nodes to advertise their VIDs, and other broadcast information; and also to listen to advertisement by neighbor nodes. Remaining slots are used for transferring data packets and other control packets between CCs and CH. From a node's perspective, assigned data slots can either be used for reception or transmission using focused beams.

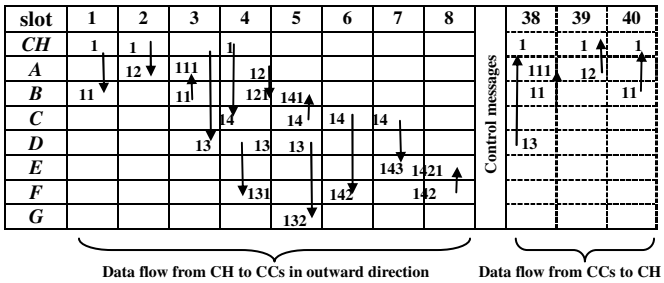
Slots that are not assigned to be either control or data slots are considered temporary slots. Nodes may use such temp slots to transmit packets when they do not have an assigned data slot yet, such as during the registration process. When a node does not have anything to send on a temp slot, it will listen for any incoming transmissions. At any time these slots can be changed to an assigned data slot by the CH.

The cluster schedule is distributed by the CH to all CCs in the cluster at the start of a frame with "beacon" packets. Each CC independently chooses the 'best' (in our case shortest, which is decided on the VID length) route to forward the beacon packet using meshed tree's routing information. Schedules for a given frame are transmitted one frame ahead to allow enough time for the beacon packets to reach nodes that are at the maximum hops from the CH.

E. Sample Schedule

A sample schedule generated by the proposed scheduler is given in Table 2. Each column is a slot; we show only 12 slots, which is a partial frame. In the first column are the node's unique IDs, which in this case are the alphabets we used for identifying the CCs in Figure 1. In each column we mark the VID of the sending and the receiving nodes, and the arrow shows the direction of transmission. For example in slot 1 CH (VID '1') sends to node A with VID '11'. The slot allocation process, proceeds by allocating slots for the CH to 1 hop nodes, followed by the 1 hop CCs sending to their 2 hop children and the 2 hop CCs sending to their 3 hop children and so on. However, due to the directional antennas used we can have simultaneous transmission between two pairs of distinct nodes; for example in slot 3 CH is sending to node D on VID 13, but node B using VID '11' is sending to node A at VID '111'. A closer look at the schedule will reveal that the flow from the outer leaf nodes to the CH is the mirror of the allocation process from CH to leaf nodes i.e., the 1st hop children are allocated the last time slots in the frame.

TABLE II. SAMPLE SCHEDULE PROVIDED BY CH



V. SIMULATION RESULTS

We conducted simulations using OPNET for 20, 50 and 75 node multi-hop networks. All nodes were randomly assigned clockwise and counter-clockwise circular trajectories, with 100 Km radius and speeds varying between 200, 250, 300, and 350 Km/h at 20000 m altitude. The circular trajectories provide a stressful test as they result in many route breaks.

The frames had 28 slots each for the 20 and 50 node scenario, and 42 slots for the 75 node scenario, because the meshing in the cluster, results in most nodes using up all 6 VIDs and hence requires more slots, which is being optimized. Each slot had a 12.5ms duration and a guard time of 1.5 ms. The cluster size was maintained at 12 with a maximum of 3 hop distance between a CC and CH, and the most VIDs a node could have was set at 6.

Nodes in the network were randomly selected to send 1 MB file simultaneously, in 2 KB packet sizes to destination nodes also randomly selected. We measured overhead, average hops, successful packet delivery rate, as well as mean packet latencies, where;

- *Success rate* was calculated as the number of packets delivered to the destination node successfully as a percentage of the number of packets that originated at the sender node

- *Overhead* was calculated as the ratio of control bits to the sum of control and data bits during data delivery.

Each simulation was run with several seeds and the average values were plotted in the graphs shown in Figures 2-7. As there are no published results for such network scenarios to the best of our knowledge, we use the graphs to highlight the performance of the proposed solution.

A. 20 Nodes Scenario

The number of such sending nodes were varied from 4, 8 to 16 nodes. In the case of 16 senders, all CCs were sending 1 MByte files to all other CCs in the network, which is stress test case.

From Figure 2, the success rate was around 97% with 4 senders and dropped to 94% with all 16 senders. With increasing traffic in the network, one notices the reduction in the overhead; this is due to the inverse relationship between the traffic load in the network with respect to the control bits generated. In Figure 3, the average packet latency increased from 0.7 second to 3 seconds, when the traffic in the network increased. We recorded the average hops encountered between sending and receiving nodes to get an indication of the distance between the communicating nodes as it affects the success rate and overhead in the network. However in this case the recorded average hops was around 3.5 hops

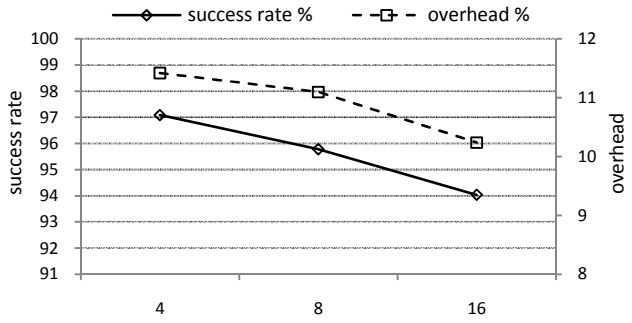


Figure 2. Success Rate and Overhead vs. Number of Senders

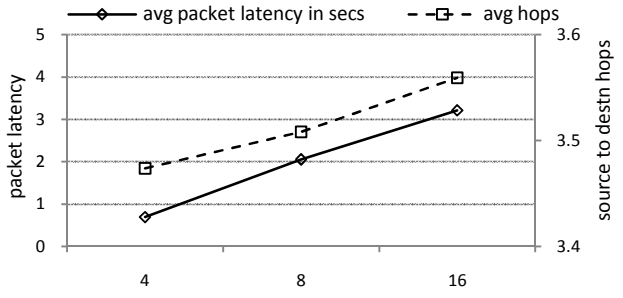


Figure 3. Packet Latency and Average Hops vs. Number of Senders

B. 50 Nodes Scenario

In this network scenario, the number of simultaneously sending nodes was varied from 10, 20 to 40. There were 10 CHs in this scenario hence with 40 sending nodes, all CCs were sending to all other CCs in the network. The success rate was 94% with 10 senders, and dropped to around 83% with 40 senders as shown in Figure 4. The overhead recorded is around 23%, which is higher than the 20 node scenario and can be attributed to the increase in route discovery maintenance across 50 nodes. In Figure 5, the average packet latency was recorded as 2.5 seconds with 10 senders and 11 seconds with 30 senders, which is reasonable to assume with the increased traffic in the network. The average hops recorded were between 6 to 7. The consistency in performance and the graph trends can be considered to be indicative of the stability of the proposed algorithms and the models.

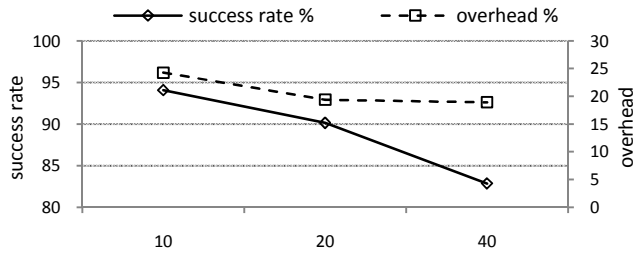


Figure 4. Success Rate and Overhead vs. Number of Senders

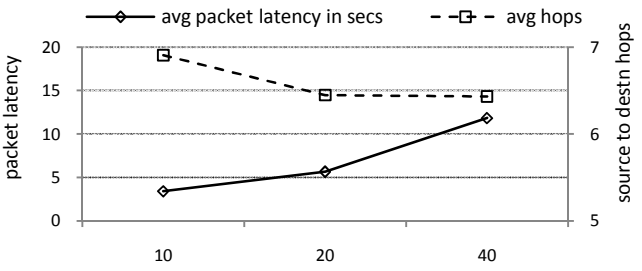


Figure 5. Packet Latency and Average Hops vs. Number of Senders

C. 75 Node Scenario

In this scenario we varied the number of senders from 15, 30 and 60 nodes. As the number of CHs was 15, in this case again we had all nodes sending traffic to all other destination nodes. The success rate shown in Figure 6 was 90% at 15 senders, which dropped to 70% with all 60 senders. The overhead was recorded to vary from 27 to 30%. In Figure 7, An increase in average packet latency can be noticed, which can be attributed to the increased traffic in the network.

D. Summary of Results

As stated earlier, due to the uniqueness of our approach we are unable to provide comparison with similar work conducted for airborne backbone networks. Furthermore, to the best of our knowledge, such stressful MANET scenario evaluations are also not available in the literature, because of which we present results, based on some targeted goals.

These being a high value of successfully delivered packets with some acceptable latencies, based on the traffic

 ISBN: 978-1-61208-133-5

in the network. High success rate is difficult to achieve in such highly dynamic MANETs, especially when the number of mobile nodes is also high – several tens in this case. This is a good performance assessment if the type of data is files.

As part of our future work, we plan to extend the work to real time services, as an airborne backbone network is expected to carry different types of traffic, which originate from its subnets, which could be ground troops, UAVs performing surveillance amongst others. This will include prioritization of traffic while forwarding at the MAC layer. Future work will also involve optimizing our slot assignment algorithm, evaluating for various slot sizes, varying number of frame sizes and cluster sizes. We also plan to investigate the impacts of meshing (intra-cluster and inter-cluster), which can be controlled by the number of VIDs and the criteria used by CCs to acquire a VID.

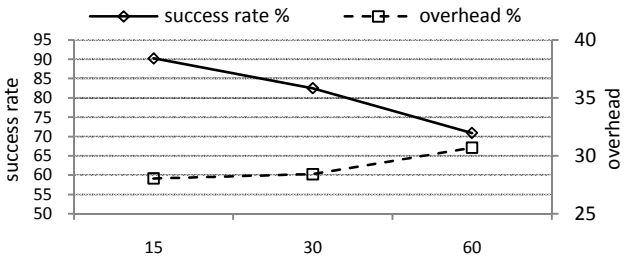


Figure 6. Success Rate and Overhead vs. Number of Senders

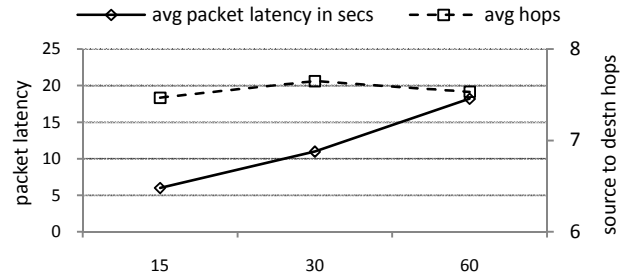


Figure 7. Packet Latency and Average Hops vs. Number of Senders

VI. CONCLUSION

We presented an integrated TDMA-Based MAC and routing protocol in this work, both of which were based on a meshed tree algorithm. The solution is unique both from the perspective of the TDMA scheduler and the routing protocol. The preliminary evaluations of this scheme show the very promising results that were obtained for airborne backbone networks. The consistent performance is also indicative of the stability of the proposed algorithms

ACKNOWLEDGMENT

The authors would like also to thank Chukwuchebem Orakwue fo his effort in OPNET model Implementations.

This work was supported partly by funding from Air Force Research Laboratories, Rome NY and Office of the Naval Research.

REFERENCES

[1] Shenoy, N.; Yin Pan; Narayan, D.; Ross, D.; and Lutzer, C.; "Route robustness of a multi-meshed tree routing scheme for Internet MANETs," Global Telecommunications Conference, 2005. GLOBECOM '05. IEEE , vol.6, no., pp.6 pp.-3351, 2-2 Dec. 2005.

[2] Qin, L. and Kunz, T., "Survey on Mobile Ad Hoc Network routing protocols and cross-layer design," *Technical Report, Systems and Computer Engineering, Carleton University*, Aug. 2004

[3] Abolhasan M., Wysocki T., and Dutkiewicz E., "A review of routing protocols for mobile ad hoc networks," *Ad Hoc Networks*, vol. 2, Issue 1, pp. 1-22, ISSN 1570-8705, Jan. 2004

[4] Lang, D., "A comprehensive overview about selected Ad hoc networking routing protocols," Technical Report, Department of Computer Science, Technische Universitat, Munchen, Mar. 2003

[5] Qadri N. and Liotta A. "Analysis of Pervasive Mobile Ad Hoc Routing Protocols," Pervasive Computing, Computer Communications and Networks, ISBN 978-1-84882-598-7, Springer-Verlag London, 2009

[6] Meghanathan N., "Survey and Taxonomy of Unicast Routing Protocols for Mobile Ad Hoc Netowrks," GRAPHHOC, vol. 1, 2009

[7] Perkins, C.; 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, Jul. 2003

[8] Johnson, D.; Maltz, D.; and Jetcheva, J., "The dynamic source routing protocol for mobile ad hoc networks," *Internet Draft, draft-ietf-manet-dsr-07.txt*, work in progress, 2002.

[9] Perkins, C.E. and Watson, T.J., "Highly dynamic destination sequenced distance vector routing (DSDV) for mobile computers," *ACM SIGCOMM_94 Conference on Communications Architectures, London, UK*, 1994.

[10] Das, S.R.; Perkins, C.E.; and Royer, E.M., "Performance comparison of two on-demand routing protocols for ad hoc networks," *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE* , vol.1, no., pp.3-12 vol.1, 2000

[11] 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, pp. 427-438, August 2001

[12] 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

[13] Guangyu Pei; 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*, pp. 11-18, Aug. 2000.

[14] Xiaoguang Niu; Zhihua Tao; Gongyi Wu; Changcheng Huang; and Li Cui, "Hybrid Cluster Routing: An Efficient Routing Protocol for Mobile Ad Hoc Networks," *Communications, 2006. ICC '06. IEEE International Conference on* , vol.8, no., pp.3554-3559, June 2006

[15] Nelson R. and Kleinrock L., "Spatial-TDMA: A collision-free multihop channel access protocol," *IEEE Transactions on Communications 33* (1985) pp 934–944.

[16] Grönkvist J.; 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.

[17] Grönkvist J., "Novel Assignment Strategies for Spatial Reuse TDMA in Wireless Ad hoc Networks," *Wireless Networks*, Springer Netherlands, ISSN 1022-0038, vol. 12, no. 2, pp. 255 – 265, 2006

 ISBN: 978-1-61208-133-5

Right-time Path Switching Method for Proxy Mobile IPv6 Route Optimization

Yujin Noishiki, Yoshinori Kitatsuji, Hidetoshi Yokota
KDDI R&D Laboratories, Inc.
2-1-15 Ohara Fujimino Saitama, JAPAN
{yujin, kitaji, yokota} [at] kddilabs.jp

***Abstract*—Proxy Mobile IPv6 provides IP mobility to a mobile node by the proxy mobility agent called a Local Mobility Anchor and a Mobile Access Gateway without requiring mobile node's participation in any mobility-related signaling. Increased demand for content-rich mobile data communications is prompting mobile network operators to deploy efficient mobility management including Proxy Mobile IPv6. The route optimization technique is applied to the data path between mobile nodes in the same Proxy Mobile IPv6 domain by bypassing the Local Mobility Anchor(s). However, when switching the data path from the default (non-optimized) route to an optimized one, the delay gap between these paths leads to performance degradation due to out-of-sequence packets, or unnecessary communication disruption. We propose a right-time path switching method for Proxy Mobile IPv6 route optimization. This method enables the Mobile Access Gateway to switch these paths with the accurate timing provided by the designated signaling messages, which prevents out-of-sequence packets as well as minimizing communication disruption during the route optimization procedure. The proposed method is evaluated in an actual testbed to show that the proposed method achieves the seamless path switch.**

Keywords - Proxy Mobile IPv6; Route Optimization; Path Switch Optimization

I. INTRODUCTION

Mobility management is an important function for mobile communication. The Internet Engineering Task Force (IETF) has standardized Mobile IP [1][2] to provide mobile nodes (MNs) with IP mobility. However, Mobile IP is a client-based mobility management scheme that requires MNs to implement protocol stacks and exchange mobility-related signaling. Therefore, the IETF has standardized Proxy Mobile IPv6 (PMIPv6) [3] as a network-based mobility management protocol. PMIPv6 brings the IP mobility to the MN without requiring its participation in any mobility-related signaling.

PMIPv6 operations are performed by two network entities, Local Mobility Anchors (LMAs) and Mobile Access Gateways (MAGs). An LMA is a home agent for the MNs and a topological anchor point for the home network prefix of MNs. An MAG is an access router that exchanges the mobility-related signaling with an LMA instead of an MN attached to the MAG via the wireless accesses.

In PMIPv6, all data traffic originating from or destined for the MNs is transferred through an LMA even if the MNs communicate with each other. Such a redundant routing increases transfer delay, which leads to performance degradation of MN's communications. In addition, data traffic transferred via the redundant route concentrates traffic on LMAs. To overcome this situation, route optimization for PMIPv6 is an attractive solution for minimizing delay and realizing traffic offload.

The route optimization for PMIPv6 is realized by using the direct tunnel established between the MAGs, to which mobile nodes are attached, with bypassing the LMA [4]. However, when switching the data path from the redundant (non-optimized) one to the direct tunnel (the optimized path), the delay gap between these paths causes performance degradation of MN's communication. If the data path is switched before finishing receipt of data packets via the non-optimized path, out-of-sequence packets occur, which decreases TCP performance of MN's communication. On the other hand, if the data path is switched too late, MNs experience communication disruption. This unnecessary disruption degrades the service (e.g., voice and video) quality of real-time applications.

In this paper, we propose a right-time path switching method for PMIPv6 route optimization. After the optimized path is ready, our proposed method initiates the path switch using signaling messages. This feature prevents out-of-sequence packets as well as minimizing communication disruption duration in the route optimization procedure.

The proposed procedure is evaluated in an experimental testbed using actual PCs. The results reveal that our proposed method prevents out-of-sequence packets while the baseline route optimization procedure causes them. In addition, performance evaluation shows our proposed method decreases communication disruption duration in the route optimization procedure.

This paper is organized as follows. Section II shows related work. Section III proposes a route optimization procedure with the optimized timing of path switch. Section IV evaluates the performance of the proposed method using the experimental testbed. Section V concludes this paper.

II. RELATED WORK

Mobile IPv6 (MIPv6) [2] supports a route optimization scheme, which allows an MN to register its binding information with a corresponding node (CN). The CN directly sends or receives data packets using the MN's care-of address after route optimization. Similarly to the route optimization for PMIPv6, when MN and CN switch the data path from the non-optimized path to the optimized path, out-of-sequence packets are caused. However, unlike PMIPv6, the route optimization procedure for MIPv6 is performed by

 ISBN: 978-1-61208-133-5

MN and CN themselves, thus they are involved in preventing out-of-sequence packets. On the other hand, in PMIPv6, MNs cannot handle out-of-sequence packets because they do not detect the timing of the path switch to the optimized path.

Lee, *et al*. [5] have proposed a route optimization scheme for PMIPv6 to prevent out-of-sequence packets. In the proposed scheme, an MAG buffers data packets originating from MN until the optimized path has been created. However, this scheme may increase communication disruption duration because the buffering MAGs cannot know the end of data forwarding through the non-optimized path. Our paper proposes a route optimization procedure that avoids out-of-sequence packets while minimizing the communication disruption duration.

In order to indicate the end of data forwarding at path switching, an end-marker approach is applied in the 3GPP standard [6]. During handover procedures, this end-marker is transferred to indicate the end of the data stream in a forwarding tunnel. The indication is included in GPRS Tunneling Protocol User Plane (GTP-U) [7], which is the transport protocol for user data packets. While this indication is deployed only for GTP-U and requires packet inspection on the user plane, our approach using PMIPv6 signaling messages is separate from user plane transport protocol.

III. PROPOSAL OF RIGHT TIME PATH SWITCHING METHOD

In this section, we introduce the basic PMIPv6 operation and the required functions for PMIPv6 route optimization. Then, the right-time path switch method for PMIPv6 route optimization is proposed.

A. PMIPv6 Operation

The PMIPv6 domain is shown in Fig. 1. In the PMIPv6 domain where mobility management is performed using PMIPv6, LMAs and MAGs are located.

Basic PMIPv6 operation is as follows. When an MAG detects the attachment of an MN, it sends a proxy binding update (PBU) message. When an LMA receives the PBU message, it registers the MN in a binding cache entry (BCE), and replies to the MAG by sending a proxy binding acknowledge (PBA) message. The PBA message includes the Home Network Prefix (HNP) of the MN. Then, the MAG notifies the HNP of the MN. After exchanging PBU and PBA messages, MAG and LMA hold the binding cache entries of MNs including the HNPs.

Once the MN is registered in the PMIPv6 domain, all data packets of the MN are transferred through the MAG and the LMA. This characteristic causes a redundant routing when MNs communicate with each other. For example, when MN1 is registered in LMA1 via MAG1 and MN2 is registered in LMA2 via MAG2, data packets from MN1 to MN2 are forwarded via MAG1, LMA1, LMA2, and MAG2. This redundant routing leads to transfer delay in communication of MNs, such as performance of real-time applications.

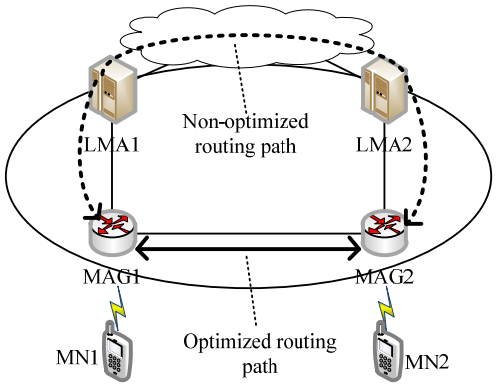


Figure 1. PMIPv6 domain with non-optimized and optimized routing paths.

B. Required Functions for Route Optimization

To overcome the redundant routing in PMIPv6, route optimization is a promising approach. Route optimization transfers data packets between two MNs attached to the same PMIPv6 domain, through an optimized routing path between MAGs bypassing the LMAs as shown in Fig. 1.

The functions required to realize route optimization are as follows.

- Detection of the target communication for route optimization: to trigger the route optimization procedures, data packets exchanged between MNs in the PMIPv6 domain must be detected. When MNs are attached to different MAGs and registered at different LMAs, this packet detection is complicated because the binding cache entries of each MN are distributed in LMAs and MAGs.
- Discovery of network entities (LMA and MAG) relating to the target MNs: to exchange signaling messages, the LMAs that register the MNs' BCEs and the MAGs that attach the MNs must be discovered. When MNs are registered at different LMAs, discovery of the LMA from another LMA is difficult since each LMA does not know the MNs in the other LMAs.
- Establishment of the optimized routing path: the optimized routing path is established between MAGs that attach MNs.

C. Baseline Route Optimization Procedure

We first explain the baseline route optimization procedure, which meets all requirements mentioned in the previous subsection.

In the case where either or both proxy mobility agent(s) (MAG/LMA) is/are shared by the MNs, the requirements of detection of the target communication and discovery of the involved mobility agents are fulfilled in a straightforward way because the shared mobility agent manages the binding caches of both MNs. However, since each MN is registered with separate MAG and LMA, none of these agents satisfy the requirements because the binding cache entries of MNs are distributed over different LMAs and MAGs. In order to cover this most generalized case, we discuss the situation

 ISBN: 978-1-61208-133-5

where MN1 is attached to MAG1 and registered at LMA1, and MN2 is attached to MAG2 and registered at LMA2 in the PMIPv6 domain shown in Fig. 1.

To fulfill the requirements for the route optimization in the above situation, the Policy Store (PS) defined in [8] is leveraged. As shown in Fig. 2, the PS is deployed in the same PMIPv6 domain and stores binding caches of MNs including the HNPs and the IP addresses of LMAs. LMAs register the binding cache with the PS when their BCEs are updated, for example, at the time of the reception of PBU by the LMA. Each LMA obtains the information of MNs registered at other LMAs by referring to this PS. In the 3GPP standard, the AAA server plays a role of the PS in the PMIPv6-based mobile core networks [9], where the LMA registers binding caches with the AAA server when the binding caches are updated.

The baseline route optimization procedure is shown in Steps 1 to 9 of Fig. 3 except Steps E1a-d and E2a-d enclosed in boxes. In this procedure, we show that the data packets of the target of route optimization are transferred from MN1 to MN2. To detect the target data packets, LMAs monitor the source IP addresses. In this procedure, LMA2 checks the data packets in Step 1a. If the source IP address is not registered at LMA2, LMA2 refers to the source IP address from the PS in Step 1b. When the source IP address is found, LMA2 begins the route optimization. In this step, LMA2 recognizes LMA1, which has the BCE of the source IP address (MN1) from the PS. To prevent a route optimization triggered by the data packets in the other direction (from MN2 to MN1), LMA notifies the beginning of route optimization for pairs of MNs to the PS. After this notification, the PS does not allow other LMAs to begin the route optimization for the same pair of MNs.

To meet the requirement of establishment of the optimized path, Steps 2 to 9 are performed. While PMIPv6 has no interface between LMAs in the IETF standardization, this paper introduces a new signaling interface between LMAs as shown in Fig. 2 to handle the situation where MNs are registered at different LMAs.

In Steps 2 and 3, LMA2 sends a Route Optimize (RO) Trigger message to LMA1, and then LMA1 sends an RO Initiate message to MAG1, respectively. The RO Trigger and RO Initiate messages include the HNPs of MN1 and MN2, and the IP address of MAG2. In Step 4, MAG1 offers MAG2 establishment of the direct tunnel by sending an RO Request message including the HNPs of MNs. When receiving this message, MAG2 creates the forwarding tunnel to MAG1 and replies with an RO Request Acknowledge message to MAG2 in Step 5. This message requests MAG1 to establish the direct tunnel from MAG1 to MAG2. After the direct tunnel is ready, MAG1 responds an RO Request Complete message to MAG2 and an RO Initiate Acknowledge message to LMA1 in Steps 6 and 7, respectively. LMA1 responds an RO Trigger Acknowledge message to LMA2 in Step 8. Finally, LMA2 updates the binding caches of MN1 and MN2 by notifying the end of the procedure to the PS in Step 9.

Steps E1a-d and E2a-d enclosed in boxes in Fig. 3 are described in the next subsection.

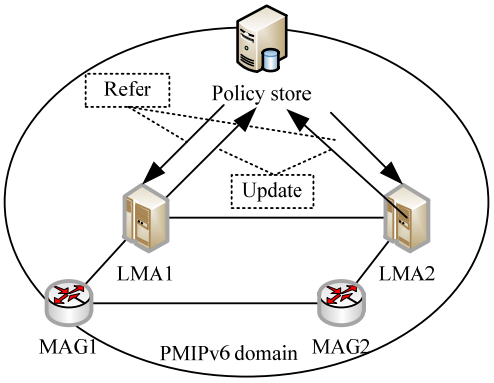


Figure 2. Proposed architecture with a policy store.

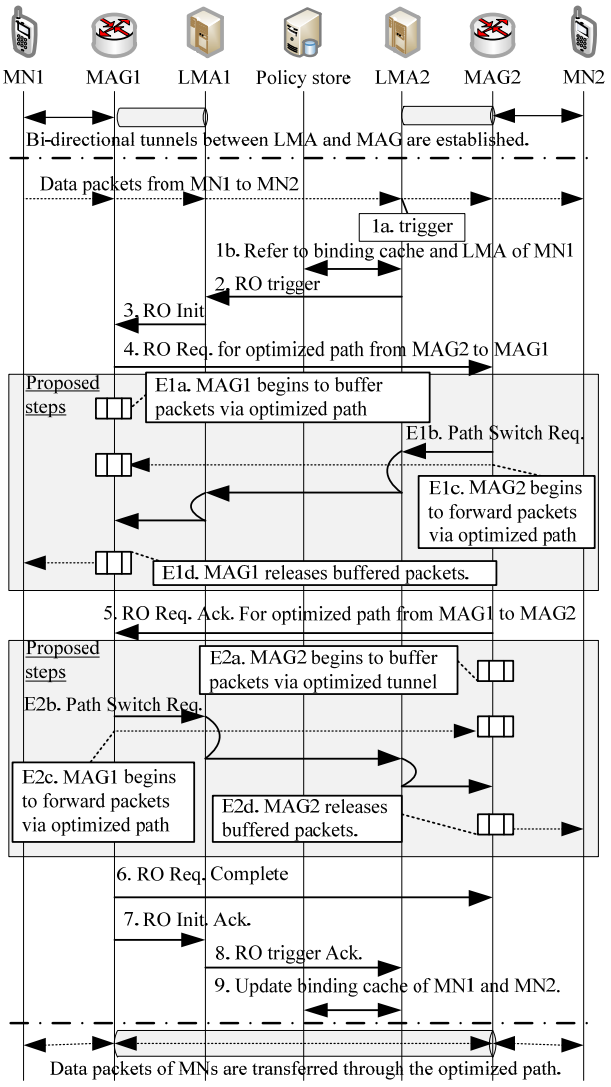


Figure 3. Route optimization procedure with optimized path switching. Signaling messages are indicated by a solid line while data packets are indicated by a dotted line. The procedure regarding the optimized path switching is enclosed by boxes.

 ISBN: 978-1-61208-133-5

D. Proposal of Right-time Path Switching Method

When the optimized path is established, the path switch from the non-optimized path to the optimized path should be performed in an appropriate timing. In this subsection, we propose the optimized path switching method for route optimization in PMIPv6.

In [5], to prevent out-of-sequence packets at the path switch, MAGs buffer the packets originating from MNs until the optimized routing tunnel is established. In this method, the sender for data packets from MN1 to MN2 on the optimized path, e.g., MAG1, buffers the data packets and decides to begin data forwarding through the optimized path. Therefore, we call this method the sender-buffering method. If this sender-buffering method is employed for the procedure shown in Fig. 3, after Step 3, MAG1 begins to buffer the packets from MN1, and then releases the buffered packets just after Step 5. Similarly, MAG2 begins to buffer the packets from MN2 after Step 4, and then releases the packets in Step 6.

However, this sender-buffering method may cause one of two drawbacks: out-of-sequence packets and relatively large communication disruption, because the sender of the optimized tunnel cannot detect when the buffered packets should be released by itself (the sender cannot know when the receiver at the tunnel receives the last data packets transferred via non-optimized tunnel). Therefore, if the data packets buffered are released too early, out-of-sequence packets will be caused at the receiver at the tunnel. On the other hand, if the buffered data packets are released too late, disruption duration for data packets occurs from the last packets through the non-optimized path to the first packets through the optimized path. Such communication disruption duration degrades MN's communication.

To switch the data path from a non-optimized path to an optimized path with accurate timing, we propose a new method shown in Steps E1a-d and E2a-d in Fig. 3. In this method, after the optimized path is established, the receiver at the tunnel buffers the data packets through the optimized path as opposed to the sender-buffering method. Then, the sender at the tunnel notifies the end of data forwarding through the non-optimized path by sending signaling messages after the sender forwards the last packets through the non-optimized path. This signaling message is transferred via the non-optimized path. The receiver that buffers the data packets recognizes the receipt of the last packets transferred through the non-optimized path, preventing out-of-sequence packets and communication disruption.

The proposed method is implemented as follows: After MAG1 sends MAG2 the RO Request message in Step 4, MAG1 begins to buffer the packets that arrive through the optimized tunnel in Step E1a. When it receives the RO request message in Step 4, MAG2 sends a Path Switch Request message to MAG1 via LMA2 and LMA1 in Step E1b. This message includes the HNPs of MN1 and MN2, and the IP addresses of LMA1 and LMA2. Just after sending the message, MAG2 begins to forward data packets from MN2 to MN1 through the optimized routing tunnel in Step E1c. While MAG1 receives the data packets from LMAs (through the non-optimized path), MAG1 buffers all the data packets forwarded through the optimized tunnel. Thus these data packets are not forwarded to MN1. In Step E1d, when MAG1 receives the Path Switch Request message, MAG1 begins to release the buffered data packets, which are transferred through the optimized tunnel. Finally, MN1 receives all the data packets in the correct order.

Similarly to Steps E1a-d, the proposed approach in Steps E2a-d is performed in the opposite direction. Thus, the proposed path switch method is applied in both directions.

IV. PERFORMANCE EVALUATION

A. Evaluation Environment

To investigate the effect of the path switching method for the route optimization procedure on the quality of service, we focus on two metrics, the total number of out-of-sequence packets in both directions and the duration of the communication disruption. These values are measured during the path switch from the non-optimized path to the optimized one. In this paper, we define the duration of the communication disruption as the time at the MN from the receipt of the last packet through the non-optimized path to the arrival of the first packet through the optimized path. All results were obtained 10 times and the average is presented.

To evaluate the number of out-of-sequence packets, the proposed procedure with the optimized timing of path switch is compared with the baseline route optimization procedure. In addition, to investigate the performance with respect to the communication disruption duration during the route optimization procedure, the sender-buffering method as described in Section III is implemented for comparison.

The performance of the proposed procedure is evaluated in an experimental testbed where actual PCs implement the proposed functions of network entities. Table I shows the hardware specifications of the network entities. The network topology of the experimental testbed is shown in Fig. 4. An MN is attached to an MAG via an IEEE 802.11g access point. LMAs and MAGs are connected to each other via gigabit Ethernet link. The expected one-way link delay between the network entities illustrated in Fig. 4 is added by a network emulator Dummynet [10]. UDP packets are transferred from MN1 to MN2 and vice versa by using an Iperf traffic generator [11]. The data rate in each direction is 500 Kbps and the packet size is 1250 bytes, that is, 50 packets per second.

Let G denote the delay gap of one-way delays between the optimized path and the non-optimized path in Fig. 4, that is, $G = d_2 + d_3 + d_4 - d_1$. This delay gap G affects the number of out-of-sequence packets because a large delay gap will cause out-of-sequence packets at path switching. Moreover, d_1, which is the link delay from MAG1 to MAG2, is a key parameter when focusing on the duration of the communication disruption. This is because the waiting time for path creation in the sender-buffering method depends on this link delay. Therefore, this paper evaluates the performance by varying the two key parameters, delay gap G and link delay d_1. Table II shows the parameter sets used in this paper.

 ISBN: 978-1-61208-133-5

TABLE I. HARDWARE SPECIFICATIONS OF NETWORK ENTITIES

	Network entities	
	MN	***LMA, MAG, and policy store***
Model	Panasonic CF-R9JWACDR	Dell PowerEdge R300
CPU	Intel Core 7 820UM 1.06 GHz	Intel Xeon L5410 2.3 GHz
OS	Fedor core 10	Cent OS 5.3
Network I/F	IEEE 802.11g	Gigabit NIC

TABLE II. ONE-WAY DELAY USED IN THIS PAPER

Notation	***Description***	***Values***
d_1	Link delay between MAG1 and MAG2	10~50 msec
d_2	Link delay between MAG1 and LMA1	10~50 msec
d_3	Link delay between LMA1 and LMA2	10~50 msec
d_4	Link delay between MAG2 and LMA2	10~50 msec
G	Delay gap ($= d_2 + d_3 + d_4 - d_1$)	20~100 msec

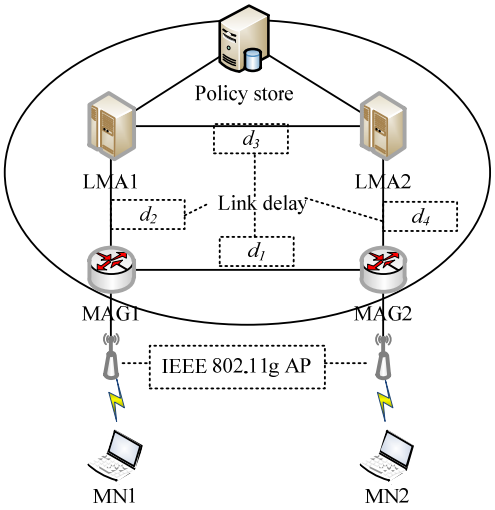


Figure 4. Network topology of experimental testbed.

B. Evaluation of Path Switching Methods

Fig. 5 plots the total number of out-of-sequence packets versus the delay gap between the non-optimized path and the optimized path, where the link delay between MAG1 and MAG2, d_1, is fixed at 10 msec. As the delay gap increases, the number of out-of-sequence packets also increases in the baseline method and the sender-buffering method. When comparing the baseline method with the sender-buffering method, the sender-buffering method decreases the number of out-of-sequence packets. This is because the sender-buffering method prevents out-of-sequence packets by buffering the data packets originating from MNs until the optimized path is ready. However, the sender-buffering method does not eliminate out-of-sequence packets when the delay gap is large. On the other hand, the proposed path switching method does not have out-of-sequence packets at any values of delay gap.

The total number of out-of-sequence packets is also shown in Fig. 6 and Fig. 7 where d_1 is 30 msec and 50 msec, respectively. In both results, the baseline method increases the number of out-of-sequence packets when the delay gap increases. The sender-buffering method prevents out-of-sequence packets, while several out-of-sequence packets occur by the large delay gap at d_1 = 50 msec (Fig. 6). When d_1 is 50 msec (Fig. 7), the sender-buffering method eliminates the out-of-sequence packets. From these results shown in Figs. 5 to 7, the sender-buffering method improves the performance when the link delay between MAGs is large. Similarly to the results in Fig. 5, the proposed method does not have any out-of-sequence packets. This means that the proposed method achieves route optimization while avoiding out-of-sequence packets by optimized path switch.

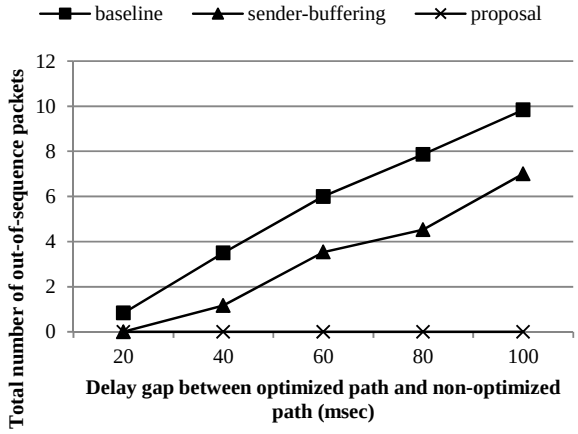


Figure 5. Number of out-of-sequence packets vs. delay gap between optimized path and non-optimized path (d_1 = 10 msec).

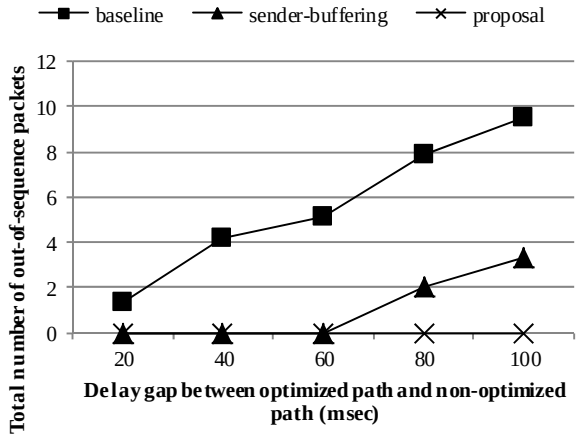


Figure 6. Number of out-of-sequence packets vs. delay gap between optimized path and non-optimized path (d_1 = 30 msec).

 ISBN: 978-1-61208-133-5

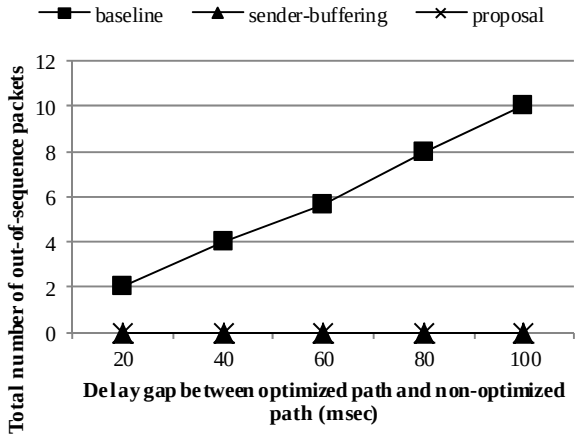


Figure 7. Number of out-of-sequence packets vs. delay gap between optimized path and non-optimized path (d_1 = 50 msec).

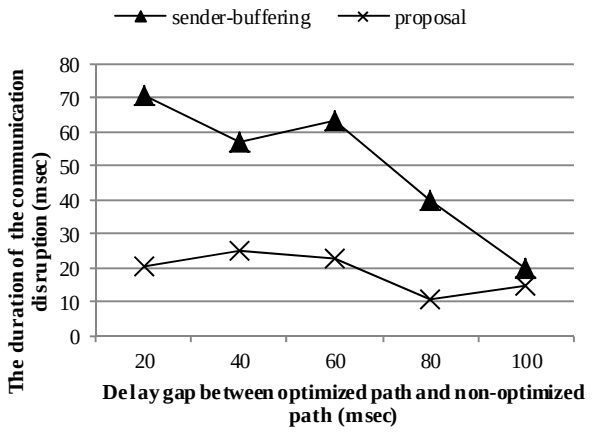


Figure 8. Average duration of communication disruption during route optimization (d_1 = 50 msec).

Fig. 8 shows the average duration of the communication disruption during route optimization procedures versus the delay gap. Here, we fix d_1 at 50 msec and compare the proposed method with the sender-buffering method. As described in the previous subsection, the packet rate in this experiment is 50 packets per second, which means that the interval of packet arrival is 20 msec. Therefore, if the result is larger than 20 msec, MNs experience unnecessary communication disruption. When d_1 is 50 msec, both methods avoid any out-of-sequence packets as shown in Fig. 7. However, from the results in Fig. 8, the sender-buffering method has longer communication disruption duration than the normal packet arrival interval. In particular, when the delay gap is small, the sender-buffering method involves long communication disruption. On the other hand, even when the delay gap changes, the proposed method maintains a small communication disruption, which is nearly equal to the normal packet arrival interval. This is because the proposed method accurately notifies of the end of data forwarding through the non-optimized path. Thus, we confirm that the proposed method realizes the route optimization by preventing out-of-sequence packets while minimizing communication disruption.

Out-of-sequence packets and communication disruption duration decrease communication performance of MNs, such as TCP throughput and service quality of real-time applications. The proposed method improves the communication performance of MNs by seamless path switching during the route optimization procedure.

V. CONCLUSION

This paper proposed a route optimization procedure in PMIPv6 with the optimized timing of path switch. The proposed procedure notifies the end of data forwarding through the non-optimized path accurately after the optimized path is established. The performance results showed that the proposed method prevented out-of-sequence packets and minimized the communication disruption time for the various values of delay parameters between network entities. With this feature, the proposed method contributes to performance improvement in TCP throughput or seamless continuity of real-time applications during the route optimization procedure.

REFERENCES

[1] C. Perkins, “IP Mobility Support for IPv4,” RFC 3344, IETF, Aug. 2002.

[2] D. Johnson, C. Perkins, and J. Arkko, “Mobility Support in IPv6,” RFC 3775, IETF, June 2004.

[3] S. Gundavelli, K. Leung, V. Devarapalli, K. Chowdhury, and B. Patil, “Proxy Mobile IPv6,” RFC 5213, IETF, Aug. 2008.

[4] S. Krishnan, R. Koodli, P. Loureiro, Q. Wu, and A. Dutta, “Localized Routing for Proxy Mobile IPv6,” Internet Draft, draft-ietf-netext-pmip-lr-01, IETF, Oct. 2010.

[5] J. Lee, H. Lim, and T. Chung, “Preventing Out-of-sequence Packets on the Route Optimization Procedure in Proxy Mobile IPv6,” Proc. IEEE International Conference on Advanced Information Networking and Applications (AINA 2008), Mar. 2008, pp. 950-954, doi:10.1109/AINA.2008.100.

[6] 3GPP, “General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access,” TS 23.401, V10.2.0, Dec. 2010.

[7] 3GPP, “General Packet Radio System (GPRS) Tunnelling Protocol User Plane (GTPv1-U),” TS 29.281, V10.0.0, Dec. 2010.

[8] F. Xia, B. Sarikaya, J. Korhonen, S. Gundavelli and D. Damic, “RADIUS Support for Proxy Mobile IPv6,” Internet Draft, draft-ietf-netext-radius-pmipv6-01, IETF, Nov. 2010.

[9] 3GPP, “Architecture Enhancements for non-3GPP accesses,” TS 23.402, V10.2.0, pp. 202-207, Dec. 2010.

[10] Dummynet, available at http://info.iet.unipi.it/~luigi/dummynet/ (last access date Mar 11, 2011)

[11] Iperf, available at http://iperf.sourceforge.net/ (last access date Mar 11, 2011)

 ISBN: 978-1-61208-133-5

A Novel Key Management Protocol in Body Area Networks

Jian Shen, Sangman Moh, Ilyong Chung
Dept. of Computer Engineering
Chosun University
Gwangju, Republic of Korea
E-mail: s_shenjian@126.com, {smmoh, iyc}@chosun.ac.kr

***Abstract*—Body Area Networks (BANs) have emerged as an enabling technique for e-healthcare systems, in which the data of a patient's vital body parameters and movements can be collected by small wearable or implantable sensors and communicated using short-range wireless communication techniques. Due to the shared wireless medium between the sensors in BANs, adversaries can launch various attacks on the e-healthcare systems. The security and privacy issues of BANs are getting more and more important. To provide secure and correct association of a group of sensors with a patient and satisfy the requirements of data confidentiality and integrity in BANs, we propose a novel key management protocol based on elliptic curve cryptography (ECC) and hash chains. The authentication procedure and group key generation are very simple and efficient. Therefore, our protocol can be easily implemented into the power and resource constrained sensor nodes in BANs. From the comparison results, furthermore, we can conclude that the proposed protocol dramatically reduces the computation and communication cost for the authentication and key derivation compared with the previous protocols.**

***Keywords*-Body Area Networks (BANs); security; privacy; sensor association; key management;**

I. INTRODUCTION

BANs can be used to continuously or remotely monitor patients' health, which have the potential to revolutionize the capture, processing, and communication of critical data for e-healthcare systems. As we know, the modern e-healthcare systems can provide new ways of hospitalization with quality health care. Wirelessly connected medical sensor nodes placed in, on, and around the body form a BAN for continuous, automated, and remote monitoring of physiological signs to support medical applications [1] [2]. In addition, a patient controller (PC) is needed to perform a multitude of functions in BANs. A PC (such as a PDA or smart phone) can sense and fuse data from sensors across the body, serve as a user interface, and bridge BANs to higher-level infrastructures.

The applications of BANs are primarily in the healthcare domain, especially for continuous monitoring and logging vital parameters of patients suffering from chronic diseases such as diabetes, asthma and heart attacks. Moreover, BAN technology is able to support other personalized applications, such as sports, gaming, entertainment, military and so on [3] [4] [5]. A wide range of applications make BANs have a promising future.

Compared with conventional sensor networks, the most distinguished difference of BANs is that a BAN needs to deal with more important medical information. Data confidentiality and integrity are the most important requirements in BANs, since wireless medium is susceptible to lots of security attacks. In this paper, we propose a novel key management protocol to provide secure sensor association and key derivation in BANs. First of all, in order to withstand security attacks from malicious insiders such as off-duty doctors or discharged patients, mutual authentication between a patient and a healthcare worker should be provided. Secondly, secure sensor association scheme should be considered to offer mutual authentication between medical sensors and a patient controller (PC) such that a healthcare worker can make sure that a group of medical sensor nodes are correctly and securely associated with an intended patient. It is worth noting that the sensor nodes also need to authenticate each other and establish a group key for subsequent communications. During the sensor association procedure, each medical sensor node can share a secret key with the PC, which is able to be used to encrypt and transmit the group key. Thirdly, a key management scheme is required to derive the group key. Our proposed protocol provides mutual authentication between a patient and a healthcare worker, mutual authentication between a sensor node and a PC, and mutual authentication between each pair of sensor nodes. We'd like to emphasize that a shared secret key between each sensor node and the PC is computed based on elliptic curve cryptography (ECC), while the authentication procedure is based upon hash chains. In addition, a group key of the sensor nodes is calculated only by the PC, since the PC is assumed to have no power and resource restriction.

The rest of this paper is organized as follows. In the next section, the related work is briefly discussed. A novel key management protocol in BANs is described in detail in Section III. Security analysis and performance analysis of our protocol are presented in Section IV and Section V, respectively. Finally, the conclusions of this paper are covered in Section VI.

 ISBN: 978-1-61208-133-5

II. Related Work

In BANs, secure sensor association is a non-trivial issue, because a healthcare worker must check whether a group of sensors are correctly and securely associated with an intended patient before any data communication happens. Lots of previous works focus only on group key agreement in sensor nodes [6], [7], [8], [9].

Recently, Keoh et al. [10] and Li et al. [11] propose some protocols considering both sensor association and key agreement. In [10], each sensor node can be securely associated with the controller using public key based authentication. However, it does not take sensor-to-patient authentication into account. It is easily for malicious nodes to join the BAN to achieve the important medical data. In [11], group device pairing (GDP) is implemented to perform authentication and establish group keys. However, the computation and communication cost of GDP is very high. In particular, in order to achieve the group key, each sensor node needs $n + 3$ times modular exponentiation operations, where n is the total number of sensor nodes in the BAN. It is a large burden for the power and resource constrained medical sensor nodes.

In our protocol, we use ECC and hash chains to perform authentication and key generation. First, a patient and a healthcare worker authenticate each other. Secondly, the authenticated healthcare worker associates the medical sensor nodes with the intended patient. Each node can establish a shared secret key with the patient controller (PC), then the LED blinking pattern can be transmitted using the shared secret key. The healthcare worker can confirm the secure sensor association when all the sensor nodes have the synchronized LED blinking pattern. Thirdly, a group key is computed by the PC, which then distributes the group key to all the sensor nodes by utilizing the shared secret keys. Note here that key distribution based on symmetric key cryptography is fast and efficient. We'd like to emphasize that ECC and hash chains are very efficient methods in cryptography. In particular, a point multiplication in ECC is more efficient than a modular exponentiation in RSA [15] [16]. In addition, hash operation is a kind of lightweight cryptographic primitive. The use of ECC and hash chains can satisfy the requirement of the resource-limited medical sensors in BANs.

III. A Novel Key Management Protocol in BANs

In this section, we elaborate the novel key management protocol in BANs. We start with describing the protocol model and threat model briefly, and then present the design of the proposed protocol in detail. By the way, we need some notations in our protocol, which are showed in Table I.

Table I
Frequently used notations

p	A prime number
Z_p	A finite field
E_p	An elliptic curve over Z_p
$\mathcal{G}$	The generator the group of points over E_p
q	The order of the group of points over E_p
s	The private key of KGC
P_{pub}	The public key of KGC
n	The number of medical sensor nodes in the BAN
$h()$	One-way hash function
$h^z()$	z cascade hash operations
ID_c	The identity of a PC
ID_d	The identity of a HWD
N_x	The identity of a medical sensor node with index x
k_c	The secret key of a PC
k_d	The secret key of a HWD
k_x	The secret key of a medical sensor node with index x
$\mathcal{K}_G$	The group key of medical sensor node.
r_c	The random number generated by a PC
r_d	The random number generated by a HWD
r_x	The random number generated by a medical sensor node x

A. Protocol and Threat Model

The protocol involves three entities: medical sensor, patient's controller (PC) and healthcare worker's device (HWD). In addition, we assume that the hospital is the key generation center (KGC) which is engaged as a trusted third party to issue important things to the patients and healthcare workers. First of all, the KGC chooses a prime number p and decides a elliptic curve E_p with order q over Z_p. Then, the KGC picks a random integer $s \in Z_p^*$ as its private key, and computes its public key $P_{pub} = s\mathcal{G}$, where $\mathcal{G}$ is the generator of the group of points over E_p. Note that the private key s of KGC should be changed periodically. At last, the KGC issues $\{p, E_p, q, P_{pub}\}$ to the patients and healthcare workers, but keeps s secret. It is worth noting that only the registered patients and healthcare workers can obtain these important materials. Our protocol works under the assumption that the medical sensor has a hash function, a random number generator and a re-writeable memory. Since a hash function is a powerful and computational efficient cryptographic tool, in the proposed protocol, we use the hash chains to perform the authentication in the BAN. In addition, a shared secret key between each sensor node and the PC is computed based on ECC, which is more efficient than RSA because the computation cost of a point multiplication is less than that of a modular exponentiation [15] [16]. Much of the details of ECC can be found in [12] [13] [14]. Furthermore, a group key of the group of medical sensor nodes associated with an intended patient needs to be calculated by the PC, which then distribute the group key to all the sensors.

In this paper, first, we'd like to emphasize that only the hospital is trusted, which is considered as the trusted third party KGC. In [10] [11], the authors assume that the patients and healthcare workers are all well-behaved.

 ISBN: 978-1-61208-133-5

However, in fact, some patients and healthcare workers may not be always trusted. For instance, an off-duty doctor or a discharged patient can perform some malicious attacks to obtain the secret keys by eavesdropping, to impersonate as a legitimate group member to join the group, or to modify the information communicated between legitimate group members so as to disrupt key authentication. In our protocol, we consider the situations mentioned above in order to deal with the malicious insiders. Second, we assume that the medical sensors can be attacked by passive attacks and active attacks. In BANs, the medical sensor nodes can be placed in, on, or around a patient's body, which are capable of sensing, storing, processing and transmitting data via wireless communications. As we know, wireless channels are susceptible to passive eavesdropping and message interception. Hence, adversaries can easily perform passive attacks. On the other hand, in active attacks, an adversary not only just records the data, but also can alter, inject, intercept and replay messages. Note that the sensor nodes do not trust each other before association and can be compromised after deployment. At last, the attackers are assumed to be able to eavesdrop on the wireless communication channel and intercept, modify, replay or inject the transmitting data.

B. Design of the protocol

The proposed protocol can be divided into three phases: initialization phase, secure sensor association phase, and key management phase. The detailed procedures are described as follows.

1) Initialization Phase: Suppose that there are n medical sensor nodes with identities $\{N_1, N_2, \cdots, N_n\}$ in a BAN. First of all, the registered PC with the identity ID_c and HWD with the identity ID_d obtain $\{p, E_p, q, P_{pub}\}$ from KGC. Then, the PC generates its own secret key k_c and a random number r_c. Similarly, the HWD derives k_d and r_d.

2) Secure Sensor Association Phase: In this phase, first, the registered PC and HWD authenticate each other so as to withstand the attacks from a malicious off-duty doctor or a malicious discharged patient. Then, a group of medical sensor nodes are securely and correctly associated to the authenticated patient.

(1) The PC and HWD need to authenticate each other before any data communication happens. Seen from Fig. 1, the PC and HWD both generate a random number $\{r_c, r_d\}$ and calculate $\{S_c = k_c P_{pub}, S_d = k_d P_{pub}\}$. After that, they compute hash values $\{A_c, A_d\}$ and exchange the messages $\{S_c, r_c, ID_c\}$ and $\{S_d, r_d, ID_d\}$. In order to authenticate each other, the PC and HWD need to check whether the equations $A_c = h(S_c \parallel r_c \parallel ID_c)$ and $A_d = h(S_d \parallel r_d \parallel ID_d)$ can hold.

(2) After the PC and HWD authenticate each other,

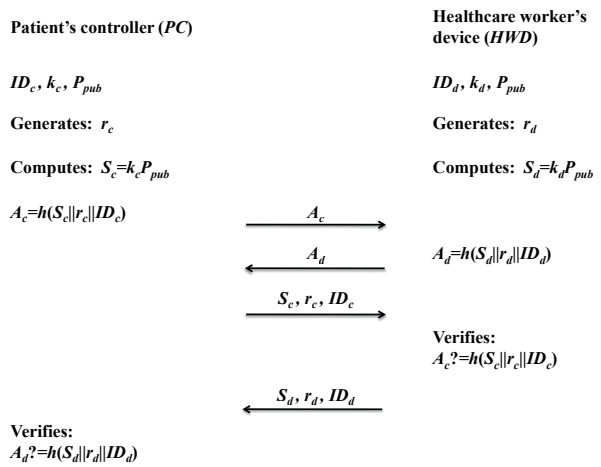


Figure 1. Mutual authentication between PC and HWD.

a group of sensor node must be securely and correctly associated with the patient. The authenticated PC first generates n secret keys $\{k_1, k_2, \cdots, k_n\}$ and n random numbers $\{r_1, r_2, \cdots, r_n\}$, and then preloads each secret key k_x and each random number r_x to node N_x, for $x = 1, 2, \cdots, n$. Next, the PC computes its own hash chain $h^z(k_c \parallel r_c)$ as well as the hash chain $h^z(k_x \parallel r_x)$ of node N_x, for $x = 1, 2, \cdots, n$. After that, the PC broadcasts all the information $h^z(k_x \parallel r_x)$ to the group of nodes. Note that z is a large constant number and $h^z(m)$ denotes the application of z cascade hash operations starting from m. For instance, $h^2(m) = h(h(m))$, $h^3(m) = h^2(h(m)) = h\left(h^2(m)\right) = h(h(h(m)))$, etc.. At last, the PC publishes $\{p, E_p, q, P_{pub}\}$. In our protocol, we assume that the broadcasting hash chain for node N_x will be updated after each successful authentication. The hash chain $h^z(k_x \parallel r_x)$ of node N_x will be replaced with $h^{z-l}(k_x \parallel r_x)$ when the node N_x have passed through authentication l times. Now, it is supposed that node N_i and PC have passed through authentication u times and v times, respectively. Then the broadcasting hash chains for node N_i and PC are $h^{z-u}(k_i \parallel r_i)$ and $h^{z-v}(k_c \parallel r_c)$, respectively. The processes of authentication and key establishment between node N_i and PC are divided into five steps, which are shown in Fig. 2.

(2-1) The node N_i generates a random number t_i as its secret key and computes the point $A_i = t_i P_{pub} = (x_i, y_i)$ over the elliptic curve E_p and $S_i = h\left(x_i \parallel h^{z-u-1}(k_i \parallel r_i)\right)$, then it sends a message $\{N_i, A_i, S_i\}$ to the PC. Similarly, the PC generates a random number t_c and computes $A_c = t_c P_{pub} = (x_c, y_c)$ and $S_c = h\left(x_c \parallel h^{z-v-1}(k_c \parallel r_c)\right)$, then it sends a message $\{N_c, A_c, S_c\}$ to the node N_i. Note here that x_i and x_c are the x-component of points A_i and A_c, respectively. For security, the secret t_i and t_c cannot be

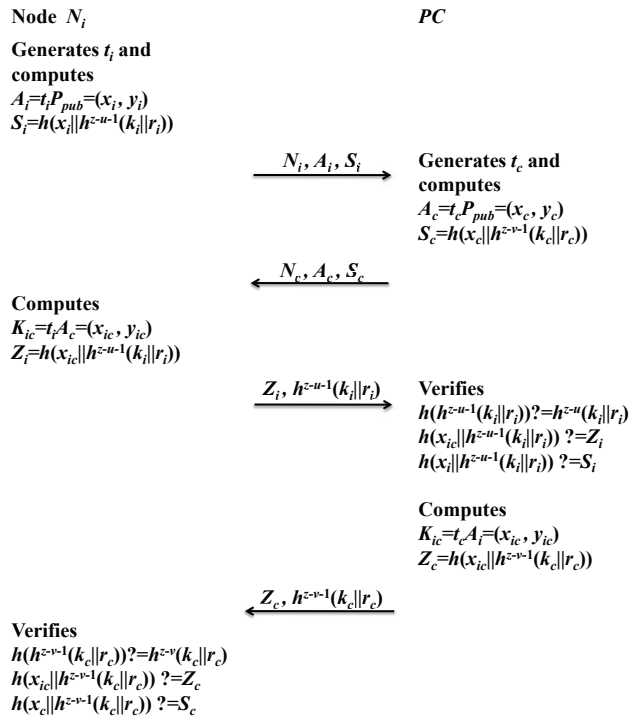


Figure 2. Mutual authentication and key establishment between node N_i and PC.

reused.

(2-2) After receiving the message $\{N_c, A_c, S_c\}$, the node N_i computes a shared secret key $K_{ic} = t_iA_c = t_it_cP_{pub} = (x_{ic}, y_{ic})$ and $Z_i = h\left(x_{ic} \parallel h^{z-u-1}\left(k_i \parallel r_i\right)\right)$. Then, it delivers a message $\left\{Z_i, h^{z-u-1}\left(k_i \parallel r_i\right)\right\}$ to the PC.

(2-3) Upon receiving the message $\left\{Z_i, h^{z-u-1}\left(k_i \parallel r_i\right)\right\}$ and the previous one $\{N_i, A_i, S_i\}$ from (2-1), the PC checks whether the conditions $h\left(h^{z-u-1}\left(k_i \parallel r_i\right)\right) = h^{z-u}\left(k_i \parallel r_i\right)$, $h\left(x_{ic} \parallel h^{z-u-1}\left(k_i \parallel r_i\right)\right) = Z_i$ and $h\left(x_i \parallel h^{z-u-1}\left(k_i \parallel r_i\right)\right) = S_i$ are satisfied. If they are satisfied, the PC can make sure that the node N_i is a authorized one and subsequently computes a shared secret key $K_{ic} = t_cA_i = t_ct_iP_{pub} = (x_{ic}, y_{ic})$. Next, the PC computes $Z_c = h\left(x_{ic} \parallel h^{z-v-1}\left(k_c \parallel r_c\right)\right)$ and then sends a message $\left\{Z_c, h^{z-v-1}\left(k_c \parallel r_c\right)\right\}$ to the node N_i. If the conditions mentioned above are not satisfied, the authentication fails and the PC beeps. Note that the shared secret key between node N_i and PC are definitely same due to $t_iA_c = t_cA_i = t_it_cP_{pub}$.

(2-4) After receiving $\left\{Z_c, h^{z-v-1}\left(k_c \parallel r_c\right)\right\}$ from the PC, the node N_i checks whether the conditions $h\left(h^{z-v-1}\left(k_c \parallel r_c\right)\right) = h^{z-v}\left(k_c \parallel r_c\right)$, $h\left(x_{ic} \parallel h^{z-v-1}\left(k_c \parallel r_c\right)\right) = Z_c$ and $h\left(x_c \parallel h^{z-v-1}\left(k_c \parallel r_c\right)\right) = S_c$ are satisfied. If they are satisfied, the node N_i can verify the authenticity of the PC. Otherwise, the authentication fails and the node N_i beeps.

(2-5) Finally, node N_i and PC update their broadcasting hash chains to be $h^{z-u-1}\left(k_i \parallel r_i\right)$ and $h^{z-v-1}\left(k_c \parallel r_c\right)$, respectively.

Now, the PC and the node N_i authenticate each other and establish a shared secret key K_{ic}, which can be used to encrypt and transmit the LED blinking pattern. The healthcare worker indicates "authentication accepted" to the controller if the LED blinking patterns of the sensor nodes are same. Furthermore, our scheme can also provide mutual authentication between nodes. The authentication for each pair of nodes is same as the authentication for the node N_i and the PC.

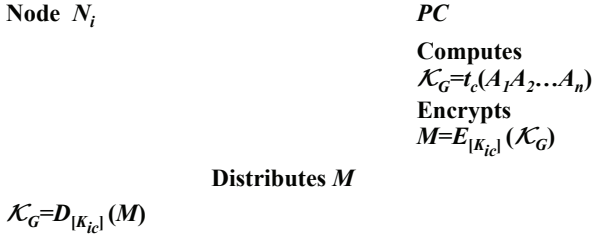


Figure 3. Group key derivation.

3) Key Management Phase: In this phase, a group key of participating medical sensor nodes is derived. The PC in the BAN is responsible for the key distribution and management since it typically has higher computation capability and storage capability. Observed from the above secure sensor association phase, the PC can receive n points $\{A_1, A_2, \cdots, A_n\}$ from nodes $\{N_1, N_2, \cdots, N_n\}$. Then, it calculates the group key $\mathcal{K}_G = t_c \cdot (A_1A_2 \cdots A_n)$, which can be subsequently distributed to each sensor node N_i by using the corresponding shared secret key K_{ic}. Observed from Fig. 3, the group key distribution is based on symmetric key cryptography. The PC can distribute the group key to all the nodes in the BAN. For instance, if the PC wants to distribute $\mathcal{K}_G$ to the node N_i, it just needs to encrypt $\mathcal{K}_G$ using the shared secret key K_{ic}. After receiving M, the node N_i can easily derive the group key by decrypting M using the same shared secret key K_{ic}.

IV. Security Analysis

In this section, we show security of the presented protocol in withstanding the attacks of passive and active adversaries.

A. Security Against Passive Adversary

A passive adversary (attacker) tries to learn information about the secret key by eavesdropping on the broadcast channel. In our protocol, an eavesdropper cannot get any information about the secret value t_i due to discrete logarithm problem in elliptic curves. Therefore, the secret value

t_i of each node N_i can be protected and the attacker is not able to learn the information of the group key.

B. Security Against Active Adversary

In active attack, an adversary not only just records the data, but also can alter, inject, intercept and replay messages. The goal of the authentication mechanism is to convince a controller that the nodes he is communicating with are indeed the nodes they claim to be. We show the analysis of the concrete security properties withstanding the active attacks that we concerned in our proposed protocol as follows:

1) Malicious Insiders Resistance: Malicious insiders mean that the misbehaved off-duty doctors or the misbehaved discharged patients. In our protocol, the KGC can issue p, E_p, q, P_{pub} to the registered doctors or patients. Note that $P_{pub} = s\mathcal{G}$, where the private key s of KGC will be updated periodically. For example, if a doctor is off-duty or a patient is discharged from the hospital, then the private key s of the KGC must be changed. Then all the subsequent authentications will fail, since the P_{pub} is changed. Hence, the malicious insiders can not perform attacks without being noticed.

2) Implicit Key Authentication: Implicit key authentication is a fundamental security property, which implies that only the users with whom *A* wants to agree upon a common key may be able to compute a key. In our protocol, the sensor nodes agreeing upon a group key are controlled by the PC. It is clear that our protocol provides implicit key authentication.

3) Known Session Key Security: Known session key security indicates that an adversary having obtained some previous session keys still cannot deduce the session keys of the current run of the protocol. In our protocol, each node N_i selects a random number t_i as secret for each session and calculates $A_i = t_i P_{pub}$. It is impossible for the adversary to derive certain secret key t_i so as to obtain the current shared secret key K_{ic} or the current group key $\mathcal{K}_G$.

4) Key-Compromise Impersonation Resistance: Key-compromise impersonation security ensures that the compromise of one user's long-term private key cannot expose the other user's long-term private key. In our protocol, each user's long-term private key k_i is individually generated by the PC. Therefore, the adversary having obtained a certain user's long-term private key cannot expose the long-term private key of other user's.

V. Performance Analysis

In this section, we will compare the performance of the proposed protocol with the protocol presented by Li et al [11]. Our protocol uses ECC and hash chains to perform authentication and key generation because ECC and hash chains are very efficient methods in cryptography. In our protocol, the most expensive operation is the point multiplication, while in Li et al.'s protocol the most expensive operation is modular exponentiation. It has been shown in [15] [16] that a point multiplication needs less computation time than a modular exponentiation unless the exponent is chosen as some specific value.

Table II
Performance comparison of Li's protocol and our protocol

	Li et al.'s protocol	Our protocol
Computations for each node to achieve authentication and compute a shared secret key	$(3+n)\,e + 9h^{(1)}$	$2p + 5h^{(2)}$
Total number of transmissions for the protocol	20	10

(1) n is the number of medical sensor nodes in the BAN, e is the modular exponentiation operation, and h is the hash operation.
(2) p is the point multiplication operation over elliptic curve.

We summarize the performance comparison of the proposed protocol with Li et al.'s in Table II. As shown in Table II, in Li et al.'s protocol [11], each node needs to perform $3+n$ times modular exponentiation operations and 9 times hash operations. Moreover, it is required to send 20 transmissions in their protocol to run the authentication and key generation. However, in the proposed protocol, each sensor node needs to perform only two point multiplications over an elliptic curve ($A_i = t_i P$ and $K_{ic} = t_i A_c$) and five hash operations. In addition, our protocol only needs 10 transmissions. Therefore, the proposed protocol is more efficient than Li et al.'s protocol. It significantly reduces the overhead of communication for sensor node to achieve secure connectivity. We'd like to emphasize that we do not compare the performance of our protocol with Keoh et al.'s protocol since Keoh et al.'s protocol uses public key cryptography to execute the authentication which is different from symmetric key cryptography used in our protocol.

VI. Conclusion

The BAN system provides a flexible, wearable infrastructure for acquisition, processing and wireless transmission of medical data and information at the human body. Using BAN, multiple vital signs can be wirelessly monitored even in mobile environments.

In this paper, we propose a novel enhanced secure sensor association and key management protocol based on elliptic curve cryptography (ECC) and hash chains in order to provide secure and correct association of a group of sensors with a patient and satisfy the requirements of data confidentiality and integrity in BANs. The authentication procedure and group key generation are very simple and efficient. Therefore, our protocol can be easily implemented into the power and resource constrained sensor nodes in BANs. Compared with the previous works in BANs, our protocol needs less computation and communication cost for the authentication and key derivation. Meanwhile, our

 ISBN: 978-1-61208-133-5

protocol can provide mutual authentication between PC and HWD, mutual authentication between PC and nodes, and mutual authentication between nodes. We believe that our protocol is attractive to the applications of BANs.

REFERENCES

[1] M. Patel and J. Wang, "Applications, Challenges, and Prospective in Emerging Body Area Networking Technologies," *IEEE Wireless Communications*, pp. 80-88, Feb. 2010.

[2] K. Lorincz, D. Malan, T. F. Jones, A. Clavel, V. Shnayder, G. Mainland, M. Welsh, and S. Moulton, "Sensor Networks for Emergency Response: Challenges and Opportunities," *IEEE Pervasive Computing*, vol.3, no. 4, pp. 16-23, Oct.-Dec. 2004.

[3] M. Hanson, H. Powell, A. Barth, K. Ringgenberg, B. Calhoun, J. Aylor, J. Lach, "Body Area Sensor Networks: Challenges and Opportunities," *Computer*, vol.42, no. 1, pp. 58-65, Jan. 2009.

[4] M. Li, W. Lou, and K. Ren, "Data Security and Privacy in Wireless Body Area Networks," *IEEE Wireless Communications*, vol. 17, no. 1, pp. 51-58, Feb. 2010.

[5] E. Jovanov, A. Milenkovic, C. Otto, and P. C. Groen, "A Wireless Body Area Network of Intelligent Motion Sensors for Computer Assisted Physical Rehabilitation," *Journal of NeuroEngineering and Rehabilitation*, vol. 2, no. 1, pp. 1-10, Mar. 2005.

[6] C. C. Tan, H. Wang, S. Zhong, and Q. Li, "IBE-Lite: A Lightweight Identity-Based Cryptography for Body Sensor Networks," *IEEE Trans. on Information Technology in Biomedicine*, vol. 13, no. 6, pp. 926-932, Nov. 2009.

[7] O. G. Morchon, H. Baldus, and D. S. Sanchez, "Resource-Efficient Security for Medical Body Sensor Networks," in *BSN'06*, pp. 80-83, Apr. 2006.

[8] T. Donovan, J. Donoghue, C. Sreenan, D. Sammon, P. Reilly, and K. A. Connor, "A Context Aware Wireless Body Area Network (BAN)," in *proc. of the Pervasive Health Conference*, pp. 1-8, Apr. 2009.

[9] K. Malasri and L. Wang, "Addressing Security in Medical Sensor Networks," in *HealthNet'07*, pp. 7-12, 2007.

[10] S. L. Keoh, E. Lupu, and M. Sloman, "Securing Body Sensor Networks: Sensor Association and Key Management," *proc. of the 7th Annual IEEE Int. Conference on Pervasive Computing and Communications (PerCom)*, Galveston, Texas, Mar. 9-13, pp. 1-6, 2009.

[11] M. Li, S. Yu, W. Lou, and K. Ren, "Group Device Pairing based Secure Sensor Association and Key Management for Body Area Networks," *proc. of IEEE INFOCOM*, San Diego, CA, Mar. 14-19, pp. 1-9, 2010.

[12] N. Koblitz, "Elliptic curve cryptosystems," *Math. Comput.*, vol. 48, pp. 203-209, 1987.

[13] H. Silverman, *The Arithmetic of Elliptic Curves*, 2nd ed. Springer, 2000.

[14] W. Stallings, *Cryptography and Network Security*, 4th ed. Prentice Hall, 2005, pp. 301-313.

[15] R. Watro, D. Kong, S. Cuti, C. Gardiner, C. Lynn, and P. Kruus, "TinyPK: Securing Sensor Networks with Public Key Technology," in *Proc. of the 2nd ACM Workshop on Security of Ad Hoc and Sensor Networks (SASN'04)*, Washington, DC, USA, pp. 59-64, Oct. 2004.

[16] D. J. Malan, M. Welsh, and M. D. Smith, "A Public-Key Infrastructure for Key Distribution in TinyOS Based on Elliptic Curve Cryptography," in *1st IEEE International Conference on Sensor and Ad Hoc Communications and Networks (SECON'04)*, Santa Clara, California, pp. 71-80, Oct. 2004.

 ISBN: 978-1-61208-133-5

Towards Knowledge-driven QoE Optimization in Home Gateways

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 the concept of using distributed knowledge components as basis for a Quality of Experience optimization process. We also present simulation results indicating the potential in using this approach for access and home networks. The main novelty of the paper is the presentation of how specific end user preference information can be combined with specific content provider service information, and used as input to an optimization process in a home gateway device. The results show that the effect of doing this is significant.**

Keywords: QoE, Home Gateway, Adaptive Services

I. INTRODUCTION

The focus on QoE (Quality of Experience) rather than just QoS (Quality of Service) has been growing in strength over the last years. The main reason for this relates to the acknowledgement of that users are not equal. The QoE approach covers not only technical metrics, but also metrics describing the uniqueness of a specific user (cf. Table 1). As a result, it represents a measure of the overall customer satisfaction with a service or vendor. This makes it more suitable for user oriented service delivery architectures [3].

Table 1. EXAMPLE QoS AND QoE METRICS

QoS metrics	QoE metrics
Bandwidth	Perception
Delay	Preferences
Packet Loss	Expectations
Jitter	Acceptance
Availability	Price
…	…

The traditional approach of assigning a fixed priority per service or service class, and then implement a QoS design may not be rich enough to support more advanced QoE optimization schemes. Even with the full range of DiffServ values/classes [12], this will be a limiting resource. Further on, the actual QoS implementation with a high number of classes would have significant complexity issues. As an alternative to this, the concept of knowledge based QoE optimization is proposed.

For content providers operating in the Over-The-Top domain it is natural to focus more on the QoE dimension rather than the QoS subset, as the latter would be partly outside of their control. In line with this statement it is easy to understand that this type of content providers would appreciate techniques enabling them to adapt their service delivery according to different users and varying network conditions. Further on, the location of effective optimization processes outside of the network operator domain is beneficial, as this would not require involvement from the network operator.

The structure of this paper is as follows. Section 2 provides an overview of state of the art in the relevant field and also defines the objectives of the research reported in this paper; Section 3 describes the role and components of the Knowledge Plane; Section 4 describes the simulation model; Section 5 presents simulation results; Section 6 presents an analysis of the results; Section 7 provides the conclusions and an outline of future work.

II. STATE OF THE ART

The framework used for QoE optimization in a home network environment is in line with related work as stated in [9][10][11] and illustrated in Figure 1.

The addition of a Knowledge Plane in network architecture as an addition to the well known control and management plane was originally proposed by the authors of [5]. The purpose of this Knowledge Plane was to give a unified view of network aspects, to analyze it – to explain it – and finally also to make suggestions on what to do in order to achieve specified objectives.

The use of a Knowledge Plane in the networking context, and the ideas from autonomic computing [8] was taken further by the MUSE Project "Advanced features for MM enabled access platform" [6]. Their work lead to a proposal of having Monitor Plane (MP) and Action Plane components distributed across a network, including the end systems.

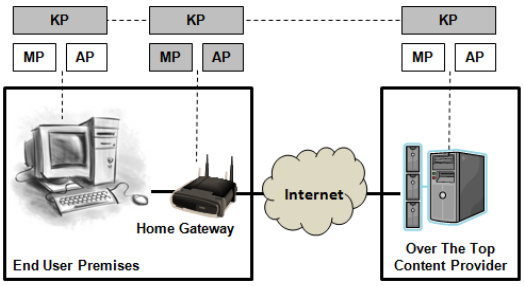


Figure 1. Optimization framework

The main difference between the optimization model used in this paper and earlier work by others is the inclusion of KP/MP/AP components from end user and content provider domains (cf. Figure 1). The KP components from

 ISBN: 978-1-61208-133-5

these domains are used as input to an optimization process in the home gateway which then is studied in this paper.

The content of the Monitor Plane and Action Plane is not the main focus of this paper, as we just assume their presence in the home gateway. More information on this can be found in previous work [9][15][17][18][19]. The type of Action Plane components applied would to a large extent depend on whether the traffic flows subject to control are of a responsive (TCP) or non-responsive (UDP) type. Related work in this area can be found in [1][7][20]. It is also important to note that the location of Action Plane components in the home gateway and not at network edge impose some challenges. Reason being that the congestion point for downstream traffic is at network edge.

The objective of the research documented in this paper is to support the statement that QoE optimization mechanisms for Internet services can be implemented in the home network domain, with the use of appropriate knowledge sources. The chosen method for providing this support is by means of simulation of a defined service usage scenario, with variable input parameters.

III. KNOWLEDGE PLANE

The Knowledge Plane is represented by information objects distributed across the platform components involved.

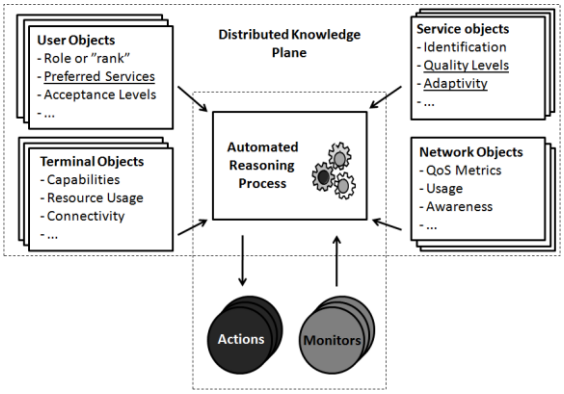


Figure 2. Knowledge Plane input to reasoning process

The use of Knowledge components in an optimization process requires a reasoning process (cf. Figure 2). This reasoning process combines and interprets the different components, allowing them to be used for some actions, and also effects to be monitored and understood.

The information objects used in the work reported in this paper are the user and service objects, and selected parameters from these (user: preferred service, service: quality level and adaptivity)

A. User Objects

The list of user preferences and associated capabilities which, could be used as input to an optimization scheme is potentially long, and depends to a large extent of the type of users being discussed (residential vs. business). What is considered as important by one user may not be of interest at all to another user, and vice versa. The thresholds for what is considered as good or bad quality are also different between users. This dynamic picture of user preferences and profiles are considered important to analyze and structure, in order to use part of it as input to optimization mechanisms.

In addition to the specific preferences of a user, there are also other differences in terms of factors contributing to the per-user QoE. Users are, e.g., different in terms of expectations concerning real quality. This may be directly related to the preferred user terminal capabilities or just basic differences in human perception. User preferences are also influenced by cost factors and assumed user rank in the specific home environment.

B. Service Objects

Many Internet-based services have certain requirements in terms of what is needed in order to be used. These requirements have traditionally been described by QoS parameters (delay, packet loss, jitter and bandwidth). This set of service information is still valid, but should be extended with additional parameters. This is especially important in light of the rapid evolution in content delivery techniques and associated technologies. The concept of adaptive streaming is an example of this. In this scenario the quality levels of a service is able to adjust itself according to end-to-end performance before and during service usage. This makes the bandwidth requirement for a certain service no longer fixed, but rather a variable parameter with some min/max thresholds and granularity. Further on, the concept of multi-source streaming from distributed and shared service platforms is also growing in popularity making it more challenging to recognize and classify services. The distribution of sources also makes the services become less sensitive for high delay, packet loss and jitter as it can pick the best performing streams and compose the service based on this.

C. Reasoning Process

In order to see the effect on using end user and content provider knowledge in the optimization process, three different schemes have been studied. These schemes are to some extent in line with the concept of a DiffServ bandwidth broker [13], but instead of priorities and policies as basis for bandwidth sharing we are using other knowledge components.

The first scheme is the basic FCFS (First Come First Served), where all knowledge use has been disabled and the home gateway operates in a regular best effort mode. The second mode is named STOPINC, where the Action Plane prevents background traffic source from increasing (if attempted) during a period where an end user preferred service is running below its maximum level. The third mode is called STEPDOWN, where the Action Plane in addition to what the STOPINC mode does - also makes a background traffic source decrease its rate according to the end user preferred service granularity. In the latter mode, the purpose is then to make it easier for the preferred service to increase its rate – one step closer to its maximum. For both the

 ISBN: 978-1-61208-133-5

STOPINC and STEPDOWN modes, when a background traffic source is either prevented from increasing or even made decrease its rate – it will be subject to this control for a certain period. This period should be enough so that the adaptive source notices that there is a chance of increasing rate.

IV. SIMULATION MODEL

The user scenario modeled in the simulator is a residential user group present in a typical home environment. The user group is connected to the Internet through a typical broadband connection. The broadband connection represents the resource shared between users and associated service.

A group of 4 users are considered, each of which operate independently of each other. Each user can start a single service at a time. There are no feedback mechanisms implemented in terms of users changing behavior as a result of good or poor performance.

Table 2. Simulation parameters

Parameter	Adaptive Service	Bkgd. Service
Max sessions	1	3
Bitrate (Kbps)	300-900	100-2800
Granularity (Kbps)	300	1
Time to first start (s)	Uniform(3,10)	
Session lifetime (s)	Uniform(10,30)	
Time to next start (s)	Uniform(3,10)	
Conn. capacity (Kbps)	1000-7000	
Control Period (s)	3	
Simulation time per run	7 days	
Number of seeds / run	100	

As can be seen from the simulation parameters, the session lifetimes are very short – much shorter than what could be expected in real life. The purpose of this was to make the simulation scenarios as dynamic as possible.

In order to see the effect of the studied QoE optimization process during different levels of congestion, the access capacity was varied while the service characteristics are kept the same.

The simulator was built using the process oriented Simula [14] programming language and the Discrete Event Modeling On Simula (DEMOS) context class [4].

A. Adaptive Service

The adaptive service (cf. Figure 3) operates between the max/min thresholds of respectively 900Kbps and 300Kbps and the granularity of increase/decrease could be set to values between 50Kbps and 300Kbps in the simulator - corresponding to fine versus course rate granularity. The granularity used in the presented results is 300Kbps and the interval of potential rate change was set to 2 sec. The reason for choosing both these values was that these are common parameter values used by live services [2][16].

The adaptive service will always try to increase its rate if possible, and will remain at max level when reached until it finishes unless if influenced by background traffic bursts. The influence from traffic bursts has been included in the model as it would be difficult to prevent, due to the location of the optimization process in the home gateway after the downstream congestion point.

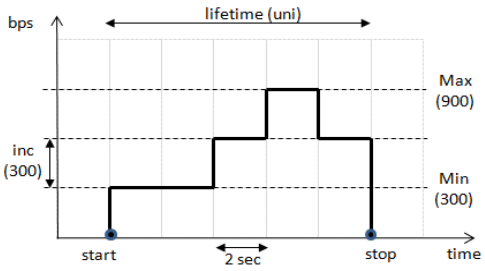


Figure 3. Adaptive service profile

The lifetime of the adaptive service session is taken from a random uniform distribution. A single adaptive service is run at a time, with repeated starts/stops during the simulation period.

B. Background Service

The background services (cf. Figure 4) used in the simulation operates in a rather simple mode, but potentially close to a worst case scenario. The sources are very bursty and pick a new target rate for each interval between a lower (100kbps) and upper threshold (2800Kbps) according to a uniform distribution. The intervals between each rate change is according to a negexp distribution (λ=1).

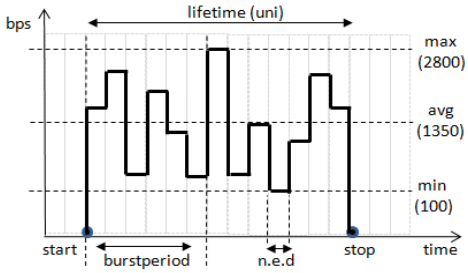


Figure 4. Background service profile

Whenever a background service starts up, it enters a burst period. During this burst period, the background services are allowed to influence the user preferred services, and in the case of congestion – they will make the adaptive service decrease its rate. The reasoning behind this is that the optimization process simulated is placed in the customer home gateway, and therefore after the access congestion point for traffic to the customer.

The duration of the burst period is decided by how fast new background traffic can be put under control by action plane components in the home gateway. Depending on the traffic type (TCP, UDP) and associated application this period will have different values. In the simulation results presented in this paper, the burst period has been varied between zero and 0.6 sec. The value of zero would represent no burst impact (ideal situation).

 ISBN: 978-1-61208-133-5

The lifetime of the background service session is taken from a random uniform distribution. Maximum three background services are run at a time, with repeated starts/stops during the simulation period.

V. SIMULATION RESULTS

The parameter studied in the simulations is the average achieved bitrate for the adaptive service as a function of access capacity. Traffic load is kept constant.

A. FCFS, STOPINC and STEPDOWN results

In Figure 5 results are presented where the burstperiod is set to 0.2 seconds, the adaptive service has increments of 300Kbps and the background services have n.e.d rate increment intervals with λ=1. The three different models FCFS, STOPINC and STEPDOWN are then compared.

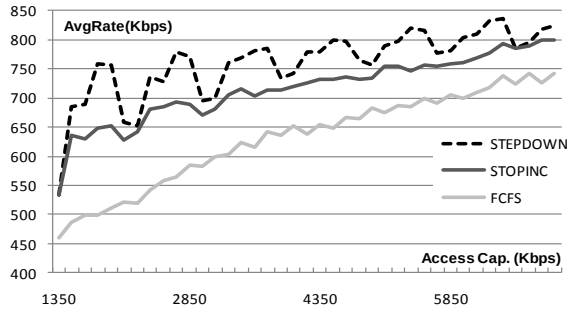


Figure 5. Comparison of optimization models

The 95% confidence intervals for the STEPDOWN model are shown in Figure 6, in order to give see how similar the results from the different simulation runs are.

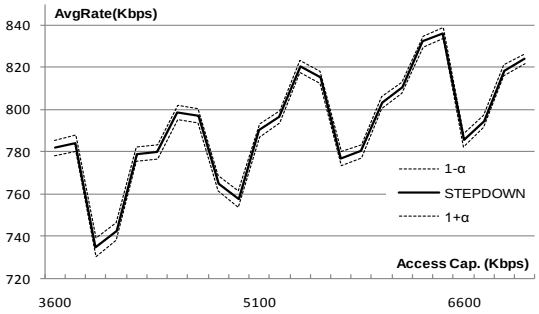


Figure 6. Confidence intervals for STEPDOWN

The confidence intervals are all in the region of +/- 2 to 7 across the studied access capacity range, which is very close to the plotted averages.

B. Effect of changing burstperiod

In Figure 7 the effect of changing the burstperiod for the background service is shown for the STEPDOWN optimization model.

The purpose of changing this parameter was to see if it had a major impact on the simulation results, and also to provide an indication on how fast the relevant Action Plane components would have to be in order to support the proposed QoE optimization process.

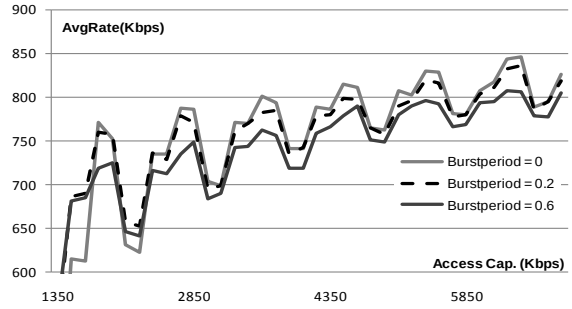


Figure 7. Change of burstperiod for STEPDOWN

The burstperiod values used are 0, 0.2 and 0.6 – whereas the value of 0 corresponds to an ideal scenario where the background services never influences the preferred adaptive service. The higher burstperiod values corresponds to scenarios where the Action Plane require some time interval in order to achieve control on the background services.

VI. ANALYZING THE RESULTS

The results presented in the previous section are considered promising, as they support the statement subject to investigation. The comparison between the FCFS, STOPINC and STEPDOWN modes of operation (cf. Figure 5) shows that for a home gateway the potential improvement in average bitrate for a preferred service is significant, if knowledge about the service granularity is made available. The simulation results show that during high congestion both the STOPINC and STEPDOWN models perform significantly better that FCFS. For the STOPINC mode there is a potential for between 10-30% higher average rate, and for the STEPDOWN mode the same potential is between 10-40%. The STEPDOWN mode performs significantly better that STOPINC for all access capacity levels (cf. Figure 8).

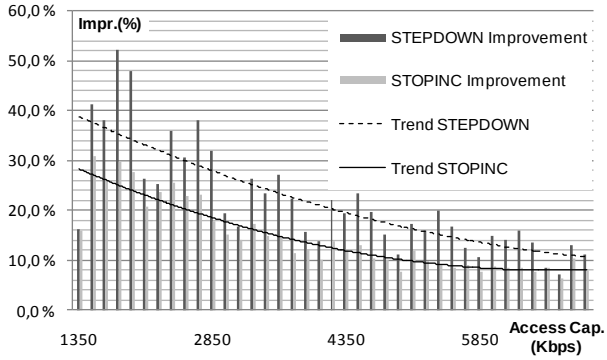


Figure 8. Rate improvements in percentage per model

The results when changing the burst period (cf. Figure 7) illustrate the importance of having an efficient Action Plane supporting the reasoning process. If services are not put

under control as fast as possible, it reduces the potential of STEPDOWN in the order of tens of Kbps.

It should be noted that there is no general 1:1 mapping between an achieved value of a QoS metric such as bitrate and a specific QoE metrics. However, it is a fair assumption that there is a weighted mapping between QoS metrics and related QoE metrics, following the preference and perception levels of a certain user. In line with this, we can say that the achieved increase in bitrate for the preferred adaptive service contributes to an increased QoE level.

VII. Conclusion and future work

Based on the analysis and simulation results presented, the statement of a potential gain in implementing QoE optimization mechanisms in the access and home network domain is strengthened. It is clear that even with just very basic knowledge components available from the user and service objects (cf. Figure 2) a significant improvement in QoE can be achieved.

The presented results may also have value for pure network oriented QoS mechanisms, if this type of stepwise service adaption becomes a success in emerging service delivery architectures. As an example, it is likely that the bandwidth broker concept of DiffServ could benefit from introducing this type of service knowledge in its operation.

As future work in this area, the plan is to investigate more complex user and service scenarios. It is also the intention to make the service models used in the simulator closer to real life traffic. Further on, the logics in the reasoning process together with efficient action plane components will be addressed.

VIII. Acknowledgements

The reported work is done as part of the PhD studies for the first author, which is an integrated part of the Road to media-aware user-Dependant self-aDaptive NETWORKS - R2D2 project. This project is funded by The Research Council of Norway.

References

[1] A. Aggarwal, S. Savage, and T. Anderson. “Understanding the performance of tcp pacing”. In Proc. IEEE Nineteenth Annual Joint Conf. of the IEEE Computer and Communications Societies INFOCOM 2000, volume 3, pages 1157–1165, 2000.

[2] 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.

[3] 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, pages 1–6, 2009.

[4] G. Birtwistle. Demos - A system for Discrete Event Modelling on Simula. School of Computer Science, University of Sheffeld,, July 1997.

[5] D. D. Clark, C. Partridge, J. C. Ramming, and J. T. Wroclawski. “A knowledge plane for the internet”. SIGCOMM’03, August 2003.

[6] H. Dequeker, T. V. Caenegem, K. Struyve, E. Gilon, B. D. Vleeschauwer, P. Simoens, and W. V. de Meerssche. Advanced features for mm enabled access platform (muse project deliverable). Technical report, Alcatel Bell and IBBT - IBCN, December 2006.

[7] H. Jamjoom and K. G. Shin. “Persistent dropping: an efficient control of traffic aggregates”. In Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications, SIGCOMM ’03, pages 287–298, New York, NY, USA, 2003. ACM.

[8] J. O. Kephart and D. M. Chess. “The vision of autonomic computing”. Computer, 36(1):41–50, 2003.

[9] S. Latr´e, P. Simoens, B. De Vleeschauwer, and V. de Meerssche. “An autonomic architecture for optimizing qoe in multimedia access networks”. Computer Networks 53, 10:1587–1602, 2009.

[10] S. Latre, P. Simoens, B. De Vleeschauwer, W. Van de Meerssche, F. De Turck, B. Dhoedt, P. Demeester, S. Van Den Berghe, and E. G. de Lumley. ”Design for a generic knowledge base for autonomic qoe optimization in multimedia access networks”. In Proc. IEEE Network Operations and Management Symp. Workshops NOMS Workshops 2008, pages 335–342, 2008.

[11] S. Latr´e, S. Verstichel, and B. D. Vleeschauwer. “On the design of an architecture for partitioned knowledge management in autonomic multimedia access and aggregation networks”. MACE, pages 105–110, 2009.

[12] K. Nichols, S. Blake, F. Baker, and D. Black. “Definition of the differentiated services field (ds field) in the ipv4 and ipv6 headers”. RFC 2474 (Proposed Standard), Dec. 1998. Updated by RFCs 3168, 3260.

[13] K. Nichols, V. Jacobson, and L. Zhang. “A two-bit differentiated services architecture for the internet”. RFC 2638 (Informational), July 1999.

[14] R. J. Pooley. An Introduction to Programming in Simula. Blackwell Scientific Publications. ISBN: 0632014229, 1987.

[15] P. Simoens, B. D. Vleeschauwer, W. V. de Meerssche, F. D. Turck, B. Dhoedta, and P. Demeester. “RTP connection monitoring for enabling autonomic access network qos management”. In Proceedings of the 12th European Conference on Networks and Optical Communications (NOC), 2007.

[16] TV2. Adaptive webtv streaming solution from TV2 Norway. Available online at www.tv2sumo.no. Last access 01.01.11.

[17] B. J. Villa and P. E. Heegaard. “Monitoring and control of qoe in media streams using the click software router”. In Norsk Informatikk Konferanse (NIK2010), Norway. ISBN 978-82-519-2702-4., volume 1, pages 24–33, Nov 2010.

[18] B. D. Vleeschauwer, W. V. de Meerssche, P. Simoens, F. D. Turck, B. Dhoedta, and P. Demeester. “Enabling autonomic access network qoe management through tcp monitoring”. In Proceedings of the First IEEE Workshop on Autonomic Communications and Management (ACNM), 2007.

[19] B. D. Vleeschauwer, P. Simoens, W. V. de Meerssche, S. Latré, F. D. Turck, and S. V. den Bergheb. “Autonomic qoe optimization on the access node”. In Proceedings of the Broadband Europe, 2007.

[20] H.-Y. Wei, S.-C. Tsao, and Y.-D. Lin. “Assessing and improving tcp rate shaping over edge gateways”. IEEE Trans on Computers, 53(3):259–275, 2004.

 ISBN: 978-1-61208-133-5

Efficient Mobile IP Location Update Mechanism for Idle Terminals in Optical Wireless Integrated Access Networks

[1&2]S.H. Shah Newaz*, [1]Raja Usman Akbar, [1]Youngmi Lim, [2]Gyu Myoung Lee**, [2]Noel Crespi, and [1]Jun Kyun Choi

[1]Korea Advanced Institute of Science and Technology (KAIST).
Daejeon, South Korea.
[2]Institut Telecom, Telecom SudParis.
Evry, France.
E-mail: *newaz@kaist.ac.kr, **gm.lee@it-sudparis.eu

***Abstract*— During an off-peak hour of a day, a Mobile Terminal (MT) can stay in idle mode for long time as there is no incoming or outgoing packet. An idle MT can move from one location to another and the wireless access network keeps tracking the idle MT. Idle MTs need to assist the wireless access network for location update; so that, on call arrival the network can route the call successfully. Mean while, Voice over IP (VoIP) got wide acceptance. To provide VoIP service in mobile environment or any IP packet based service, Mobile IP is very important protocol undeniably. However, Proxy Mobile IP was developed by IETF without considering the idle mode condition of MTs. Consequently, an idle MT needs to conduct Mobile IP binding (Layer 3 location update) whenever it moves to new area of a Foreign Agent although it does not have any incoming or outgoing packets during idle period. This phenomenon unnecessarily increases location update signaling cost. Here in this paper, based on optical wireless integrated access network we propose a mechanism that allows only Layer 2 location update when an MT is in idle mode and the Layer 3 location update is conducted after call arrives for an idle MT. Our numerical results show that proposed mechanism out performs than the existing solutions.**

Keywords- Mobile IP; Energy saving; Converged; Passive Optical Networks; Location Update .

I. INTRODUCTION

Energy consumed by network equipments is huge [4]. In the future, it can be anticipated that network will have more expansion, hence number of network node/equipment supposed to have an astronomical growth. To reduce amount for carbon footprint, energy saving in network has become an important research issue. Therefore, some researchers have considered idle mode condition in network devices/nodes. Authors in [5, 6] consider about idle mode in Optical Network Units (ONU). Adopting there mechanisms it is possible to reduce energy consumption in ONUs significantly. Similarly, it is also possible to reduce energy consumption by minimizing unnecessary signaling messages. It is because to produce signaling message one node needs to spend its computational power. Furthermore, by eliminating such unnecessary signaling message it is also possible to improve bandwidth utilization.

During idle period an MT needs to listen the paging signals periodically and updates its location [1]. All wireless access technologies (e.g., IEEE802.16e) require location update for idle MTs. So that, on call arrival they can be paged at the right location. Mean while Voice over IP (VoIP) got wide acceptance. To provide VoIP service in mobile environment, Mobile IP (MIP) is very important protocol undeniably. However, even the latest Proxy Mobile IP (PMIP) developed by IETF does not consider the idle mode situation MTs [2]. As a result, an idle MT needs to conduct PMIP binding (Layer 3 location update) whenever it moves to new area despite that it does not have any active session. Undeniably, this phenomenon increases location update signaling cost. Note that the location update signaling cost can be defined as amount of resources spent for managing mobility of an MT [2].

For convergence of optical and wireless network; for example, the integration of PON and WiMAX or WLAN, some possible cheap solutions have been proposed already by some researchers. Some of these solutions consider IP mobility in Optical Wireless Integrated Access Networks (OWIAN). Gangxiang Shen et al. in [3] proposed four access architectures for integrating Ethernet PON (EPON) and WiMAX. To support IP mobility, they suggest implementing a handover coordinator in the Optical Line Terminal (OLT). This handover coordinator communicates with a Wimax base station (BS), which is integrated with ONU, through a dedicated control channel. In [7], authors suggest PMIP for IP mobility among different ONUs. Besides, these days manufacturers are considering Layer 3 functionality in ONUs [19]. To reduce Layer 3 location update signaling cost, there are several existing solutions [2, 13-14]. However, they were not developed considering optical wireless integrated access network environment. To the best of our knowledge none of these existing works consider MTs' idle mode situation in OWAIN. Therefore, in OWIAN for an idle MT it is needed to conduct Layer 3 location update despite that there is no incoming or outgoing call for that idle MT.

Hence, we develop a protocol considering the characteristics of shared network like EPON. In the OWIAN, we consider EPON and WiMaX integration similar to [3, 18]. Here, our objective is to introduce a mechanism using

 ISBN: 978-1-61208-133-5

which Layer 3 location update becomes obsolete for an idle MT in OWIAN.

In our work, the terms 'Layer 3 location update' and 'PMIP binding' are used interchangeably.

In this paper, Section II discusses related work. In Section III, we present proposed mechanism. Numerical analysis is presented in Section IV. Finally, Section V draws the conclusion and states future research direction.

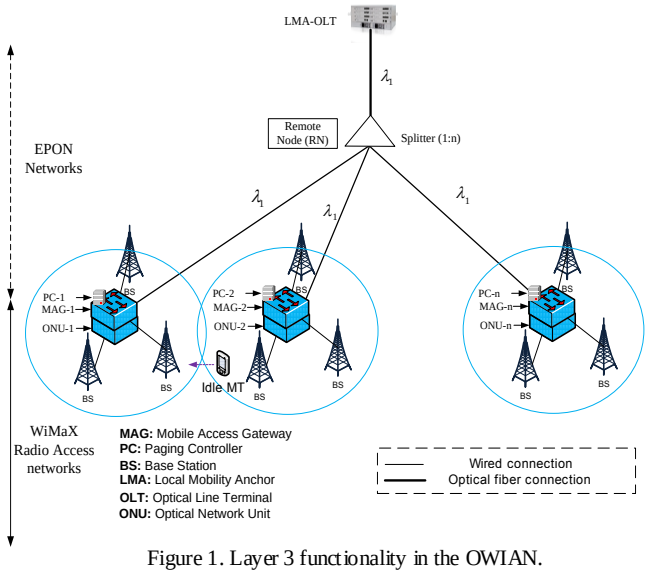


Figure 1. Layer 3 functionality in the OWIAN.

II. RELATED WORK

In this part, we first discuss some possible solutions that can be used in OWIAN to manage the IP mobility for the idle MTs and their limitations. Then we briefly explain how PMIP location update works. We also summarize MTs' idle mode consideration in wireless access networks. Finally, we present the existing works that put effort for reducing Layer 3 location update signaling cost when an MT is in idle Mode.

A. *Discussion:*

Fig. 1 states how the PON network topology looks when Layer 3 functionality for IP mobility management is considered on top of the Layer 2 based PON networks.

Note that the behavior of a particular protocol of an OSI Layer can have significant impact on other OSI Layers' protocols. Let's consider the Fig. 1 where PMIP mobility agents are considered in OWIAN similar to [3, 7]. Now, if an idle MT moves from one MA to another MA, the PMIP binding is needed to be conducted [8]. To conduct the PMIP binding, collocated MAG needs to send proxy binding update (PBU) message to the OLT and OLT needs to pass this message to the LMA (see Fig. 1). Now, if during that time ONU is idle mode, the PBU cannot be transmitted to the LMA which is collocated with the OLT. Question might be raised on why does not the MAG trigger the ONU to wakeup. Even the MAG does so, an ONU cannot transmit. The reason comes from the fact that PON is a shared media where ONUs can transmit during a dedicated uplink transmission slot only.

It is also possible that an ONU can buffer that PBU message and then transmit to the OLT when dedicated uplink time slot comes. However, this solution can bring another problem. As the downstream of PON is broadcast and select mechanism, OLT needs to mark the link Layer frames with Logical Link Identifier (LLID), so that destination ONUs can recognize their frames after reading the LLID. Therefore, when the PBU is buffered inside the ONU, in the mean time OLT might starts sending packets on call arrival for an idle MT to the previous serving MAG as the Layer 3 binding table has not been updated yet at the LMA. Consequently, it can cause long startup latency because in such case LMA needs to retransmit those packets for that idle MT through the PON network. Otherwise, the ONU of that MA somehow needs to forward to the serving ONU (collocated MAG). An ONU can forward those packets when architecture like [7] is deployed.

B. *Proxy Mobile IP (PMIP) Location Update:*

In PMIP, there are two important functional units: LMA and MAG. Any packet for an MT must come first at the LMA. A LMA knows the serving MAG for that MT.

① *LMA Working Description:*

Packets send by other nodes to the destination MT are received at the LMA within and beyond the PMIP's domain [8]. LMA creates and maintains a binding table, which keeps the record of each MT and the corresponding MAG. On packets arrival for an MT, the LMA encapsulates the incoming packets with Proxy-CoA of the serving MAG and then forwards them towards the MAG. LMA also decapsulates the outgoing packets from MT by removing LMA address (LMAA) and forwarding the packet towards its destination. It is important to mention that PMIP standard does not consider the status of an MT in LMA (idle or active).

② *MAG Working Description:*

Any incoming and outgoing packet of an MT should pass through MAG. MAG plays the role of the router on point-to-point link [8]. MAG encapsulates the packets with LMAA before sending it to LMA and also decapsulates the packet when it receives from the LMA.

C. *Ethernet Passive Optical Network in Brief:*

In EPON, an OLT is the centralized intelligence. It receives the packets from core network and then delivers to the destination ONU through optical fiber. OLT marks the frames with LLID so that ONUs can recognize their corresponding frames [15, 16, 18]. If the LLID matches, an ONU retrieves the frames and then processes. While it discards if the LLID does not match. For unicasting point of view, each ONU used to have single unique LLID, however, for broadcasting or multicasting they can have common LLIDs. In EPON, when an OLT transmits a frame conveyed by a wavelength λ_1 it is splitted by a passive splitter which

 ISBN: 978-1-61208-133-5

situates at a remote node as shown in Fig. 1. And then the same frame is directed to the all ONUs through fiber. Usually the splitting ration is *1: n*, where *n* can be 16 or 32 or 64.

D. Idle Mode Operation in WiMaX:

When there is no incoming or outgoing packet for a certain amount of time T_{th}, an MT moves from active mode to idle mode to save battery power. Fig. 2 states the state transition diagram. WiMAX (e.g., IEEE 802.16e) has three mobility entities for managing idle mode and paging operation [1, 17]. These include: paging agent (PA), paging controller (PC), and paging grouping (PG). PC can be collocated in a MAG or in a BS [1]. PC performs the task of observing the activities of all MTs located in particular Paging Group. An idle MT periodically wakes up and listens whether it has any downlink packet or not.

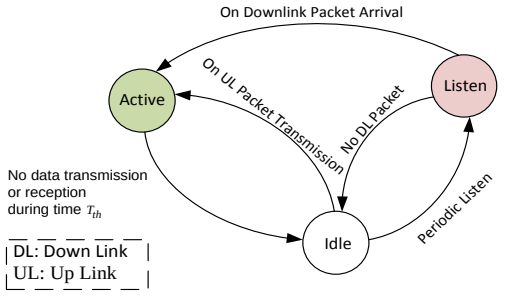


Figure 2**:** MT's mode (active/idle) transition diagram.

E. Existing Solutions for Location Update:

Many of the researchers have considered idle-mode condition of MTs in Mobile IP environment. They propose to reduce Layer 3 location update cost when an MT is in idle mode by adopting different approaches.

For example, in [13], the authors suggest to delay the PMIP location update till a call arrives for an idle MT. And when a packet arrives, LMA forwards it to new-MAG via old-MAG. It is because this scheme does not allow Layer 3 location update as long as the MT is in idle mode. However, when the distance between the old-MAG and the new-MAG is long, MT experiences a long startup latency and additional signaling burden at MAGs.

In [14], the authors indicate that Layer 3 location update should be performed at every MAG. We named this scheme in this paper as per-MAG location update (PML). This scheme reduces startup latency because Layer 3 routing information is refreshed whenever an MT visit a new MA. However, it also increases location update signaling cost.

In [2], a distance based approach is mentioned. After crossing a predefined distance (*d*) PC, which is the Layer 2 mobility agent of IEEE 802.16e, invokes the MAG to conduct Layer 3 location update. This mechanism can reduce the number of Layer 3 location update significantly. However, it has some drawbacks. For example, with the increase of predefined *d*, startup latency also increases. This is because the startup latency is a function of distance between o-MAG and n-MAG [2].

As a matter of fact, existing solutions proposed in [2, 13, 14] cannot be used for reducing Layer 3 location update signaling cost for idle MTs in OWIAN. This is because those solutions were made considering a wired network architecture. Adopting those in OWIAN, there might be improper operation. For example, solution in [2] needs retunneling between old-MAG and new-MAG. In fact, it is not possible in typical PON architecture as an ONU here communicates with other ONUs through the OLT. Therefore, if aforementioned procedure is followed startup latency supposed to be increased.

III. PROPOSED LOCATION UPDATE MECHANSIM IN THE OWIAN

In this paper, our objective is to develop a protocol in such a way that can eliminate Layer 3 location update during the inter-call arrival time (MT's idle period) and can forward the packets of an idle MT on call arrival properly in the OWIAN.

We consider that when time between two call arrivals for an MT crosses T_{th}, OLT assumes that the MT is in idle mode. We further assume that in that EPON network domain another kind of LLID is used for only the idle MTs' packets. In other words it can said that the OLT marks all the idle MTs' frame with a special LLID (SLLID) and this SLLID is known to all ONUs in the OWIAN. Therefore, whenever an ONU receives any frame marked with that SLLID, it opens that frame. Note that the reason behind the SLLID is if the OLT marks the packets based on serving MAG, OLT might forward to the wrong ONU. The reason is that after Layer 3 location update an MT may not stay in the same MA.

On the other hand, in the proposed solution an idle MT only performs Layer 2 location update in the wireless access network domain with the PC, which is collocated with the MAG, through the serving BS as shown in Fig. 1. And Layer 3 location update at LMA is postponed as long as the MT is in idle mode. So that signaling cost for Layer 3 location update can be saved. The following paragraphs explain how proposed mechanism works.

(i) The idle MT moves from one MA to another MA, without performing any Layer 3 location update.

(ii) When call comes for any MT at the OLT, OLT resolves status of the MT (idle/active) from T_{th}. If the MT is in idle mode OLT sends those packets after marking with that SLLID. Otherwise, OLT assumes that MT is active and it is in the same MA from where last Layer 3 location update was conducted. In fact, there is still chance that active MT might move to another MA within that t_{move}, where $t_{move} < T_{th}$. However, we avoid such scenario in this paper.

(iii) If the MT is active, the OLT resolves the corresponding LLID of that ONU from the Proxy-CoA of the serving MAG and then forwards the packets after marking with that LLID. The following algorithm states how the OLT makes decision on arrival of a call.

 ISBN: 978-1-61208-133-5

(iv) On the other hand, when an ONU receives a frame marked with SLLID, it extracts the packets inside. Then, it invokes collocated PC asking all the idle MTs' lists of its MA. PC provides the list in reply. If the ONU finds any packet which is destined for any one of these idle MTs, it requests PC to page that MT to wakeup. After successful paging, idle MT switches from idle mode to active mode. Then, those destined packets are forwarded through the serving BS to that MT.

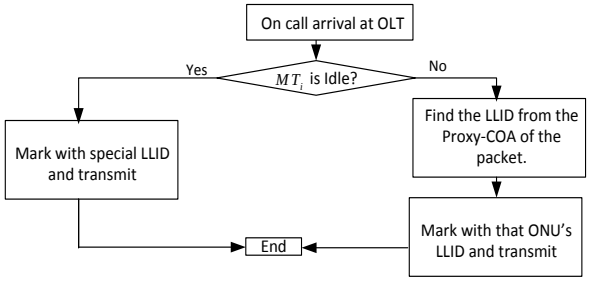


Figure 3: LLID selection and forwarding procedures in OLT.

(v) When the idle MTs wakeup and start receiving packets, ONU can collect all the PBU messages for Layer 3 location update from the collocated MAG, and then can transmit to the OLT during dedicated uplink transmission slot.

(vi) On arrival of PBUs at OLT, OLT sends those to collocated LMA. And finally LMA updates its binding table (Layer 3 location update is performed).

IV. NUMERICAL MODELING AND PERFORMANCE ANALYSIS

In this part, we are interested to know how these location update protocols proposed in [2, 14] perform between two call arrivals when they are dispensed in an OWIAN. Similar to [2, 10], we consider that location update cost is the amount of resources spent for managing mobility of an MT. It might be possible that network operators can consider relative cost spent at a mobility agent (e.g., LMA, MAGs, OLT) for tracking an MT. For example, resources spent for processing a signaling message, amount of bandwidth consumes for transiting that signaling message.

We draw the previous OWIAN architecture depicted in Fig. 1 again in Fig. 4, for numerical modeling. Similar to [10-12], the parameters are defined as follows for numerical analysis:

- λ : Call arrival rate follows Poisson process.
- a_l : Location update processing cost at the LMA.
- a_o : Location update processing cost at the OLT.
- a_m : Location update processing cost at the MAG node (the node where MAG and PC exists) with the collocated PC.
- $l_{OLT,ONU}$: Distance between the OLT and an ONU.
- $l_{m,mt}$: Average hops between the MAG and MT via a BS.
- D: Diameter of an MA.
- η_m^{-1} : Is the mean residence time follows exponential distribution, of an MT in an MA.

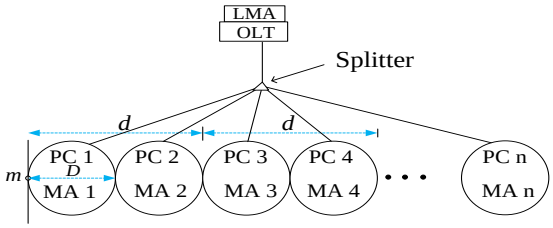


Figure 4: OWIAN architecture for numerical modeling.

We assume that trip of an idle MT starts from the point *m* as shown Fig. 4. Similar to [10], it is also considered here that signaling cost for transmitting a signaling message is proportional to the distance between source and destination node. And the proportional is δ_u. The time between two call arrivals or inter-call arrival time can be presented as $T_c = 1/\lambda$. And then number of MA crossed by an MT during that T_c is $x = \lambda^{-1}\eta_m$. Based on the parameters defined earlier, the equation (1) and (2) shown below are developed for measuring location update cost at the LMA and MAG respectively. Note that LMA location update should include location update processing cost at LMA and OLT.

$$C_l = a_l + a_o + 2(l_{OLT,ONU} + l_{m,mt})\delta_u. \quad (1)$$

$$C_m = a_m + 2l_{m,mt}\delta_u. \quad (2)$$

Using (1) and (2), location update signaling cost between two call arrivals can be expressed as follows for PML [14].

$$L_{cost}^{PML} = \lambda^{-1}\eta_m(C_m + C_l). \quad (3)$$

As mentioned before DBLU does not suggest Layer 3 location update always, rather it makes Layer 3 location update after crossing a predefined distance *d*. Therefore, based on the inter call arrival time, DBLU can have three different location update signaling cost. First, if a call comes at every MA, then DBLU must do Layer 3 location update at all MAs [(a)]. Second, calls come at the time when idle MT crosses predefined distance [(b)]. And finally, a call comes in MA where DBLU does not suggest performing Layer 3 location update as the predefined distance has not been crossed by that idle MT [(c)].

$$L_{cost}^{DBLU} = \begin{cases} x(C_m + C_l) & \left(\lambda^{-1}\eta_m = x\right) < 1 \quad \text{(a)} \\ xC_m + xDd^{-1}C_l & \left(\lambda^{-1}\eta_m = x\right) when\ x \in \text{integer} \quad \text{(b)} \\ xC_m + xDd^{-1}C_l + C_l. & \left(\lambda^{-1}\eta_m = x\right) > 1, when\ x \notin \text{integer} \quad \text{(c)} \end{cases} \quad (4)$$

In the proposed mechanism, it is mentioned earlier that there would not be any Layer 3 location update with the LMA for an idle MT as long as any call arrives. And during

T_c an idle MT must conduct Layer 2 location update with a PC through a serving BS. Then location update signaling cost in proposed scheme can be presented as:

$$L_{cost}^{Proposed} = \lambda^{-1}\eta_m C_m + C_l. \quad (5)$$

Between an MT and a MAG, an intermediate hop can be a BS [1]. So we assume $l_{m,mt}$ = 1Km. Besides, similar to [10, 12] we assign values for the parameters in table 1. We vary the T_c from 0 to 4000 sec. while the $1/\eta_m$ is kept 640 sec [11].

TABLE I. PARAMETER VALUES FOR LOCATION UPDATE COST EVALUATION

Parameter	Value	Parameter	Value
a_l	25	δ_u	0.1
a_m	10	$l_{OLT,ONU}$	20 Km
a_o	25	D	3Km
d	6 Km	r	9 Km

Fig. 5 states that as T_c increases the location update signaling cost, which encompasses Layer 2 and Layer 3 location update signaling, increases gradually in all three schemes: PML, DBLU, and proposed. However, among those proposed solution performs best. It is because this scheme can successfully ignore Layer 3 location update during the T_c (time between two call arrivals) without increasing the startup-latency.

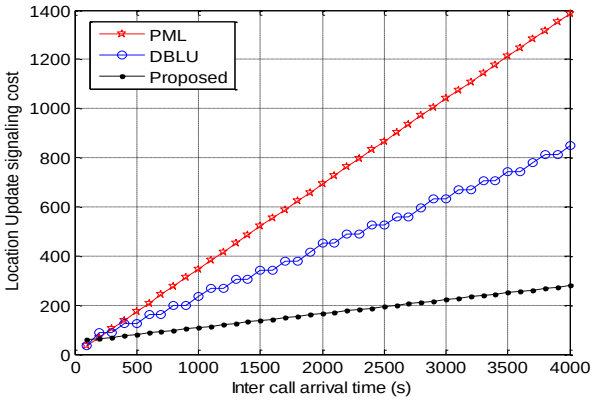


Figure 5: Location update signaling cost between two call arrivals in an OWIAN.

V. CONCLUSION AND FUTURE WORKS

Our proposed mechanism helps to fully utilize the advantage of idle mode. In this mechanism, an idle MT does not need to wakeup for conducting Layer 3 location update; therefore they can stay in idle mode only by performing Layer 2 location update in the wireless access network domain. Besides, mainly it contributes a great deal of reduction of Layer 3 location update signaling messages in OWIAN. In our future work, we would like to evaluate the performance of proposed mechanism at different residence time of an idle MT.

ACKNOWLEDGMENT

This work was supported by the IT R&D program of MKE/KEIT (KI001822, Research on Ubiquitous Mobility Management Methods for Higher Service Availability; and A1100-0801-3015Development of Open-IPTV Technologies for Wired and Wireless Networks).

REFERENCES

[1] Y. Zhang and H.-H. Chen, "Mobile WiMAX: toward Broadband Wireless Metropolitan Area Networks," Auerbach Publication, 2008.

[2] S. Jin, C. Yoon, and S. Choi, "A Simple Remedy for Idle Mode via Proxy MIP," *IEEE Comm. Lett.*, pp. 423-425, June 2008.

[3] G. Shen et al., "Fixed Mobile Convergence Architectures for Broadband Access: Integration of EPON and WiMAX," *IEEE Communications Magazine*, Vol. 45, Issue 8, pp. 44-50, Aug. 2007.

[4] M. Gruber, O. Blume, D. Ferling, D. Zeller, M.A. Imran, and E.C. Strinati, "EARTH — Energy Aware Radio and Network Technologies,"2009 IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications, pp. 1-5, Sept. 2009.

[5] S. Wong, L. Valcarenghi, S. Yen, D.R. Campelo, S. Yamashita, and L. Kazovsky, "Sleep Mode for Energy Saving PONs: Advantages and Drawbacks," *IEEE GLOBECOM* Workshops, pp. 1-6, Nov. 2009.

[6] P. Chowdhury, M. Tornatore, S. Sarkar, and B. Mukherjee, "Building a Green Wireless-Optical Broadband Access Network (WOBAN)," *Journal of Lightwave Technology*, pp. 2219-2229, Aug. 2010.

[7] S. Lee, J. Kim, and M. Kang, "Route optimization for proxy mobile IPv6 using local customer internetworking in PONs," *Optical Internet (COIN)*, pp. 1-2, Oct. 2008.

[8] S. Gundavelli, "Proxy Mobile IPv6," RFC 5213, IETF, August 2008.

[9] S.H.S. Newaz, S. H. Jeon, J. H. Lee, J. Lee, J.K. Choi, and M. H. Kang, "An approach of multipath transmission from intelligent OLT in optical-wireless converged networks," *Optical Internet (COIN)*, 9th International Conference on , pp. 1-3, July 2010.

[10] J. Xie and I.F. Akyildiz, "A Novel Distributed Dynamic Location Management Scheme for Minimizing Signaling Costs in Mobile IP," *IEEE TMC*, pp. 163- 175, Sep. 2002.

[11] S. Jin, K. Han, and S. Choi, "Idle Mode for Deep Power Save in IEEE 802.11 WLANs," *JCN*, Oct. 2010.

[12] Y. Li, Y. Jiang, H. Su, D. Jin, Li Su, and L. Zeng, "A Group-Based Handoff Scheme for Correlated Mobile Nodes in Proxy Mobile IPv6," in *Proc. IEEE GLOBECOM'09*, pp. 1-6, Dec. 2009.

[13] S. Jin, K. Han, and S. Choi, "A novel idle mode operation in IEEE 802.11 WLANs," in Proc. *IEEE ICC'06*, pp. 4824-4829, June 2006.

[14] J. Na, Y. Chung, H. Noh, S. Lee, and S. Kim, "Two Alternative Registration and Paging Schemes for Supporting Idle Mode in IEEE 802.16e Wireless MAN," *IEEE VTC'06 Fall*, pp. 1-5, Sep. 2006.

[15] G. Kramer and B. Mukherjee, "Supporting differentiated classes of service in Ethernet passive optical networks," *Journal of Opt. Netw.*, vol. 1, no. 8/9, pp. 280–298, Aug. 2002.

[16] IEEE Std. 802.3ah-2004, "Part 3: Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications," *Institute Computer Society*, Sep. 2004.

[17] IEEE Std. 802.16-2009, May 2009.

[18] S.H.S. Newaz, Y. Bae, M.S. Ahsan, and J. K. Choi, "A study on PMIP deployment over EPON," *Optical Internet (COIN)*, 9th International Conference on , pp. 1-3, July 2010.

[19] <http://www.madisontech.com.au/fibre-optics/product/ftth-onu> 06.05.2011.

 ISBN: 978-1-61208-133-5

Versatile Configuration and Deployment of Realistic Peer-to-Peer Scenarios

George Milescu, Răzvan Deaconescu, Nicolae Țăpuș
Automatic Control and Computers Faculty
University POLITEHNICA of Bucharest
Emails: {george.milescu,razvan.deaconescu,nicolae.tapus}@cs.pub.ro

***Abstract*—With the advance of Peer-to-Peer solutions, research and commercial players have shown interest in enhancing local client and overall swarm performance in order to improve content distribution and user satisfaction. Protocol measurements and careful client and swarm behavior analysis are required to provide valuable information on improving performance. In this paper, we present a Peer-to-Peer testing infrastructure that enables easy deployment of complete and controlled BitTorrent swarms. The infrastructure allows a variety of realistic scenarios to be run with the ability to configure characteristics such as client type, bandwidth management, churn rate and number of connections.**

***Index Terms*—Peer-to-Peer, BitTorrent, infrastructure, automation**

I. Introduction

The last 20 years have seen the birth and expansion of the Internet from a small network of academic and government institutions to a global network spanning borders, cultures and homes. With the ever increasing network bandwidth, file and data transfer is the Internet service that is responsible for the largest chunk in the Internet backbone. HTTP and Peer-to-Peer systems are nowadays the main bandwidth consumers in the Internet, with video content as the most common type of traffic going through the Internet links [1].

Peer-to-Peer systems have emerged as the most suitable solution to capitalize on the huge unexploited network bandwidth available on the Internet. Since the inception of Napster in the late '90s, Peer-to-Peer systems have evolved to a variety of solutions and applications that continuously stir the interest of institutions (be them academic or commercial) across the world.

The most eloquent example of Peer-to-Peer systems' success story is the BitTorrent protocol, currently responsible for the larges chunk in Internet Peer-to-Peer traffic [1]. With simple, yet highly effective features such as optimistic unchoking, tit-for-tat and rarest-piece first, the BitTorrent protocol is one of the best suited solutions for large data distribution. Recent research focus has been in integrating features such as social networking, reputation management, video streaming as core features or overlays on top of the protocol.

In this paper we present a Peer-to-Peer software testing infrastructure that provides flexibility, control and automation. The infrastructure allows deployment of realistic P2P scenarios, gives full control to the experimenter and automates the interaction with Peer-to-Peer clients. Our solution provides the means to define an array of input variables for scenarios: number of peers, leechers, seeders, bandwidth limitation, client type, number of connections, intervals of activity (churn rate). Client output information is automatically retrieved as log files and rendered through statistical processing.

The testing infrastructure uses shell scripts and configuration files to setup and manage BitTorrent client swarms. The use of shell scripts allows easy integration with BitTorrent clients, takes advantage of the SSH (Secure Shell) protocol for remote system control and provides interaction with tools for parsing and processing output information.

We have successfully deployed and used the testing infrastructure both on a physical environment (consisting of 10 hardware nodes) and on a virtualized environment (consisting of 100 OpenVZ [4] containers) running on top of the physical environment. We have been able to run scenarios containing 100 hosts, each running a BitTorrent client instance. The use of OpenVZ allows lightweight virtualization and easy simulation of complete systems on top of a small number of hardware nodes.

II. Related Work

Current research focus regarding Peer-to-Peer systems and protocols uses carefully crafted experiments and network simulators.

A survey of the use of Peer-to-Peer simulations has been undertaken by Naicken et al. [13]. The authors surveyed papers and collected information regarding the use of simulators for Peer-to-Peer systems. Five criteria had been used to evaluate the simulators: simulation architecture, usability, scalability, statistics and underlying network simulation. A large number of custom simulators were detected to have been deployed, the main cause for that being assumed to be the lack of proper statistics output. The authors criticize the use of NS-2 as a simulator for Peer-to-Peer systems and provoke discussion to help build a consensus on the common platform for Peer-to-Peer research.

One of the best places to look for deploying network experiments, also heavily used by Peer-to-Peer researchers, is Planet Lab [6]. With more than 1000 nodes and 500 sites spread all over the world and healthy documentation, PlanetLab offers a suitable environment for Peer-to-Peer experiments. As user nodes are virtualized through the use

 ISBN: 978-1-61208-133-5

of Linux-Vserver, experimenters have complete control over their system and its resources. The user may deploy a given set of tests or use PlanetLab as an underlying layer for a testing infrastructure (such as the one presented in this article) and be able to deploy a realistic environment for various scenarios.

NS-2 [7] is one of the most popular network simulators. Thorough documentation, continuous development over the past two decades and a rich set of features have ensured NS-2 as a prime candidate for network experiments. However, as Naicken et al. [13] conclude, NS-2 is particularly useful for detailed modelling of the lower network layer, a characteristic that is of little interest to Peer-to-Peer researchers, though it has been often used in Peer-to-Peer experiments.

We consider PlanetLab [6] and NS-2 [7] to be located at separate poles when discussing about the purpose of Peer-to-Peer experiments. PlanetLab and virtualized environments allow deployment of realistic scenarios, and collected valuable realistic information, but lack scalability. On the other hand, NS-2 and network/P2P simulators allow simulation of large number of nodes (even to the degree of millions) while failing to provide accurate data about client behavior and detailed statistics. We consider that, given the nature of the BitTorrent protocol as a solution for content distribution, realistic (or even real) environments are appropriate for experiments regarding BitTorrent swarms.

Dinh et al. [11] have used a custom network simulator (dSim) for large scale distributed simulations of P2P systems. The authors have been able to simulate approximately 2 million nodes for Chord and 1 million nodes for Pastry. Similar work has been presented by Sioutas et al. [16]. Video streaming in Peer-to-Peer networks has been simulated as described by Bracciale et al. [9] using a custom simulator dubbed OPSS.

With respect to BitTorrent simulators and closer to the purpose of this article, Pouwelse et al. [14] have undertaken a large BitTorrent measurement study spanning over several months on real BitTorrent swarms (provided by the Suprnova tracker). Data was collected through HTML and BitTorrent, (ab)using scripts, from the central tracker and BitTorrent clients. A similar approach has been employed by Iosup et al. [12]. The authors have designed and implemented MultiProbe, a framework for large-scale P2P file sharing measurements on the BitTorrent protocol. MultiProbe has been deployed in real swarms/environments and collected status information from BitTorrent peers and subject it to analysis and dissemination.

Our testing infrastructure is deployed on a hardware experimental setup (similar to a local PlanetLab) presented in an earlier paper [10]. Instrumented BitTorrent clients, logging facilities and an OpenVZ lightweight virtualization solution are basic block on top of which the software testing infrastructure was developped and used.

III. Design and Architecture

A. Design goals

The use of network simulators for creating controlled environments has been an easy solution for achieving BitTorrent measurements. However, real BitTorrent clients behave differently from simulators and the network protocol stack has an important influence on the outcome of a scenario.

Considering the decreasing cost of hardware and the improvements in virtualization solutions, running network emulations with hundreds of nodes, each having a dedicated instance of an operating system, is an achievable objective. In this sense, Rao et al. [15] concluded that results gathered from BitTorrent experiments performed on clusters are realistic and reproductable.

The proposed infrastructure for controlling peer-to-peer clients aims at providing an extensible and adaptable tool for experiment setup, execution and analysis. It has four primary goals, allowing it to be used in a large variety of scenarios.

The first goal is to provide an **extensive tool for managing both clients and log files** during experiments. Running scenarios that include a large number of clients (up to a few hundred) requires a control mechanism for starting, monitoring and stopping clients in a short time-frame. Most of the scenarios result in a collection of log files, at least one log file per client or per machine. Collecting and analysing these log files, considering the large number of remote machines, has to be automated.

The second goal is to use a **common interface for accessing remote systems**. The nodes on which clients run must consist of various Linux or Unix distributions, and, most likely, the machines are not administrated by the user running the scenarios. Also, the nodes could be hardware or virtual machines. A common access interface to this heterogeneous node infrastructure is needed, and the interface must not require administrative privileges for accessing the remote nodes.

The third goal is to offer **support for bandwidth control**. Cluster computers are generally connected with 1Gbit/s or faster network connections. These types of connections are not common for end-users. In order to provide realism to the experiments, the infrastructure needs to offer a mechanism for controlling the amount of bandwidth each client can use. Having the bandwidth control integrated in the infrastructure offers the advantages of fine-tuning the scenarios and recreating a wide range of network environments.

The last goal is to **allow the user to introduce churn in the environment**. Starting and interconnecting P2P clients is only the first step towards reproducing a real-life scenario. Two of the elements that characterize real swarms are churn and population turnover. Both translate into clients joining and leaving the network at different time intervals. Controlling the periods when each client is connected to the network gives the user the freedom of creating a variety of scenarios, from a controlled flash-crowd to a swarm close to extinction.

As mentioned, the proposed infrastructure provides a tool for experiment setup, execution and analysis. It is the experimenter's task to design the experiment parameters and to validate the used models against simulated results or other real-life measurements.

 ISBN: 978-1-61208-133-5

B. Design elements

From a design point of view, the infrastructure uses four concepts: campaign, scenario, node and client.

A **campaign** consists of a series of experiments, each experiment being independent of others and having associated a specific type of data processing. The difference between a campaign and an experiment resides in the fact that results from an experiment may be plotted on a single graph, while results from a campaign need a deeper analysis. Multiple experiments may be included in a campaign. If an experiment needs to be run multiple times (to retrieve significant results), it can be included multiple times in the same campaign.

A **scenario** coresponds to a single experiment. It is associated with a specific type of data processing and its results are generally presented on a single graph.

A **node** is one of the infrastructure machines. It can be a virtual or a hardware machine. The user running the experiments needs to have access to the nodes both for experiment deployment and execution and for bandwidth control.

A **client** is a single instance of a peer. The infrastructure is designed to run a single client on each node, in order to reproduce the real-life execution context for P2P clients.

Campaigns and scenarios each use configuration files that include a complete specification of the experiments. The campaign configuration file specifies the scenarios included in the campaign. The scenario configuration file includes all nodes that are part of the infrastructure used to execute the experiment; for each node, the configuration file defines access parameters, client types, churn and the bandwidth limitations.

C. Architecture overview

The local machine is used to control the infrastructure. It stores the infrastructure scripts, configuration files, and campaign output. It may also store code or executable files for P2P clients. The infrastructure scripts copy required files from the local machine to remote nodes, set up the enviroments and start the clients. After the experiment ends, the results (log files) are brought back from the remote node to the local machine.

The testing infrastructure uses a modular architecture. Some of the modules are *generic* (such as the module that parses the configuration files); other modules are *node or client specific* (for example the module that parses the log files obtained from a client). From a different point of view, part of the modules are executed on the *local machine*, others on the *remote host*.

The infrastructure architecture is depicted in Figure 1. The *run_campaign* component reads the campaign configuration file and executes each of the specified scenarios. After a scenario is executed, its results are processed, and the next scenario is run. At the end of the campaign, campaign results may be published as a web-page for preliminary analysis.

run_scenario, the central point of the intrastructure, is responsible for managing all activities related to the execution of an experiment. Its specific components will be detailed in the following section.

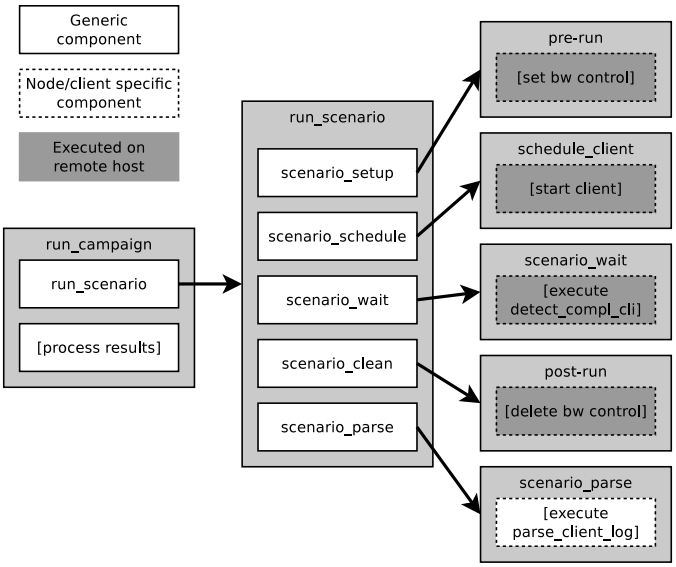


Figure 1. Infrastructure design overview. The components use "_" between the component names. The actions, that are not directly included in a component, are placed between [ ]

D. Architecture details

Figure 1 presents the architecture overview. The central point of the infrastructure is *run_scenario*, the component responsible for executing a scenario. This section details its components and explains the mechanisms it uses to deploy and execute scenarios.

After the scenario configuration file is parsed, each of the nodes will be prepared for the experiment by *scenario_setup*. This component is detailed in Figure 2. The first step is to synchronize the local infrastructure scripts with the remote host. The synchronization phase cleans up the remote host and ensures that consecutive scenarios do not influence each other.

A local node-specific configuration file, including parameters related to that node, is created for each of the nodes specified in the scenario configuration file. The node-specific configuration file is then copied to the remote host. This file is used for inter-component communication between the local-executed and the remote-executed components.

The *pre-run* component prepares remote host environments for the experiment. This component parses the node-specific configuration file and applies settings required for the scenario. The *pre-run* component also handles bandwidth limitations.

The *schedule_client* component schedules client executions on the remote host. Based on the node-specific configuration files stored on the remote host, *schedule_client* starts and stops the client to simulate the specified churn. The client lives until the *scenario_wait* component detects completion of the experiment, after which it may be stopped. The client will not be immediatelly stopped, as the infrastructure waits for all the clients to complete the experiment before stopping them.

After all clients complete the experiment, each node will be cleaned up by the *scenario_clean* component, as presented

 ISBN: 978-1-61208-133-5

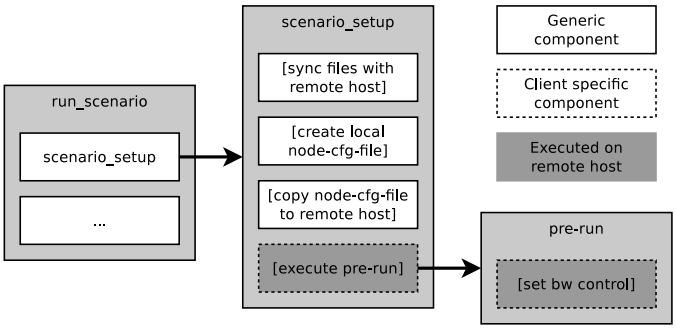


Figure 2. Detailed scenario_setup components. The components use "_" between the component names. The actions, that are not directly included in a component, are placed between [ ]

in Figure 3. This component stops the client and retrieves the remote log files. A *post-run* component is then executed reverting all settings applied by *pre-run* to ensure that consecutive scenarios do not influente each other. In the end, the remote node-specific configuration file is deleted and local infrastructure scripts are synchronized to the remote host to clean any temporary file.

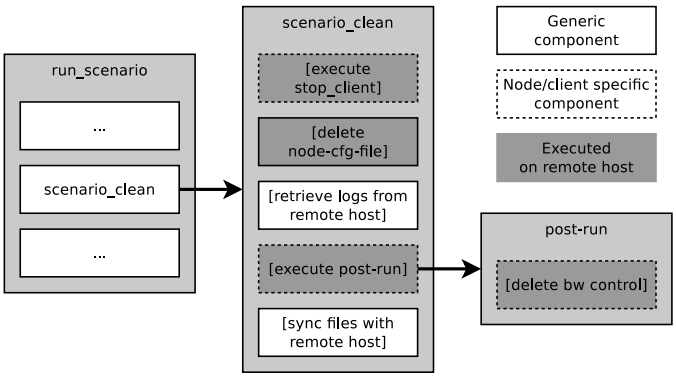


Figure 3. Detailed scenario_clean components. The components use "_" between the component names. The actions, that are not directly included in a component, are placed between [ ]

Information from clients is stored in log files. The last stage of the scenario execution, *scenario_parse*, translates the client-specific log format to an unified format used by the processing stage.

Log files are used to analyse the evolution of various client parameters during each scenario by storing periodic status information, such as download speed, number of connections or ratio. Specialized log files could also be created, if the clients are instrumented, and gather detailed periodic information (for example information consisting of instant per-peer download speed and upload speeds).

Specially designed R scripts are invoked in the post-processing phase. Using information stored in the unified log files format as input, the R scripts output graphical representation of the evolution of client parameters such as download speed.

IV. Infrastructure Implementation

A. Node and client specific components

Part of the components presented in Section III are node or client specific and will be detailed in this section.

The first client- and node-specific component is *pre-run*. One of its main tasks is to configure the bandwidth limitations on the remote host. Three solutions have been tested to enforce the limitations (the solutions will be detailed in IV-D):

- controlling bandwidth at the operating system level
- controlling bandwidth at the process level
- controlling bandwidth within the P2P client

pre-run is both client and node-specific. Some clients do no offer bandwidth control, while bandwidth used by some virtualization solutions can not be limited at the operating system level.

The interface between the infrastructure and the BitTorrent clients is composed of three client-specific components: *scenario_schedule*, *scenario_wait* and *scenario_clean*. The only infrastructure requirements for BitTorrent clients are to provide a CLI (Command-Line Interface) interface to run on top of a Linux system and to offer runtime-generated log messages..

scenario_schedule is responsable for starting the clients on the remote nodes. The *start client* script is client-specific. This script also prepares the running environment prior to starting the client.

After a client starts, the *scenario_wait* component monitors it to detect the experiment completion. The detection phase is dependant on both the goal of the experiment, and on the type of client used. A remote client is considered completed either by reaching a run-time state defined by the scenario or when the churn configuration implies a final stop action (see Section IV-C). The runtime-based completion detection requires the infrastructure to detect the completion of the experiment based on the messages the client logs while it runs. As each client has a different log format and specific experiments require special log messages, *scenario_wait* is adapted to user needs. Given the generic architecture, the infrastructure may be used for multiple types of experiments, targeting download performance, epidemic protocol measurements, user behavioral patterns, etc.

The *scenario_wait* component causes the command station to wait for all remote clients to complete the experiment. Subsequently, log files from remote clients are retrieved to the command station and parsed. The parsing process results in an unified generic format (consisting of table and matrix structured files) that is used as input for statistical analysis.

After all clients have completed the experiment, the *scenario_clean* component stops them and cleans up the remote host. The script used to stop the client is paired with the script used to start it, and is client-specific. The *post-run* component is used for the clean-up phase. Similar to *pre-run*, it has to revert the settings prior to stopping the experiment; this stage includes deleting the bandwidth limitations. *post-run* is node specific.

 ISBN: 978-1-61208-133-5

The last client-specific component is *scenario_parse*. Each client uses a particular log format that should be transparent to the results processing stage. As mentioned before, a translation is required, from the client-specific log format to an unified format used by the processing stage.

B. Employed technologies

The testing infrastructure implementation is based on shell (Bash) scripts. With support in any Linux operating system, and no requirements for additional software, shell scripts provide an ideal environment for easy deployment and exploitation. Shell scripting offers access to a flexible set of tools for parsing client output logs and automating tasks.

The common interface used to access remote systems is based on the SSH protocol. Although file transfers are available via SCP (Secure Copy), the rsync protocol was preferred for folder synchronization between different hosts, transferring only the information that was updated.

Statistical analysis in the testing infrastructure is achieved through the use of automated R language scripts. A powerful tool for processing large amounts of data, R can also do graphical post-processing.

With the exception of scripts used to run a campaign or a scenario or for post-processing, all other scripts are run on the remote systems. The scripts running a campaign or a scenario parse the configuration files on the command station and use SSH to command the scripts on the remote systems. The remote system scripts prepare the node for the experiment and manage the P2P clients (start, monitor, stop).

C. Churn simulation

One of the main goals of the proposed infrastructure is to allow the user to introduce churn in the environment by controlling the periods when each client is connected to the network. An array of intervals included in the scenario configuration file specifies the on-off behaviour for each of the clients. The *schedule_client* control script uses the UNIX signals SIGTOP and SIGCONT to suspend and resume the client processes at the specified moments of time. The churn model (specified the array of time intervals) has to be provided by the user.

D. Bandwidth limitation

As mentioned in IV-D, three solutions regarding bandwidth limitation have been tested.

The first solution is using the `tc` [2] (traffic control) Linux tool, allowing a variety of limitation algorithms implemented at the kernel level. Due to particularities of the OpenVZ implementation, `tc` cannot be currently used as a bandwidth limiter between containers.

In order to bypass this issue, client level limitation (also known as rate limiter) was also tested. hrktorrent and transmission clients offer implicit limitation functionality. This approach does have its downsides, as it is less flexible and is process-centric – one cannot limit the total amount of traffic sent by a client (e.g. a combination of P2P and HTTP traffic).

If the client offers no implicit rate limiter, bandwidth control may still be enabled through the use of the `trickle` [8] tool. `trickle` uses a form of library interposition to hook network related API (Application Programming Interface) calls and limit per-process traffic. It has two drawbacks: it is not actively maintained and issues arise when using the `poll` library call; in case of Linux, `epoll` support is absent.

V. Running Experiments

One of the main goals of the testing infrastructure is to relieve the experimenter of the burden of experiment management and monitoring, providing an extensive tool for managing both clients and log files. As much of the experiment as possible should be run in "background" with little input from the user.

By use of the proposed testing infrastructure, the activity of managing clients, sending commands and collecting information is completely automated, leaving the experimenter with only three tasks to accomplish, sequentially:

1) create the client-specific scripts
2) create the campaign configuration and the scenario configuration files
3) run the campaign startup script

After filling the required information in configuration files, the user running the experiment starts the campaign through the use of a control script that receives, as argument, the name of the campaign configuration file. The script parses the configuration file and creates and manages a swarm for each scenario accordingly. In order to limit the possibility of the user accidentally stopping the campaign control script, it is recommanded to detach the running terminal using tools such as `screen`, `nohup` or `dtach`.

After completion of campaign experiments, all output information and R processed graphic files are stored locally, in a campaign-specific folder. This folder contains a sub-set of folders, one for each scenario, that store log and graphics files.

In terms of scalability, we have successfully deployed and used the testing infrastructure on a 100-node virtualized infrastructure [10] running on top of the physical environment. Thus, we were been able to run scenarios containing 100 hosts, each running a BitTorrent client instance.

Figure 4 presents the outcome of such a scenario, comparing the evolution of peer download speed with respect to download percentage in a 90 peer swarm consisting of 50 seeders and 40 leechers. All peers were limited to `8Mbit/s` upload and download speed and shared a 700MB file. As the figure depicts, the clients reached the maximum allowed transfer rate.

VI. Conclusion and Further Work

This article presented a new approach to building an automated infrastructure that allows easy deployment of experimental scenarios involving Peer-to-Peer clients. Main design goals for the infrastructure were providing an extensive tool for managing both clients and log files, using a common interface for accessing remote systems, offering support for

 ISBN: 978-1-61208-133-5

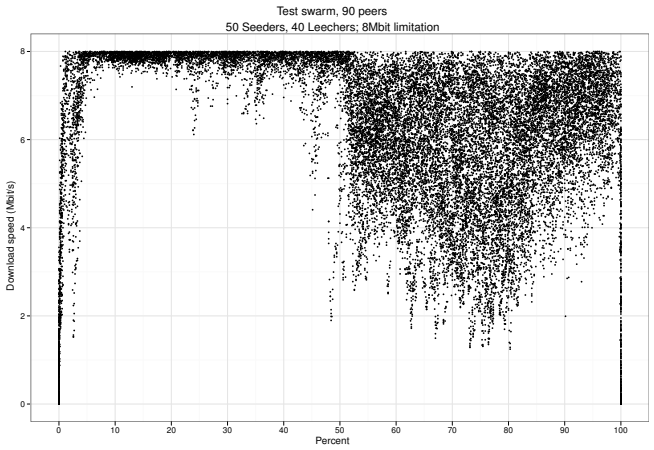


Figure 4. Scenario output: download speed evolution

bandwidth control and allowing the user to introduce churn in the environment. The infrastructure uses a hierarchical set of configuration files and run scripts and has been deployed for a variety of Peer-to-Peer experiments.

The main advantage of the proposed infrastructure when compared to other solutions is automation coupled with easy deployment. The use of a single commanding station, shell scripts, SSH and rsync allows the user to rapidly deploy a given scenario. The possible use for deployment of an OpenVZ virtualization allows consolidation – a small number of hardware nodes are used to create a complete virtualized framework capable of running sandboxed BitTorrent clients. With the use of Linux specific networking tools, the user may define bandwidth limitation and network topology characteristics in order to simulate realistic scenarios.

Given the flexibility of the client-interface and the provided churn and bandwidth-control features, any given Peer-to-Peer scenario can be deployed using the proposed infrastructure. The definition of the scenario and the validation of the Peer-to-Peer models used to design it are however the experimenter's task.

As of this writing the infrastructure has been up and running for one year. Tracker interaction scripts have been added to allow deployment of experiments consisting of multiple trackers. Various BitTorrent clients (hrktorrent, nextshare, swift) have been configured and deployed to provide valuable information regarding performance. Since the initial implementation new scripts have been added for client monitoring and data processing, proving the flexibility of the infrastructure.

Future plans include heavy usage of the infrastructure in various BitTorrent experiments. Bandwidth limitation is currently limited to client features; we aim to identify how this can be migrated to container level – how can one configure upload/download speed limitation for each container. A medium-time goal is "porting" the proposed infrastructure to run on top of Linux Containers (LXC [3]).

ACKNOWLEDGMENTS

This paper is suported from POSCCE project GEEA 226 - SMIS code 2471, which is co-founded through the European Found for Regional Development inside the Operational Sectoral Program "Economical competivity improvement" under contract 51/11.05.2009, and from the Sectoral Operational Programme Human Resources Development 2007-2013 of the Romanian Ministry of Labour, Family and Social Protection through the Financial Agreement POSDRU/6/1.5/S/19.

This work is part of the EU FP7 P2P-Next project [5], aiming to deliver the next generation Peer-to-Peer content delivery platform.

The authors would like to thank Alex Herișanu for providing access to the NCIT cluster systems we have been using throughout our experiments.

REFERENCES

[1] ipoque Internet Studies. http://www.ipoque.com/resources/internet-studies/internet-study-2008_2009, accessed 2010.

[2] Linux Advanced Routing & Traffic Control HOWTO. http://lartc.org/, accessed 2011.

[3] Linux Containers - LXC. http://lxc.sourceforge.net/, accessed 2011.

[4] OpenVZ. http://wiki.openvz.org/, accessed 2011.

[5] P2P-Next. http://www.p2p-next.org/, accessed 2011.

[6] PlanetLab. http://www.planet-lab.org/, accessed 2011.

[7] The Network Simulator – ns-2. http://www.isi.edu/nsnam/ns/, accessed 2011.

[8] trickle. http://monkey.org/ marius/pages/?page=trickle, accessed 2011.

[9] L. Bracciale, F. L. Piccolo, S. Salsano, and D. Luzzi. Simulation of peer-to-peer streaming over large-scale networks using opss. In *ValueTools '07: Proceedings of the 2nd international conference on Performance evaluation methodologies and tools*, pages 1–10, ICST, Brussels, Belgium, Belgium, 2007. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

[10] R. Deaconescu, G. Milescu, B. Aurelian, R. Rughiniș, and N. Țăpuș. A Virtualized Infrastructure for Automated BitTorrent Performance Testing and Evaluation. *International Journal on Advances in Systems and Measurements*, 2(2&3):236–247, 2009.

[11] T. T. A. Dinh, G. Theodoropoulos, and R. Minson. Evaluating large scale distributed simulation of p2p networks. In *DS-RT '08: Proceedings of the 2008 12th IEEE/ACM International Symposium on Distributed Simulation and Real-Time Applications*, pages 51–58, Washington, DC, USA, 2008. IEEE Computer Society.

[12] A. Iosup, P. Garbacki, J. A. Pouwelse, and D. H. Epema. Correlating Topology and Path Characteristics of Overlay Networks and the Internet. October 2005.

[13] S. Naicken, B. Livingston, A. Basu, S. Rodhetbhai, I. Wakeman, and D. Chalmers. The state of peer-to-peer simulators and simulations. *SIGCOMM Comput. Commun. Rev.*, 37(2):95–98, 2007.

[14] J. Pouwelse, P. Garbacki, D. Epema, and H. Sips. The Bittorrent P2P File-Sharing System: Measurements and Analysis. *Peer-to-Peer Systems IV*, pages 205–216, 2005.

[15] A. Rao, A. Legout, and W. Dabbous. Can realistic bittorrent experiments be performed on clusters? In *Peer-to-Peer Computing (P2P), 2010 IEEE Tenth International Conference on*, pages 1 –10, 2010.

[16] S. Sioutas, G. Papaloukopoulos, E. Sakkopoulos, K. Tsichlas, and Y. Manolopoulos. A novel distributed p2p simulator architecture: D-p2p-sim. In *CIKM '09: Proceeding of the 18th ACM conference on Information and knowledge management*, pages 2069–2070, New York, NY, USA, 2009. ACM.

 ISBN: 978-1-61208-133-5

Deploying a High-Performance Context-Aware Peer Classification Engine

Mircea Bardac, George Milescu, and Adina Magda Florea
Automatic Control and Computers Faculty, Computer Science Department
University POLITEHNICA of Bucharest, 313 Splaiul Independentei, Bucharest, Romania
Email: {mircea.bardac, george.milescu, adina.florea}@cs.pub.ro

***Abstract*—This research presents a generic approach for context-aware entity classification with emphasis on integration and use of contextual information in Peer-to-Peer systems. The designed peer classification engine isolates high-latency update processes in order to minimize the latencies of the lookup queries. By using a key-value data-store with support for sorted sets, high complexity context-classifying functions can be executed asynchronously without impacting the lookup queries. The performance of the system is evaluated through experimental and complexity analysis, identifying directions for improving and scaling the peer classification engine. As ubiquitous computing evolves and becomes part of everyday life, the designed context-aware classification engine provides a basis for deploying the next-generation network-based services.**

***Keywords-classification*; *context-awareness*; *peer-to-peer*; *BitTorrent*.**

I. Introduction

Contextual information is becoming more important as ubiquitous computing evolves and integrates into everyday life. An increasing need for context-aware network services has led the research community to investigate means of managing contextual information within the scope of existing network protocols. Whether it is data location, data availability or some other characteristic of the data, the context can play an important part in defining the interaction model and how data is being accessed.

Rapid content distribution and various optimizations have boosted the use of Peer-to-Peer (P2P) solutions. Scientific research has focused on improving existing P2P protocols for delivering better performance and providing higher availability for various network topologies, overlays and swarms.

The Peer Classification Engine presented in this paper provides a generic approach for context-aware classification for any type of entity or data inside a network. Entities can be peers in Peer-to-Peer networks, sensors in Wireless Sensor Networks, computer nodes in a cluster, or any other networked entities. The evaluation of this solution has been performed considering the Peer-to-Peer paradigm, using BitTorrent interactions for contextual evaluation. BitTorrent was chosen as it is the most widespread Peer-to-Peer protocol implementation currently in use, accounting for most of the Internet traffic [1].

The solution is designed to ensure fast responses for *lookup queries*, while asynchronously processing the slow *update queries* in the background.

A. Context-Awareness in P2P systems

In a world where providing differentiated services based on various conditions is becoming important both from the technical and economical points of view, context awareness plays an integral part of designing the next-generation network-based services. Various models such as [2] have been developed to accommodate the needs for contextualized data retrieval in P2P networks. Even so, contextualized information is still difficult to include into testing and deployment platforms such as the ones described in [3] and [4], and has not been yet considered for performance evaluation [5].

Monitoring platforms such as [6] and [7] are the best suited for extracting contextual information but, even so, this information must be used by the service-providing layers.

Existing P2P context-aware applications usually rely on location data [8] as the most easy to determine context. Search also benefits from contextual information [9]. P2P applications running on mobile devices also take into account device capabilities in order to facilitate an adapted service discovery mechanism for the peers [10], or for processing mobile data over large mobile network environments [11].

A system that simplifies context-aware classification is needed for supporting the increased demand of contextualized services. The system must scale to a large number of contexts without affecting the performance of the consumer applications within the P2P network. These are the basic requirements that led to the development of the proposed solution.

B. Storage and performance constraints

One of the problems in classifying peers in a large-scale network with numerous contexts is the underlying data structure used for storing the classification information. Regular databases have been considered as they are commonly used for storing and retrieving information easily. Unfortunately, as the entire system aims for high performance lookup queries and the classification information has a fixed format, the complexity of a relational-database was considered an impediment and other solutions were evaluated.

Structured storage solutions with simple key-value mappings, modeled after the NoSQL paradigm [12], are considered the best choices in terms of design complexity and provided performance. Therefore, the storage does not have fixed table schemas and most operations scale horizontally, making this

 ISBN: 978-1-61208-133-5

suitable for describing large numbers of contexts. An in-memory storage solution brings even a higher read/write throughput. If needed, this solution can scale beyond a single machine.

The structured storage is the core part of the Peer Classification Engine. The peers in the P2P network are the ones providing updates for the Peer Classification Engine, and they are also the ones benefiting from the classified results. Depending on the desired functionalities and supported contexts, updates are being processed by the system, and peers, content and other information are classified into several classes, as described in Section II.

In order to provide a high-performance peer classification solution, the major design goal of the system is to *minimize latencies for lookup queries* while doing most peer classification computations during the updates. Previous designs [13] have shown that the computationally intensive processes should be isolated, and common operations, in this scenario - lookup queries, should be optimized.

This paper is structured as follows: Section I presents an overview of P2P, context-awareness in P2P and the storage and performance requirements that needed to be solved, Section II details the design decisions taken for deploying the Peer Classification Engine, Section III presents a complexity analysis of the underlying operations implemented in the proposed solution. Section IV describes the experimental analysis of the Peer Classification Engine, focusing on system latency, resource consumption and presenting some scalability issues. Section V makes an overview of the planned future work as identified after the initial deployment of the Peer Classification Engine, and Section VI presents the conclusions.

II. Peer Classification Engine Design

This section presents the design decisions that led to the development and analysis of the Peer Classification Engine. Based on the requirements detailed in Section I, the design covers the impact of the type of queries being handled, the software components of the system and the classification and ranking algorithms.

A. Queries

Depending on whether or not a query changes the peer classification within the system, two types of queries have been identified: update and lookup queries. The performance requirements of the system have been elaborated based on these two operations and their underlying effects.

In a BitTorrent-based P2P system as the one being considered in evaluating the Peer Classification Engine, update queries would be similar to updates received from the peers by the tracker entities in the BitTorrent network. Lookup queries would be similar to the scrapes performed by the peers for gathering information about the swarms. These scrapes are used to retrieve the lists of active peers, but, in a more complex system, a monitoring entity could use the lookup queries to retrieve much more context-rich information, such as the content availability information within a certain geographical region.

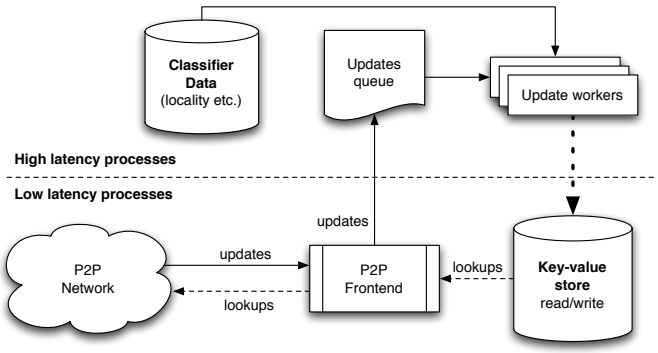


Fig. 1. Peer Classification Engine Design

1) Updates: An *update* coming from a peer pushes a new state to the Peer Classification Engine. The new state may contain various pieces of information about the peer such as its availability (whether it has become online or offline) or content availability (which content is provided by the peer, what parts of the content are provided, etc.). In BitTorrent, *announce messages* are being used by the peers to push updated state information such as number of uploaded/downloaded/corrupt bytes, whether a download has been completed or not, etc.

In the Peer Classification Engine, the new state is used to update the classes the peer belongs to, a high-latency operation by design, given the requirements of the system.

2) Lookups: Queries that retrieve information from the Peer Classification Engine are considered *lookups*. The engine is designed to provide low-latency lookup replies by having the peer classification pre-computed during the updates. In BitTorrent, the *scrape messages* that are used to retrieve the list of peers providing a specific content only contain the hash of the data (torrent) being downloaded.

In the proposed solution, lookup queries can also aggregate information across multiple classes using different weights in order to provide a multi-contextual response for a peer, as detailed in Section III - Complexity Analysis.

B. Components

As detailed in Figure 1, the classification engine relies on the existence of an in-memory *key-value data-store* for managing the peer classes. As previously presented in Section I, a relational database does not meet the performance needs required for high-performance lookup operations. Cassandra [14] has been initially considered as a structured data solution but its complexity added too much overhead to the system. In the end, *Redis* [15] was chosen as a storage solution.

Update queries are being received directly from the peers by the *P2P frontend*, and pushed for later processing to an *updates queue*. This allows asynchronous processing of the updates by one or more *update workers*, which consume the updates from the updates queue. This also implies that, once an update arrives, its effect might not be initially visible in the lookup results. Depending on the number of update workers and on the complexity of the update operations, the time required for the update to reach the key-value store may vary. Lookup queries

 ISBN: 978-1-61208-133-5

are also received through the *P2P frontend*, but results are retrieved directly from the key-value data-store.

Figure 1 also presents the separation between the low-latency processes and the high-latency processes. This separation is given by the design requirements as detailed in Section I. Most classifying processes are considered high-latency and therefore require the use of a queue, while the lookup queries are low-latency and should not suffer any performance penalty from other computations except for the data-store reads.

C. Classes and Ranking

The entire functionality of the Peer Classification Engine relies on identifying classes associated with various information coming from the peers inside the P2P network. A class is associated with a context (such as being in a certain geographical region). This association between information and a certain class is done using special functions, called class association functions (CAFs). A class association function ($f_{context\_type}(data\_packet)$) characterizes only *one type of context* (such as locality, or data availability) – for multiple types of contexts, multiple functions are being used.

In order to provide the best-suited replies for specific contexts, a class association function can also return a rank of a specific piece of information within a class. A class association function will therefore return:

- a *class*, which characterizes the information (for example: the geographical region of a peer);
- an *ID* to identify the information being classified (for example: the peer ID);
- a *rank* of the information in the class (for example: the uplink bandwidth of the peer).

The returned *ID* and *rank* are optional, as there are situations where ranking is not necessary, such as for lookup queries. Classification can be used both for the update and for the lookup queries, but with different purposes:

- *update queries* contain information on various changes in a peer's state. These changes affect the peer's classification within several classes, and several CAFs can be used. Classification within a class may require ranking, depending on the context described by the class;
- *lookup queries* may be classified in order to identify which context is addressed within the query (for example: which region a query is related to or which data is being looked for).

As the design of the proposed solution requires providing a low-latency lookup response, the class association functions used on lookup queries must be fast and take preferably constant time, extracting the class information directly from the lookup query without using external data sources.

Table I presents an overview of the class association functions and their return values depending on the type of a query being processed.

Typical applications for using classes include:

1) selecting which peers are the most suited for serving a requesting peer based on the context of the requesting peer and the ranking of the other peers for that specific context;
2) selecting which peers are the most suited for providing data to other peers;
3) determining if peer queries should be directed to other storage shards, as described in Section IV, Subsection Scalability Analysis.

TABLE I
CLASS ASSOCIATION FUNCTIONS FORMS

Query type	Class Association Function form
update	$f_{context}(update\_message) = (class, ID, rank)$
lookup	$f_{context}(lookup\_message) = (class)$

Several class association functions as seen in Table II have been defined in order to evaluate the proposed solution for BitTorrent P2P systems. They have been listed based on the types of queries being processed.

TABLE II
CLASS ASSOCIATION FUNCTIONS FOR BITTORRENT P2P QUERIES

Update Queries
$locality\_update(update) = (locality\_class, peer\_ID, rank)$
$availability(update) = (t\_availability\_class, peer\_ID, rank)$
Lookup Queries
$locality\_lookup(lookup) = (locality\_class)$

A $locality\_class$ can be specific to a certain geographic region (one class per region), and multiple locality functions can be used, each one considering a different region size for example - the more functions are used, the longer updating the classes will take. Dynamically choosing the number of classes and size of the classes is a future goal, as described in Section V. The two $locality$ functions differ by their output: the $locality\_lookup$ function does not have to compute a rank; it only has to return the class of the peer issuing the query, so that peers could be examined within that certain class.

The $t\_availability\_class$ (torrent availability class) is torrent specific and contains all the peers that share the specific content identified by the torrent. The $rank$ is given by the completion status (how much of the torrent has been downloaded by $peer\_ID$, in percent).

In order to minimize the latency for lookup replies, the results of the $locality\_lookup$ function applied on the lookup queries can be cached, thus reducing even more the overhead of the functions and relying only on the deterministic behavior of the data-store operations.

III. COMPLEXITY ANALYSIS

The maximum performance of the provided solution is lower bounded by the theoretical limit of the underlying components, mainly the key-value store and the update workers.

The Peer Classification Engine relies on the use of *sorted sets* in Redis for storing information in classes together with the ranking. One sorted set is used for each class. Using normal

 ISBN: 978-1-61208-133-5

sets is also an option, providing a better complexity but no ranking: adding an item to a normal set is $O(1)$, retrieving the members of a set is $O(n)$, intersecting sets is $O(n \cdot m)$.

The theoretical complexity boundaries achievable using the proposed underlying storage solution are presented below. They are grouped by the type of operations being performed, and they take into account the operations needed for accessing the data structures in the key-value store.

A. Update operations

As mentioned in Section I, most computations should be done on updates in order to minimize the lookup times. Classes are meant to be populated after computing the ranking of peers and the class for each peer. Adding an entry to a sorted set is being done with the ZADD Redis command.

In a BitTorrent P2P network, given the class association functions above, processing an update would require the following operations:

1) identifying the $locality\_class$ of a peer - very costly in terms of complexity and time consumed, it requires querying a MySQL database with GeoIP information
2) ranking the peer for the identified $locality\_class$; if the rank is considered to be upload bandwidth and the peer reports it, this can be ignored as complexity
3) adding the peer to the $locality\_class$ with the given $rank$ in the key-value store (in the sorted set associated with the $locality\_class$): $O(log(n))$ complexity, where n is the size of the set.
4) calculating the $rank$ for the peer in the torrent availability class can be ignored as complexity, if it is calculated from the values received from the peer
5) adding the peer to the $t\_availability\_class$ with the previously calculated $rank$ in the key-value store (in the sorted set associated with the $t\_availability\_class$): $O(log(n))$, where n is the size of the set.

B. Lookup operations

Lookup queries can be classified as presented in Table II. This means that, for a given lookup query, the location of the peer must be identified. This can be a costly operation, and, in order to provide a low-latency lookup reply as intended in the design, a default class can be assigned on the first query, and the resulting value of the class association function can be cached to be used on subsequent lookup queries.

Lookup replies can be formed by taking into account one context or multiple contexts. This is equivalent to verifying one or more classes in the Peer Classification Engine, which results in retrieving data from one sorted set, or by intersecting multiple sets. The computation becomes more complex as the number of sets being used increases. This might look expensive on the first lookup, but sequential lookups in the same classes can be fed from a cache with $O(1)$ lookup complexity, if nothing changes in those classes.

Lookup into one single class using the ZREVRANGE Redis command for retrieving the best m items (in a sorted set with n items) is $O(log(n)) + O(m)$ complexity.

Intersecting multiple classes using the ZINTERSTORE Redis command is $O(n \cdot k) + O(m \cdot log(m))$ complexity, with n being size of the smallest input sorted set, k being the number of input sorted sets, and m being the number of elements in the resulting sorted set.

IV. Experimental analysis

A. Peer Classification Engine Deployment

The implementation of the context classes with ranks using sorted sets is partially inspired by the data-structures supported by Redis. Redis is not a simple key-value store but also has support for storing/updating/deleting lists, sets, sorted sets, hashes with atomic operations providing low complexity. Atomic operations and support for transactions also allow keeping data consistent between updates and lookups.

Redis version 2.2.0-rc3 has been compiled and installed on Core(TM)2 Quad Q9550 @ 2.83GHz system with 1GB of reserved RAM in order to perform an experimental evaluation of the Peer Classification Engine. The Redis store was filled with torrent information as received from a tracker and passed through the Peer Classification Engine. Measurements were taken for determining the slow and fast paths within the system, ensuring that the design requirements are being met.

B. Experimental Results

The experiments concentrated on evaluating the following performance metrics in order to determine the impact of the solution on both the system and on the context-aware applications being deployed using the system:

- *update query times*, considering the impact of the location classifying functions;
- *lookup query times*, for single and multi class lookups;
- *memory usage*, as it can bring performance limitations with the extensive use of contexts or in systems with many participating peers.

The measurements were realized considering systems with 256, 512, 768 and 1024 active torrents, and 25 locality contexts. As there is one *torrent availability class* for each torrent and one *locality class* for each locality context, the experiments tested the use of the system for $x + 25$ classes, where x is the number of torrents in the system.

As determined from the complexity analysis, most operations are impacted by the number of peers inside a class at a certain time. Therefore, the experiments tested the system with 128, 256, 512, 1024 and 2048 peers for each torrent. Figure 2 shows that a peak memory use of about 500 MB of RAM is reached for the 1024 torrents with 2048 peers each.

Figure 3 shows the time required for an update to be performed in the system, taking into account the time spent determining the location (identifying the locality context) using a GeoIP database. The average time spent for retrieving the geolocation information is 13.75 msec, making the time spent writing the update to the key-value store insignificant. Under these conditions, the times of an update query do not vary much depending on the number of torrents or peers for a torrent, being 13.88 msec on average.

 ISBN: 978-1-61208-133-5

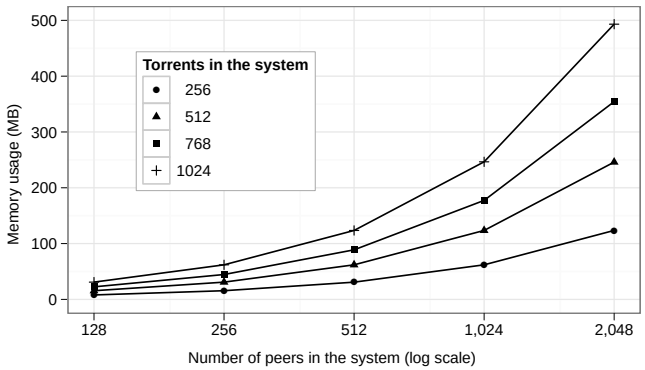


Fig. 2. Memory usage of the key-value data-store

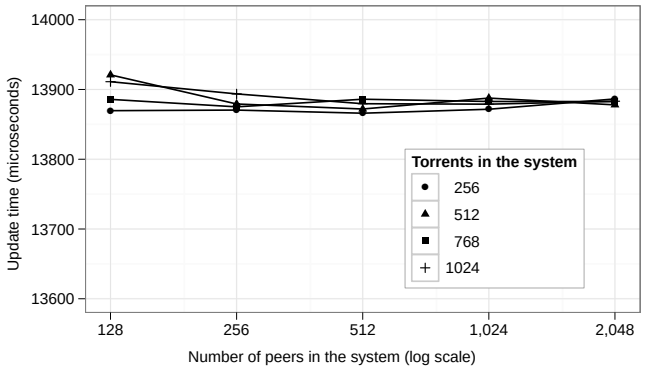


Fig. 3. Update query time

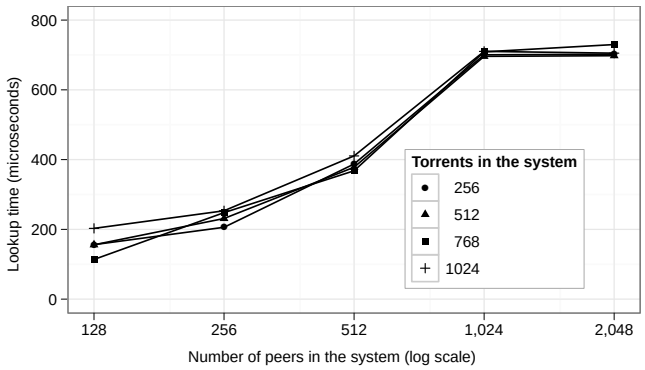


Fig. 4. Lookup query time for one single class

The complexity analysis shows that lookup times are directly impacted by the size of the classes where information is being looked for. As lookup operations are more frequent and results are most likely to be cached, the experimental analysis concentrated on measuring the lookup times only for fetching updated information from the key-value store. Figure 4 shows the lookup times using one single class (checking one single context). As expected from the complexity analysis, this shows that the lookup time increases as the number of peers inside a class grows. Multiple experiment runs have shown that lookup times for the 1024 and 2048 peers are similar.

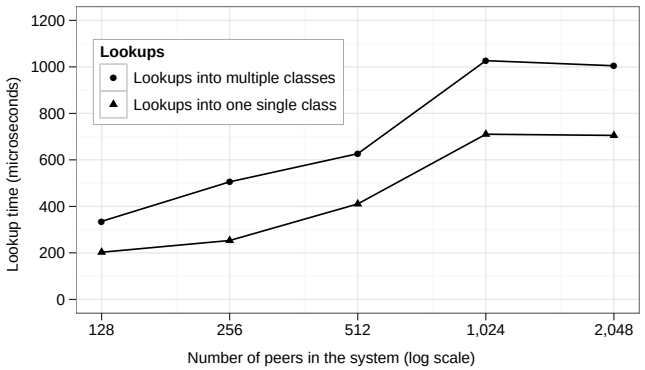


Fig. 5. Lookup query time in the 1024 torrents tests

Redis has internal data-size thresholds used for triggering data reorganization, and the elevated lookup times in the 1024-peers experiments might appear after reaching these thresholds.

Multi-contextual lookups require looking for data into multiple classes, performing intersections between the sorted sets at the level of the key-value store. The time spent looking for data into multiple classes presented a similar trend, although with an added overhead. Figure 5 shows the difference between lookup queries performed into one class versus lookup queries performed into multiple classes (for the 1024 torrent tests). In the experiment, the locality and the torrent availability class were considered. As noticed, intersecting multiple classes adds a significant overhead to the lookup queries, but the impact grows slower as the overhead is limited by the number of results retrieved from the intersection and the size of the smallest input sorted set (which, in this situation, was the locality class set with 25 entries).

The lookup query times ranged from 113 μsec to 1138 μsec, while the update query times averaged at 13883 μsec.

C. Scalability analysis

1) Data sharding: Depending on the content being monitored, number of peers, number of contexts (classes) and other information that can be stored in the in-memory key-value data-store, the Peer Classification Engine might run out of physical memory, and using swap is not an option for a low-latency service. Data sharding can overcome this problem, at the expense of using more machines.

A Redis-based key-value data-storage requires extra middleware logic to implement sharding. Depending on the functions describing the contexts, various data-clustering algorithms [16] can be applied to identify which data needs to be placed inside the same shard. Once the clusters have been identified, data sharding can be used as a solution for dividing data across different machines or even different locations.

In a BitTorrent-based P2P network, peers would make the most use of having data sharded by:

- *torrent hash* - this is the easiest and the most practical method, as content is identified using a hash, and most lookups for information are performed on specific torrent hashes;

 ISBN: 978-1-61208-133-5

- *locality information* - this might bring lower latencies for the lookup replies for local-peers, but brings higher complexity for creating replies for the queries coming from peers which are non-local.

2) Data replication: Most queries on the Peer Classification Engine are expected to be lookups. An increased number of lookup requests or updates in the P2P swarm might increase load on the system, and therefore the latency on the lookup-reply path might become unacceptable. In order to reduce this latency, as reads are very easy to scale, multiple read-only replica storages and/or P2P front-ends can be deployed.

V. Future work

The designed system takes into account that class association functions return a predefined set of classes. Therefore, the number of contexts available in the system can be easily predicted. In continuously evolving systems with a variable number of peers, the use of a fixed number of contexts might lead to uneven resource allocation. A system with 20 peers can benefit more by using a locality association function returning 2 contexts (in this case geographical regions), compared to a 2000-peers system where using 2 locality contexts might lead to an inefficient use of the locality information.

With the given system architecture, *on-the-fly peer-reclassification* can be implemented as an asynchronous high-latency process. The lookup queries will not be affected by an on-going peer reclassification, and, on completion, the new classification might be put in effect. The same behavior can be implemented for sharding. When the load on the system goes over a predefined threshold, an *on-the-fly class-resharing* process can be executed in the background to (1) determine which are the classes (contexts) that can be moved to other data shards, and (2) perform the changes in the data-store.

The key-value store also has support for expiring entries, making it easy to have *contexts that automatically expire* if not updated. This allows outdated information to be automatically removed. Other cleanup operations can be implemented as high-latency processes, without impacting the lookup queries.

VI. Conclusions

The research presented in this paper focuses on designing, implementing and evaluating a Peer Classification Engine that allows contextual information to be used by the peers in a P2P network. The solution provides a flexible approach for defining contexts through class association functions that also support ranking the entities within classes.

The design takes into account that the update and lookup queries impact differently the performance of the entire system. The solution minimizes lookup latencies by offloading the high-latency computations to the update-processing phase. This isolation ensures a predictable behavior of the classification engine as it is being used by the entities in the network.

The solution is not only designed for high-performance lookup queries, but can also provide *high-availability* contextual services through data replication and data sharding. These developments have been explored during the experiments, and, in order to provide an adaptive contextual scaling of the system, dynamic context-based content resharding and on-the-fly peer reclassification are being considered as future work.

Moreover, even though the contexts chosen for evaluating the Peer Classification Engine are Peer-to-Peer specific, the solution can be easily deployed in multiple other environments such as Wireless Sensor Networks, computer nodes in a cluster, or any other networked entities. The proposed solution simplifies the integration and use of contextual information and provides a basis for deploying the next-generation network-based services.

Acknowledgment

This research was supported by CNCSIS - UEFISCSU, project number PNII - IDEI 1315/2008, and by the Sectoral Operational Programme Human Resources Development 2007-2013 of the Romanian Ministry of Labour, Family and Social Protection through the Financial Agreement POSDRU/6/1.5/S/19.

References

[1] ipoque GmbH, "The Impact of P2P File Sharing, Voice over IP, Instant Messaging, One-Click Hosting and Media Streaming on the Internet," accessed in March 2011. [Online]. Available: http://www.ipoque.com/resources/internet-studies/internet-study-2008_2009

[2] M. Bardac, G. Milescu, and R. Rughinis, "A Distributed File System Model for Shared, Dynamic and Aggregated Data," *17th International Conference on Control Systems and Computer Science (CSCS17)*, vol. 1, pp. 45–51, 2009.

[3] R. Deaconescu, G. Milescu, B. Aurelian, R. Rughinis, and N. Tapus, "A Virtualized Infrastructure for Automated BitTorrent Performance Testing and Evaluation," *Internation Journal on Advances in Systems and Measurements*, vol. 2, pp. 236–247, 2009.

[4] M. Bardac, R. Deaconescu, and A. M. Florea, "Scaling Peer-to-Peer Testing using Linux Containers," in *Proceedings of the 9th RoEduNet IEEE International Conference*, 2010, pp. 287–292.

[5] G. Milescu, M. Bardac, and N. Tapus, "Swarm metrics in peer-to-peer systems," in *Proceedings of the 9th RoEduNet IEEE International Conference*, 2010, pp. 276–281.

[6] M. Bardac, G. Milescu, and R. Deaconescu, "Monitoring a BitTorrent Tracker for Peer-to-Peer System Analysis," in *Intelligent Distributed Computing*, 2009, pp. 203–208.

[7] R. Deaconescu, M. Sandu-Popa, A. Draghici, and N. Tapus, "Using Enhanced Logging for BitTorrent Swarm Analysis," in *Proceedings of the 9th RoEduNet IEEE International Conference*, 2010, pp. 52–65.

[8] K. Harumoto, S. Fukumura, S. Shimojo, and S. Nishio, "A Location-Based Peer-to-Peer Network for Context-Aware Services in a Ubiquitous Environment," pp. 208–211, 2005.

[9] W. Thiengkunakrit, S. Kamolphiwong, T. Kamolphiwong, and S. Saewong, "Enhanced Context Searching Based on Structured P2P," pp. 309–313, April 2010.

[10] C. Doulkeridis, V. Zafeiris, and M. Vazirgiannis, "The role of caching and context-awareness in P2P service discovery," *International Conference On Mobile Data Management*, pp. 142–146, 2005.

[11] K. F. Yeung, Y. Yang, and D. Ndzi, "A Context-Aware System for Mobile Data Sharing in Hybrid P2P Environment," pp. 63–68, 2009.

[12] M. Stonebraker, "SQL databases v. NoSQL databases," *Communications of the ACM*, vol. 53, no. 4, pp. 10–11, Apr. 2010.

[13] A. Thusoo, Z. Shao, S. Anthony, D. Borthakur, N. Jain, J. Sen Sarma, R. Murthy, and H. Liu, "Data Warehousing and Analytics Infrastructure at Facebook," pp. 1013–1020, 2010.

[14] A. Lakshman and P. Malik, "Cassandra: a decentralized structured storage system," *ACM SIGOPS Operating Systems Review*, vol. 44, no. 2, pp. 35–40, Apr. 2010.

[15] Redis Community, *Redis Documentation*, accessed in March 2011, http://redis.io/documentation.

[16] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," *ACM Computing Surveys*, vol. 31, no. 3, pp. 264–323, Sep. 1999.

 ISBN: 978-1-61208-133-5

Parallel Measurement Method of System Information for 3GPP LTE Femtocell

Choong-Hee Lee
School of Electrical and
Computer Engineering
Ajou University
San 5 Wonchon-dong,
Yeongtong-Gu, Suwon, Korea
Email: hedreams@ajou.ac.kr

Jae-Hyun Kim
School of Electrical and
Computer Engineering
Ajou University
San 5 Wonchon-dong,
Yeongtong-Gu, Suwon, Korea
Email: jkim@ajou.ac.kr

***Abstract*—In the 3rd Generation Partnership Project (3GPP) Long Term Evolution (LTE) system, User Equipments (UEs) have to measure system information of target cells, if the target cells are Closed Subscriber Group (CSG) cells. Home evolved NodeB (HeNB) in the LTE femtocell is CSG cell. Femtocell is one of promising cellular network technologies to enhance both of cell coverage and capacity. This paper introduces six measurement methods of system information in the 3GPP LTE system. And, we also analyzed performance of the proposed measurement methods in aspect of service interruption time and measurement delay. We find out that the autonomous and parallel method shows the smallest measurement delay and the methods which use small gaps have small service interruption time.**

***Keywords*-LTE; Femtocell; System Information Measurement**

I. INTRODUCTION

In 3rd Generation Partnership Project, Long Term Evolution release 8 standardization has been finalized [1]. And also, release 9 standard is on the final step. LTE release 10, in other words LTE-Advanced is being discussed actively. One of the key features of LTE/LTE-Advanced system is heterogeneous deployment. Therefore, femto technology is adopted to expand cell coverage and enhance cell capacity, in the 3GPP LTE release 8. HeNB is femtocell eNB of 3GPP LTE network. HeNBs operate only CSG mode in the release 8. Therefore, only authorized users can access the CSG HeNBs. In the LTE release 9, HeNB is extended to be able to operate in open mode. HeNB can operate in hybrid mode in the LTE-Advanced standard. Hybrid mode HeNB operates not only as CSG HeNB, but also as an open HeNB. Therefore, it provides limited service even though the user who tries to access is not a member of its CSG.

UE does not need to measure system information from target cell in general LTE handover procedure [2], [3], [4]. CSG inbound handover technology is essential to support femto technology in LTE/LTE-Advanced system. And, CSG inbound handover requires system information measurement from target HeNBs instead of getting the information from serving eNB. However, there is not detailed description for it in the 3GPP technical specifications. Therefore, we propose four methods to measure the system information of CSG femtocell in 3GPP LTE/LTE-advanced system.

The remainder of the paper is organized as follows. Section II introduces the backgrounds about System Information in 3GPP LTE system. Section III introduces and analyzes System Information Measurement methods which are proposed in [5] . Section IV proposes the Autonomous Measurement with Parallel Small Gaps method. Section V shows the evaluated performance comparison results and section VI concludes this paper.

II. SYSTEM INFORMATION IN 3GPP LTE SYSTEM

The System Information is exchanged in a type of System Information Blocks (SIBs) between eNBs and UEs. SIB messages are kinds of Radio Resource Control (RRC) messages [2]. SIB messages include Master Information Block (MIB), System Information Block Type 1 (SIB1) and System Information message. MIB message contains essential information such as downlink bandwidth, Hybrid Automatic Retransmission Request (HARQ) channel configuration and System Frame Number (SFN) information. SIB1 message contains information related to cell selection and scheduling information of SIB2 - SIB13 [3].

SIBs except SIB1 are not a RRC message but message elements that are carried by SI message. SIB2 is composed of configuration information of common and shared channels. SIB3 contains cell reselection information, mainly related to the serving cell. SIB4 contains information about the serving frequency and intra-frequency neighboring cells relevant for cell reselection including common parameters for a frequency as well as cell specific reselection parameters. SIB5 contains information about other Evolved Universal Terrestrial Radio Access (E-UTRA) frequencies and inter-frequency neighboring cells relevant for cell reselection. SIB6 contains information about UTRA frequencies used in Universal Mobile Telecommunications System (UMTS), and UTRA neighboring cells relevant for cell reselection. SIB7 contains information about GSM/EDGE Radio Access Network (GERAN) frequencies relevant for cell reselection. SIB8 contains information about CDMA2000 frequencies and CDMA2000 neighboring cells relevant for cell reselection.

 ISBN: 978-1-61208-133-5

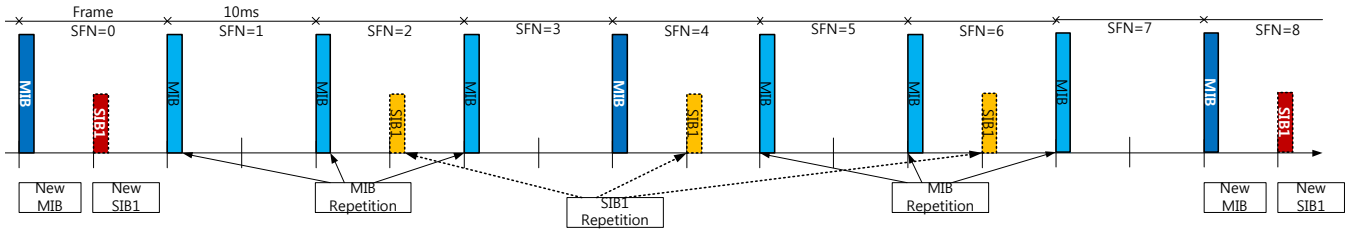


Fig. 1. System Information broadcasting scheduling in time domain

SIB9 contains a home eNB identifier (HNBID). SIB10 contains an Earthquake and Tsunami Warning System (ETWS) primary notification. SIB11 contains an ETWS secondary notification. SIB12 contains a Commercial Mobile Alert System (CMAS) warning notification. SIB13 contains information related to Multimedia Broadcast and Multicast Service (MBMS).

Downlink channel of the 3GPP LTE system uses Orthogonal Frequency Division Multiple Access (OFDMA). Therefore, resources are divided into time and frequency axis. Measurement of MIB and SIB1 is essential steps before handover decision to CSG femtocell. MIB and SIB1 messages are broadcasted from eNBs periodically. System information broadcasting scheduling in time domain is shown in Figure 1. Length of a subframe is 1ms. And, 10 subframes compose a frame. Therefore, length of a frame is 10ms. MIB message packets are generated for every 4 frames (40ms) and replicas are transmitted for every frame (10ms). The packet generation time and transmission duration of SIB1 are double of those of MIB, respectively. Therefore, SIB1 packets are generated for every 8 frames and its replicas are transmitted for every 2 frames. Another system information might be scheduled aperiodically or use other transmission duration. The scheduling information of SIBs in the system information messages is contained in SIB1 messages [6].

III. System Information Measurement in 3GPP LTE System

UE measures system information from neighboring cell after neighboring cell detection. UE has to disconnect the link with serving eNB to measure system information from target eNB. The service interruption time for measurement is called “Measurement Gap”. System information measurement could be performed autonomously or by scheduling from serving eNB. In the autonomous method shown in Figure 2, UE determines the measurement gap by itself. During the measurement gap, packet drop can be occurred since serving eNB cannot know whether the UE is disconnected or not. In the scheduled method shown in Figure 3, UE requests for measurement gap to serving eNB and serving eNB allocate measurement gaps to the UE. Therefore, packet drop does not occur in scheduled methods.

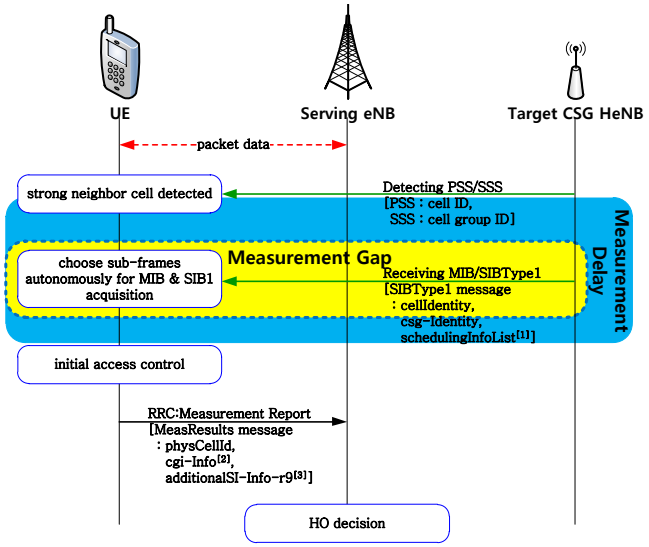


Fig. 2. Message flow of the autonomous measurement with a large gap method

A. Autonomous Measurement with a Large Gap

The Autonomous Measurement with a Large Gap (AMLG) method is shown in Figure 2. In this method, a UE disconnect with serving eNB when candidate neighboring eNBs are detected, and measures MIBs and SIB1s . The UE reconnect with serving eNB after successive MIB and SIB1 measurement. The measurement gap and the measurement delay are given by

$$T_{gap} = T_{MIB+SIB1} \tag{1}$$

and

$$\begin{aligned} T_{delay} &= T_{MIB+SIB1} + T_{RRC} + \cdots \\ &= 6\left(T_{MIB+SIB1} + T_{RRC}\right), \end{aligned} \tag{2}$$

respectively. The measurement delay is the duration from cell detection to measurement completion. And, It is represented as dark (blue) area in Figure 2. The measurement gap is represented dotted (yellow) area. $T_{MIB+SIB1}$ is the service interruption time to receive both of MIB and SIB1 packets of target CSG HeNB at once. $T_{MIB+SIB1}$ is 25ms in the worst case, because maximum time for MIB is 10ms and maximum distance between MIB and SIB1 in time domain is 15ms. And T_{RRC} is the time during a RRC message is transmitted. We assume that T_{RRC} is about 10ms because it is the length of a frame. Figure 4 shows the timing diagram of AMLG.

 ISBN: 978-1-61208-133-5

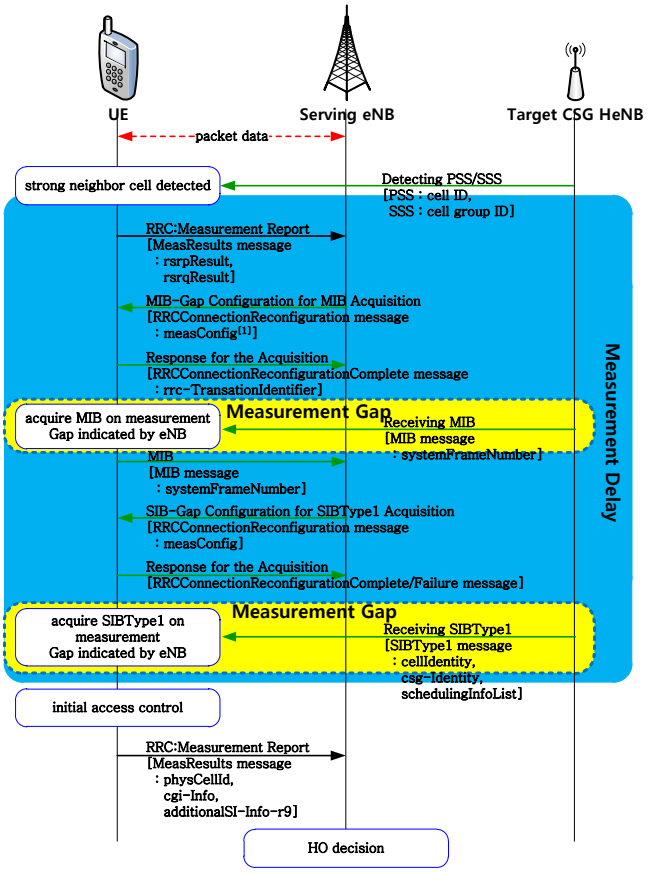


Fig. 3. Message flow of the scheduled measurement with sevaral small gap

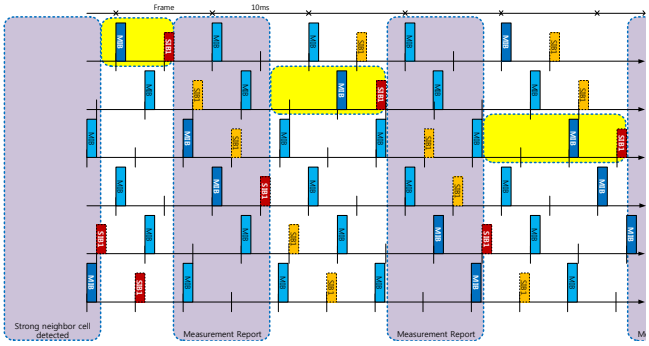


Fig. 4. Measurement Gaps of the Autonomous Measurement with a Large Gap Method

B. Scheduled Measurement with a Large Gap

Figure 5 shows the Scheduled Measurement with a Large Gap (SMLG) method. In the SMLG method, UE request for measurement gap to serving eNB. Then, the serving eNB allocate a large measurement gap to the UE. Therefore, SMLG need more measurement delay than AMLG to exchange scheduling messages. The measurement gap and delay are

$$T_{gap} = T_{MIB+SIB1} \tag{3}$$

and

$$\begin{aligned} T_{delay} &= T_{MIB+SIB1} + 3 \cdot T_{RRC} + T_{MIB+SIB1} \\ &\quad + 3 \cdot T_{RRC} + \cdots + T_{MIB+SIB1} + T_{RRC} \\ &= 6 \cdot T_{MIB+SIB1} + 19 \cdot T_{RRC}. \end{aligned} \tag{4}$$

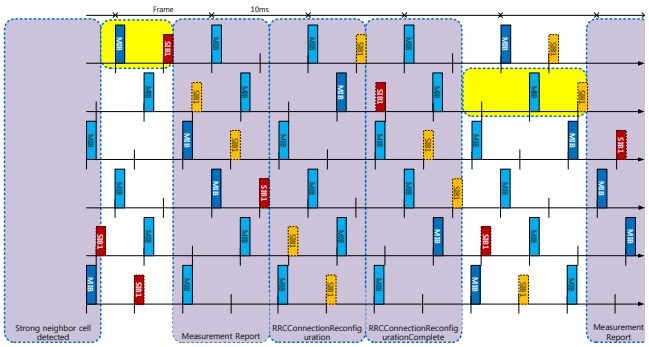


Fig. 5. Measurement Gaps of the Scheduled Measurement with a Large Gap Method

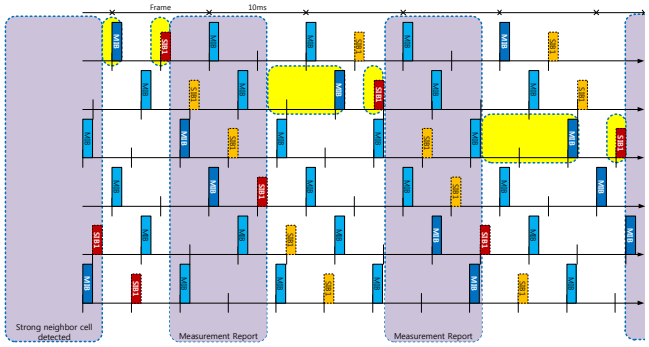


Fig. 6. Measurement Gap of the Autonomous Measurement with several Small Gaps

UE transmits measurement result message to its serving eNB to be assigned a measurement gap. Then, the serving eNB send back RRC connection reconfiguration message to the UE. Finally, the UE acknowledge with RRC connection reconfiguration complete message.

C. Autonomous Measurement with several Small Gaps

Measurement gap of the Autonomous Measurement with several Small Gaps (AMSG) method is not a single large gap, but two or more small gaps shown in Figure 6. UE interrupts connection with its serving eNB until successive

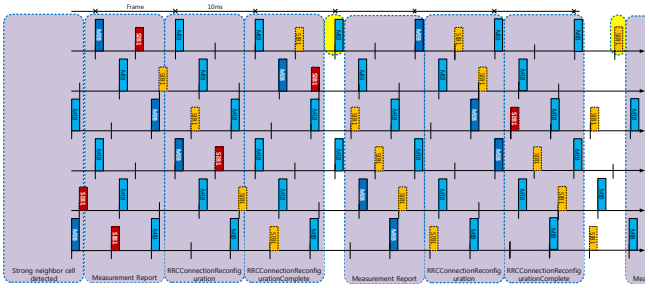


Fig. 7. Measurement Gap of the Scheduled Measurement with several Small Gaps method I

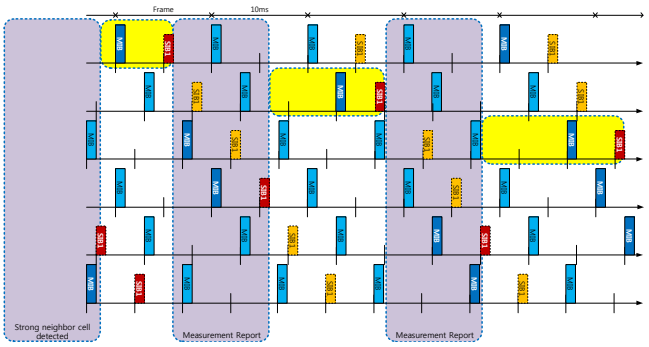


Fig. 8. Measurement Gap of the Scheduled Measurement with several Small Gaps method II

measurement of MIB after detect strong neighboring cell. The UE can predict the time when the target cell transmit a SIB1, because the UE get SFN information from MIB measurement. Therefore, the UE need only one more tiny measurement gap to measure SIB1. The measurement gap and delay are

$$T_{gap} = T_{MIB} \text{ or } T_{SIB1}, \tag{5}$$

and

$$\begin{aligned} T_{delay} &= T_{MIB} + W_{SIB1} + T_{SIB1} + T_{RRC} \\ &\quad + T_{MIB} + W_{SIB1} + T_{SIB1} + T_{RRC} + \cdots \\ &= 6\left(T_{MIB} + W_{SIB1} + T_{SIB1} + T_{RRC}\right) \end{aligned} \tag{6}$$

where T_{MIB} is time to measure MIB and T_{SIB1} is time to measure SIB1. T_{MIB} is less than 10ms and T_{SIB1} is about 1ms without channel error. W_{SIB1} is waiting time until the SIB1 packet is transmitted.

D. Scheduled Measurement with several Small Gaps I

The Scheduled Measurement with several Small Gaps I (SMSG1) method is almost same with AMSG method except the part of exchanging RRC messages to schedule the measurement gaps. The flow chart and timing diagram are shown in Figure 3 and Figure 7, respectively. The measurement gap and delay are given by

$$T_{gap} = T_{MIB} \text{ or } T_{SIB1}, \tag{7}$$

and

$$\begin{aligned} T_{delay} &= 3 \cdot T_{RRC} + T_{MIB} + 3 \cdot T_{RRC} + W_{SIB1} \\ &\quad + T_{SIB1} + 3 \cdot T_{RRC} + \cdots + W_{SIB1} \\ &\quad + T_{SIB1} + T_{RRC} \\ &= 6\left(T_{MIB} + W_{SIB1} + T_{SIB1}\right) + 37 \cdot T_{RRC} \end{aligned} \tag{8}$$

respectively. The SMSG1 method naturally has hybrid characteristic of scheduled methods and methods those use several small gaps.

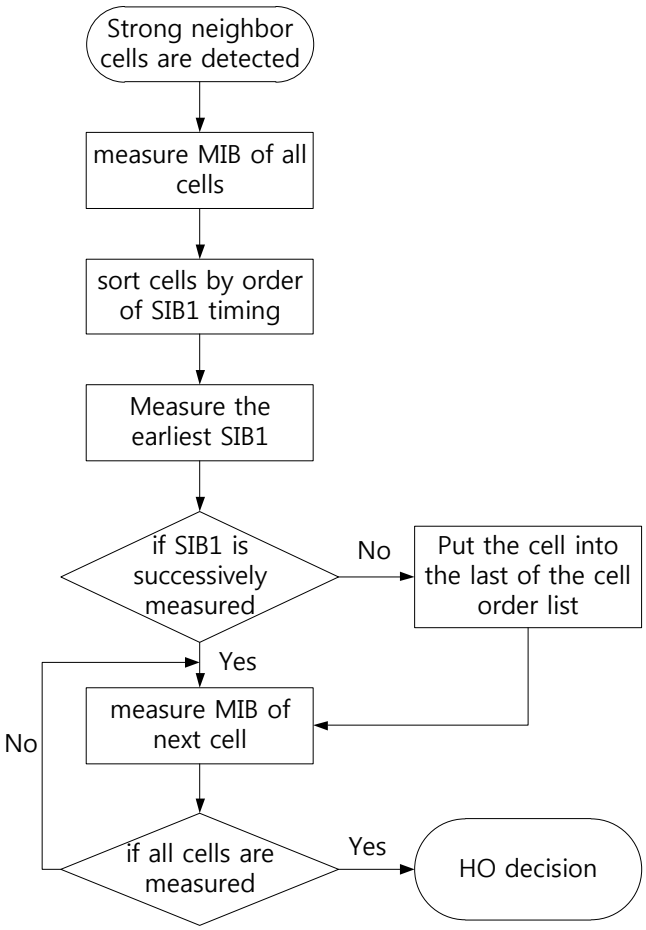


Fig. 9. Flow chart of the Autonomous Measurement with Parallel Small Gaps method

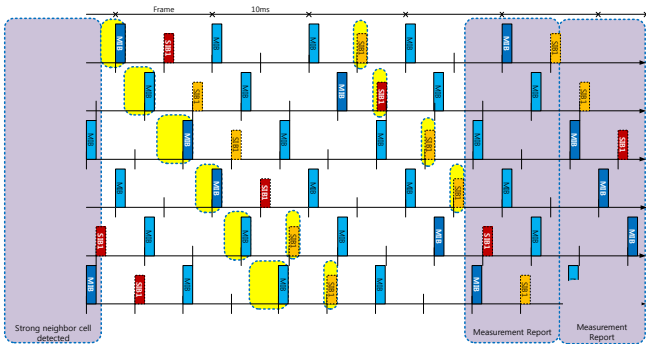


Fig. 10. Measurement Gap of the Autonomous Measurement with Parallel Small Gaps method

E. Scheduled Measurement with several Small Gaps II

The SMSG2 method is hybrid method. UE which uses this method, request measurement gap for MIB to serving eNB. After the UE measure MIB, the UE measure SIB1 autonomously, because the UE knows the timeslot when the SIB1 packet will be transmit and the serving eNB knows the maximum measurement gap for a single cell is 25ms. The timing diagram are shown in Figure 8. The measurement delay

 ISBN: 978-1-61208-133-5

and gap are given by

$$T_{gap} = T_{MIB} \text{ or } T_{SIB1}, \quad (9)$$

and

$$\begin{aligned} T_{delay} &= 3 \cdot T_{RRC} + T_{MIB} + W_{SIB1} + T_{SIB1} \\ &\quad +3 \cdot T_{RRC} + \cdots + W_{SIB1} + T_{SIB1} + T_{RRC} \\ &= 6\,(T_{MIB} + W_{SIB1} + T_{SIB1}) + 19 \cdot T_{RRC} \quad (10) \end{aligned}$$

respectively.

IV. Autonomous Measurement with Parallel Small Gaps

We propose the Autonomous Measurement with Parallel Small Gaps (AMPSG) method to reduce measurement delay. All methods which are proposed in section III, are serial methods. 'Serial methods' means that the UE measures system informations of target cells cell-by-cell in order of Reference Signal Received Power (RSRP). First, the UE which operates with AMPSG method, measures MIB packets of target cells cell-by-cell in order of RSRP. Then, the UE sorts the target cells in order of distance in time domain. Second, the UE measures the earliest SIB1 packet of target cells and next SIB1 packets until all the SIB1 packets of target cells are received. Figure 9 is the flowchart of the proposed AMPSG method. The measurement delay and gap are given by

$$T_{gap} = T_{MIB} \text{ or } T_{SIB1}, \quad (11)$$

and

$$T_{meas} = 6 \cdot T_{MIB} + 6 \cdot T_{SIB1} + T_{RRC} \quad (12)$$

respectively.

V. Performance Evaluation

In this section, we present the results of mathematical analysis and simulation. In the mathematical analysis, we analyze the maximum measurement delay of measurement methods in worst case. The equations which are used in mathematical analysis, are equation 1-12. In the simulation we use OPNET simulation tool and the parameters which are used, are presented in Table I. And, the network deployment in the simulation is shown in Figure 11. The UE perform system information measurements of 6 target HeNB in a simulation execution. We repeated simulation 200 times with different random seed value for each measurement method.

Figure 12 and 13 show numerical results of mathematical analysis and simulation. In the graph, the x axis represents four methods those are proposed in this paper, and the y axis represents time in the unit of milliseconds. In the mathematical analysis, the maximum service interruption time of both the AMLG and the SMLG methods is about 25ms, while that of the AMSG, SMSG1, SMSG2 and the AMPSG methods is about 10ms. And, the maximum measurement delay of the autonomous methods is 210ms, while the scheduled methods have extra delays about 30ms

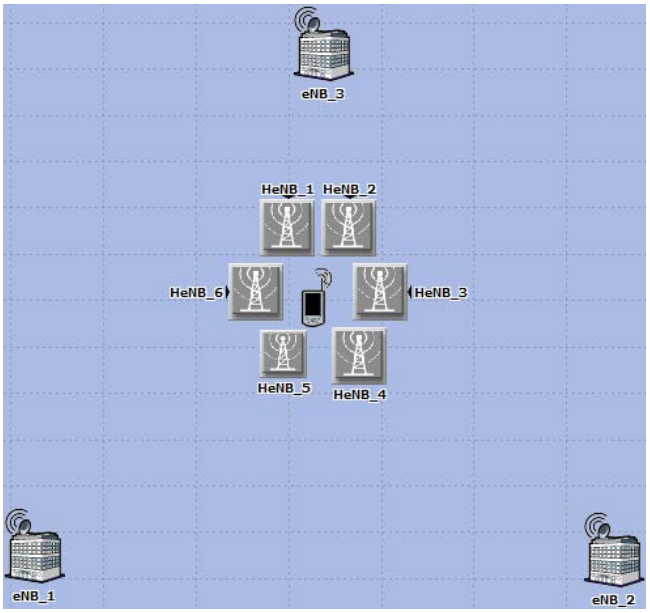


Fig. 11. Network deploy model of simulation

TABLE I
Simulation Parameters

Number of macro cell	3 eNBs
Number of HeNB	6
Number of UE	1
Inter-site distance	500 m
System frequency	2GHz
System bandwidth	FDD:10+10MHz
Propagation loss model	Inside the same cluster $L = 127 + 30 \log_{10} R$ For other link $L = 128.1 + 37.6 \log_{10} R$
Shadowing model	Lognormal shadowing
Shadowing standard deviation	10dB for Link between HeNB and HeNB UE 8dB for other links
Penetration Loss	Inside the same cluster: 0dB All other links: 20dB

or 60ms for scheduling message exchange. The AMPSG method shows the shortest measurement delay compared with other methods. In the simulation results, the graph shows similar trend with that of mathematical analysis but, the scale is not exactly matched. The differences between two graphs are due to that the mathematical analysis performed with assumption of worst case. And, the AMPSG shows more delay compared with worst case analysis because overlapping of system information broadcast can occur in the simulation.

As a result, the methods with several small gaps show better performance in aspect of service interruption time. But, all of four methods have smaller than 25ms service interruption time. Therefore, the measurement does not influence on video or VoIP services critically, if the channel quality is properly good to prevent errors in system information message packets. And, autonomous methods show much better performance than that of scheduled methods in the aspect of measurement

 ISBN: 978-1-61208-133-5

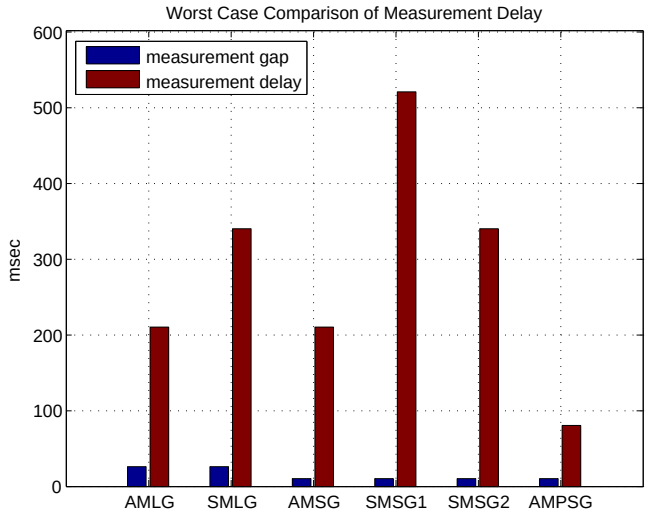


Fig. 12. Worst Case Anlysis Results of Measurement Gaps and Measurement Delay

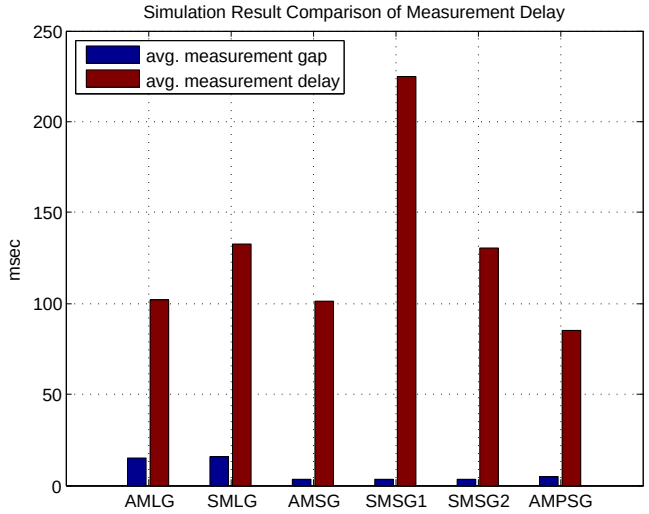


Fig. 13. Simulation Results of Measurement Gaps and Measurement Delay

delay. However, autonomous methods have possibility of packet drop, because the serving cell cannot know whether the UE is disconnected or not. If we want to use the AMLG, AMSG or AMPSG method, we need extra mechanism to prevent possible packet drops. Moreover, parallel method shows better measurement delay performance than that of serial methods. Consequently, the AMPSG method is best solution when the system requires small measurement delay.

VI. Conclustion

In this paper, we propose for system information measurement methods for the 3GPP LTE CSG cell. And we evaluated the performance of those methods in the terms of service interruption time (measurement gap) and measurement delay by mathematical analysis and simulation. As a result, the measurement gaps of methods with a large gap are larger than those of methods with several small gaps. And, the measurement delays of autonomous methods are much shorter than those of scheduled methods. Also, the parallel method shows best performance in aspect of measurement delay.

Acknowledgment

This work was supported by the IT R&D program of MKE/KEIT [KI001822, Research on Ubiquitous Mobility Management Methods for Higher Service Availability] and partly supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology (2009-0066642).

References

[1] T. Nakamura, *3GPP LTE Radio Access Network*. GSMA Americacas Conference, June 2010.

[2] *E-UTRA and E-UTRAN overall description, 3GPP Technical Specification 36.300, version 8.12.0*, April 2010.

[3] *E-UTRA RRC Protocol specification,3GPP Technical Specification 36.331, version 8.10.0*, June 2010.

[4] *E-UTRA UE procedures in idle mode, 3GPP Technical Specification 36.304, version 8.8.0*, January 2010.

[5] C.-H. Lee, J.-H. Kim, and H.-D. Bae, "Analysis of service interruption time due to system information measurement in 3gpp lte femtocell," in *Proc. ICUIMC 2011*, Febrary 2011.

[6] S. Sesia, I. Toufik, and M. Baker, *LTE: The UMTS Long Term Evolution from Theory to Practice*. West Sussex: John Wiley & Sons, 2009.

 ISBN: 978-1-61208-133-5

A Performance Study of Conventional and Bare PC Webmail Servers

Patrick Appiah-Kubi
Department of Computer and Information Sciences
Towson University
Towson, MD 21252, U.S.A.
appiahkubi@towson.edu

Ramesh K. Karne
Department of Computer and Information Sciences
Towson University
Towson, MD 21252, U.S.A.
rkarne@towson.edu

Alexander L. Wijesinha
Department of Computer and Information Sciences
Towson University
Towson, MD 21252, U.S.A.
awijesinha@towson.edu

***Abstract*-In this paper, we compare the performance of several Webmail servers and a bare PC Webmail server running without an operating system. The conventional Webmail servers used in the study are Icewarp, MailTraq and Hexamail running on Windows, and Atmail and Afterlogic running on Linux. Server performance is compared with respect to the typical email transactions (login, compose, and read), CPU utilization, and throughput. Performance under increased loads is measured by using a stress tool, and experiments are conducted in both LAN and WAN environments. The results indicate that bare PC Webmail server performance is consistent and predictable, whereas conventional Webmail server performance varies considerably depending on message size, server load and transaction.**

Keywords – Operating Systems; Bare Machine Computing; Webmail servers; Performance; Email.

I. INTRODUCTION

Web-based email or Webmail enables users to access email from any computer located anywhere using a Web browser such as Internet Explorer, Firefox or Google Chrome. Although popular Webmail systems such as gmail provide useful services, they are not designed for high performance or security. While there are some email/Webmail servers and systems that are designed for high-performance and/or security, they require an operating system (OS) to run. The performance of such systems is limited by the capabilities of the underlying OS and they are also susceptible to attacks that exploit the vulnerabilities of this OS.

In contrast, bare PC or bare machine servers and applications do not use an OS. Each application contains only essential functionality and has its own interfaces to the hardware. This eliminates OS overhead and enables the system to be optimized for performance by fully exploiting the capabilities of the underlying hardware. Moreover, bare PC applications are immune to conventional attacks that target a specific OS such as Linux or Windows. Many bare PC applications have been developed including Web servers, email servers and Webmail servers. Performance studies of bare PC applications serve to verify that the application provides the desired performance benefits and that it outperforms its OS-based counterparts when running on compatible systems. In this paper, we compare the performance of a bare PC Webmail server with several OS-based Webmail servers. The rest of this paper is organized as follows. Section II discusses related work and Section III provides a brief overview of the bare PC Webmail server. Section IV contains the performance results and Section V concludes the paper.

II. RELATED WORK

Atmail [2], MailTraq [17], Axigen [3], Afterlogic [1], Squirrelmail [23], Facemail [7], Adaptive Email [4], Petmail [21], Icewarp [12], Roundcube [22], Emailman [8], WinWebmail [24], and Hexamail [11] are just a few of the numerous Webmail systems in existence today. Some of these systems are designed for high performance, while others such as Webex [6] are designed for high reliability and availability. An email architecture to address problems associated with scalability and dependability due to conventional design approaches is proposed in [15]. There appear to be no studies that evaluate the performance of Webmail systems or servers. Techniques to improve performance of the Open Webmail system are discussed in [5]. In [13], an email server architecture, which is based on a spam workload and optimized with respect to concurrency, I/O and IP address lookups, is shown to significantly improve performance and throughput. The design and implementation of an email pseudonym server providing anonymity to reduce server threats and capable of reducing risks due to OS-based vulnerabilities is presented in [18]. The notion of semantic email is discussed in [19].

An email server that runs on a bare PC is the focus of [9] and [10]. However, the server does not support Webmail. Many bare PC applications including Web servers [16] and VoIP clients [14] have been previously developed. In [20], the design and implementation of a bare PC Webmail server are described and some preliminary performance results are presented. The present paper differs from earlier work in that it compares the performance of a bare PC Webmail server and several OS-based Webmail

 ISBN: 978-1-61208-133-5

servers in both routed LAN and WAN environments, and also under stress conditions.

III. THE BARE PC WEBMAIL SERVER

Only a brief overview of bare PC Webmail server internals is given here as its design and implementation were discussed in detail in [20]. Since bare PC applications run directly on the hardware without the support of an OS, they are self-supporting. The Webmail server includes lean implementations of the HTTP/TCP/IP/SMTP/POP3 protocols (that are intertwined with the server application), and an Ethernet driver.

CPU task and memory management and the concurrent processing of requests from multiple clients are done by the application itself, which is written in C++ except for some low-level assembly code. There are only 4 task types in the Webmail server application: The Main task, consisting of a loop that runs whenever no other task is running; the Rcv task that receives incoming packets and is used for Ethernet, IP, and TCP processing; and multiple Get and Post tasks that manage the processing of client requests. A given request or packet is processed as a single thread of execution. Once activated, a task runs to completion unless it has to wait for an ack or a timeout. Delay and Resume lists are used to efficiently manage the suspension and resumption of tasks. Get/Post are modeled using state transitions.

The application is initially booted from a USB flash drive and does not use a hard disk. The USB is also used for persistent storage of email messages and user information, but a separate server could be used for auxiliary storage in the future. The Webmail server currently runs on an ordinary PC (not a server machine). The main data structure used by the Webmail server (and all bare PC applications) is the TCB (Transmission Control Block) table that contains entries to enable the management of concurrent requests, associated data, and TCP/application state information. Get/Post tasks are placed in the Resume list when requests arrive and their active status is indicated by a flag in the TCB table. The Webmail server application includes a lean PHP parser that interprets client Get/Post data.

IV. PERFORMANCE RESULTS

A. LAN Setup

For the LAN studies, a dedicated test network consisting of five Ethernet switches (S1-S5) interconnected linearly by four Linux routers (R1-R4) was set up. The client (C) and Webmail server (WMS) were connected to the ends of the network so that messages between the client and Webmail server are routed along the following path:

C--S1--R1--S2--R2--S3--R3--S4--R4--S5--WMS

All switches were gigabit switches except for the 100Mbps switch (S1) used to connect the client to the network. The clients ran Windows XP and the OS-based Webmail servers ran Windows XP or Linux (CentOS). All machines were Dell Optiplex GX520s. OS-based Webmail server details are as follows: Afterlogic MailSuite Pro (Linux), MailTraq Server (XP), Atmail Server 6.20.3 (Linux), Icewarp Server 10.2.1 (XP), and Hexamail Server 4.0.1.002 (XP).

B. LAN Results

Fig. 1 below is derived from the Wireshark timestamps for each message in the sequence of messages exchanged during a login Get request. The difference in timestamps for a pair of consecutive messages such as (Get, Ack) or (Data, 200_OK) gives the delay between the pair. As expected, the performance for all servers during the initial TCP handshake is the same. There is a rise between the client Get request and the server Ack due to the server delay in processing the request. All servers show little variation in processing time for subsequent message pairs.

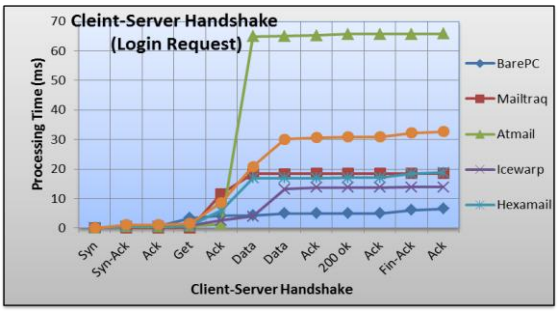


Figure 1. Login Get request message times

Similarly, Fig. 2 compares the processing time for a login Post request. The (Post, Ack) behavior for all servers except for Hexamail and MailTraq is the same. The (Ack, 302_Found) delay is visible for all servers except MailTraq. The bare PC server processing times for both Get and Post requests are minimal. Since different servers may do the work to process the requests during different steps, only the overall processing time should be compared. Fig. 3 shows the processing time for compose with varying message sizes. The varying behavior of the servers reflects the combination of TCP, HTTP and the mail server application. Hexamail has stable behavior for large message sizes, while Icewarp shows the most variation. The bare PC server has the highest processing delay for a message of 10,000 bytes, but shows a general reduction for larger sizes except for a

 ISBN: 978-1-61208-133-5

small rise at 20,000 bytes. Fig. 4 shows the processing time for receiving an inbox with 6 messages. While all servers complete processing in about 1.1 milliseconds on the average, the bare PC server requires less than 0.1 ms.

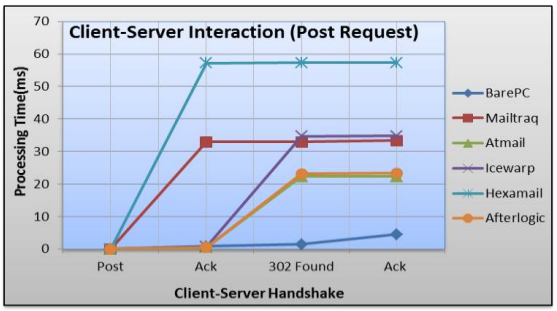


Figure 2. Login Post request message times

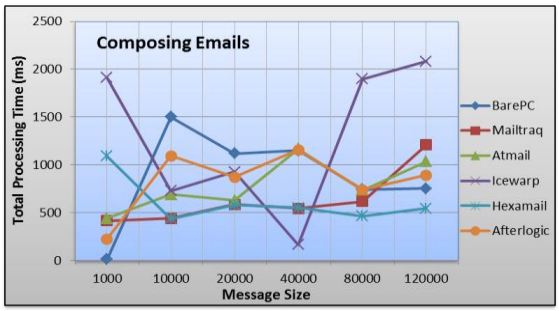


Figure 3. Processing time for compose (varying message sizes)

Figure 4. Processing time for an Inbox request (6 messages)

Fig. 5 shows the processing time to retrieve messages of sizes 1000-120,000 bytes. Once a message is retrieved into an inbox, it takes less time to process and transmit the message to the client. Processing time for the bare PC is minimal, and it has the smallest increase in processing time. Among the OS-based servers, Icewarp followed by Afterlogic have the lowest processing times (except for a 1000-byte message), while Icewarp and Hexamail have the smallest increase in processing time (the latter actually has the highest processing time). Fig. 6 shows the throughput measured during compose for increasing message sizes. The bare PC server throughput is highest and approximately twice the throughput of the best OS-based server Afterlogic. The low throughput of Icewarp reflects its large processing time in Fig. 3.

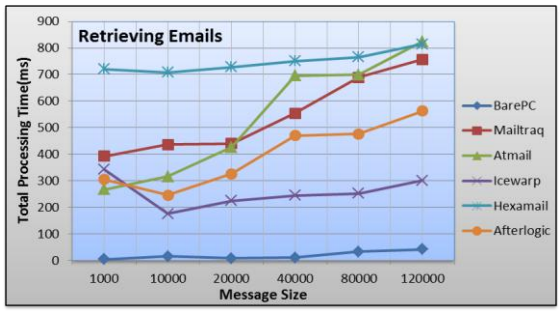


Figure 5. Processing time for read (varying message sizes)

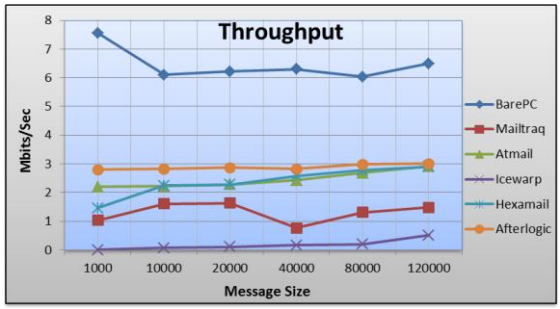


Figure 6. Throughput for compose (varying message sizes)

Further performance tests were conducted on the Webmail servers using the Web stress tool [25]. The tool was used to increase the number of users from 1 to 10 and determine the resulting impact on performance. Each test was run for 10 minutes and each user makes 100 requests/s. Fig. 7 illustrates the variation of server CPU utilization over time for a maximum of 10 users. The average CPU utilization of the Linux-based Afterlogic and Atmail servers and the bare PC server is less than 4%, while that of the Windows-based Mailtraq, Icewarp and Hexamail servers is between 8-12%. It is evident that more CPU processing is required by the Windows-based servers when processing concurrent requests. The figure also indicates that the CPU utilization of the bare PC server shows some slight initial variability compared to that of the Atmail server.

Fig. 8 shows the variation of server bandwidth over of time for 10 users. It can be seen that the bandwidth of all servers is relatively stable after the initial increase during the first 5 seconds. However, while there is little difference between the bandwidth of the OS-based servers (average < 60 kbps, maximum <700 kbps), the bandwidth of the bare PC server is significantly higher (average and maximum exceed 100 kbps and 12 Mbps respectively).

 ISBN: 978-1-61208-133-5

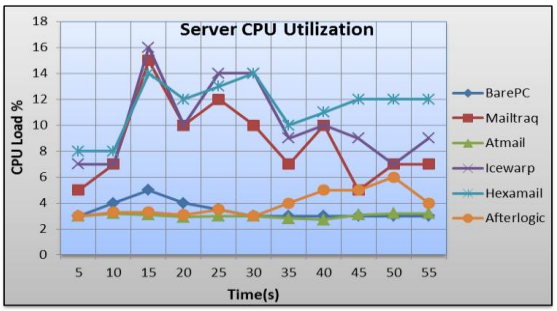


Figure 7. CPU utilization for 10 users

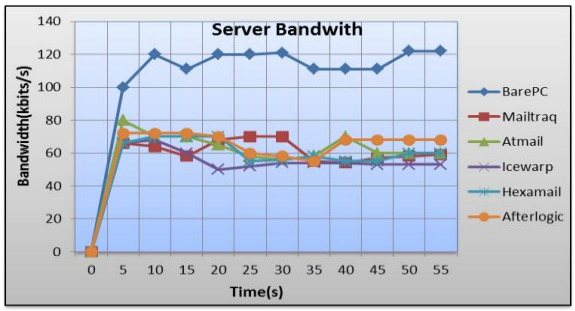


Figure 8. Bandwidth variation for 10 users

Figure 9 above is the average time (delay) to complete a Post request for a varying number of users. Even two users cause the delay to increase significantly compared to one user, but the delay stabilizes for three or more users. The bare PC has the least delay, while the Windows-based servers have the highest delay in this case. It can be seen that the delays for the bare PC server and Linux-based servers differ by almost 100 ms.

Fig. 10 shows the amount of time a user waits for the server to establish a connection in the presence of multiple users. The Linux and bare PC servers perform much better than the Windows servers, but the performance advantage of the bare PC server compared to the Linux servers is reduced since the performance of the latter improves significantly when there are 6-10 users. Figs. 11 and 12 show respectively the Webmail server processing times for a read request and the throughput for a compose request with and without stress. To create stress, the tool is used to generate 100 concurrent requests/s from 10 users and an additional client is used to generate the read or compose request involving an email message of 120,000 bytes. Although performance degrades under stress for all servers as expected, the bare PC server's performance with and without stress is significantly better than the performance of the OS-based servers. However, no simple relationship exists between throughput and processing time for the OS-based servers (for example, Icewarp and Mailtraq have the lowest throughput, but respectively low and high times).

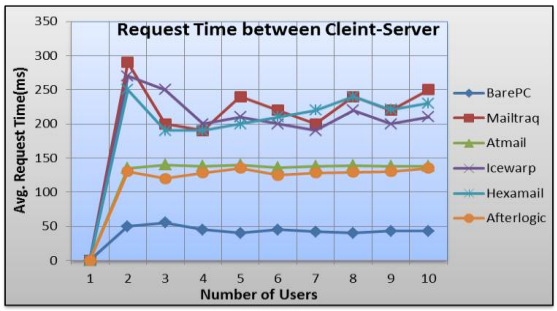


Figure 9. Post request completion time

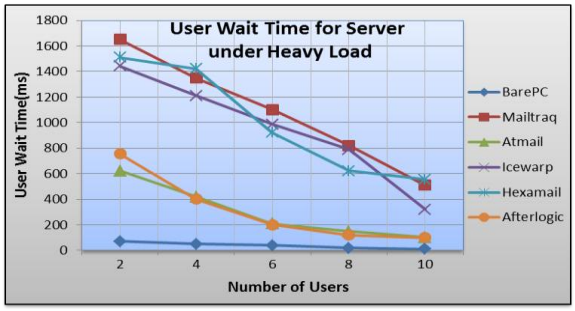


Figure 10. User wait time (increasing number of users)

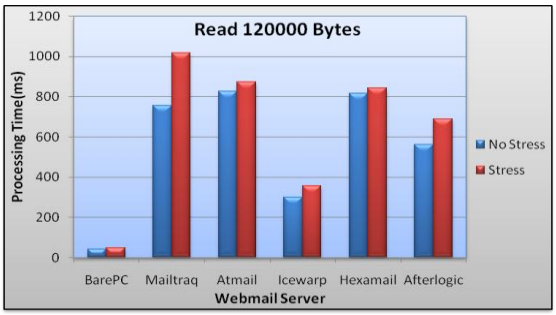


Figure 11. Message read time (120000-byte message)

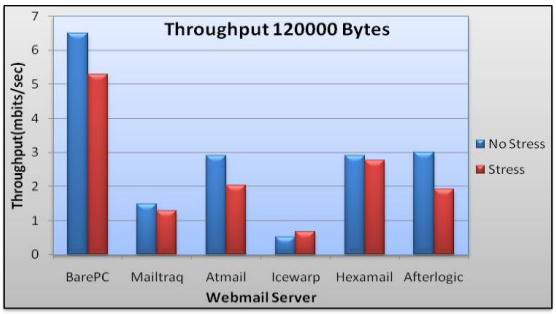


Figure 12. Throughput for compose (120000-byte message)

C. WAN Setup

For the WAN tests, an Internet connection was established between each Webmail server and a client PC located approximately 50 miles away with about 30 hops between the two destinations. To ensure consistency, tests were performed during a contiguous time and repeated several times to ensure that the results were stable and independent of varying network conditions. As before, a Wireshark

packet analyzer was used to capture the data. The machines used were the same as for the LAN studies.

D. WAN Results

Figs. 13 and 14 are derived from the Wireshark timestamps for each message in the sequence of messages exchanged over the WAN during a login Get or Post request (they correspond to Figs. 1 and 2 for the LAN tests). As before, the difference between cumulative processing times for a pair of consecutive messages such as Get-Ack for Get or Post-Ack for Post gives the delay between the pair. It can be seen that these delays for Get and Post requests are significantly less for the bare PC server than for the OS-based servers.

A closer examination of Figs. 13 and 14 reveals that the Get and Post delays for the OS-based servers vary considerably across message pairs. For example in Fig. 13, MailTraq has the highest Get-to-Ack time, Atmail has the highest Ack-to-Data time, and Icewarp has the highest Data-to-200_OK time. However, Atmail has the lowest Get-to-Ack and Data-to-200_OK times among the OS-based servers. Similarly, compared to the Windows servers, Afterlogic has lower Get-to-Ack and Data-to-200_OK times, but a higher Ack-to-Data time. In case of a login Post request (Fig. 14), it can be seen that MailTraq and Atmail have respectively the highest and lowest (next to the bare PC) Post-to-Ack time, whereas Atmail has the highest and MailTraq has the lowest (next to the bare PC and Icewarp) Ack-to-302_Found time.

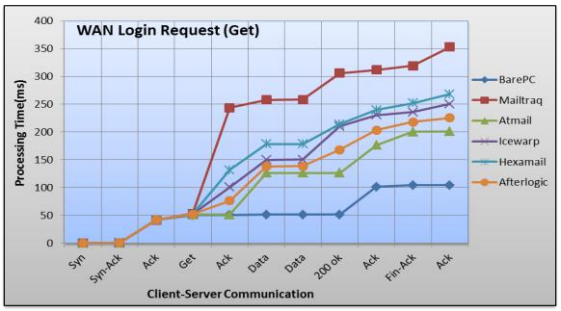


Figure 13. Processing time for login Get request

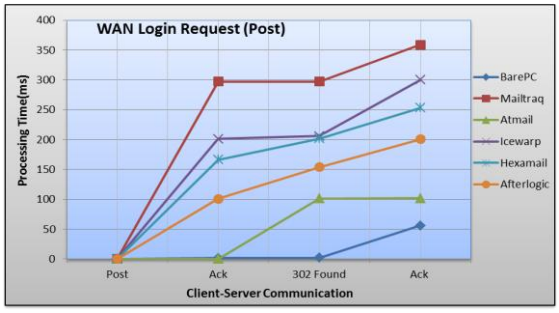


Figure 14. Processing time for login Post request

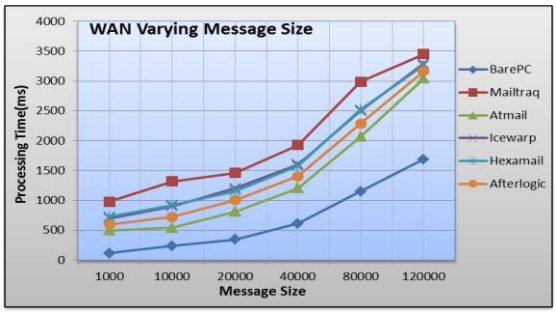


Figure 15. Processing time for compose (varying message sizes)

Figs. 15 shows the processing time over the WAN for compose, with message sizes varying from 1000 to 120,000 bytes. Processing time increases in an approximately linear manner as the message size increases. This was not the case for the corresponding LAN result in Fig. 3. However, it can be seen that the processing time for all servers increases at a higher rate for message sizes from 40,000-120,000 bytes.

Fig. 16 shows the processing time over the WAN for receiving an inbox containing 6 messages. The bare PC receives the inbox in 0.392 milliseconds, while the other servers require an average time of about 230 milliseconds. Fig. 17 shows the processing time on the WAN for reading individual emails of varying message sizes. The processing time for the bare PC server is stable up to 40,000 bytes and increases slowly thereafter for larger messages. The processing times on the other servers are stable up to 20,000 bytes, but rise sharply to 1400 milliseconds.

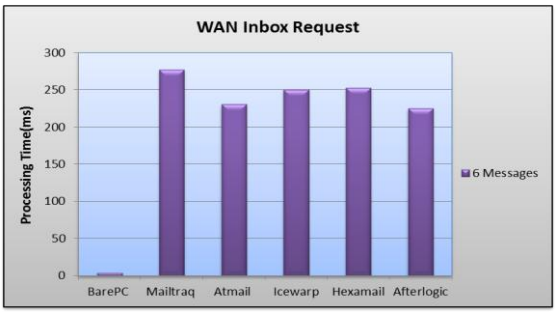


Figure 16. Processing time for an Inbox request (6 messages)

Figure 17. Processing time for read (varying message sizes)

 ISBN: 978-1-61208-133-5

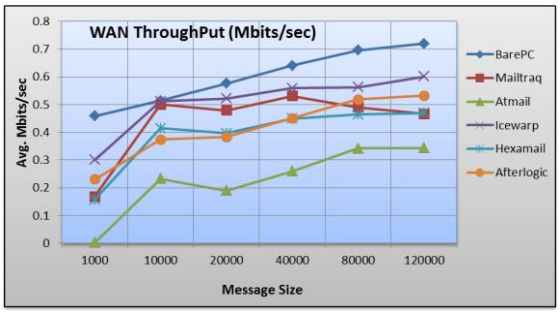


Figure 18. Average throughput (varying message sizes)

The throughput for varying message sizes was also captured during the Internet test and compared in Fig. 18. The average throughput for all servers is about 1.28 times better than the throughput of the Atmail server, whose performance in general on the previous tests was better than the other OS-based servers.

V. CONCLUSION

We conducted a performance study of six Webmail servers including a bare PC server with no operating system. Server processing times, CPU utilization and throughput in LAN and WAN environments and under stress conditions were compared with respect to common email transactions. The results show that performance of the OS-based servers is variable and no single server performs consistently better than the others on all tests. There appears to be no simple relation between LAN and WAN results even for the case of a single server. With a few rare exceptions, only a small drop in performance was seen under stress conditions. However, a detailed study under real workloads and conditions would be needed to determine the ability of servers to handle stress. As expected, the bare PC server performs significantly better on all tests with a few minor exceptions. This suggests that some of its novel design features could serve as a baseline for designing secure high-performance Webmail servers in the future.

REFERENCES

[1] Afterlogic Webmail Server, www.afterlogic.com Retrived on May 20th, 2010

[2] Atmail-Linux Webmail Server, www.atmail.com Retrived on October 5th, 2009.

[3] Axigen Email Sevrer, www.axigen.com Retrived on January 23th, 2009.C. Yeh and C. Mao, Adaptive e-mail intention finding mechanism based on e-mail words social networks. Workshop on Large Scale Attack Defense, 2007.

[4] C. Tung and S. Tsai, Tuning Webmail performance-The design and implementation of Open Webmail. National Cheng Kung University, www.openWebmail.org Retrived on May 20th, 2010.

[5] Cisco WebEx Mail: High availability design for your most mission critical application. www.ciscowebexmail.com. Retrived on May 20th, 2010.

[6] E. Lieberman and R.C. Miller, Facemail: showing faces of recipients to prevent misdirected email. 3rd Symposium on Usable Privacy and Security, 2007.

[7] Emailman Webmail Server, www.emailman.com Retrived on May 21th, 2010.

[8] G. Ford, R. Karne, A. L. Wijesinha, and P. Appiah-Kubi, The design and implementation of a bare PC email server, 33rd Annual IEEE International Computer Software and Applications Conference (COMPSAC), 2009.

[9] G. Ford, R. Karne, A. L. Wijesinha, and P. Appiah-Kubi, The performance of a bare machine email server, 21st International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD), 2009.

[10] Hexamail Webmail Server, www.hexamail.com Retrived on June 10th, 2010.

[11] IceWarp Mail Server, www.icewarp.com Retrived on May 20th, 2010

[12] A. Pathak, S. A. R. Jafri, and Y. C. Hu, The case for high-performance spam-Aware mail server architecture. 29th IEEE International Conference on Distributed Computing Systems, pp. 155-164, July 2009.

[13] G. H. Khaksari, A. L. Wijesinha, R. K. Karne, L. He, and S. Girumala. A peer-to-peer bare PC VoIP application. IEEE Consumer and Communications and Networking Conference, Las Vegas, Nevada, January 2007.

[14] E. Kageyama, C. Maziero, and A. Santin, A pull-based email architecture. ACM Symposium on Applied Computing, Fortaleza, Ceara, Brazil, 2008, pp. 468-472.

[15] L. He, R. Karne, and A. L. Wijesinha, "The design and performance of a bare PC Web server", International Journal of Computers and Their Applications, vol. 15, June 2008, pp. 100-112.

[16] Mailtraq email server-the complete email server, www.mailtraq.com Retrived on June 10th, 2010.

[17] D. Mazieres and M. F. Kaashoek, The design, implementation and operation of an email pseudonym server. 5th Conference on Computer and Communications Security, 1998, pp. 27-36.

[18] L. McDowell, O. Etzioni, A. Halevy, and H. Levy, Semantic email. 13th International Conference on World Wide Web, New York, NY, USA, 2004, pp. 244 – 254.

[19] P. Appiah-Kubi, R.K. Karne, and A. L. Wijesinha, The design and Performance of a bare PC Webmail Server. 12th IEEE International Symposium of Advances on High Performance Computing and Networking. September 2010.

[20] Petmail Webmail Server, www.petmail.lothar.com Retrived on June 12th, 2010

[21] RoundCube Webmail Server, www.roundcube.com Retrived on November 30th, 2010

[22] SquirrelMail Webmail, www.squirrelmail.com Retrived on November 30th, 2010

[23] WinWebmail Server, www.winWebmail.net Retrived on October 20th, 2010

[24] Web Stress Tool 7, www.paessler.com Retrived on May 20th, 2010

 ISBN: 978-1-61208-133-5

Performance of Soft Reservation-based Soft Frequency Reuse Scheme for Cellular OFDMA Systems

Hye-Joong Kang, Jin W. Park, and Chung G. Kang
School of Electrical Engineering, Korea University
{dreamftr, jwpark, ccgkang}@korea.ac.kr

***Abstract*—The conventional soft frequency reuse (SFR) scheme has been considered as a useful means of inter-cell interference coordination (ICIC) in the downlink of cellular OFDMA systems. It is based on hard reservation, which partitions the resource regions into two orthogonal portions, one solely dedicated to users in the cell center and the other solely dedicated to those in the cell edge. In this paper, we consider the variants of SFR scheme, which are based on a notion of soft reservation. As they allow for sharing a whole resource region among all or some users, the wider resource region leads to the multi-user diversity gain, whiles still maintaining a feature of interference mitigation by power control and dynamic interference avoidance by opportunistic scheduling in each cell. We demonstrate that a soft reservation-based SFR scheme can be the best means of trading off the average system throughput and edge-user throughput.**

Keywords – inter-cell interference coordination; soft fractional reuse; soft reservation; multi-user diversity; OFDMA

I. INTRODUCTION

Link adaptation has been a common means of dealing with co-channel interference (CCI) under time-varying and location-dependent situations in cellular OFDMA systems. However, it cannot be a means of improving the spectrum efficiency of users at the cell edge as it cannot reduce the CCI itself. Thus, various types of inter-cell interference coordination (ICIC) schemes are considered for combating the CCI problem in the cellular OFDMA network [1] . In general, ICIC may involve a complex optimization procedure with exchanging inter-cell information.

Meanwhile, inter-cell frequency reuse scheme is one particular type of ICIC from the previous works, which does not require for sharing any inter-cell information. It divides the frequency band into orthogonal channels, a subset of which will be allocated to the different cells so that the inter-cell interference can be mitigated by the sufficient frequency reuse distance. In the cellular OFDMA system, for example, a fractional frequency reuse (FFR) scheme is one particle example of realizing the frequency reuse strategy, which deals with the orthogonal frequency bands. In the frequency reuse scheme, the most important design criterion is to trade-off the performance gain with interference mitigation subject to the larger reuse distance and performance degradation with the reduced amount of resource. In other words, the overall system throughput must be maximized by trading off these two aspects.

There are two different types of FFR schemes: Partial Frequency Reuse (PFR) scheme [2] and Soft Frequency Reuse (SFR) scheme [3]. Dividing a resource region into two orthogonal portions, the PFR scheme employs the different frequency reuse factors (K) for the different portion in each cell, e.g., K = 1 for the resource allocated to the cell-center users and K = 3 for the resource allocated to the cell-edge users, so that all users can be loosely protected by the sufficient reuse distance. As opposed to the PFR scheme, the SFR scheme can fully use the frequency band in each cell by mitigating the inter-cell interference with power control. This particular scheme has been introduced in the early standardization stage of 3GPP LTE system. Even if no explicit specification has been provided, the ICIC parameters in the current LTE standard can be used to support the SFR scheme.

We find that the conventional PFR/SFR scheme is based on hard reservation, which partitions the resource region into two orthogonal portions, one solely dedicated to users in the cell center and the other solely dedicated to those in the cell edge. This kind of resource allocation which restricts resource region to each user can ensure an SINR gain in average manner. As described in [4] and [5], however, it hurts a multi-user diversity gain and fairness of resource allocation, especially when the users are not uniformly distributed throughout the coverage. These problems can be handled by a soft reservation mechanism, which allows for sharing a whole resource region among all or some users. By employing soft reservation, the wider resource region leads to a more multi-user diversity gain and fairer resource allocation [4][5]. However, it is not straightforward to warrant the average SINR gain.

In this paper, we consider the variants of SFR scheme, which are based on a notion of soft reservation. Our objective is to analyze their system throughput so that their performance can be characterized under the varying conditions. Ultimately, we are expecting to identify the best means of trading off the average system throughput and edge-user throughput.

The remaining of this paper is organized as follows. In Section II, we describe the SFR scheme and present its system model. A notion of soft reservation is introduced in Section III, leading to the variants of the SFR schemes. The performance characteristics of those variants are investigated in Section IV and the concluding remarks are presented in Section V.

 ISBN: 978-1-61208-133-5

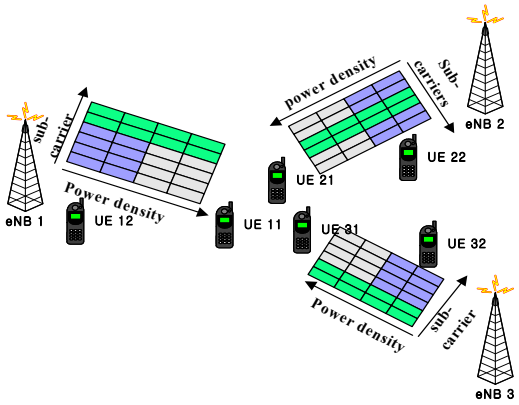


Fig. 1. Soft Frequency Reuse Scheme: Illustration [3]

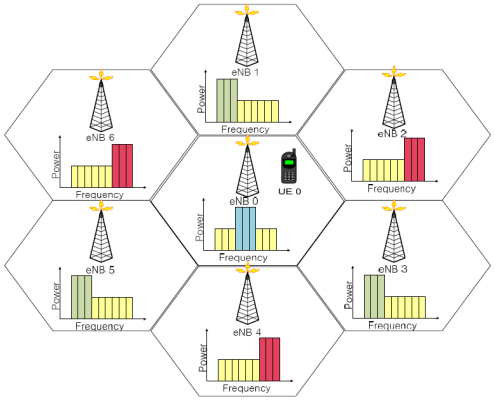


Fig 2. Power allocation pattern

II. SOFT FREQUENCY REUSE SCHEME: SYSTEM MODEL

In the soft frequency reuse (SFR) scheme, the users are classified into the different groups, depending on their location within the cell. Without loss of generality, we consider a special case of two user groups. More specifically, the users in each cell site belong to the center-user group $\mathcal{U}_C$ or edge-user group $\mathcal{U}_E$. The various design criteria can be employed to determine $\mathcal{U}_E$ and $\mathcal{U}_C$ [6]. For example, the users with their average SINR greater than the given threshold η_S are grouped into $\mathcal{U}_C$ while those with the average SINR less than η_S are grouped into $\mathcal{U}_E$ [3]. The different power levels, P_H and P_L, are employed for $\mathcal{U}_E$ and $\mathcal{U}_C$, respectively ($P_H > P_L$). As illustrated in Fig. 1, SINR of the cell-edge users will be enhanced with the higher power level and furthermore, with the partial frequency reuse, i.e., allowing no other edge users in the adjacent cells to reuse the same resource. Meanwhile, all remaining resources are allocated to the center-user group, with the lower power level, which allows for the full frequency reuse among all the cells.

In the current design, let us consider the SINR trade-off between $\mathcal{U}_E$ and $\mathcal{U}_C$ for the illustrative example in Fig. 2. Here, we assume that there is the power difference of α dB between P_H and P_L, i.e., $\alpha = P_H - P_L$ (dB). Consider a serving base station (BS) that is surrounded by 6 adjacent cells as shown by the typical hexagonal cell structure in Fig. 2. In case that all BSs are using the same power P, the corresponding signal-to-interference (SIR) for UE 0 is given as

$$SIR_{no} = \frac{P\,|\,h_0\,|^2}{P\sum_{i=1}^{6}|\,h_i\,|^2} \tag{1}$$

where $\{h_i\}_{i=0}^{6}$ are the link gains between individual BSs and reference user, UE 0, with i =0 denoting the serving BS. Employing the power allocation pattern for a cluster of three cells as in Fig. 1, SIRs for the users in $\mathcal{U}_E$ and $\mathcal{U}_C$ are respectively given as

$$SIR_H = \frac{P_H\,|\,h_0\,|^2}{P_L\sum_{i=1}^{6}|\,h_i\,|^2} = 10^{\alpha/10}\,SIR_{no} \tag{2}$$

$$\begin{aligned} SIR_L &= \frac{P_L\,|\,h_0\,|^2}{P_L\sum_{i=1}^{3}|\,h_{2\times i-1}\,|^2 + P_H\sum_{i=1}^{3}|\,h_{2\times i}\,|^2} \\ &= \frac{P_H\,|\,h_0\,|^2}{P_H\sum_{i=1}^{6}|\,h_i\,|^2 + (10^{\alpha/10}-1)\,P_H\sum_{i=1}^{3}|\,h_{2\times i}\,|^2} \\ &\le SIR_{no} \end{aligned} \tag{3}$$

where $SIR_{no} = |\,h_0\,|^2 / \sum_{i=1}^{6}|\,h_i\,|^2$. The above results shows that the users in $\mathcal{U}_E$ can achieve with the SIR gain of α dB, while those in $\mathcal{U}_C$ suffer from the SIR loss in a range of $[0, \alpha]$ dB, depending their relative positions from the base stations.

III. VARIANTS OF SFR SCHEMES: HARD RESERVATION VS. SOFT RESERVATION

In this section, we consider the variants of SFR schemes, which differ by how the resources are shared between the users in $\mathcal{U}_E$ and $\mathcal{U}_C$, for performance comparison between hard and soft reservation. An obvious reference scheme is the one that does not employ the FFR scheme while allocating the power level of P_H to all users (referred to as "Normal"). The conventional SFR scheme is characterized by hard reservation, which means that the resources are completely partitioned into two regions, one for $\mathcal{U}_E$ and the other for $\mathcal{U}_C$. Depending on whether the partitioned regions can be changed by adapting their boundary to the user distribution, it can be either dynamic or static. In the current comparative studies, we consider the static case in which the partitioned regions are fixed, i.e., referred to as "Static SFR."

Meanwhile, we consider two different variants of the static SFR schemes, which allows for sharing the reserved resource of each user group whenever necessary, e.g., when

 ISBN: 978-1-61208-133-5

one resource region is overloaded by the non-uniform user distribution. In fact, there are two different extreme cases: one case of allowing the center users to share the resources reserved for the edge users and the other case of allowing the edge users to share the resources reserved for the center users. More specifically, the center users can employ K = 1 with the power level of P_L, i.e., allowed to borrow the resource reserved for the edge users (referred to "SFR-FCR: SFR with Full Center Reuse"). In other words, the resources for the edge users are only *softly* reserved as they are shared with the center users. As the lower power level is employed by the center users, the edge users are not suffered from the additional interference caused by soft reservation in this case.

On the other hand, the edge user can employ K = 1 with the power level of P_H, i.e., allowed to use the resource reserved for the center users (referred to "SFR-FER: SFR with Full Edge Reuse"). In this case, some edge users will be suffered from the additional interference from the edge users that share the same resource reserved for the center users in some adjacent cells. In fact, SFR-FCR and SFR-FER schemes are two extreme cases for SFR scheme subject to soft reservation.

The most generalized form of soft reservation in SFR is to share all the resources among all users in the system. As a whole resource can be used by both center and edge users, it is just the same as the conventional scheme, except that two different power levels are employed, depending on their position. The advantage of the soft reservation-based SFR (SFR-SR) scheme would be to improve a multi-user diversity gain by extending the allocation region to a whole band. As mentioned earlier, however, SFR-SR cannot maintain a fixed SINR gain, as implied by (3). Fig. 3 illustrates the various types of SFR schemes in the perspective of resource sharing and power allocation.

In order to understand the characteristics of all these schemes, a notion of *collision* must be addressed from a viewpoint of inter-cell coordination. Collision can be roughly understood as an event of failing the interference coordination that is intended by scheduling and power allocation, mainly due to the excessive other cell interference incurred by the high power level for the same resource shared between the adjacent cells. In this work, we define a rather quantitative notion of collision in the course

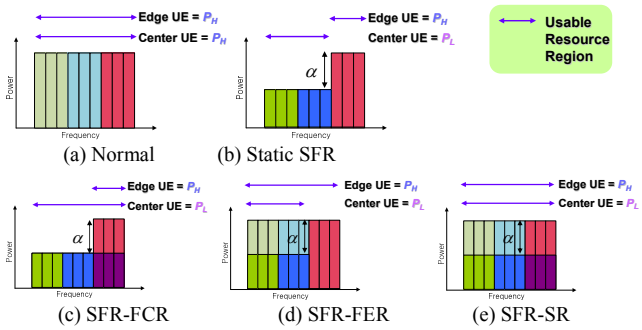


Fig. 3. Variants of SFR schemes: Comparison

of link adaptation. It is defined as an event that collision occurs when a user in $\mathcal{U}_E$ allocated to a certain resource region undergoes degradation in MCS level because its highest- interfering cell allocates P_H to the corresponding resource region.

IV. PERFORMANCE ANALYSIS

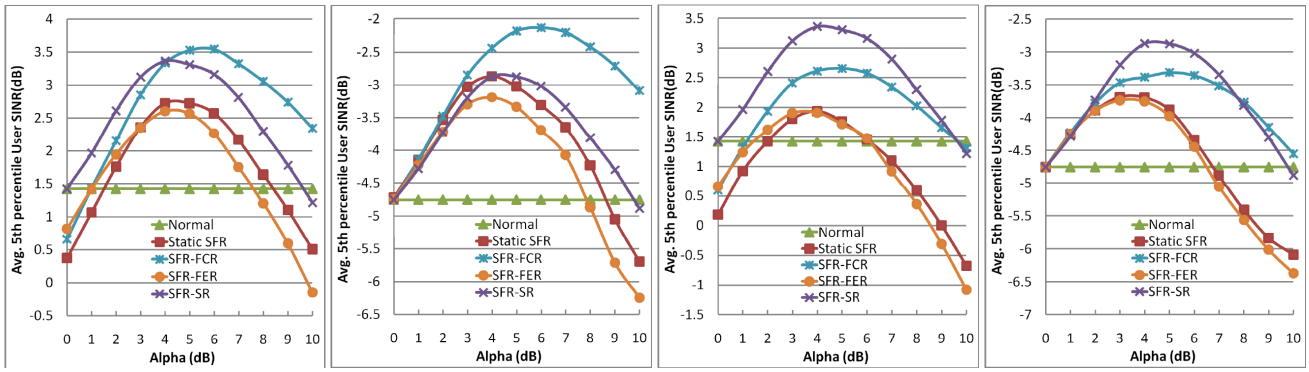
Various types of the SFR schemes are evaluated with system-level simulation for 3GPP LTE system. Table 1 summarizes the simulation parameters used in the current studies, including the system parameters for LTE specification.

In order to capture the performance characteristics of individual SFR scheme, we consider four different scenarios that are varying with the fading characteristics and user-grouping criterion. We consider two different user-grouping criteria with using the different thresholds: SINR threshold and ratio threshold. More specifically, UE can be grouped into an edge user set if its SINR drops below the given SINR threshold η_S (SINR threshold-based), or if its SINR is smaller than the lowest x-th percentile SINR (load threshold-based). The SINR threshold-based grouping may suffer from the situation that may turn too many UEs into the edge- or center-user group, i.e., incurring overload to one of two partitioned regions. Such an overload problem can be solved by the load threshold-based grouping, which sets the threshold by traffic load for edge users. In the following evaluation, we employ the load threshold-based user grouping. Meanwhile, we consider both fading and non-fading channels. The performance under a non-fading

Table 1. Simulation Parameters

System Parameter		Note
The number of cells	19	3 sectors per cell
Inter-system distance	500m	-
The number of UEs per sector	10	-
Antenna configuration	SIMO	1x2 MRC
Carrier frequency	2GHz	-
Bandwidth	10MHz	FFT: 1024, 50RBs/slot
Hybrid ARQ	Chase Combining	The maximum number of retransmissions: $N_r(max) = 3$
		The number of HARQ process channels: $N_h = 8$
BS Tx power	Max: 40W (46dBm)	-
Log-normal shadowing	STD: 10dB	-
Channel model	ITU-R Pedestrian B (3km/h)	-
Noise figure	9dB	
Scheduler	Proportional fairness	T=1000
Traffic model	Full buffer	-
Link-to-system interface	Effective SNR: Mutual Information-based	IEEE802.16m EMD [7]
Link adaptation	Adaptive modulation & coding	-
CQI type	Subband CQI	The number of subbands: 9
CQI report period (T_{CQI})	2ms	Actual period for each subband: 18ms
Power ratio (α)	Variable	$\alpha = P_H - P_L$
SINR threshold for edge UE (η_S)	0dB	UE to be grouped into an edge user set if η_S <0dB
Load threshold for edge UE (x)	30%	UE with the lowest x% SINR to be grouped into the edge user set

 ISBN: 978-1-61208-133-5



(a) Case i (b) Case ii (c) Case iii (d) Case iv

Fig. 4. 5th percentile user SINR as varying the power ratio for different scenarios

channel would be useful for solely investigating the effect of resource allocation while eliminating the effect of opportunistic scheduling. Then, evaluations are carried for the following four different scenarios:

Case i) SINR-based grouping subject to fading
Case ii) SINR-based grouping subject to no fading
Case iii) load-based grouping subject to fading
Case iv) load-based grouping subject to no fading

A. Limitation of hard reservation: Static SFR

Fig. 4 shows the average 5th percentile user SINR while Fig. 5 shows the average 5th percentile and cell throughput at the same time, as the power difference $\alpha = P_H - P_L$ varies. As shown in Fig. 4, SINRs of all schemes increase with α up to a certain point, beyond which they decrease because the extremely high power for cell edge user would hurt the center users in the neighbor cells. In Fig 5, meanwhile, there is an obvious trade-off between 5th percentile and total average throughput as α increases up to a certain point, beyond which both 5th percentile and total average throughput decrease with α. These aspects are attributed to SINR degradation, as depicted in (3). More specifically, as the SINR of cell-center users is degraded with extremely large α, some of the center users tend to have 5th percentile SINR, which causes degradation in both 5th percentile and total throughput. The observations from Fig. 4 and Fig. 5 imply that the range of α must be carefully chosen for our comparative studies, so as to make sure that both SINR and throughput are not degraded at the same time over the given range. Note that Static SFR has neither SINR gain nor loss at $\alpha = 0$, as clearly in (2) and (3). In Fig. 4(a), however, its SINR is worse than that in the scheme that employs any ICIC ("Normal"), which is attributed to the hard reservation feature of the Static SFR. In fact, it is the feature that reduces the multi-user diversity gain in hard reservation. It is clear from Fig. 4(b), which demonstrates that there is no performance difference between Normal and Static SFR.

The same reason for aforementioned 5th percentile SINR degradation in hard reservation can explain the difference in throughput gain for each scheme between case i) and case ii) in Fig. 5. However, the difference between case i) and case iii) cannot be clearly understood by the same reason. In fact, the different grouping criterion is applied to case i) and case iii), respectively. For the SINR threshold-based grouping in case i), the number of users in each group varies with the user distribution, which may overload one resource region over the other and thus, lead to degradation in 5th percentile throughput. In case iii), meanwhile, no such degradation is expected as the number of users in the cell-edge group is pre-determined.

We note that Static SFR in case iv) outperforms that in case ii) or case iii) for 5th percentile throughput performance in Fig. 5. It is due to the fact that case iv) inherits both features in case ii) and case iii).

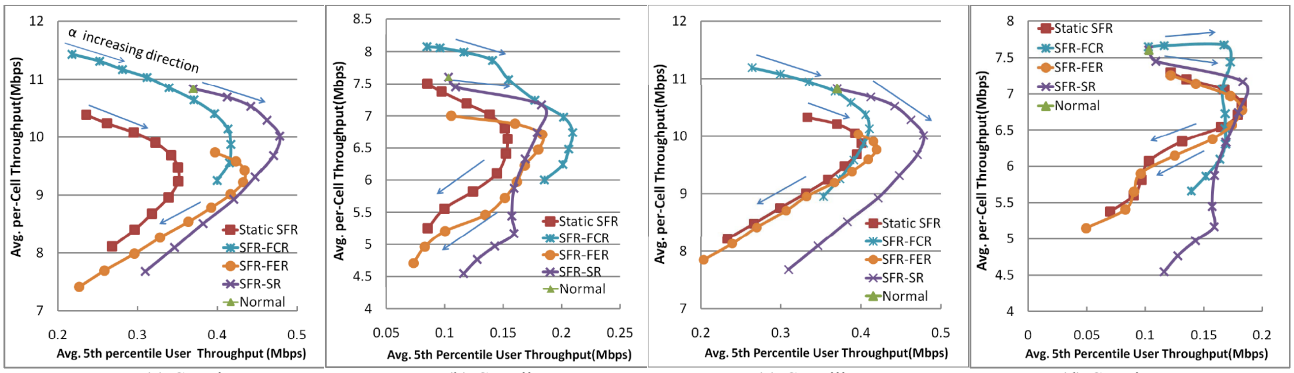
To summarize, any scheme based on hard reservation, including Static SFR, suffers from degradation in throughput performance by losing the multi-user diversity gain as well as the efficiency in resource allocation.

B. SFR-FCR vs. SFR-FER schemes

SFR-FCR and SFR-FER are SFR schemes that employ soft reservation, allowing one resource region to be shared by the other user group. Comparison between these two schemes may be useful for characterizing the performance of soft reservation.

In case i) of Fig. 4, SFR-FCR and SFR-FER have the 5th percentile SINR gain of 0.3dB and 0.5dB over that of Static SFR with $\alpha = 0$. These 5th percentile SINR gains are attributed to the multi-user diversity gain obtained by soft reservation, which is supported by observation that neither SFR-FCR nor SFR-FER show any SINR gain with $\alpha = 0$. Note that SFR-FER has the better SINR gain with $\alpha = 0$, which is due to the fact that wider resource region is used by those in the cell-edge user group, improving the multi-user diversity gain. As α increases, however, the SINR gain for SFR-FCR improves while that for SFR-FER decreases. The performance difference between SFR-FCR and SFR-FER is attributed to the situation that the low power P_L can be allocated to resource region for the cell-edge users with SFR-FCR , warranting the SINR better than (3) for the cell-center users, while the high power P_H can be allocated to resource region for the cell-center users with SFR-FER, degrading the SINR worse than (3) for the cell-center users

 ISBN: 978-1-61208-133-5



(a) Case i (b) Case ii (c) Case iii (d) Case iv

Fig. 5. Trade-off relation: 5th percentile vs. total throughput

in each cell.

Furthermore, SFR-FCR and SFR-FER show the opposite characteristics in the average total throughput performance. In general, SFR-FCR tends to improve the total average throughput, while SFR-FER tends to improve the 5th percentile throughput. The difference can be explained as follows. The total average throughput for SFR-FCR improves as more resource is allocated to $\mathcal{U}_C$ by soft reservation of the resource allocated to $\mathcal{U}_E$. As shown in case i) of Fig. 5, therefore, it outperforms the Normal scheme in the total average throughput, while reducing the 5th percentile throughput. By the similar reason, meanwhile, the average throughput for $\mathcal{U}_E$ improves with SFR-FER, which subsequently improves the 5th percentile throughput as the low throughput users tend to belong to $\mathcal{U}_E$. The total average throughput of SFR-FER is much worse than Normal and Static SFR, as the resource allocated to $\mathcal{U}_C$ is reduced.

In conclusion, we find that soft reservation only for one user group improves the performance of the user group subject to hard reservation. Furthermore, soft reservation of the resource for the lower SINR user group is more effective for throughput trade-off.

C. Performance of soft reservation: SFR-SR

In Fig. 4 and Fig. 5, the SFR-SR scheme turns out to provide the best trade-off performance. In Fig. 4(a), the difference among the largest SINRs of Static SFR, SFR-SER, and SFR-SR is 0.64dB, while this is 0.01dB in Fig. 4(b). It implies that SFR-SR has the larger multi-user diversity gain than SFR-FER and Static SFR, while maintaining almost the same level of SINR gain with power control for each user group as that with hard reservation in (2). Note that a soft reservation feature in SFR-SR cannot warrant the SINR gain in (2) only by the power control, without any means of inter-cell interference control. Nevertheless, it demonstrates a rather acceptable performance gain, which implies that performance degradation caused by collision is not significant. In fact, Fig. 6 shows the collision probability as varying α. Note that the collision probability of SFR-SR is twice as large as that of Static SFR, demonstrating their maximum difference of 15%. The SINR loss by the collision can be at most 2.4dB in all SFR schemes, which corresponds to reduction in the bandwidth efficiency by 1/2 or less in the low SINR region. Therefore, the collision incurs reduction in bandwidth efficiency with the probability of 0.15 or less, which is also true for the cell-edge user group.

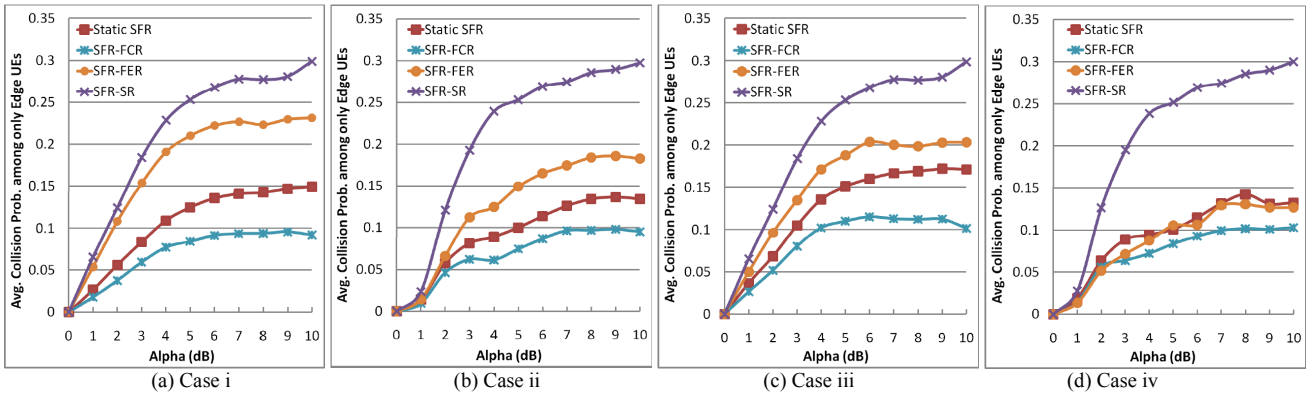
In general, downlink interference is usually dominated by a small number of base stations only [10]. In other words, a change in total interference is mainly governed by that in the dominant interferers. When α is small as compared to the change in the channel gain, therefore, the effect of α on total interference can be negligible. When the gain of interference channel is large enough, interference is still large for any level of power over each scheduling period. Therefore, other users with the lower gain of interference channel must be allocated to the corresponding resource. The current explanation can be supported by the results in Fig. 6, which shows that the collision probability increases with α.

Meanwhile, we note that SFR-SR is free from any imbalance in resource allocation between each user group, which is caused by hard reservation, since the required resource for each user group is dynamically determined by user scheduling. Such a desirable feature supports the fact that SFR-SR provides the best trade-off relation for average throughput as shown in case iv) of Fig. 5.

V. CONCLUSION

We have shown that a soft reservation-based SFR scheme can be a useful alternative to the variants of the hard reservation-based conventional SFR scheme. More specifically, it can be more flexible and adaptive to the varying user distribution in practice. Performance analysis with our simulation for LTE system has demonstrated that the proposed soft-reservation approach performs fairly well, even if there is no explicit coordination among the neighbor cells. In fact, a virtual coordination is realized by avoiding the inter-cell interference dynamically, yet in a rather long-term basis. Furthermore, it fully exploits a multi-user diversity over a whole frequency band, as opposed to the

 ISBN: 978-1-61208-133-5



(a) Case i (b) Case ii (c) Case iii (d) Case iv

Fig. 6. Collision probability among only edge UEs

hard reservation-based SFR scheme, in which a multi-user diversity gain is limited by partitioning the resource regions in an orthogonal manner. As the overall performance depends on how frequently interference measurement is made in each cell, our future work will investigate the validity of the soft reservation-based SFR scheme upon the system dynamics and user distribution.

ACKNOWLEDGMENT

This work is supported in part by the Seoul R&BD Program [WR080951, Establishment of Bell Labs in Seoul / Research of Services Application for Broadband Convergent Networks and their Enabling Sciences] and in part by the IT R&D program of MKE[KI001822]/KCC[08913-05001], Research on Ubiquitous Mobility Management Methods for Higher Service Availability.

REFERENCES

[1] Gary Boudreau, John Panicker, Ning Guo, Rui Chang, Neng Wang, and Sophie Vrzic, “Interference Coordination and Cancellation for 4G Networks,” IEEE Communi-cations Magazine, April 2009

[2] R1-060135, Siemens, "Interference Mitigation by Partial Frequency Reuse," TSG-RAN WG1 Ad Hoc Meeting on LTE, 23 - 25 January 2006

[3] R1-050841, Huawei, “Further Analysis of Soft Frequency Reuse Scheme,” 3GPP TSG RAN WG1#42, August 29 – September 2 2005

[4] Gábor Fodor, Chrysostomos Koutsimanis, András Rácz, Norbert Reider, Arne Simonsson, and Walter Müller, “Intercell Interference Coordination in OFDMA Networks and in the 3GPP Long Term Evolution System,” Journal of Communications, Vol 4, No 7, pp. 445-453, Aug 2009

[5] Guoqing Li, and Hui Liu, “Downlink Radio Resource Allocation for Multi-Cell OFDMA System,” IEEE Transaction on Wireless Communication, Vol 5. No. 12, Dec. 2006

[6] R1-080331, Nokia Siemens Networks and Nokia, "Performance analysis and simulation results of Uplink ICIC," 3GPP TSG RAN WG1 #51bis, January 14-18, 2008

[7] IEEE802.16m-07/004r4, “IEEE802.16m Evaluation Methodology Document (EMD),” November 2008

[8] TS 36.211, Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channel and Modulation (Release 8), Dec. 2009

[9] TS 36.213, Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures (Release 8), Dec. 2009

[10] R1-072444, Ericsson, “Summary of Downlink Performance Evaluation,” 3GPP TSG-RAN WG1 #49, May 7 – 11, 2007

 ISBN: 978-1-61208-133-5

A Distributed Cooperative Trust Based Intrusion Detection Framework for MANETs

Sureyya Mutlu
Computer Engineering Department
Turkish Air Force Academy
Istanbul, Turkey
s.mutlu@hho.edu.tr

Guray Yilmaz
Computer Engineering Department
Turkish Air Force Academy
Istanbul, Turkey
g.yilmaz@hho.edu.tr

***Abstract*— Mobile Ad Hoc Network (MANET) is a collection of nodes, which form an infastructureless topology. There is no central access point or centralized management. For their nature, MANETs present a number of unique problems for Intrusion Detection Systems (IDS). This paper introduces an intrusion detection framework for MANETs, which is based on trust relationship. In our proposed framework, intrusion detection system relies on local and global determination of attacks within network and carried out in a distributed fashion with cooperation among nodes. Trust in this manner is an important issue. The nodes watch suspicious activities of neighboring nodes. An intrusion detection alert message is disseminated throughout the network to report the anomaly. Reputation of intrusion detection alert messages is used for trust assessment. The proposed framework aims to utilize a distributed and cooperative trust based intrusion detection system to cope with the disadvantages drawn from mobility of nodes and the probability of selfishness, which are unique to MANETs.**

Keywords*— *Mobile Ad Hoc Networks; Trust Management; Intrusion Detection Systems.

I. INTRODUCTION

Mobile Ad Hoc networks (MANETs) have received considerable attention in recent years. A mobile ad hoc network is a collection of autonomous nodes, which form an infastructureless topology. The network topology dynamically changes as nodes join and move out of the network. There is no central access point or centralized management. Routing and other network operations are carried out by individual nodes. Each node acts as a wireless router and routes packets to neighbor nodes to reach intended destination. Therefore, ad hoc networking has been proven to be a promising solution to increase the radio coverage of broadband wireless systems in an infastructureless fashion.

MANETs are ideally suited for applications where such infrastructure is either unavailable or unreliable. Typical applications include military communication networks in battlefields, emergency rescue operations and environmental monitoring [1].

The wireless characteristics of transmission medium imply limited bandwidth and high error rate in radio transmission. Thus, the key point in designing a protocol for MANETS is the effective use of bandwidth. On the other hand, mostly, MANETs are formed of battery-powered devices such as laptops, PDAs and so on in which power consumption is critically important. Moreover, the availability of an individual node cannot be assured and therefore, services cannot rely on a central entity and must be provided in a distributed and adaptive manner [2]. MANETs need well-organized distributed algorithms to determine network organization, link scheduling, and routing.

Because network topology can change at any point of time, conventional routing will not work in MANETs. Ad hoc routing protocols can be classified into two types; proactive and reactive. In proactive protocols nodes in a wireless ad hoc network keep track of routes to all possible destinations. The route is identified in advance. In case of a topology change, this modification needs to be disseminated throughout the entire network. On the other hand, reactive protocols will figure out the routes when required by the source node, as needed. When a node needs to send packets to several destinations but has no route information, it will start a route detection process within the network. Routing protocols in ad hoc networks need to deal with the mobility of nodes and constraints in power and bandwidth [3].

Due to their nature, MANETs are more vulnerable to security attacks than wired networks. Security in wireless ad hoc networks is principally difficult to maintain, particularly because of the limited physical protection of each individual node, the irregular characteristics of connectivity, the lack of certification authority, centralized monitoring or management. Unlike wired networks where an adversary must gain physical access to the network or pass through several lines of defense at firewalls and gateways, attacks on a wireless network can come from all directions and target at any node. That is why every node must be prepared for attacks directly or indirectly. Additionally, an attack from a compromised node within the network is far more damaging and much harder to detect [2].

MANETs are subject to passive and active attacks. The passive attacks typically involve only eavesdropping of data, whereas the active attacks involve actions performed by adversaries such as replication, modification and deletion of exchanged data. Specifically, attacks in MANET can cause

 ISBN: 978-1-61208-133-5

congestion, propagate incorrect routing information, prevent services from working properly or shutdown them completely [4].

MANETs present a number of unique problems for Intrusion Detection Systems (IDS). Monitoring traffic by promiscuously within wireless radio range is a limiting factor in IDS on MANETS. Another problem is the mobility of the nodes. Additionally, a node in ad hoc network is more vulnerable to compromise. Also, because of the dynamic network topology of MANETs, an IDS may not be able to obtain enough sample data for accurate intrusion detection [5].

In this paper, we proposed a trust based distributed and collaborative intrusion detection framework for MANETs. Section 2 summarizes related work and proposed IDSs on MANETS relevant to our approach. Our proposed model is presented in Section 3, and finally, conclusion and future work are given in Section 4.

II. RELATED WORK

Because of their own characteristics, IDSs for traditional wired networks do not suit well for MANETs. There have been several proposals for Intrusion Detection Systems on MANETs [6-18].

One general approach for IDS on MANETs is distributed and cooperative architecture. In this architecture all nodes in MANET have their own local IDS system. Nodes come to a decision in a distributed fashion cooperatively. Upon determination of an intrusion, nodes share this information, asset attack risk degree and take necessary actions to eliminate the intrusion using active or passive precautions.

Other IDS architectures in MANETs are stand-alone and hierarchical IDSs. In stand-alone architectures every node performs IDSs locally without collaborating and respond locally. This IDS architecture has a drawback for network attacks. In hierarchical IDS architectures, MANETs are grouped into clusters or zones. One of the nodes in a zone/cluster is responsible for IDS. IDS is carried out in a distributed fashion and with collaboration with other clusters/zones. The main advantage of this architecture is effective use of constraint resources but has a drawback for highly mobile MANETs for establishing zones and detecting responsible nodes in clusters.

A. *Distributed and Cooperative IDS*

The first IDS for MANETs proposed by Zhang and Lee is a distributed and cooperative IDS [1][6]. In this architecture, each node detects intrusions locally and come to a decision globally if the local evidence for a network attack is inadequate. Respond may be local or global depending on the coordination among neighborhood nodes.

Statistical anomaly-based detection is preferred since rules cannot be updated in a wireless ad hoc environment over misuse base detection. The statistical anomaly-based detection composes the local data for IDS.

A multi-layer intrusion detection and response is proposed allowing different attacks at the most effective layer. It is believed to achieve a higher detection rate with a lower false positive rate.

Ad Hoc On-Demand Distance Vector [19] (AODV), Dynamic Source Routing [20] (DSR) and Destination-Sequenced Distance Vector [21] (DSDV) algorithms are used to have a better rate and false alarm rate metrics.

The system is reliable if the majority of the nodes are not compromised [1]. Additionally, the collaborative detection mechanism is susceptible to denial of service and spoofed intrusion attacks.

B. *Zone Based Intrusion Detection System*

Sun B et al. proposed a non-overlapping zone-based IDS [22]. In this architecture, the network is divided into zones based on geographic partitioning to save communication bandwidth while improving detection performance by obtaining data from many nodes. The nodes in a zone are called *intrazone nodes*, and the nodes which work as a bridge to other zones are called *interzone (gateway) nodes*. Each node in the zone is responsible for local detection and sending alerts to the interzone nodes. Their framework aims to allow the use of different detection techniques in each IDS agent.

Intrazone nodes carry out local collection and correlation, while gateway nodes are responsible for global collection and correlation to make final decisions and send alarms. Therefore, only gateway nodes participate in intrusion detection. The alerts sent by interzone nodes simply show an assessment of the probability of intrusion; the alarms generated by gateway nodes are based on the combined information received. In their aggregation algorithm, gateway nodes use the following similarities in the alerts to detect intrusions: classification similarity (classification of attacks), time similarity (time of attack happening and time of attack detection) and source similarity (attack sources). Source similarity is the main similarity used, so the detection performance of aggregation algorithm could decrease with increasing number of attackers.

The advantages of an aggregation algorithm using the data from both partial and full victims are emphasized: lower false positive and higher detection rate than local IDS achieves. Nevertheless, its performance can decrease with the existence of more than one attacker in the network. They also conclude that communication overhead is increased in proportion to mobility where local IDSs generate more false positives and send more intrusion alerts to gateway nodes. In addition, aggregating data and alerts at interzone nodes can result in detection and response latency, when there is sufficient data for intrusion detection even at intrazone nodes.

C. *General Cooperative Intrusion Detection Architecture*

A cooperative and dynamic hierarchical IDS architecture, which uses multiple-layering clustering, is proposed by Sterne et al. in [23]. At the beginning, the nodes are assigned to clusters and first level clusters act as a management focus for IDS activity of immediate surrounding nodes. Then,

 ISBN: 978-1-61208-133-5

these first level clusters form a second layer clusters. This process goes on until all nodes are assigned to a cluster. To avoid single point of failure, they propose choosing more than one cluster-head for the top-level cluster. The selection of cluster heads is based on topology and other criteria including connectivity, proximity, resistance to compromise, and accessibility by network security specialists, processing power, storage capacity, energy remaining, bandwidth capabilities and administratively designated properties.

In this dynamic hierarchy, data flow is upward, while the command flow is downward. Data are acquired at leaf nodes and aggregated, reduced and analyzed as it flows upward. The key idea is given as detecting intrusions and correlating with other nodes at the lowest levels for reducing detection latency and supporting data reduction, even as maintaining data sufficiency. It supports both direct reporting by participants and promiscuous monitoring for correlation purposes.

This architecture targets military applications with high scalability and reduced communication overhead through hierarchical architecture [23]. However, the cost of configuration of the architecture in dynamic networks should also be considered.

D. Intrusion Detection Using Multiple Sensors

Kachirski and Guha propose an IDS solution based on mobile agent technology [24], which reduces network load by moving computation to data. This is a significant feature for MANETs that have lower bandwidth than wired networks.

Proposed IDS structure distributes the functional tasks by using three mobile agent classes: monitoring, decision-making and action-taking. The advantages of this structure are given as increased fault tolerance, communication cost reduction, improved performance of the entire network and scalability.

Hierarchically distributed IDS architecture divides the network into clusters. Cluster-heads are selected by voting for a node, which is based on its connectivity. Each node in the network is responsible for local detection. Only cluster-heads are responsible for detection using network-level data and for making decisions. Cluster nodes can respond to the intrusions directly if they have strong evidence locally. If the evidence is insufficient, they leave decision-making to cluster heads by sending anomaly reports to them.

In this proposal although, a scalable and bandwidth-efficient IDS is proposed by using mobile agents, but security issues for mobile agents are need to be investigated.

E. DEMEM: Distributed Evidence Driven Message exchanging ID Model

DEMEM [25] is a distributed and cooperative IDS in which each node is monitored by one-hop neighbor nodes. In addition to one-hop neighbor monitors, 2-hop neighbors can exchange data using intrusion detection (ID) messages. The main contribution of DEMEM is to introduce these ID messages to help detection, which they term evidence-driven message exchange.

Evidence is defined as the critical information (specific to a routing protocol) used to validate the correctness of the routing protocol messages, for instance, hop count and node sequence number in AODV. To minimize ID message overhead ID messages are sent only when there is new evidence, (it is called evidence-driven). DEMEM also introduces an ID layer to process these ID messages and detect intrusions between the IP layer and the routing layer without modifying the routing protocol, so it can be applied to all routing protocols.

DEMEM uses the specification-based IDS model. There are nodes called Multipoint Relays (MPRs), which serve to reduce the flooding of broadcast packets in the network. These nodes are selected by their neighboring nodes called MPR selectors. The packets of an MPR node's MPR selectors are only retransmitted by that MPR node. Topology control messages are sent by each node periodically to declare its MPR selectors.

DEMEM cannot detect collaborative attacks. For example, two attackers who falsely claim that they are neighbors might not be detected by the above constraints.

DEMEM introduces three authenticated ID messages for Optimized Link State Routing Protocol [20] (OLSR).

- The first message is ID-Evidence, which is designed for two-hop-distant detectors to exchange their evidence concerning one-hop neighbors, MPRs and MPR selectors on OLSR.
- The second message, ID-Forward, is a request to forward any held ID-Evidence messages to other nodes. This means that a node can request the holder of evidence to forward it directly, rather than sending it itself, so reducing message overhead.
- The last message, ID-Request, is designed to tolerate message loss of ID-Evidence with low communication overhead. The false positives and delay detection due to message loss are decreased by an ID-Request message. Moreover, they specify a threshold value to decrease false positives due to temporary inconsistencies resulting from mobility. When a detector detects an intrusion, it automatically seeks to correct the falsified data.

III. DICOTIDS: DISTRIBUTED COOPERATIVE TRUST BASED INTRUSION DETECTION ARCHITECTURE FOR MANETS

In this section, we propose a distributed cooperative trust based intrusion detection architecture for MANETs. The architecture is based on running Local Intrusion Detection engines in each node independently. The objective is to monitor all network activity within wireless range to detect misbehaving nodes on promiscuous mode. That means, if node A is in wireless range of node B, it can watch communication activity to and from B even node A is not involved in. Accruing intrusion detection data in this manner has significant advantage. First, it allows local data collection without consuming any additional communication overhead. Second, it provides first hand observations, which

 ISBN: 978-1-61208-133-5

means no need to rely on observations from other nodes, which might be false.

Moreover, intrusion detection is distributed throughout the network in case of weak or inconclusive evidence of anomaly. A global investigation is initiated to support local intrusion detection.

Flooding algorithm is used to share IDS alert messages. Flooding is the mechanism by which a node receives a flooded message for the first time, it rebroadcasts that message once. Each node is responsible to deliver the message to its neighbor within wireless transmission range.

DICOTIDS mainly focus on detecting compromised modes in network. A compromised node can disseminate false IDS alert messages or drop the IDS alert message flooded by other nodes. Therefore, a trust mechanism is established in the network. Trust management can mitigate nodes' selfish behaviors', such as dropping messages or unwillingness for cooperation. Reputation mechanism is used as a dynamic rating system.

Once, a node detects misbehavior of a neighbor node or suspicious activity, it starts a distributed IDS algorithm by broadcasting IDS alert messages. Nodes periodically share their respective data by flooding algorithm and then start a diagnostic phase. After the diagnostic phase in which all collected data from other nodes are compared, trust evaluation phase starts. If a trustworthy node broadcast an IDS alert message, intrusion response is activated even if the relevant node is not directly involved in IDS assessment. Trust management is maintained by watching the neighbor nodes activities whether they rebroadcast the IDS alert messages or not. A reputation mechanism is used to evaluate the trust level of a node. Figure 1 depicts the components of the framework.

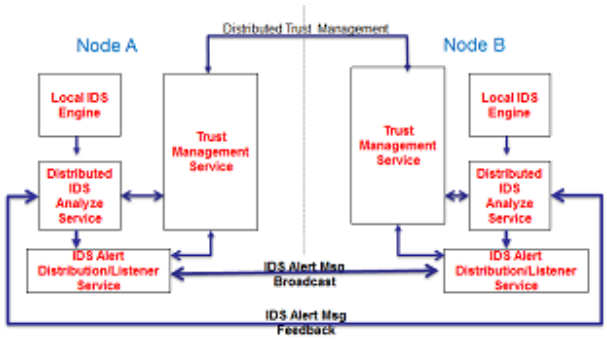


Figure 1. Components of DICOTIDS

The details of the framework are described as follows:

A. *Local IDS Engine:*

The first phase of the intrusion detection process starts at Local Intrusion Detection engine. It sniffs the neighbor nodes network activity in promiscuous mode. The engine runs a popular network-based IDS, which is the open-source Snort [26]. Snort is able to sniff the network activity in promiscuous mode and configured with a rule set it can function as a real-time IDS. A Snort rule set is a file of attack signatures. A match to a signature means that an attack is recognized. Each node assumed to have the database of these rule sets and functions as a real-time detection system.

Once an intrusion attempt or a suspicious activity is determined, all relevant data is passed to distributed IDS analyze service.

B. *Distributed IDS Analyze Service:*

IDS analyze service will use outputs of the Local IDS engine as well as IDS alert messages disseminated from other nodes. If there is enough evidence for intrusion, this service will put intrusion prevention measures into effect and forward the related information to IDS alert distribution service to inform the other nodes in the network. If there is weak or inconclusive evidence of anomaly IDS analyze service will request global analysis. Only the replies from the trusted nodes will be taken into consideration.

The functional diagram of Distributed Analyze Service is depicted in Figure 2. The service will also try to verify the attack by additional IDS Alert messages originated from other nodes in the network.

If the evidence comes via IDS alert message from another node in the network, first the trust level of the sender node is checked and;

- If the message is from a trusted node and there is more than one trusted node disseminating IDS alert message, than there is strong evidence for an intrusion attempt.
- If the IDS alert message is from an untrustworthy node, the IDS message is ignored.
- If the message is from a node, which the trust level has not been evaluated yet, then special interest is performed.
- If the intrusion alert is supported more than a single (trust level undecided) node or an intrusion is also approved by local IDS, the service may conclude of an intrusion.

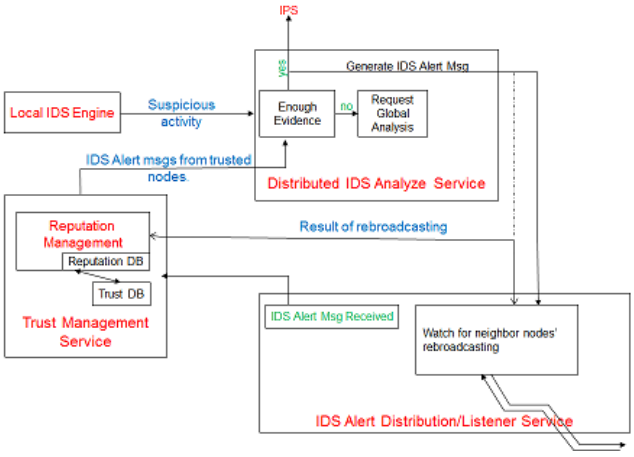


Figure 2. Functional Diagram of Distributed Analyze Service

Once the service concludes for an intrusion, first it will inform the Intrusion prevention module to take necessary actions in order to prevent intrusion. The next is to pass this information to IDS alert distribution service. The information

 ISBN: 978-1-61208-133-5

includes nodes involved in the intrusion attempt, type of attack, priority, strength, timestamp etc.

The last thing is to inform the trust management service to downgrade or set the trust level of the involving node to untrustworthy.

C. IDS Alert Listener / Distribution Service:

This service is responsible to broadcast the IDS alert messages within wireless radio range and watches for the neighbor nodes if they rebroadcast the message within a time frame. Each message will have a unique message number and detected intrusion related information. IDS alert message contains:

- Originator ID and Originator Message ID (null if produces for the first time)
- Sender Node ID
- Sender Message ID
- Compromised/Attacker node's ID/IP
- Attack Type
- Classification
- Priority
- Date/time

Immediately after, this service will inform the trust management service to evaluate reputation values. If the neighbor nodes rebroadcast the IDS alert message without any modification, trust management service will perform the reputation update procedures accordingly. In addition, if this does not occur in a limited time frame or the rebroadcasted IDS alert message is corrupted then reputation and trust assessment is evaluated as described below.

IDS listener service sniffs the neighbor node's activities in promiscuous mode for the rebroadcasted messages. Upon receipt of an IDS alert message, the message is passed to distributed IDS analyzer and trust management service.

D. Trust Management Service:

Trust management service is responsible to maintain relationships among nodes in the network. This service will mitigate misbehaving of nodes and enforce cooperation. Projected trust management is derived form a reputation based scheme proposed by Jiangy hu [27]. Figure 3 depicts the components of the service.

Trust in a node is associated with its reputation value. There are three trust levels and we use a trust value T, to represent the trustworthiness of a node. A node considers another node B either

- Trustworthy, with T = 1,
- Untrustworthy, with T = -1, or
- Trustworthy undecided, with T = 0

A trustworthy node is a well-behaved node that can be trusted. An untrustworthy node is a misbehaved node and should be avoided in distributed IDS evaluation process. A node with undecided trustworthiness is usually a new node in the network and special interest should be taken in IDS evaluation process.

Each node keeps a reputation table, which associates a reputation value with each of its neighbors. It updates the table on direct observation only. Reputation value of a neighbor node will not be distributed globally and will be stored locally. Reputation values will be shared only if requested by other nodes.

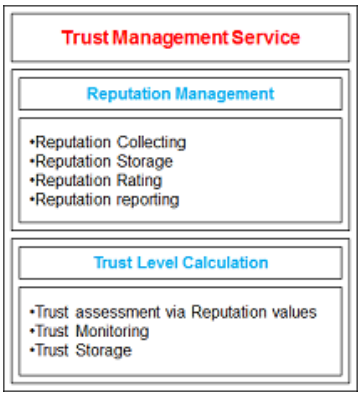


Figure 3. Components of Trust management Service

Reputation values R are between a range of $0 \leq R \leq 1$ and there is one threshold R_t,

$R \geq R_t$ for trustworthy and

$R < R_t$ for untrustworthy.

For a new node N with reputation value R and trust value T,

- $T = 1$, if $R \geq R_t$
- $T = -1$, if $R < R_t$
- $T = 0$, if $R < 0$

Reputation values depend on the behaviors of the node. If a node broadcasts an IDS alert message, then it sniffs the neighbor nodes in promiscuous mode. If that node rebroadcasts the IDS alert message, the originator node promotes the reputation value for that node; otherwise, the reputation value is downgraded. If the rebroadcasted message is modified the nodes trust value will be in untrustworthy state. *R* is the proportion of the total number of forwarded messages to the total number of sent messages.

Each node keeps track of the neighbor nodes and establishes reputation values directly. If a node needs to query a specific node that is beyond the wireless radio range, it will ask for reputation values to all the trusted nodes in the network. The average of the replies will set the reputation value for the requested node.

Another factor for a node that will affect it is trust level is the correctness of the IDS alert message. All the nodes that receive an IDS alert message will also monitor the evidences. If there is not enough evidence, the IDS message is concluded to be false. So that the trust level for the disseminating false messages node will be untrustworthy.

E. Pseudo Code for DICOTIDS :

- ***Local IDS Engine (LIDS)***

```
Watch for Neighboring Node's Network Traffic
Compare Net Traffic with IDS Signature Database
If Network Activity Matches IDS Signature
  Create IDS Alert Msg
  Pass IDS Alert Msg to Dist. IDS Analyze Service
```

 ISBN: 978-1-61208-133-5

Inform Trust Management Service
Endif

- ***Distributed IDS Analyze Service***

For each IDS Alert Msg received form LIDS do
If there is strong evidence
Activate IPS
Forward IDS Alert Msg to Distribution Service
Inform Trust Management
Else
Request Global Analyze
Endif
For each IDS Alert Msg received from the network
Check Trust Level of the sender
If the sender is a trusted node
Activate IPS
Forward IDS Alert Msg to Distribution Service
Else
Ignore message
Inform Trust Management
Endif
If the sender's trust level is not assigned
If there is more than one sender
Activate IPS
Forward IDS Alert Msg to Distribution Service
Inform trust Management
Request Global Analyze for confirmation
Else
Request Global Analyze
Endif

- ***Distribution/Listener Service***

Broadcast IDS Alert Message
Listen for the neighboring nodes to rebroadcast
Inform Trust Service for successful rebroadcasts
Inform Trust Service for unsuccessful rebroadcasts

- ***Trust Management Service***

Evaluate the reputation value for each neig nodes
For each neighboring node
If reputation value is greater than the threshold
Assign node's Trust Level as Trustworthy
Else
Assign node's Trust Level as Untrustworthy
Endif
Update databases respectively

IV. SIMULATION AND PEFFORMANCE ANAYSIS

The objective of simulation and performance analysis is to determine the feasibility of DICOTIDS in MANETs where there are a number of malicious nodes. The metrics to evaluate the performance are described below.

A. Metrics

IDS Alert Message Delivery Ratio: The ratio of the IDS alert message delivered to the destination nodes. The delivery ratio is directly affected by uncooperative behavior, number of malicious nodes, and packet loss.

Message Overhead: The ratio of redundant messages to the total number of messages relevant to the IDS instance.

The number of IDS instances to evaluate the trust level of nodes: The reputation rates are directly involved in evaluating the trust level of a node in coherence with the reputation threshold value.

B. Simulation and Results

We have partially simulated the DICOTIDS in Network Simulator (NS-2) [28]. For all the metrics we want to evaluate, we used fixed parameters for network environment. Also, we assumed that every node has a Local Intrusion Detection System (LIDS) with updated signatures.

We ran the simulation for two scenarios.

Scenario #1: Few ($n<3$) malicious nodes with a total number of 15 nodes.

Scenario #2: More ($n>7$) malicious nodes with a total number of 15 nodes.

C. Analysis

As the number of malicious nodes in the network increase, the IDS alert delivery ratio is decreased proportionally. The layout and the mobility of the nodes have an impact in the ratio also. However, mostly this ratio satisfied the requirements of the whole system.

In some cases, especially with a high dense node layout, several nodes initiated the distributed IDS analysis process for the same instance. Because the reputation values and trust levels are stored and evaluated locally, the disharmony among nodes resulted in the increase of redundant message. However, this did not have a crucial effect on the total performance.

In order to determine untrusted nodes and successfully identify malicious nodes, a number of intrusion instances are required. On the first occasion of an intrusion attempt, nodes need to rely on local intrusion detection system (LIDS). But, as the number of instances increase, accurate reputation values and trust levels are evaluated respectively.

Additionally, the reputation threshold value (R_T) should be set to lower values for fixed or low mobile networks rather than the value for highly mobile networks.

The proposed framework should be feasible for networks with nodes with low mobility. On the other hand, we assumed that all nodes have the same emitting power. That means with different emitting powers, reputation mechanism may fail for the event that node B is in the range of node A, but node A is not in the range of node B.

V. CONCLUSION

A trust based distributed intrusion detection framework is proposed in order to protect nodes from performing misbehavior or selfish behavior in MANETS. Trust, in the framework, is mainly based on direct observation, but indirect observations are also applied. The proposed infrastructure provides robustness against the propagation of false trust information by malicious nodes.

A dynamic and collaborative ad hoc intrusion detection system has been proposed. Our approach does not modify or restrict the network discovery or routing protocols. The concepts discussed in this paper are in broad sense that they can easily be integrated to existing routing protocols.

 ISBN: 978-1-61208-133-5

We aim to fully simulate the framework in NS-2 [28], an open source network simulator. The message overhead and resistance to intrusion in relevant to the number of compromised nodes in the network is critical. In addition, the effects of the mobility of the nodes in the network need to be observed. Additionally, testing the model using different routing protocols will conclude valuable data.

REFERENCES

[1] Y Zhang and W.Lee, Intrusion Detection In Wireless Ad Hoc Networks. In proc of the 6th Int Conf on Mobil Comput and Netw (MOBICOM), 2000, pp. 275-283

[2] S. Sen and J.A.Clark, Intrusion Detection in Mobile Ad Hoc Networks, Guide to Wireless Ad Hoc Networks, Computer Communications and Networks, 2009, pp. 441-442

[3] K. Pathan and C.S. Hong, Routing in Mobile Ad Hoc Networks, Guide to Wireless Ad Hoc Networks, Computer Communications and Networks, 2009, pp.63-66.

[4] L. Zhou and Z. Haas, Securing Ad Hoc Networks, IEEE Network Magazine, vol. 13, no. 6, 1999. pp.21

[5] P. Brutch and C. Ko, Challenges in Intrusion Detection for Ad Hoc Networks, IEEE Workshop on Security and Assurance in Ad hoc Networks, Orlando, FL, January 28, 2003.

[6] Y. Zhang and W. Lee, Y-A. Huang, Intrusion detection techniques for mobile wireless networks, Wireless Networks, vol. 9, 2003, pp. 545-556.

[7] R. Ramanujan, A. Ahamad, J. Bonney, R. Hagelstrom, and K. Thurber, Techniques for intrusion-resistant ad hoc routing algorithms (TIARA), IEEE MILCOM 2000, Los Angeles, 2000, pp. 660-664.

[8] S. Buchegger and J-Y. Le Boudec, Performance analysis of the CONFIDANT protocol (Cooperation of Nodes: Fairness in Dynamic Ad-hoc Networks), 3rd ACM Int. Symp. on Mobile Ad Hoc Networks and Computing, Switzerland, 2002, pp. 226-236.

[9] M. Kuchaki Rafsanjani, A. Movaghar, and Faroukh Koroupi, Investigating Intrusion Detection Systems in MANET and Comparing IDSs for Detecting Misbehaving Nodes, World Academy of Science, Engineering and Technology 44, 2008, pp.351-355.

[10] R. Puttini, J-M. Percher, L. Me, and R. de Sousa, A fully distributed IDS for MANET, 9th Int. Symp. on Computers and Commun. (ISCC 2004), 2004, pp. 331-338.

[11] G. Vigna, et al., An intrustion detection tool for AODV-based ad hoc wireless networks, Annual Computer Security Applications Conf. (ACSAC 2004), Tuscon, 2004, pp. 16-27.

[12] A. Pirzada and C. McDonald, Establishing trust in pure ad-hoc networks, 27th Australian Conf. on Computer Science, Dunedin, New Zealand, 2004, pp. 47-54.

[13] Y. Rebahi, V. Mujica, and D. Sisalem, A reputation-based trust mechanism for ad hoc networks, 10th IEEE Symp. on Computers and Communications (ISCC 2005), 2005, pp. 37-42.

[14] A. Karygiannis, E. Antonakakis, and A. Apostolopoulos, Detecting critical nodes for MANET intrusion detection, 2nd Int. Workshop on Security, Privacy, and Trust in Pervasive and Ubiquitous Computing (SecPerU 2006), 2006, pp. 7-15.

[15] H. Yang, J. Shu, X. Meng, and S. Lu, SCAN: self-organized network-layer security in mobile ad hoc networks, IEEE J. on Sel. Areas in Communications, vol. 24, 2006, pp. 261-273.

[16] T. Chen and V. Venkataramanan, Dempster-Shafer theory for intrusion detection in ad hoc networks, IEEE Internet Computing, vol. 9, 2005, pp. 35-41.

[17] D. Subhadrabandhu, S. Sarkar, and F. Anjum, A framework for misuse detection in ad hoc networks - part II, IEEE J. on Sel. Areas in Communications, vol. 24, 2006, pp. 290-304.

[18] Y. Zhang and W.Lee, Intrusion Detection Techniques for Mobile Wireless Networks, Wireless Networks 9(5), 2003, pp. 545-556.

[19] C. E. Perkins, E. M. Royer, and S. R. Das, Ad hoc on-demand distance vector (AODV) routing. July 2000.

[20] P. Jacquet, P. Muhlethaler, T. Clausen, A. Laouiti, A. Qayyum, and L.Viennot, Optimized link state routing protocol for ad hoc networks, in: IEEE International Multi Topic Conference, 2001, pp. 62–68.

[21] C. Perkins and P. Bhagvat, Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers. ACM Computer Communications Review, 1994, pp. 234-244.

[22] Sun B and Wu K et al. Zone-Based Intrusion Detection System for Mobile Ad Hoc Networks. Int J of Ad Hoc and Sens wireless Networks 2:3,2006

[23] D. Sterne and R. Balasubramanyam et al. A general Cooperative Intrusion Detection Architecture for MANETs. In Proc of the 3rd IEEE IWIA, 2005, pp.57-70

[24] O. Karchirski and R. Guha, Effective Intrusion Dtection Using Multiple Sensors in Wireless Ad Hoc Networks. In proc of the 36th IEEE Int Conf on Syst Sci (HICSS), 2003, pp.244-248

[25] C. Tseng and S.Wang, DEDEM: Distributed Evidence Driven Message Exchange Intrusion Detection Model for MANET. In proc of the 9th Int Symp on Recent Adv in Intrusion Detect LNCS 4219, 2006, pp.249-271

[26] Snort, <www.snort.org> 10.04.2011

[27] J. Hu and M. Burmester, Cooperation in Mobile Ad Hoc Networks, Guide to Wireless Ad Hoc Networks, 2009, pp.43-53.

[28] NS2, Network Simulator – 2, <www.isi.edu/nsnam/ns> 10.04.2011

 ISBN: 978-1-61208-133-5

Packet Tracer as an Educational Serious Gaming Platform

Ammar Musheer, Oleg Sotnikov and Shahram Shah Heydari
University of Ontario Institute of Technology, Oshawa, Canada
{musheer.ammar@gmail.com, oleg.sot@gmail.com, shahram.heydari@uoit.ca}

***Abstract*— Serious gaming is quickly becoming an important trend in education, as it provides an interactive as well as enjoyable environment for learning various technology, business and scientific topics. The main objective of this research is to explore the possibility of using Cisco packet tracer's multiuser capability to develop several interactive, multiuser games for teaching networking topics and assessing students skills in the class. We present a list of primary principles for design of such games, and design two multiuser educational games as in-class activities to showcase the benefits of our approach. We demonstrate how students' skills in important network topics such as routing, remote access and security help them succeed in these games. We also provide test results to evaluate the performance of packet tracer in handling multiuser traffic in this scenario.**

Keywords- Packet Tracer; Serious Gaming; Multiuser; Collaborative; Cisco Networking Academy; Emerging Teaching Methods; Simulation; Online Teaching

I. INTRODUCTION

A. Background

The maturing of the video gaming industry has triggered significant interest in developing simulated and interactive serious gaming platforms. The main advantage of serious gaming platforms is that they combine in-depth educational topics with goal-oriented and realistic scenarios. To date there have been many applications of serious gaming. These games have spanned over many industries, from urban planning to business, military, healthcare training etc. These applications have provided the means to an individual, or a group of individuals, to engage in an artificial conflict, assess, and learn the complexities associated with each of their individual area of work. Some real world applications of serious games include INNOV8 Business Process Management simulator by IBM [1], physical conditioning games for the general public, e.g. Fitness Shape Evolved for the new Microsoft Kinect hardware platform [2], Foldit for biologists [3], and flight simulator applications for the military [4]. Serious gaming has allowed users of the application to attain skills and knowledge related to a specific activity. The subject can be as difficult as addressing physical and physiological disorders or be as simple as promoting physical activity. The main goal of serious gaming is to provide a powerful means of encouraging people to learn and provide them with a more entertaining way to obtain the skills and knowledge addressed by the specific serious gaming activities.

During the research project we focused on creating serious games for the purpose of teaching topics in information networking at the University of Ontario Institute of Technology (UOIT) in Ontario, Canada. UOIT has an undergraduate program in networking and IT security, and also serves as a regional Cisco Network Academy. With growing interest in the networking field and the hands-on nature of the topic, the objective of this project was to make the classroom lecture hours more interactive.

The research team began by analyzing and utilizing the tools available through the Cisco Networking Academy. Cisco Packet Tracer's is most commonly used as an emulation platform for Cisco networking devices and IOS network operating system. The new version includes multiuser capabilities that enable designers to create large group based interactions during classroom hours. The Packet Tracer includes many necessary elements needed for CCNA-level education.

The research team explored the area of using packet tracer as an educational serious gaming platform by developing two gaming activities, Domination & Relay Race. Domination & Relay Race utilize the multiuser capabilities of packet tracer to the greatest extent. The two games provide students with a great educational value and help them hone their networking expertise. Packet Tracer provides a great interface to gain experience with practical configuration scenarios. Our aim was to combine the powerful capabilities of packet tracer with serious gaming in order to provide a simple and entertaining environment for teaching basic and advanced networking topics.

The rest of this paper is organized as follows: In Section II we examine some of the current serious games that focus on networking education. A basic description of the multiuser capability in Packet Tracer and our methodology is presented in Section III. We describe the design and technical details of our serious games in Sections IV and V. Some system performance results, e.g. memory, network and CPU requirements are presented in Section VI. We present our conclusions and future work plan in Section VII.

II. CURRENT NETWORKING SERIOUS GAMES

Cisco has developed a number of serious games for networking education. Most recently, Cisco ASPIRE Beta 3 has been released. This game utilizes the Packet Tracer platform to provide the players a role-playing/SimCity-like serious game focused on the business of networking. Players take on the role of a network engineer who applies his/her entrepreneurial and networking skills to complete several contracts that arise during the game. They must purchase the correct hardware and apply the correct configuration schemes in order to complete the contract. For the correct completion of contracts players receive credits which they can spend toward improving their network. Complete with a mobile device that rings when new

 ISBN: 978-1-61208-133-5

contracts are available, Cisco's Aspire immerses players into the networking world while providing an entertaining way to practice your networking and entrepreneurial skills.

Cisco Systems has also released Cisco myPlanNet 1.0 [5]. Players are put into the shoes of a service provider CEO. They must direct their business through various technological ages ranging from dial-up all the way up to the broadband/mobile connected ages. Along the way players learn about the various technologies that make the networking world possible. The game provides a SimCity-like overlay where issues arise in the city and must be resolved. There is a credit system that players must use to purchase new technologies to advance their businesses and provide better services to the civilians in their cities. There is also a leaderboard to encourage players to attain higher scores and improve their skills.

Additionally, Cisco provides many smaller serious games that allow users to improve their skills, such as: Wireless Explorer, Unified Communications Simulation Challenge, Cisco Mind Share, and Edge Quest. These serious games provide players with a means to enjoy learning the complicating subjects related to the networking field, albeit on a smaller scale.

III. GAME DESIGN

A. Brief Overview of the PT Multuser environment

One of Packet Tracer's greatest features is the ability to connect two or more lab sessions together, known as a Multiuser Connection. By utilizing Packet Tracer's multiuser capabilities very extensively throughout the term of the research project and within the completed serious gaming activities, the creation of two very unique and interactive Packet Tracer game activities was possible. Packet Tracer allows users to use a simple drag and drop cloud icon to connect to a peer cloud. Each multiuser cloud supports one-to-one, many-to-one and many-to-many peer-connection configurations. The activities that are created operate on a model similar to the client/server architecture observed widely today. The central file with the game scenario is hosted on an instructor PC and the student side file is used by each student to connect to the central Packet Tracer game file hosted on the instructor PC. This instructor/student architecture allows for easy management, control and provides an overall view for the instructor operating the game. The multiuser capabilities of Packet Tracer during testing allowed the connection of up to 60-75 users simultaneously to a single activity over LAN. The number of connections possible is directly limited by the hardware performance of the instructor PC. An in-depth examination of the multiuser environment was conducted and reported in a research paper submit last year, Multiuser Collaborative Practical Learning Using Packet Tracer [7], or refer to the research paper being submitted to ICNS this year, Building Interactive multi-user in-class learning modules for computer networking.

B. Project Objectives

The simulated-based games were created to provide students with a very different approach to networking. Goals and objectives had to be met in order to satisfy the CCNA course's academic requirements and the serious game aspects of the activities. The main objectives include:

- Easy deployment in large user environments
- Scalability to ensure reusability
- Provide an educational experience
- Can be used to assess student progression
- Have to provide a psychological perception of having won or loss

The psychological perception of having won or loss is one of the key elements that make up a serious game. The enthusiasm of a student is very crucial in a serious gaming aspect. It is an emotional driver that encourages students to improve his/her networking abilities in order to better compete against fellow classmates. Ensuring that the perception of having won or loss exist, improves our chances of encouraging students to participate and learn the skills and knowledge needed to complete the activities and course material more efficiently.

C. Methodology

Using the tools and interactive features provided by Packet Tracer 5.3, Domination and Relay Race simulation-based games were created to help teach and observe the students progression in the first year CCNA courses. The games covered topics learned throughout the whole year. Each activity tested the configuration and troubleshooting abilities of the players while providing an educational experience that was novel and entertaining for the students.

Each of the games presented students with artificial conflicts that resembled several real world networking problems that had to be fixed before achieving the game's end goal. DHCP mechanisms were used to provide each connected player with a set of unique IP address ranges for management and game play structure purposes. A detailed explanation of each game follows.

IV. DOMINATION GAME

A. Topology Structure

The game topology is broken down into four sections. Each section consists of layer 3 switches, multiuser clouds, and clusters. The Section switches all connect to a central Main Domination Switch. Each section consists of 15 cluster clouds and 15 multiuser clouds, all of which are connected to a Section Domination Switch. Each cluster cloud contains an identical network topology that presents students with a network problem. Each of the clusters have been assigned a /24 subnet, from which the first host IP is assigned to the default gateway. The DHCP IP addressing scheme has been kept simple, allowing for easy scalability and management. Student side Packet Tracer instances connect into the multiuser clouds assigned to them by the instructor. Multiuser clouds are distinguished by the Peer

 ISBN: 978-1-61208-133-5

Network Name property within each cloud. The game is initiated by the instructor when all players have obtained a connection to the peer multiuser cloud. Once all students clouds are active students begin by using the telnet protocol to obtain access into their specific cluster. Fig 1 shows an overview of the Domination topology.

B. Game-play

The student's main goal is to fix the network problems presented to them within their clusters. The problems within the cluster can vary in complexity, but initially the clusters have been set up with a basic problem. Once the cluster problems are solved the students must quickly telnet into the directly connected section switch and shutdown all other interfaces except the interface directed toward the main domination switch. Once a student has managed to dominate their sections switch, they must quickly telnet into the Main Domination Switch and shutdown the 3 other ports allowing other section switches to telnet into the main switch. The first person to quickly dominate their section's switch and the main switch is the winner. The other 3 that have managed to dominate only their sections switch are runner up's. It is possible for 1 person to dominate all 5 switches but for that the player will have to be really fast.

C. Domination Technical Details

The domination game comes with only one student side file that will work with all of the Multiuser clouds within the instructor file. The student file consists of a single workstation and a multiuser cloud to allow connectivity to the instructor file.

By providing a standard student file with basic configurations already applied to the workstation allows for easy deployment in large user environments. Furthermore, IP addresses are dynamically assigned by a DHCP located in each cluster. Having DHCP enabled on each cluster eliminates the need for students to configure an initial IP address to their student side Packet Tracer Workstation.

The cluster problems can be increased in complexity easily as needed. Additional devices can be added within a single cluster and be duplicated easily across the clusters because of the easy IP addressing scheme. The EIGRP protocol is running between the 5 switches to allow telnet capabilities to the students. IP addressing schemes were designed to be as simple as possible so to encourage future development within the activity. The number of vty lines available within each switch limits the section sizes to 15.

D. Educational Value

The domination game allows students to experience the pressure sometimes put on network engineers in the real world environment. It forces students to apply all of their learnt networking knowledge to a problem. By gaining the ability to combine the hands-on and theoretical knowledge learnt throughout a Cisco Network Academy course, students will gain a better understanding of how to apply these skills and tools in the networking world. Moreover, the game can also be used as an evaluation tool to see if the students understand the concepts delivered in the course.

V. RELAY RACE GAME

A. Topology Structure

Adapted from the original Relay Race game presented by Cisco, this Relay Race game incorporates new and old features. The topology is broken down into 4 sections, each consisting of 5 Main Line Routers, 4 network clusters and 5 multiuser clouds. The whole topology is brought together at a central Finish Line Router. Each cluster contains problematic network scenarios that students must correct in order to allow a Runner, designated in each team, to move closer to the Finish Line Router. The network scenarios within the clusters progressively become harder the closer the cluster is to the Finish Line Router. The 5 Main Line Routers act as doorways, locked, preventing the runner from moving forward. Fig 2 shows an overview of Relay Race topology.

B. Game-play

Although slightly more complicated, the concept of the game remains somewhat similar to the Domination game. Relay Race consists of 4 Teams each consisting of 5 team members. Each team consists of 1 Runner and 4 members responsible for solving the problematic network clusters. Once the problem within the cluster is solved they must move forward to their Main Line Router and no shutdown their routers Serial 0/0/0 interface. This will allow the runner of the team to telnet into their routers in order to move forward towards the Finish Line Router. Many ACL's have been put in place so that only the appropriate hosts can telnet into the appropriate devices. For example, none of the team members responsible for solving cluster problems can telnet into the Finish Line Router, only the runner will be able to telnet into that router. The ACL's also ensure that telnets to other Main Line Routers will only work from the Runner's Workstation on the student Packet Tracer file. The Goal of the game is to have the fastest team to no shutdown all of their S0/0/0 interfaces within the Main Line Routers. The Runner must then run (telnet) into the Finish Line Router before any of the other teams and shutdown all of the other interfaces allowing the opposing teams access to the Finish Line Router.

C. Relay Race Technical Details

Only one student file is included with this game because DHCP has been implemented in a very unique manner allowing the file to be used among any multiuser peer-connection within the game. The instructor file contains the Relay Race game that students must telnet into. The students connect to the multiuser clouds according to the name of the cloud in correspondence to the role of the student within each team. By providing this instructor/student architecture we ensure easy deployment of the game during play time. The clusters problem can be easily changed if need be.

D. Educational Value

Relay Race pits students together in a team environment where they must communicate and apply their skills to a

 ISBN: 978-1-61208-133-5

problem. The activity will help hone communication skills and also improve the ability of the students to solve network problems. The game will force students to work rapidly and correctly to solve their problems the fastest in order to win. This environment emulates the fast paced environments of the networking world today, where problems arise quickly and must be solved rapidly. The game can also be used as an evaluation tool to see how far student's configuration and troubleshooting skills have progressed.

VI. PERFORMANCE RESULTS

With Packet Tracers ability to manage and create Multiuser Packet Tracer sessions, it was still unknown to the team whether or not Packet Tracer could handle large scale Multiuser environments that consisted of 60+ Multiuser sessions. We conducted various stress tests to confirm that it could handle the 60+ user load. The tests included: CPU Average Utilization, Memory Utilization, Total Network Utilization, Instruction Offline File Size, Time to Create Offline File, and Memory Utilization by offline File.

A. The Test Bed

Two sets of tests were conducted on each of the two activities instructor files. The first was done using the Domination instructor file and its companion student file. The second was done using the Relay Race instructor file and its companion student file. During the testing procedure for the Domination game a total of 60 multiuser clouds were connected in increments of 5 users to the instructor file. Each time 5 users connected to the game, new data was collected. The Relay Race game has a maximum capacity of 20 users playing at the same time therefore, 20 multiuser clouds were connected to the Relay Race instructor file in increments of 5 users.

B. CPU Utilization

The results for this test were very satisfying for both of the activities. Figure 3 shows both the CPU utilization for Domination and Relay Race. It is evident that the as more hosts connect to the instructor files, the CPU utilization increases in a logarithmic form. This indicates that the packet tracer software is capable of handling higher levels of stress if need be. If we compare the Relay Race and Domination results it's evident that continuous even if we add up to 60 users.

C. Memory Utilization

Test results for both the activities indicated a linear growth rate in memory utilization. Figure 4 shows that growth rate for both the Domination activity and the Relay Race activity is linear as the number of users increases. Although memory is not an issue with 50 users connected to an activity, if we were to connect up to 100 or more users' memory could potentially become an issue. For our university use cases memory does not pose a threat to the functionality of our activities.

D. Network Utilization

The test results indicated that the multiuser environment does not put a huge or rather places a minimal burden on the computers network resources. It indicates that we can have a large number of users connected to a single instructor file without worrying about the burden it puts on the network. The network utilization test increased in a linear fashion and these values are bound to increase as the number of users increase. An important issue to note is that these values are bound to change depending on how actively the students interact with the instructor side network.

E. Offline File Save Tests

Packet Tracer allows the ability to save an offline file. The offline file saves can be described as a snapshot of the entire network at an instance of time. The file consists of all of the peer connected multiuser clouds and each device in that peer cloud. The offline file saves also includes all of the devices current state, configuration, and connection status. The offline file proves to be a great tool to use when assessing the results of an activity or grading students depending on the configuration status of the devices they were responsible for. This gives the instructor the ability to save offline files after the completion of the games and assess the students after the activity is completed.

Offline saving times of greater than 5 minutes for a 30 user Multiuser environment could render useless a 20-minute activity and a class room limited to an 80 minute class. Therefore it is important to test how long it actually takes to create the offline file. The results indicated that for the Domination activity the time to create the offline file grew exponentially. The Relay Race activity indicated a linear growth rate. The exponential growth rate for the Domination file could prove to be troublesome if more than 60 users are connected to the Domination activity. Currently it takes about 25 seconds for a 50 user multiuser environment offline file to save. Although it serves our purpose to provide 60 students a reliable platform to use the activity, connecting more than 100 users to a single instructor file could take much longer. These values are bound to change if the complexities of the problems in each cluster are increased.

VII. CONCLUSION AND FUTURE WORK

By using packet tracer as an educational serious gaming platform we can provide a simple and entertaining solution to present complex networking scenarios to networking students enrolled in our courses. We can train students to not only improve their configuration and troubleshooting skills but also improve communications skills in an IT environment. We can also increase the speed at which they tackle these problems and provide them with the expertise and knowledge to excel in the fast paced networking environments of today.

With Cisco constantly updating their packet tracer platform there is an endless possibility of how many different types of game activities we can create. Currently the activities are limited to only CCNA topics, but future packet tracer improvements by Cisco could allow us to

 ISBN: 978-1-61208-133-5

cover CCNP topics as well. Cisco has already taken the potential of serious gaming seriously and began with creating their packet tracer based educational serious gaming platforms with Cisco Aspire Passport21 and Cisco myPlanNet.

Using Cisco's Packet Tracer 5.3 as our serious gaming platform, the development of multiuser serious gaming activities was made possible. By utilizing the familiar environment known to all Cisco Network Academy students, we ensured that the activities could be easily understood and collaboration between multiple or groups of students was possible using the multiuser capabilities. Combining the concept of serious gaming with a network simulator allows us to approach teaching methodologies for Cisco network oriented courses in an entirely new light.

With the future for serious gaming looking bright, it can be surmised that we will see the development of many networking educational serious gaming platforms. This research has shown the great potential of the topic and the educational values its holds. Lastly, Cisco's packet tracer has proven to be a power platform that can be used to provide an educational serious gaming platform.

VIII. ACKNOWLEDGMENT

This research was supported through a 2010 Teaching Innovation Funding (TIF) grant from the office of the Associate Provost, Academic of the University of Ontario Institute of Technology. Packet Tracer is a product of Cisco Networks and is provided free of charge to Cisco Networking Academy students and instructors.

REFERENCES

[1] A. All, Serious Games Entertain, Educate Employees. IT Business Edge, August 5, 2009. http://www.itbusinessedge.com/cm/community/features/articles/blog/serious-games-entertain-educate-employees/?cs=34730 (Accessed March 20, 2011).

[2] C. Kohler, Xbox Kinect Games Give You a Serious Workout, Wired Magazine, June 15, 2010. http://www.wired.com/gamelife/2010/06/kinect-hands-on/ (Accessed March 20, 2011).

[3] J. Bohannon, Unravel the Secret Life of Protein, Wired Magazine, issue 17.05, 20 April 2009, http://www.wired.com/medtech/genetics/magazine/17-05/ff_protein?currentPage=all (Accessed 20 March 2011).

[4] Macedonia, M. 2005. Games, simulation, and the military education dilemma. http://www.educause.edu/ir/library/pdf/ffpiu018.pdf . (Accessed 20 March 2011).

[5] M. Torrieri., Cisco's MyPlanNet Simulation Game Touches on Broadband Growth and Other Hot Communications Topics, TMCnet, Nov. 4, 2009. http://4g-wirelessevolution.tmcnet.com/broadband-stimulus/topics/broadband-stimulus/articles/68180-ciscos-myplannet-simulation-game-touches-broadband-growth-other.htm (Accessed March 20, 2011)

[6] Behrens, J. T., Frezzo, D. C., Mislevy, R. J. Kroopnik, & Wise, D. (2007). Structural, Functional and Semiotic Symmetries in simulation based games, and assessments. In E. L. Baker, J. Dickieson, W. Wulfeck, & H. F. O'Neil (Eds.) Assessment of Problem Solving Using Simulations. Mahwah, NJ: Erlbaum, pp. 59-80.

[7] A. Smith and C. Bluck, "Multiuser Collaborative Practical Learning Using Packet Tracer," in Networking and Services (ICNS), 2010 Sixth International Conference on, pp. 356-362, 2010.

 ISBN: 978-1-61208-133-5

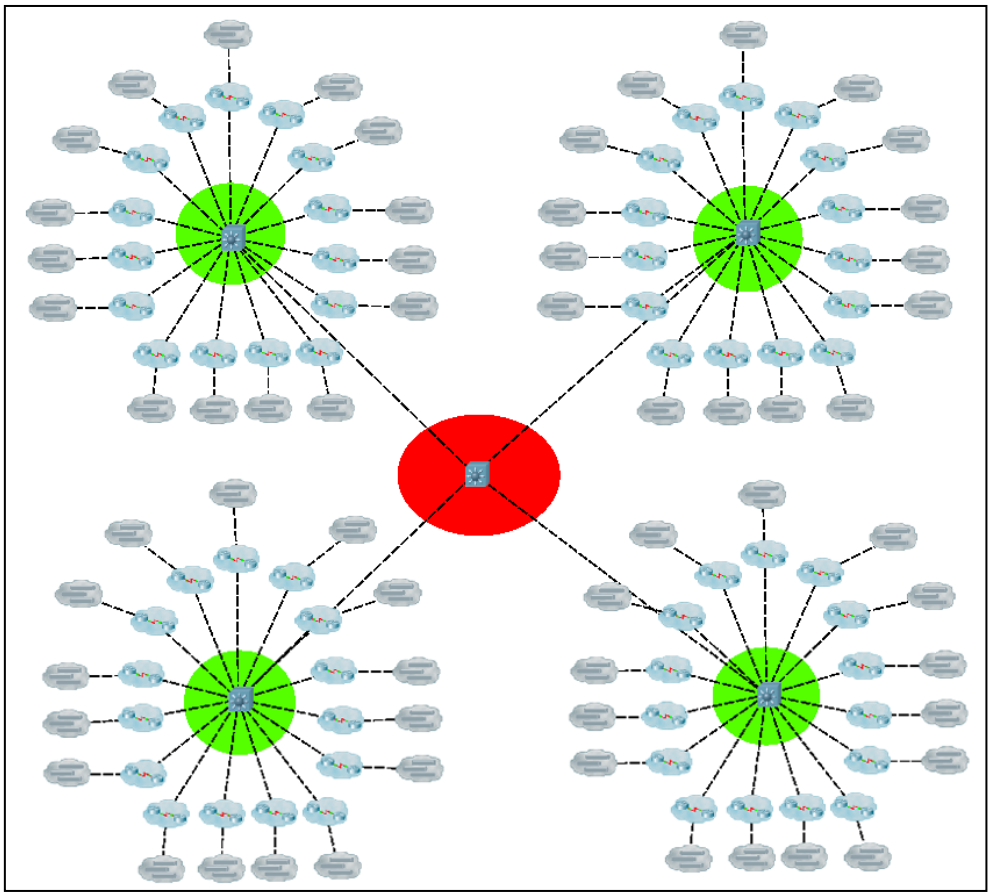


Figure 1. Domination Game Screen Shot

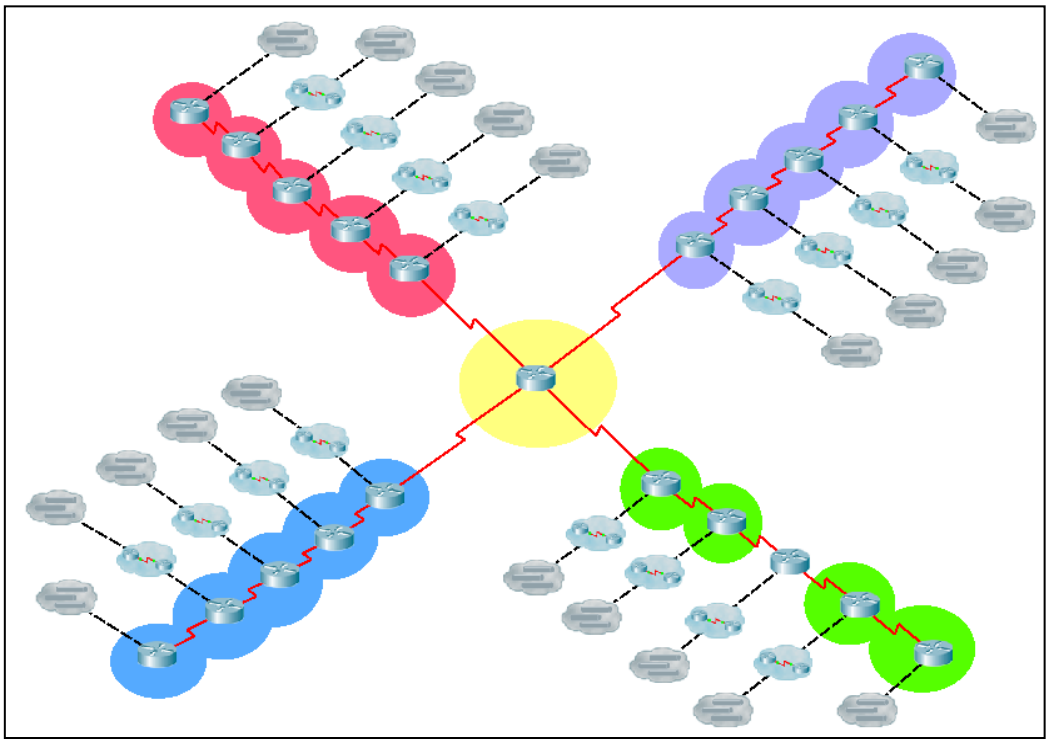


Figure 2. Relay Race Game Screen Shot

 ISBN: 978-1-61208-133-5

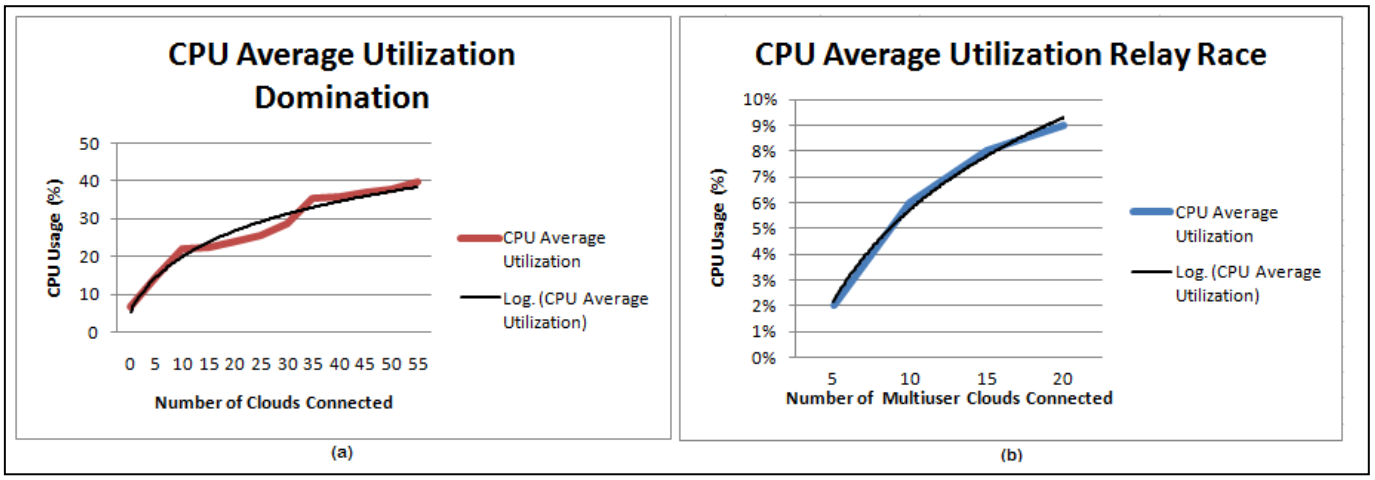


Figure 3. CPU Utilization: (a) Domination CPU Utilization (b) Relay Race CPU Utilization

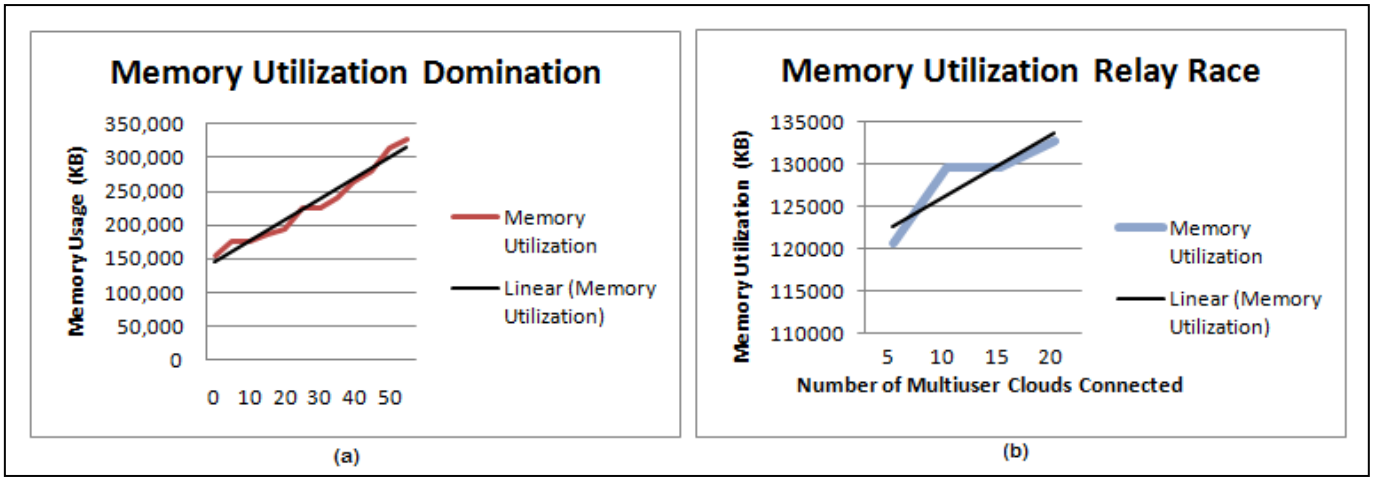


Figure 4. Memory Utilization (a) Domination Memory Utilization (b) Relay Race Memory Utilization

 ISBN: 978-1-61208-133-5

Classroom-based Multi-player Network Simulation

Experiences of interactive scenarios using Packet Tracer

Andrew Smith
Faculty of Mathematics, Computing and Technology
The Open University
Milton Keynes, United Kingdom
a.smith@open.ac.uk

***Abstract*— The delivery of group-based remote collaborative work in the practice based learning domain of computer networking, historically has presented challenges in scale, management, security and technological resource to support delivery, assessment and learning. In partnership with the Packet Tracer development team at Cisco Systems, this paper explores the outcomes of a series of class based experiments, supporting the research into the development of a 'simulated' Internet using Packet Tracer.**

Keywords- packet tracer; practice based; collaborative; Cisco; virtual labs; Networking Academy; CCNA; Supported Open Learning; gaming;

I. INTRODUCTION

The Open University in the United Kingdom has offered the Cisco Certified Networking Associate (CCNA) and Cisco Certified Networking Professional (CCNP) via blended distance learning to over 3000 students since 2005, as discussed by Moss and Smith [1].

Research into the creation of learning solutions for students taking the CCNA and CCNP programmes via blended distance learning has already taken place with work on the NDG [2] system in the management of remote tutorials using Skype by Smith and Moss [3] and the utilization of the virtual laboratory by the setting of course assessment items, Prieto-Blázquez, J. et al. [4]. The utilization of the Packet Tracer environment led to the initial development of a remote relay server using a one:many community of practice, where each participant connected their remote simulated network (the client) to a central simulated network (the server), as described by Smith and Bluck [5].

A conclusion of the research by Smith and Bluck is that there was the potential to develop group-based activities, where learners in an online situated learning environment, Lave, J. et al [6] could work on a simulated practical to create a large infrastructure, based on a set 'local' task to create a Wide Area Network (WAN) connection with a Local Area Network (LAN).

This paper explores how the research was extended into classroom-based experiments, how these were designed, the rationale for group selection.

A. *The group-based scenario*

The group-based activity is presented to students in two parts. A reflection on previous work accomplished with packet tracer was that many students as well as their instructors did not as yet understand the full feature and function of the inbuilt 'multiuser' tool. To overcome this, the student group would commence the activity with a formative warm-up exercise, where students are paired with the task of creating a simple network of two hosts and being able to send a 'virtual' ping from one Packet Tracer instance to the other across the academic network as illustrated in Fig. 1.

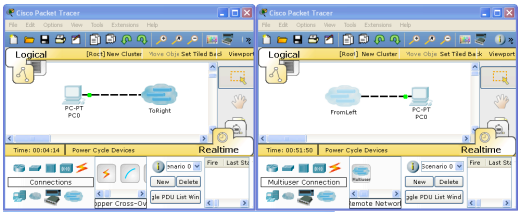


Figure 1. Peered example of Packet Tracer Multi-user communication.

This short exercise sets the scene and ensures all participants are working from the same stand-point in their ability to use the software. All of the participants are already low-level users of Packet Tracer, by virtue of their membership as students of the Cisco Academy programme, CCNA discovery and exploration curricula [7], in the integrated use of this application in their studies. By introducing the participants to the 'multiuser' tool, their understanding of the additional tools available in Packet Tracer is increased.

Following the formative activity, the students participate in the large-scale activity to build a large-scale simulated WAN, with multiple individual simulated LAN's inside the confines of the typical academic network. Having no physical connection or contact with the underlying 'real' network environment, by virtual of the simulated nature of Packet Tracer and its use of the Packet Tracer Messaging Protocol (PTMP) [8].

The structure of the activity is a duplication of the experimentation explored in the paper by Smith and Bluck [5], with the relay, no longer a remote server, but the

 ISBN: 978-1-61208-133-5

teachers computer. This assists the learning process experienced by the students and observing instructor, discussed by Laurillard [9] as in each session, the teacher's computer is attached to a classroom data projector. Each student is able to see their own multiuser connection locally as well as their remote connection on the teachers Packet Tracer instance. This reinforces the assurance they are correctly participating in the practical task and students are also successfully building a remote (otherwise normally unseen) connection.

The relay instance of Packet Tracer, run a router with a series of switches all connected to a core switch (Fig. 2). The simulated protocol selected is the Extended Interior Gateway Routing Protocol (EIGRP), in technical terms; has a lower device configuration overhead. The simulated WAN uses a class A, IP address, of 10.x.x.*n* and each simulated LAN is a Class C, with each student having 192.*n*.x.x. For each system, *n*, is a unique number issued to each participating student. The relay has been designed to cope with 120 participants, the activity will by virtual of the second class C IP address octet support 254 participants.

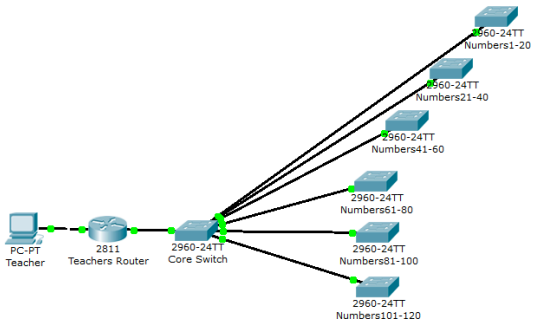


Figure 2. Teacher relay-server structure

Routing protocols as with many network technologies can be configured many different ways to achieve the same goal. To remove any confounding variance, all students are presented with an instruction sheet, with the commands they must use, and where they need to be applied. Fig 3, lists the routing commands on the teachers relay-server, which all students are able to see during the activity. Fig. 4 illustrates the routing commands the students have to apply onto their own Packet Tracer instances.

```
interface FastEthernet0/0
no shutdown
ip address 10.255.255.254 255.0.0.0
interface FastEthernet0/1
no shutdown
ip address 200.200.200.200 255.0.0.0
router eigrp 123
network 10.0.0.0
network 200.200.200.0
no auto-summary
```

Figure 3. Teacher relay-server routing commands

```
interface FastEthernet0/0
no shutdown
ip address 10.0.0.n 255.0.0.0
interface FastEthernet0/1
no shutdown
ip address 192.n.0.1 255.0.0.0
router eigrp 123
network 10.0.0.0
network 192.n.0.0
no auto-summary
```

Figure 4. Student Packet Tracer instance routing commands

The students own instance of packet tracer is a self-constructed system which when assembled should resemble the illustration in Fig. 5. The system is deliberately simplified to reduce any potential confounding variance, by ensuring the students had two specific devices and cable types to implement.

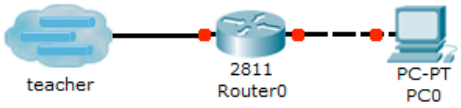


Figure 5. Student Packet Tracer instance

B. *Selecting the groups*

The challenge in any activity requiring volunteers is the recruitment of the volunteers themselves. In the time span since the initial research and paper there have been three successful interactive scenarios. The timing of these have been principally driven by centre availability and timetabling constraints, collaborating teacher availability and ensuring the session fits in with the students learning, where the activity is in harmony with their studies and their current curriculum.

All participating students need to have a minimum of the equivalent to Cisco CCNA exploration or discovery first course behind them to have a comprehension of the networking terminology and technology.

The student age range was kept to a small window, with participating students being either second years on a high school equivalent technical vocational programme [10] or first years on an undergraduate honours degree programme. This gave a range of 17-20 years of age with the majority in the 18/19 year-old age group.

 ISBN: 978-1-61208-133-5

The sessions were in May, November and December 2010, reflecting the academic calendars of each group and availability. The May and November sessions were with two groups of 18 and 11, 17-19 year olds at a college of further education. The December session was with a group of 30, 18-20 year-old year one undergraduate students at a London University.

Group selection was based on the class/group available at the time suited to the demands of timetable, and availability of the willing volunteer teacher and the researcher.

II. MANAGING THE ACTIVITIES

The sessions were scheduled for a three hour ½ day block, according to timetabling. The first was a morning session 09:00 to 12:30 with break, the second session was 10:30 to 15:00 with intervening lunch break and the third session was a PM session from 14:00 to 17:00 with a short break.

Each session used the majority of the time, with an average of thirty minutes remaining to enable the students to complete an optional challenge activity.

In each session the researcher acted as teacher/facilitator whilst the normal session teacher/instructor acted as class-room assistant and secondary observer.

Each session was facilitated as a in-class teaching session, where each of the student participants were aware that they were helping to test the multiuser functionality of Packet Tracer and get in return additional networking skills (learning).

Using a data projector connected to the teacher's computer running the relay instance of packet tracer, provided an essential conceptual and visual tool for students already discussed by Janitor and Kniewald [11]. Enabling them to see how their own LAN and WAN was behaving in relation to the greater WAN infrastructure.

Typical of an academic network, each computer running had the same hardware specification and operating system installation, including local policy constraints and user rights, therefore ensuring that each student participating had the same technological advantage/disadvantage, as all others during the lifecycle of the activity.

The activity was managed in a systematic follow-the-leader step-by-step format, keeping all students to the same position in the process. By having the additional facilitator, the students were able to remain engaged and have their questions/misunderstandings answered.

III. OBSERVATIONS

Qualitative feedback was collected from each cohort, the intention was to understand their personal viewpoint of their experience participating in the sessions as well as participating in the activities.

At the end of each session, before departure, the students were asked to complete a short anonymous questionnaire, with questions listed in Table 1.

TABLE I. QUESTIONAIRE

Question Number	Question
1	Has this exercise enhanced your practical understanding of IP addressing? (Y/N)
2	Have you used the Packet Tracer Multiuser tool before this session? (Y/N)
3	In your own view, has this given you some understanding of routing protocols? (Y/N)[a]
4	Would you consider continuing to use Packet Tracer in the way demonstrated today? (Y/N)

a. this was contextualized for some students, describing their use of EIGRP

The questionnaire results are summarized in Table 2. As the groups are small, and the questionnaire short, there are no missing responses. The students were able to drop the forms into a box on departure. There was no additional personal information requested.

TABLE II. QUESTIONAIRE RESULT DATA

Question Number	Feedback					
	May (18)		*November (11)*		*December (30)*	
	Y	***N***	***Y***	***N***	***Y***	***N***
1	14	4	11[a]	0	21	9
2	0	18	0	11	2	28
3	15	3	11	0	*24*	6
4	16	2	11	0	26	4

a. this is earlier in the academic year for this cohort

From the results in Table 2, the dominant feedback implies that the students believed that using the simulated practical was a personal benefit, where the responses to questions one and three indicate a high percentage (Table 3) of positive responses to the enquiry.

TABLE III. QUESTIONAIRE RESULT PERCENTAGE

Question Number	Feedback as a percentage					
	May (18)		*November (11)*		*December (30)*	
	Y	***N***	***Y***	***N***	***Y***	***N***
1	77.7	22.3	100	0	70	30
2	0	100	0	100	6.7	93.3
3	83.3	16.7	100	0	*80*	20
4	88.8	11.2	100	0	86.6	13.4

It is notable that for the November cohort, the groups of students were in the early stages of their learning for the

 ISBN: 978-1-61208-133-5

academic year, whereas the May and December cohorts were either at the end of their respective academic year or semester.

Questions two and four explored the student's experience of Packet Tracer. Apart from two outliers (reason unknown), question two indicated that the majority had not used the multiuser tool beforehand. With Question four, the response indicates an interest held by the students to continue using the multiuser tool in packet tracer. This may have been stimulated by their feelings regarding the preceding session.

In engaging with the practical activities, the students could be seen to link constructivist personal concepts as described by Piaget [12] and readily connect their own private concepts to a visual, simulated physical network environment.

IV. NEXT STEPS

The centres involved are willing to host future sessions, inviting the researcher back to continue the same exercise as well as different scenarios.

Other centres are interested in participating in the research and are willing to engage in the activities described in this paper, as well as working towards more complex scenarios. The challenge for these centres as for the original participants is finding the right group, at the right time in their year as well as in the study week.

More complex systems have already been designed, where the students will participate in an activity to create a relay-mesh, with the students working in a group to build one system around a local relay. The local relay will connect to a central relay, illustrated in Fig. 6, as a relay-relay.

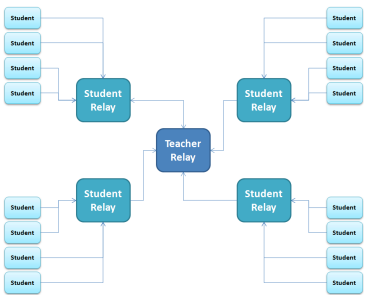


Figure 6. A relay-relay model

Once the centres, the students and also the teacher/instructors become familiar with the technology and the constructivist learning experience, the plan is to move the activity to a remote 'central' relay server model. This is reflective of the work carried out by Smith and Bluck [5] in 2009 and able to support a larger collaboration of participants.

V. CONCLUSIONS

The activities have demonstrated that once the student has been given an introduction to the multiuser tool, they are able to engage in a structured activity to build a complex simulated network environment, reflecting the model of situated learning discussed by Lave and Wenger [6].

Formal questionnaire feedback as well as the in class anecdotal experience of the researcher, reflects an enthusiasm from the learners to continue studies using the Packet Tracer application in this mode.

Research discussed in this paper, allied with prior research suggests that the structured development of a system to create a simulated Internet would provide an alternate learning methodology for in class as well as remote distance based learners.

ACKNOWLEDGMENT

The author would like to thank Paul Wallin of Kingston University (UK) and Ken Lamb, Milton Keynes College (UK) and their students for their support, hard work and enthusiasm.

REFERENCES

[1] Moss, N. and Smith, A. (2010). Large scale delivery of Cisco Networking Academy Program by blended distance learning. *In: ICNS'10 Proceedings of the 2010 6th International Conference on Networking and Services, Cancun, Mexico.*

[2] NDG: http://www.netdevgroup.com, last accessed 15/03/2011

[3] Smith, A. and Moss, N. (2008). Cisco networking: using Skype and Netlab+ for distance practical learning. *In: IADIS e-Learning 2008 conference, Amsterdam, Netherlands, 22-25 July 2008,*

[4] Prieto-Blázquez, J. et al. (2008). An Integrated Structure for a Virtual Networking Laboratory. *In IEEE transactions on Industrial Electronics*, Vol 55, no 6, pp 2334-2342

[5] Smith, A. and Bluck, C. (2010). Multiuser Collaborative Practical Learning using Packet Tracer. *In: ICNS'10 Proceedings of the 2010 6th International Conference on Networking and Services, Cancun, Mexico.*

[6] Lave, J. and Wenger E. 1991, *Situated Learning: Legitimate Peripheral Participation*, Cambridge University Press, Cambridge, UK, p.91

[7] Cisco Networking Academy: http://cisco.netacad.net, last accessed 15/03/2011

[8] Wang, M. (2008), Packet Tracer 5.0, Packet Tracer Messaging Protocol (PTMP), Specifications Document: Cisco Systems

[9] Laurillard, D. (2002) Rethinking University Teaching: a framework for the effective use of educational technology (2nd edition) London, Routledge Falmer, P154.

[10] BTEC in Information Technology : http://www.edexcel.com/quals/nationals10/it/Pages/default.aspx, last accessed 15/03/2011

[11] Janitor, J and Kniewald, K (2010). Visual Learning Tools for Teaching/Learning Computer Networks. *In: ICNS'10 Proceedings of the 2010 6th International Conference on Networking and Services, Cancun, Mexico.*

[12] Piaget, J. (1978). *Success and Understanding.* London: Routledge & Kegan

 ISBN: 978-1-61208-133-5

Low-Cost Pre-Evaluation of New Educational Programs

Bowen Hui, Bruce Hardy
Function Four Ltd.
Winnipeg, Canada
{bowen,bruce}@functionfour.ca

Yvonne Pratt
Communications
University of Calgary
Calgary, Canada
ypratt@ucalgary.ca

Rob Kershaw
Center for Digital Storytelling
Toronto, Canada
rob@storycenter.org

***Abstract*—Educational programs are developed to accommodate for new pedagogical findings and evolving curriculums. Methods to evaluate the effectiveness of these programs are typically labour intensive and time consuming — requiring the recruitment of program participants, the execution of the program, the collection of qualitative and quantitative data, and the analysis of results by expert researchers. To directly address the cost of such evaluations, we propose a pre-evaluation method that estimates the *expected value* of a new educational program before implementing it in practice. This approach allows educators, researchers, and stakeholders to obtain a preliminary assessment of new programs with minimal investment. To demonstrate our approach, we describe a case study that evaluates the impact of a digital storytelling workshop in a rural community.**

***Keywords*-program evaluation; decision theory; ICT skills**

I. INTRODUCTION

As new programs are introduced into the learning curriculum, systematic methods of evaluation are needed to assess the value of these programs with respect to their intended learning objectives. Educational programs are often evaluated through qualitative methods, such as case studies, content analysis, and grounded theory [1]. Quantitative methods adopted from behavioural psychology have also been applied to evaluate educational programs. One such approach is the use of pre-tests and post-tests; that is, through a pilot execution of the new educational program, students complete a skill test before and after the program so that changes in the test results (in comparison to the performance of a *control* group[1]) are credited toward the pilot program. In all of these cases, evaluation relies on data collected from executing the new program (in a pilot setting or in its full capacity).

Unfortunately, the execution of new educational programs comes with high costs. The evaluation study needs to be designed and conducted in a controlled and reliable way, so that the resulting data can be used to validate the effectiveness of the program. In particular, researchers and trained assistants are required to run the study, collect "clean" data, and analyze the results. Moreover, student participants are needed to undergo the new program in these evaluations. As a result, these methods are typically time consuming and labour intensive. In practice, evaluators are typically operating under limited resources — budget, turnaround time, and staffing. Thus, more cost-effective ways of program evaluations are needed to provide an early assessment of potential outcomes.

In this paper, we propose a simple *pre-evaluation* technique that directly addresses the cost of program evaluation. Our approach stems from decision theory in Economics and provides a normative evaluation of programs. We assume the availability of a standard assessment tool, such as a survey or an aptitude test, which we use as the benchmark to measure the current learning levels of the population of interest. Using this tool, we conceptually estimate the (hypothetic) change expected to be observed in the assessment of the population undergoing the newly proposed program. We use this information to obtain the *expected utility* of the program without actually executing it in practice. In this way, a program that has positive expected utility will likely yield an improvement in learning skills (in the overall population), according to the benchmark assessment tool. In contrast, a program that has an estimated non-positive expected utility is not worth further investment of time and effort. We elaborate on the details with examples below.

Our technique is particularly useful in decision scenarios where educators and funding agencies must choose a learning program for their schools or communities from multiple, available programs. By following the steps outlined in our approach, the stakeholders are able to assess the potential value of each program with respect to a pre-established benchmark. Thereafter, the program offering the highest expected utility according to these estimations would be chosen for further evaluation.

We emphasize that the purpose of our approach is to provide an early estimate of potential value in new educational programs before putting them in place. This work is not meant to replace existing evaluations; rather, it is designed to give an earlier, faster, and cheaper assessment. As such, this pre-evaluation technique compliments existing program evaluation methods. Moreover, it can be used in conjunction with any program evaluation methods.

[1]Students in the control condition take the same tests but do not participate in the pilot program.

 ISBN: 978-1-61208-133-5

The rest of this paper is organized as follows. Section II describes the pre-evaluation technique with illustrative examples. Our interest focuses on the community adoption of information and communication technology (ICT). As such, we describe a standard survey assessment for ICT called the *E-Index* in Section III. To demonstrate our method, we present a case study in Section IV, with emphasis on ICT skills. Lastly, we report the lessons learned in Section V.

II. Decision-Theoretic Program Pre-Evaluation

For comparison purposes, we assume that a typical program evaluation process consists of five steps as illustrated in Figure 1. In Step (1), the population of interest is identified and participants are selected (e.g., via a stratified sampling procedure [3]). In Step (2), participants take part in a pre-test that is deemed to be appropriate and sufficient for measuring the performance of the intended learning objectives of the pilot program. At this point, participants would be split into two groups: group A undergoes the pilot learning program (i.e., the *test condition*) and group B undergoes the regular learning program (i.e., the *control condition*). Generally speaking, the grouping of the participants should be done randomly, and in a way that allows the two groups to have (approximately) equal numbers. However, researchers may want to control for certain grouping variables, such as age, gender, and pre-test performance. In this case, we view groups A and B as having subgroups, and that the participants for each subgroup are selected randomly.

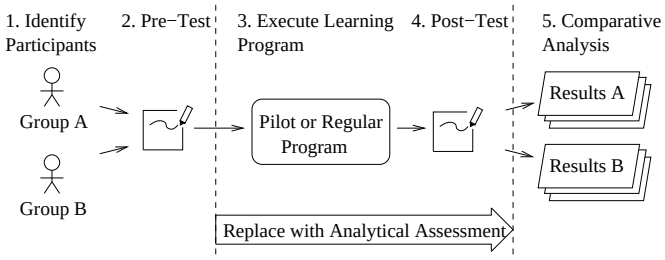


Figure 1. Process for program pre-evaluation.

Next, the learning programs takes place in Step (3), and the participants take a post-test in Step (4) upon completion. Variations of the pre-test may be used as a post-test, so long as the variables being measured by these tests are the same in both tests. Finally, in Step (5), the results are separated according to the original participant groupings and a comparative analysis (e.g., ANOVA [5]) is conducted to measure: (i) the performance difference between each group before and after the learning program, and (ii) the performance difference between the groups of the two programs.

In contrast, the pre-evaluation process that we propose replaces Steps (3) and (4) with an analytical calculation that estimates the projected post-test results. For example, if the assessment tool used is a survey, then an unbiased expert who understands the objectives of the pilot program needs to mine through each survey question, compare the pre-test scores obtained in the participant groups, and project a post-test score for each group. Moreover, if subgroups are used (e.g., group A male, group A female, group B male, group B female), then an estimated expected performance should be expressed for each subgroup.

We demonstrate the analytical assessment procedure with an example. Suppose a community's governing body is considering implementing a new technology training program that has demonstrated to be effective in other communities. How does one evaluate the effectiveness of such a program in this community without first executing it?

Table I
Example scores for two types of questions.

Q1	**Pre-Test Results**	**Post-Test Projections**
A (90/100)	30%	40%
B (75/100)	45%	50%
C (60/100)	10%	5%
D (55/100)	10%	5%
E (30/100)	5%	0%
Q2	**Pre-Test Results**	**Post-Test Projections**
Yes (1)	60%	75%
No (0)	40%	25%

To elaborate on the example, we suppose that a group of community members took a pre-test (e.g., survey) consisting of two types of questions with scores summarized in Table I. Suppose question $Q1$ is a type of question that assess performance and assigns a score between $0-100$ along with a summary letter grade. Through a hypothetical pre-test, we have 30% of the participants scoring a grade of A, 45% scoring B, and so on. On the other hand, if a question $Q2$ is of yes/no type (e.g., "do you know how to send emails?"), then we simply tally up the number of participants for each response in the pre-test. Obtaining pre-test results marks the end of Step (2) in the pre-evaluation process.

Next, we use the pre-test as a guide to estimate the maximal impact of a new educational program for this community. Going through each question and the participant scores one by one, a knowledgeable expert projects the expected change in the results if the training program were to be conducted followed by a post-test. For example, the pre-test evaluate one's telephone skills, computer skills, and Internet skills, while the new program promises to train participants on computer skills only. Thus, the skills that we would expect to see improvement in pertain only to computer knowledge. Example estimates for post-test scores are shown in the right-most column in Table I.

To calculate the expected change in a question, we calculate an average score for each scenario and subtract the difference. In particular, we use the median value for each score/response (e.g., 90 for A, 75 for B, etc.). An average score for the pre-test for $Q1$ is calculated by multiplying the percentage of participants who obtained each score and

 ISBN: 978-1-61208-133-5

summing up all the possible cases as follows: $(30\% \times 90) + (45\% \times 75) + (10\% \times 60) + (10\% \times 55) + (5\% \times 30) = 70$. Following the same calculating using the projected post-test percentages, the average score is 79.25. Thus, the expected value of the learning program is $79.25 - 70 = 9.25$. One may further reference the grading scheme of the question to check whether this value improves the grade (say, from a B to an A). While such a number may seem abstract, ensuring that all the learning programs using the same benchmark assessment tool will enable a fair comparison in the same scale. An analogous calculation can be used for $Q2$, where a value of 1 is used for "yes" responses and a 0 otherwise.

III. E-INDEX: AN ICT ASSESSMENT TOOL

The current approach in community assessment is to conduct door-to-door surveys which are time consuming and labour intensive. Results in the quantitative components of these surveys are typically summarized as average participant responses. Here, we describe the community assessment survey called the E-Index that consists of questions about ICT adoption [4]. To date, the E-Index has been applied in 43 rural communities across Canada [2].

For a participating community, A housing list is used to establish the sampling frame. A random sample of households is drawn from this frame. As part of the rural community development initiative, local residents of the community are trained and certified as E-Index surveyors. These surveyors are responsible for conducting the survey with a member of their assigned households (as determined based on birth dates). Throughout the project, surveyors are supported by E-Index researchers remotely.

In total, there are four sections in the E-Index survey: demographics, ICT infrastructure, ICT skills, and ICT utilization. Example questions are: "Do you have access to $Tech$ at Loc", "Do you know how to use $Tech$", and "How often do you use $Tech$". where $Tech$ is replaced by radio, television, fixed phone, mobile phone, fax, computer, and Internet, and Loc is replaced by home, work, public areas.

Among other data collected in the E-Index, we focus on the quantitative results only. Responses for the questions in each section are averaged across all the respondents to obtain an infrastructure score, a skills score, and a utilization score for each of the seven technologies surveyed. These averages simply represent the percentage of respondents who indicated they have a technology at home, a skill for a technology, or a use for a technology. These percentages are then scaled to obtain a grade score using *goalposts* — a numeric score expressing the expected proportion of individuals who would indicate a positive response. In effect, if the goalpost is 100, the calculated percentage and the grade score are the same. However, if the goalpost is, say, 60, then only 60% of the population is expected to use the Internet. In this case, the grade score will be scaled to a number that is higher than the calculated (raw) percentage.

Although these scores are summary statistics, they may be used to help community leaders and policy makers understand their communitys adoption rates on various ICTs as a whole. For example, these scores may indicate that the infrastructure for Internet is very high but the actual skills to use it is very low. In this case, leaders may decide to invest in better training programs for Internet to increase knowledge and utilization of it. On the other hand, these scores may indicate that the community has very high skills in a technology but not enough infrastructure to support it. Such a result would suggest that leaders need to direct their investments to create broader access for that technology.

A. Assessment on Fishing Lake, Alberta

In 2008, Fishing Lake Métis Settlements participated in the E-Index project (version 2.4). A total of 148 households were used to establish the sampling frame and 84 households were sampled. With a response rate of 89.3%, a total of 75 surveys were completed successfully.

B. Project Findings

The overall E-Index score for Fishing Lake is 64.5% (a letter grade of B). Table II shows a further breakdown of the category and technology grade scores. In comparison to 17 other communities who participated in this version of the E-Index, Fishing Lake has similar scores in about half of the cases, with Infrastructure, Utilization, Fixed phone, Mobile phone, and Computer scoring slightly below the average.

Table II
E-INDEX SCORES FOR FISHING LAKE IN 2008.

Category/Technology	Grade Score	e2.4 Average
Infrastructure	79.7 (B)	88.7 (A)
Skills	91.6 (A)	90.1 (A)
Utilization	27.6 (D)	52.4 (C)
Radio	83.5 (A)	84.8 (A)
Fixed Phone	87.8 (A)	92.8 (A+)
Fax	44.4 (C)	52.3 (C)
Mobile Phone	68.8 (B)	79.6 (A)
Television	74.0 (B)	80.1 (A)
Computer	50.8 (C)	59.9 (B)
Internet	42.1 (C)	55.8 (C)

IV. CASE STUDY: DIGITAL STORYTELLING

We report our experience with a pilot study in Fishing Lake that was conducted in the Fall of 2010. This study is centered around a digital storytelling (DST) workshop, where DST experts are invited into the community to train youths on various technologies and teach them about digital story making. After a period of hands-on training, the participants are required to create a digital story in teams. In this study, seven types of software and hardware were used in the digital storytelling workshop. These technologies are:

- Word processing software (WP): MS Word, Notepad
- Movie editing software (ME): Final Cut Studio, Windows Movie Maker

 ISBN: 978-1-61208-133-5

- Storage device (SD): USB key, CD, DVD
- Digital camera (DC)
- Camcorder (CC)
- Scanner (SC)
- Audio recorder (AR)

To assess the specific technology skills, we extended the E-Index questionnaire with detailed questions pertaining to the knowledge (with yes/no questions) and utilization of these technologies. The objectives of the workshop is to expose new technology to participants, equip them with the necessary skills to use the technology, and inspire them to explore technology use in their daily lives.

A. Calculating the Expected Impact

A total of 5 youths participated in this pilot project. Participants completed the extended E-Index questionnaire as the pre-test. Results are shown in Figure 2. With emphasis on the technology skills questions only, Table III shows the corresponding distributions and average technology scores.

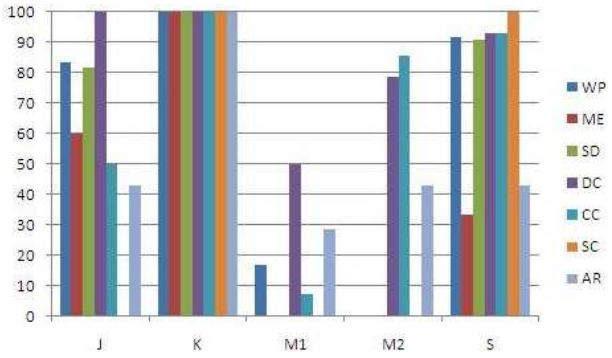


Figure 2. Pre-test results.

Based on the participants' engagement level, the complexity of the technology, and their perceived enjoyment level (as reported by the participants), we augmented the distributions and calculated the projected averages shown in the last row of Table III. Averaging across the seven technologies, we expect an improvement of 14.3 points from the workshop.

Table III
EXAMPLE SCORES FOR TWO TYPES OF QUESTIONS.

Grade	WP	ME	SD	DC	CC	SC	AR
A	0.6	0.2	0.6	0.6	0.6	0.4	0.2
B	0	0	0	0.2	0	0	0
C	0	0.2	0	0	0	0	0
D	0	0	0	0.2	0.2	0	0
E	0.4	0.6	0.4	0	0.2	0.6	0.8
Pre-Test Averages	66	48	66	80	71	54	42
Projected	81	74	76	87	84	63	62

B. Preliminary Results

About two months later after the workshop was over, participants were asked to fill out the E-Index again as a post-test. To minimize the effort in answering the same questions twice, we created an online survey tool that automatically populated the post-test responses using the responses provided from the pre-test. Post-test results (actual scores) are shown in Figure 3.

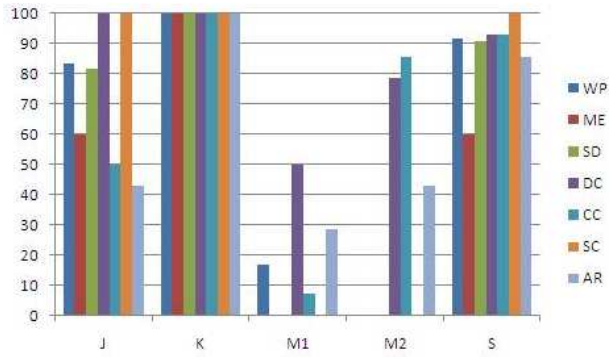


Figure 3. Post-test results.

Due to the small number of participants and a lack of a control group, we cannot conduct rigorous analyzes on the quantitative results to measure the impact of the workshop nor to assess the value of our pre-evaluation process. Through casual observations, community members noticed that after the workshop, the student participants began to spend more time in the technology laboratory at the community centre.

V. CONCLUSION

We presented a simple pre-evaluation process to address the cost of program evaluations. This technique is designed to compliment existing program evaluation approaches with an earlier, faster, and cheaper estimate assessment of new programs. Through analytical examples and preliminary data, we demonstrated the details of the technique. Further empirical evidence is needed to assess the value of this pre-evaluation process.

REFERENCES

[1] D. Ary, L.C. Jacobs, C. Sorensen, and A. Razovieh. (2010) *Introduction to Research in Education.* Wadsworth Publishing.

[2] B. Hui, S. Irwin, and P. Robins. (2010) *Cultural Community Comparison: First Nations, Métis Settlements, and Non-Aboriginal Communities.* DVA reports, Function Four Ltd.

[3] L. Kish. (1995) *Survey Sampling.* Wiley-Interscience.

[4] W. Kelly. (2008) *Knowledge Planning: Community Development in the Knowledge-Based Economy.* Department of Rural Development, Brandon University, Brandon, Canada.

[5] P.S. Maxim. (1999) *Quantitative Research Methods in the Social Sciences.* Oxford University Press.

 ISBN: 978-1-61208-133-5

Solutions for virtual laboratory

Peter Feciľak, Katarína Kleinová, František Jakab
Dept. of Computers and Informatics, Faculty of Electrical Engineering and Informatics
Technical University of Košice
Letná 9, 04001 Košice, Slovakia
Email: {Peter.Fecilak,Katarina.Kleinova,Frantisek.Jakab}@cnl.sk

***Abstract*—This paper deals with the solutions for next genaration virtual laboratory providing services for the operation of remote laboratory. The paper addresses problems related to requirement of flexible, secure and easy remote access to laboratory equipment. This paper also presents a unique concept for logical topology building with the usage of AToM, Q-in-Q tunneling and Frame relay technology, automated password-recovery procedure and knowledge evaluation.**

Keywords-Virtual laboratory; Q-in-Q tunneling; AToM, Knowledge evaluation; Automated password-recovery; Logical topology; Remote access

I. INTRODUCTION

Virtual laboratory in terms of this paper represents environment which is used for blended distance learning in Networking Academy program. Main purpose of virtual laboratory is to provide remote access to physical devices in network laboratory with the goal of doing exercises on real devices placed in virtual laboratory environment as well as combining them with virtual devices and providing services for wide range of applications used in complex labs (like Authentication server, Virtual Private Network server, Active directory (domain) server, etc.).

Main areas which must be reflected by modern virtual laboratory are:

- Remote access to real or virtualized devices (defined interface)
- Devices maintainance (password-recovery procedure, power on/off)
- Reservation system
- Content for exercises (labs)
- Physical/Logical topology re-configuration (dependent on exercise)
- Knowledge evaluation system

This paper describes several solutions for areas listed above with the contribution to the topic of virtual laboratories mainly with the unique approach to combination of real and virtualized devices as well as to logical topology building with the usage of technologies like Q-in-Q [5] tunneling, Frame relay and AToM [7]. Paper also describes physical lab environment components that we have used in our virtual laboratory at Regional Cisco Networking Academy at Košice.

II. DRAWBACKS OF EXISTING SOLUTIONS

In this section we will pass through each area of modern virtual laboratory (chapter I) and describe some drawbacks of solutions that are used in virtual laboratories.

A. Remote access to real or virtualized devices

Depending on devices used in virtual laboratory, it is necessary to define an interface for remote access to equipment. In case of remote laboratory for computer networks we usually use network devices like routers and switches that can be managed over telnet/ssh protocol or by serial or auxiliary interface. In general, it is TCP/IP or serial communication interface (RS232) that is using 9600 bits per second speed by default.

The cheapest way to access devices remotely is by using their own TCP/IP interface for remote access like using telnet or ssh protocol. This solution has its weaknesses in the need of correct configuration of TCP/IP stack at used devices. In case that IP address will be re-configured by user of virtual laboratory, then device of virtual laboratory will be no more accessible remotely at defined IP address. Even in case that we will put some kind of warning such as ”do not re-configure interface, etc.”, we are unable to guarantee accessibility of virtual laboratory as it might be malfunctioned by user. Therefore it is more stable if remote access is based on terminal server that has defined interface for accessing the remote devices connected to terminal server. Usually it is telnet or ssh protocol used for remote access.

There are a lot of laboratories that are using terminal server with telnet access on different ports for each device connected to terminal server. Terminal server based on Cisco integrated services router (terminal server) with 8 to 32 serial asynchronous interfaces is mostly used. There is also need for possibility of direct access to terminal server settings, like clearing frozen sessions and changing port speeds. If there is no other user interface to communicate settings directly to terminal server, then there is no way to access devices with different speeds of console port than default. This can be lack of this solution. As soon as user will change the speed of console port during lab exercise configuration, device becomes unusable for next reserved sessions.

 ISBN: 978-1-61208-133-5

Direct access (even relayed through terminal server) to devices using telnet protocol might be problematic in networks with too restrictive security policy. Due to security weakness of this protocol it is usually blocked in computer networks or service provider networks. Therefore there is strong need to provide secure way of communication with terminal server.

User interaction in remote network topology is also an important part of virtual laboratory. There is lack of virtual laboratories which combines remote laboratory equipment and user equipment with possibility of connecting own equipment (like user computer) into remote network topology. Usually remote network topology contains intermediate devices like routers and switches and there is lack of end devices like computers, IP phones and printers that can be controlled remotely. There is also strong need for terminal services not only for serial communication, but also for virtualization of operating systems and emulation of other network devices.

B. Devices maintainance

There are several actions that can be done by user and that can completely malfunction virtual laboratory. Therefore there is need for virtual laboratory equipment maintainance. These actions include:

- Re-configured passwords for console access or privileged exec mode *will cause inaccessibility of virtual laboratory for next users trying to access device*
- Changed speed of console or auxiliary port on device *will cause virtual lab device inaccessibility due to need for speed change at terminal server*
- Enabled security features that are blocking password-recovery procedure *will cause lab equipment to be unreachable due to impossibility of automatic password-recovery procedure*
- Erased flash memory *will cause device fails to boot and due to this problem it will not be accessible for lab training*

In modern virtual laboratories there is need for command authorization that cannot be done on Cisco ISR terminal server. Therefore a lot of virtual laboratories that are using Cisco terminal server are facing problems listed above and are solving them by person manually checking devices after each lab reservation. It is also possible to authorize commands on IOS application level with AAA server using tacacs or radius protocol. The solution using an authorization server has its weakness in that it relies on correct device configuration and its connectivity to AAA server. It is also limiting in case that authentication, authorization and accounting is part of exercise.

C. Reservation system

Each virtual laboratory has its own reservation system. Usually there is lack of easiness during reservation process. Some reservation systems are based on manual account creation (on devices or on terminal server) allowing user to access devices remotely. This process can be also partially or fully automated, which means that during reservation of lab session there is process including:

- Receival of lab reservation request from community using virtual laboratory. There are different forms of request receival - e-mail, web form, phone call to maintainer, etc.
- Approval and/or direct reservation
- Creating account and defining access rules
- Notification of person wishing to reserve lab equipment

Electronical requests (done via web form) can be almost fully automated, but there is also possibility of other non e-form requests to lab equipment maintainer. Therefore there is request for easy and fast process of equipment reservation integrated into traditional work user interfaces without spending too much time logging into reservation system, filling form items like e-mail of requester, date and time of lab reservation and notifying requester back.

D. Physical/Logical topology re-configuration

Sometimes it is necessary to re-configure network topology depending on the exercise that user wants to perform on virtual laboratory. There are a lot of virtual laboratories that do not allow topology re-configuration and all labs are based on the same topology or allow topology re-configuration only by technical staff physically changing network topology. There are also some approaches to automated change of physical topology based on connection matrices that are physically interconnecting wires by relay circuits. Some virtual laboratories are using logical topology change on ethernet network instead of physical topology re-configuration. There is issue for using solution based on VLANs for creating interconnection on L2 device for exercises related to L2 protocols like CDP, STP, VTP.

E. Content for exercises and knowledge evaluation

Every virtual laboratory has its technical limitations. Based on technical limitations there is limited set of exercises that can be done on set of equipment in virtual laboratory. Laboratories that did not solve technicaly topological re-configuration are usually dedicated to specific areas and therefore there is lack of scalability and possibility of doing wide range of excercies is typically missing. If laboratory is more static than dynamic in terms of topology creation, then there is usually no option for content creation (like connecting of devices together by web-oriented application of type similar to packet tracer [2] application).

Important part of modern virtual laboratory is a system for knowledge evaluation. Based on exercise that is user doing in virtual laboratory there should be system for configuration collection and configuration files evaluation. There are number of virtual laboratories that are evaluating

 ISBN: 978-1-61208-133-5

solution of exercise only by comparing solution file against user solution. Percentage of the difference between two solutions (template and user) is inverse percentage to 100%. There are still some problems with this solution as it is not so exact and also there is almost no variability in exercises (like IP address needs to be the same) and therefore exercise needs to be written so precisely that there is no other solution for the task.

III. Solutions for next generation virtual laboratory

In this section we will focus on solutions used in our virtual laboratory operated at our Regional Cisco Networking Academy. We have some unique approaches to logical topology management and knowledge evaluation that we will introduce in this chapter. Our virtual laboratory components are shown on Figure 1.

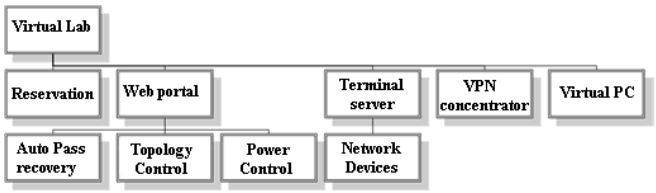


Figure 1. Virtual laboratory components

Our virtual laboratory is equipped with the following devices:

- 1x Virtual lab server with 8port PCI UART
- 1x APC switched rack PDU, 8 ports
- 1x Buttoner - mechanical MODE button pushing
- 3x WS-C3560-24-TS (1x MLS, 2x as SW1,SW2)
- 2x WS-C2960-24-TTL
- 4x Cisco 2811 router
- 2x ISDN B-link exchange
- 1x Frame relay switch (Cisco 2600 series router)

Features that are supported by our virtual laboratory are:

- Remote access to devices like routers and switches through the usage of terminal server with support of command authorization by regular expressions
- Speed change for each console connected to devices via Web user interface
- Automated password-recovery for routers and switches
- Automated topology re-configuration depending on exercise
- Content creation environment integrated to UI
- Knowledge evaluation system based on regular expression lookup
- Reservation system based on iCalendar events
- Virtualization of operating systems (end devices with MS Windows and OS Linux environment)
- Virtualization of routers with dynamips
- Secure connection to virtual laboratory by using VPN technologies with possibility of user interaction in remote topology
- HTTP tunneling for remote devices

A. Remote access via terminal server

History of our virtual laboratory started with specific hardware that was used as terminal server. It was TO108 hardware that had 8 serial ports and communication was multiplexed on one serial port used as management interface that was connected to virtual laboratory server. Main idea on this specific hardware was to have possibility of command authorization [6]. After some time of using this solution we have found one limitation of this terminal server. It was speed change limitation. Terminal server that we had used did not allow to change speed of console port (hardware limitation due to crystal oscilator and ICs used). Device became unavailable as soon as user in virtual laboratory applied command ”speed 115200” in line-console mode of the device. Therefore one of our pre-requisites for terminal server was ability to change speed of each console. As we wanted to avoid some situations causing virtual laboratory to malfunction (see chapter II-B) we have defined the need for command authorization as the second pre-requisite for terminal server.

Solution that we have used is based on multi-serial PCI card that we have inside of our virtual laboratory server. It is PCI8S950LP - 8 Port Low Profile RS232 PCI Serial Card with 16950 UART (Figure 2).

Figure 2. 8 Port low profile RS232 PCI RS232

As we have each serial console represented in UNIX-like system as /dev/ttyS*, it is possible to change speed of each console via setserial. This is really important for automated password-recovery procedure (chapter III-E). We are also using software based terminal server transforming telnet connection on specified port (one for each device) to physical serial interface of server. For this purpose we have modified serial to network proxy application (ser2net) [9]. We have modified this application to support regular expression based filtering for executed commands. This allows us to filter commands that can cause virtual laboratory failure (like erase flash and reload or changing speed of console port). Configuration of regular expression based filter is of Cisco ACL [1] style with definition of type of action - allow or deny. Table I shows an example of

 ISBN: 978-1-61208-133-5

ser2net filter configuration when we want to block specified commands.

Table I
SER2NET FILTERED COMMANDS

Command	*Regular expression statement*
H# erase startup-config	deny .*[#] erase star.*
H# write erase	deny .*[#] wr.* er.*
H(config)# line console 0 H(config-line)# speed 115200	deny .*[(]config-line[)][#] speed .*$
Allow everything else	allow .*

B. VPN concentrator

The goal of VPN concentrator is to allow secure connectivity to virtual laboratory with the option of user computer interaction with remote laboratory. For this purpose we have used openvpn [4] solution modified to allow connection via VPN client authenticated by username and password with pre-build package for end user. Depending on the time of lab reservation it allows connection for specified user and disconnect this user immediately after end of reservation. Due to unsecure nature of telnet protocol that is used on terminal server we do not allow direct telnetting to terminal server and we support direct telnetting (e.g., from user computer) only over VPN. In all our services we are using single sign-on to provide easy way to login to any service inside of virtual laboratory. Usage of VPN is only an option for those users that wants to access devices directly by telnetting from end user computer or to interact with own equipment like connecting end user computer to remote network topology. For this purpose we are using interface bridging so we are bridging VPN interface at end user computer with VLAN in network topology (802.1q and brctl is used on server side).

C. Reservation system

Key idea of reservation system is fast reservation with integration to daily used tools for time-management and communication like e-mail and calendar events. There is no need to develop special environment that is handling communication (messaging) and reservation of time/date based events as there are existing forms like iCalendar events that are handling time based events. One-click reservation system represents drag-and-drop action to visually select time slot of reservation and typing e-mail address of requester (see Figure 3). Reservation system on behind of this reservation process automaticaly parses information from iCalendar event, generates login and password information that is used for different services (VPN access, webGUI, Authentication server, virtual machines access (ssh/rdesktop/vnc), etc.). This information is sent to lab reservation requester automatically. Calendar systems have already solved problems related to read/write access to event reservation and automatic adding of events in case of non-coliding events (lab is free in defined timeslot) are added to calendar.

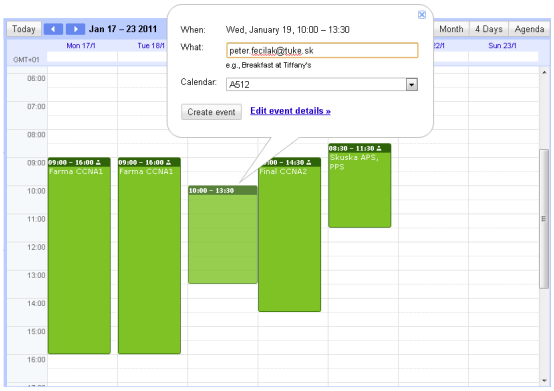


Figure 3. Reservation process using google calendar

D. Virtualized environment

The goal of virtualization is to equip virtual laboratory environment not only with real devices but to power options of virtual laboratory by operating virtual devices like end stations (computer is needed for testing of some features like port-security). From the early beginning of virtual laboratory operation we have been trying different virtualization techniques starting from linux-vserver, through XEN and finaly VMWARE ESXI [8]. Virtualized end stations with MS Windows or UNIX/Linux operating system are reachable via VPN or WebUI and VNC. By this we are powering options of virtual laboratory where user is able to interact with remote devices from user perspective (e.g., when testing port-security). Virtual machines created under ESXI are shown in Figure 4.

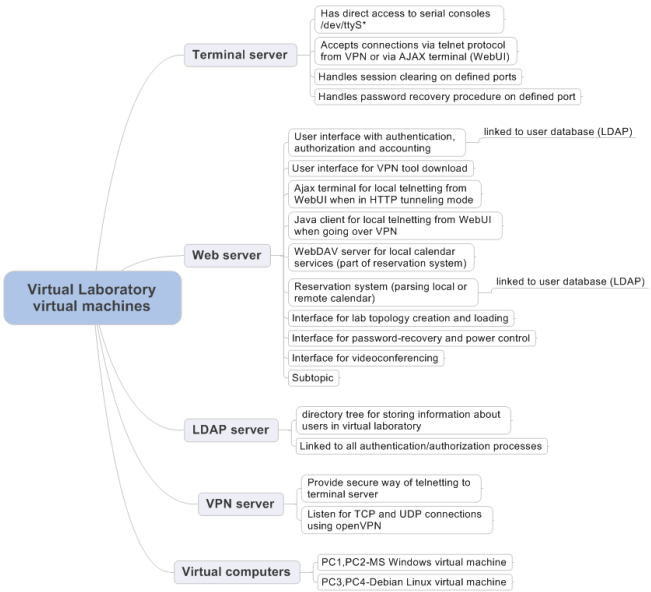


Figure 4. Virtual machines under VMWARE ESXI

E. Automated password-recovery

There are two different password-recovery actions that need to be supported by virtual laboratory. Password-recovery for routers is a bit different than for switches. There is strong need for speed change ability on each serial interface for successful password-recovery on router and mechanical push of MODE button for password-recovery on Catalyst switches. Key to password-recovery on routers is in control+break sequence generation. Practically this break sequence is generated by slowing speed down of console port lower than speed currently used and by sending 10 spaces. Therefore password-recovery on routers is done in following steps:

1) Power cycle the router (off/on)
2) Change speed of console port to 1200 bits per second
3) Send 10 spaces (0x20)
4) Change speed of console port back to default (9600 bits per second)
5) Configure config-register to 0x2142
6) Reload the router
7) Change config register back to default (0x2102)

Password-recovery procedure on Catalyst switches requires to manually push MODE button. The easiest way of how to do this is by shorting MODE button circuit by contact relay managed from server. As we want to keep warranty on our virtual lab equipment, we have developed a unique prototype for manual pressing of MODE button. "Buttoner" device is managed by SNMP and is mounted in rack on the top of catalyst switch. For the purpose of power control we have used SNMP managed power distribution unit (Figure 5).

Figure 5. APC switched rack power distribution unit

F. Topology and its re-configuration

Our physical topology of virtual laboratory is shown on Figure 6.

This network topology is physically static, we do not use any connection matrices as there is no need for doing this. We have decided to manage topology more logically than physically by using provider technologies like Q-in-Q tunneling and any transport over MPLS (AToM). This technologies allows us to logically create interconnections between each devices by using separated VLANs and to tunnel layer 2 protocols like CDP/DTP/STP by using Q-in-Q tunneling. Also interconnections between devices using serial interfaces (WIC-2T) can be done by frame relay circuits or by using encapsulation (tunneling) to MPLS

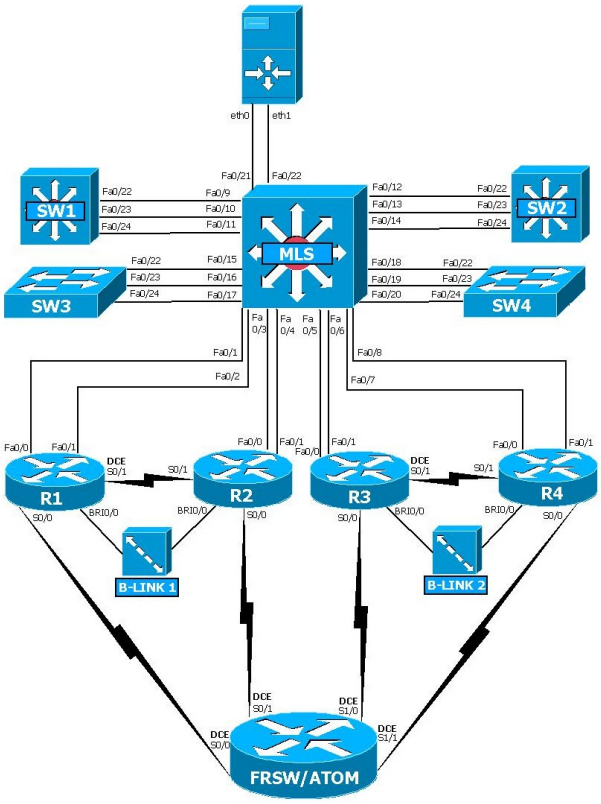


Figure 6. Virtual laboratory topology

(AToM). Table II shows configuration of virtual laboratory components when building logical topology from physical topology (Figure 6) shown on Figure 7.

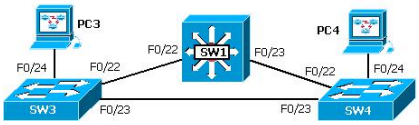


Figure 7. Example of virtual laboratory topology

Table II
EXAMPLE OF MLS CONFIGURATION FOR INTERCONNECTION SW1-SW3

```
MLS(config)# interface range Fa0/9, Fa0/15
MLS(config-if-range)# switchport mode dot1-tunnel
MLS(config-if-range)# l2protocol-tunnel cdp
MLS(config-if-range)# l2protocol-tunnel stp
MLS(config-if-range)# l2protocol-tunnel vtp
MLS(config-if-range)# switchport access vlan 10
MLS(config-if-range)# description SW1-SW3
```

G. Knowledge evaluation

Beyond technology there is content for exercises and knowledge evaluation. In our virtual laboratory we are able to practise exercises on CCNA and CCNP level without

 ISBN: 978-1-61208-133-5

any limitation. We have also some special labs on CCIP and CCIE R&S. Important part of virtual laboratory and its content is knowledge evaluation. For this purpose we have used our own regular expression based system for knowledge evaluation. Generally, we are collecting configuration files and different ”show” outputs from each device in virtual laboratory and comparing user solution against solution template.

Each template for specified lab exercise defines:

- start and end of block of evaluation by regular expression (e.g., only interface Fa0/0 configuration)
- line that should be evaluated within start-stop block by regular expression. This definition can contain fixed parts, variable parts and number-range parts
- scoring information - points for each occurence, minimal score per regular expression (important when penalty points are used), maximum number of occurences, penalty points per each occurence over the allowed maximum

Knowledge evaluation system used in our virtual laboratory is more described in [3].

H. Web user interface

Web user interface (Figure 8) acts as interface for communication with user. It is the central element for putting all the virtual laboratory pieces together. We have used some technologies like AJAX terminal that allows us to tunnel communication in case of limited access from user environment, PHP and Java technologies for running terminals from end station in non-firewalled environment.

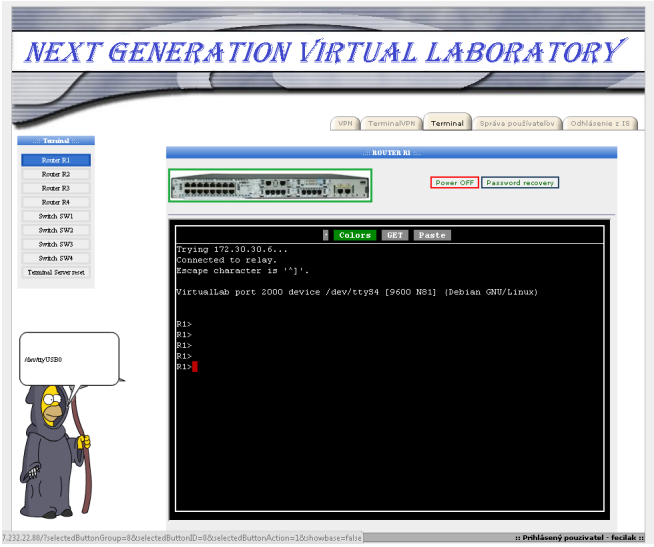


Figure 8. Web user interface of virtual laboratory

IV. Conclusion

In this paper we have presented several solutions for virtual laboratory operating at Regional Cisco Networking Academy at Košice. Our future work will be devoted to videoconferencing as a part of training in virtual laboratory and to special IPv6 training labs because we have joined to 7rp 6deploy project and as the only institution in this project from Slovakia we will be more focusing on IPv6 deployment.

Acknowledgment

THIS WORK WAS SUPPORTED BY THE SLOVAK CULTURAL AND EDUCATIONAL GRANT AGENCY OF MINISTRY OF EDUCATION OF SLOVAK REPUBLIC (KEGA) UNDER THE CONTRACT NO. 3/7245/09(60%). THIS WORK IS ALSO THE RESULT OF THE PROJECT IMPLEMENTATION: DEVELOPMENT OF THE CENTER OF INFORMATION AND COMMUNICATION TECHNOLOGIES FOR KNOWLEDGE SYSTEMS (PROJECT NUMBER: 26220120030) SUPPORTED BY THE RESEARCH & DEVELOPMENT OPERATIONAL PROGRAM FUNDED BY THE ERDF (40%).

References

[1] Cisco Access Control Lists: *Overview and Guidelines*, [on-line 11.3.2011], URL: http://www.cisco.com/en/US/docs/ios/12_2/security/configuration/guide/scfacls.html

[2] Cisco Packet Tracer, [on-line 11.3.2011] URL: http://www.cisco.com/web/learning/netacad/course_catalog/PacketTracer.html

[3] Feciľak, P. – Kleinová, K. – Jakab, F. – Bača, J.: *Automation in Knowledge Evaluation Process*, Proceddings of 6th International Conference on Emerging eLearning Technologies and Applications (ICETA 2008), Stará Lesná, Slovakia, 11. - 13. September, 2008, Košice, elfa, s.r.o., 2008, 1, 6, pp. 389-394, 978-80-8086-089-9

[4] Feilner, M.: *Building and Integrating Virtual Private Networks*, PACKT publishing 2006, ISBN:1-904811-85-X

[5] IEEE 802.1Q-in-Q VLAN Tag Termination, [on-line 11.3.2011] URL: http://www.cisco.com/en/US/docs/ios/lanswitch/configuration/guide/lsw_ieee_802.1q.html

[6] Jakab, F. – Janitor, J. – Nagy, M.: *Virtual Lab in a Distributed International Environment* - SVC EDINET, The Fifth International Conference ICNS 2009 on Networking and Services, LMPCNA 2009, Valencia, 20.-25. April 2009, Valencia, Spain, IEEE Computer Society, 2009, 5, ISBN 978-0-7695-3586-9

[7] Lobo, L. – Lakshman, U.: *MPLS Configuration on Cisco IOS Software*, Cisco Press 2006, ISBN: 978-1-58705-199-9

[8] Olegsby, R. – Herold, S. – Laverick, M.: *VMware Infrastructure 3: Advanced Technical Design Guide and Advanced Operations Guide*, BrianMadden.com Publishing Group 2008, ISBN: 0971151083

[9] Serial port to network proxy project, [on-line 11.3.2011] URL: http://sourceforge.net/projects/ser2net/

 ISBN: 978-1-61208-133-5

Enhancing Cisco NetAcad Student Learning Experience with an Integrated Learning Platform

Mihai Logofatu
University of Bucharest
Credis Cisco Academy
Bucharest, Romania
mihai.logofatu@credis.ro

Cristian Logofatu
University of Bucharest
Credis Cisco Academy
Bucharest, Romania
cristian.logofatu@credis.ro

Abstract: **The paper presents the integrated eLearning platform run by the Cisco academy at the University of Bucharest. The platform significantly adds to the functionality of the Networking Academy platform, for an improved student learning experience and better student retention rates.**

Keywords: learning platform; productivity; student experience; management; retention; online; satisfaction.

I. INTRODUCTION

The Credis Academy has been part of the Netacad Program for 12 years starting with 1999 as a local academy, and also becoming a regional academy starting in the year 2000. Managing more than 11 geographically dispersed local academies as well as offering Cisco's entire curricula to more than 2500 students only with the help of the NetAcad site had become a time consuming activity that needed to be re-engineered.

We started the process in 2006, developing in-house a course delivery platform coupled with an Enterprise Resource Planning (ERP) system [1], which is currently in use to manage over 2500 students spread over the entire country.

Our course delivery platform and ERP is currently integrated with NetAcad, NetLab, Webex, Google Apps, Project Management tools and social services like Twitter and Facebook.

This paper presents the solutions adopted in order to be able to scale up while maintaining a high degree of student satisfaction and achieving our motto: 'best learning experience".

The paper starts by justifying the need of the platform and then presents the three main parts of it. After that we have presented the achieved results and some of the future developments of the platform.

II. REASONS FOR AN ADDITIONAL PLATFORM

The official NetAcad website is the command center for the daily management activities of Cisco academies. This system is used to coordinate student activities and part of the instructor activities. However, the site does not seem to have been designed to manage a large number of returning students. We believe that is the reason why we encountered increasing difficulties dealing with the following issues:

- It is impossible to know the class history for a particular student,
- It is impossible to send custom e-mails to all students in one particular class,
- There is no scheduling/viewing system for practical hand-on activities,
- There is no tracking system for student's progress through practical activities,
- The user enrollment process is relatively difficult (manually/by username),
- It is difficult to contact students, which have no public profile.

Working with thousands of students has prompted us to implement a management system that would enable to overcome these problems. As a result, we designed and implemented our own online learning platform for the Cisco academy, to be used in conjunction with the NetAcad website. With the years this platform has become the central point of command for the entire academy, accommodating additional educational programs, like the Microsoft IT Academy, Adobe Training Center, Linux Professional Institute and Cisco Entrepreneur Institute.

Our learning platform is divided in three views: student experience, instructor experience and finance module. The student experience module is an implementation of our vision of fully flexible student services in the learning process. The instructor experience module is designed like an ERP, aiming to ease the work of academy instructors and administrative staff.

All the financial aspects of running the academy are integrated within the platform, in order to increase the productivity of the employees while considerably reducing the time a student needs for administrative purposes (registration details, accounts, invoices, etc.).

The platform is well integrated with the academy's public website, Facebook and Twitter accounts, giving us the opportunity to easily communicate with our student community.

III. STUDENT EXPERIENCE MODULE

The student experience module is a GUI with a left-hand side main menu and a content area, which is, in turn, divided into two tabs: general news and personal relevant information (Figure 1).

 ISBN: 978-1-61208-133-5

A. Main Menu

The left-hand side main menu gives students access to the following options:

- Course materials ("Curriculum" tab): access the course materials online,
- Student profile – update current contact information,
- Discussion forum- discuss with instructors and peers,
- Activity report– this is an area where students can see their own learning history throughout their entire period as students of our academy. The history shows courses taken, class dates and instructor contacts for each course.
- Legal agreement– every student must agree with the legal agreement setting on NetAcad as well as Credis policies on course completion, attendance, etc. The document clearly states out the rights and obligations of both student and academy.
- Support– this link gives students access to an online ticketing system. Student's requirements and complaints are registered and sent to the helpdesk. If the problem exceeds the helpdesk attributions, instructors and then the manager are contacted. During this period, students are constantly given feedback on the status of their tickets.
- Social networks tab– this tab allows students to follow the academy on social networks like Facebook, Twitter and YouTube. We also send out newsletters and an RSS feed.
- Netacad login– this area allows the student to beam directly to the Netacad website, upon password confirmation.

B. General News Area

The general news area is centrally located in the upper part of the browser content window. This area contains all the news regarding the academy's activity, including job offers, internship offers, new courses, sweepstakes, events, etc.

C. Personal Information

The personal information area displays information relevant to each student given their own learning context: all the classes that one is registered for, name and contact data for the corresponding instructor, etc.

Students can see the course's instructor led activities and their scheduling, can make reservations for such activities, and can access deeper information of such activities, including lab topics or lab prerequisites.

Every class has an associated study plan, class rules, additional learning materials and certification information. All courses are displayed as part of a career path graph.

D. Certification Area

In the certification area students have access to information regarding the possible certifications that the course opens the way to and are shown a certification path that helps them better understand and manage their learning process in relation to their goals.

IV. INSTRUCTOR EXPERIENCE MODULE

The instructor experience on our learning platform is, obviously, different from the students. Instructors' main menu contains the following tabs:

- Instructor home (Figure 2),
- Instructor profile,
- Study materials,
- Classes,
- Accounts,
- Activities,
- Online courses,
- Feedback,
- Forum,
- Legal agreement,
- Inventory,

We will only concentrate our presentation on the academic aspects: the instructor experience module is a result of our need to also manage a large number of instructors that are geographically dispersed (we currently have over 80 instructors around the country), while keeping the highest level of quality possible, both in order to comply with Cisco's Quality Assurance Plan (QAP) and to keep our good name with students.

In effect, the instructor experience module acts as an academic Customer Relationship Management (CRM) [2] system, giving academy staff the tools to manage students and their learning process.

Instructors and academy staff also have access to their own profiles just as students do, in order to update contact data, review their legal agreements (different than the ones signed by students), etc.

The news area and the NetAcad login are also available in the instructor experience module.

A. Scheduling Area

The main difference in the schedule area comes from the possibility of instructors to schedule new activities. When accessing this area, instructors see the entire schedule of activities grouped by location, classrooms, instructors and subjects. This way, any instructor scheduling new activities can only announce a lab in a free time slot.

Instructors schedule lab activities for a long time in advance, in order to book classroom timeslots, but have an option not to publish these activities to students. In this respect, the tool is used for classroom provisioning.

Only at the time of publishing of such a lab activity to the student experience module, do students receive emails asking them to register and confirm their participation.

Another use of this tool is to give management a good overview of the activities in the academy and the physical presence of instructors in the office.

 ISBN: 978-1-61208-133-5

B. Study Materials Area

The study materials area is used to give students access to a local copy of the curricula. We do that because we can increase the speed of access to the curricula like that.

In addition to the official curricula, instructors have the opportunity to publish materials that are relevant to the course in general or just to a particular class, thus customizing students' learning experience (Figure 3).

Also in this area, Cisco Press books [3] are recommended for further study. Starting with 2010, we have a partnership with Pearson Education Reseller for Romania and we are offering our students the opportunity to buy Cisco Press books with a discount to market price.

We believe that blending the Cisco Press titles with the learning materials increases the value of our learning proposition to our students.

V. FINANCIAL MODULE

The financial module is only available to academy staff and acts as a back-end of the application. This area uses the same database with the front-end (student and instructor experience modules), allowing for invoices to be easily and quickly issued to students, for tracking of students' financial status, to calculate instructor's wages based on a revenue sharing scheme, etc.

We are the academy with one of the largest number of instructors worldwide. These instructors are geographically spread across the entire country. We offer them the entire infrastructure to be able to offer Cisco curricula to their students. As a result, we use a revenue sharing model with most of our instructors and we use the financial area of the application to automatically calculate obligations, to generate all accounting and legal documents needed and to keep records of our activity.

We run a cash flow management system, where we record all our activity in our system, in order to be able to quickly have reports, find out the financial standing of the academy, the results of promotional campaigns, and numerous other data-analyses.

VI. RESULTS ACHIEVIED WITH THE IMPLEMENTATION OF THIS PLATFORM

With the constant development of our learning platform we have achieved several important improvements of our activities:

A. Improved student experience

The philosophy behind eLearning and the NetAcad curricula is that students should be able to learn anytime, anywhere, at any pace. This flexibility was not possible with the practical, hands-on lab activities. We have taken one step forward with the implementation of the scheduling system.

Instructors schedule labs repeatedly, at different hours and in different days, including weekends. With this approach, students are given a wide choice of options for their participation in the mandatory lab activities, being able to manage their learning schedule according to their needs.

Furthermore, at the end of each lab, instructors confirm the actual presence of the registered students. This way, we can easily track student participation and performance in the practical activities.

We can demonstrate a relatively high degree of customer retention, with one third of students in each year being "returning" students (that is students that have been our students in the previous year too).

TABLE I. STUDENT RETENTION RATES

Year	Returning students	New Students	Total / year	Retention Rate
2008	490	1109	1599	31%
2009	682	1537	2219	31%
2010	849	1676	2525	34%

Retention Rate is calculated as the number of Returning students divided by the Total number of Students per each year.

We also believe that the improved user experience has translated into higher number of students registering for continuation of their studies. There is drop off rate between CCNA 1 and 4, with numbers constantly decreasing with each new course, but we also can show increasing numbers and percentages in this area:

TABLE II. STUDENT PROMOTION RATES

	2008		2009		2010	
	#	%	#	%	#	%
CCNA 1	369		525		473	
CCNA 4	61	17%	78	15%	91	19%

Number of students registered in CCNA 1 courses and CCNA 4 course, and the percentages thereof.

B. Improved student management

By duplicating the classes in the NetAcad website in our own platform and database we are now able to easily and automatically extract any kind of information about student participation and involvement in the program.

For example, during 2010 we have had a number of over 6000 activities scheduled. For these activities we have clear information about students that have registered and students that have actually attended. This is useful to generate track reports of students in their interaction with the academy. This feat is impossible to achieve with the NetAcad site.

At any time during a course we can tell how the activities are progressing, the attendance rates, etc. This allows us to keep track of all the numerous instructors spread throughout the country and be able to ensure quality in the learning process.

Using the NetAcad site we were not able to register students into classes, but the other way around. We had to create a class and then search for the students that we needed to register in that particular class. With growing numbers of students, this operation became cumbersome and time consuming. In the present, we can select students and enroll them in whatever class they have required, automatically sending the full-HTML emails with relevant information.

C. Improved Financial Management

With the integration of the financial module, the time students spent dealing with the front-desk has significantly

 ISBN: 978-1-61208-133-5

decreased. In the present, we easily know the entire student history both from an academic view point and from the financial one.

Furthermore, classes are treated as projects in the financial module, allowing us to easily create all the reports and legal documents that support our revenue sharing model with the instructors.

D. Improved Reporting

Using our own database allows us to easily create any number of statistics that we need, empowering us with the tools for data analysis.

Using our own database we were able to create a map of the city of Bucharest related to the geographical distribution of our students, allowing us to identify the areas of the city where we need to increase our marketing efforts.

Another good example of useful reports is the "customer loyalty report", which has shown us the students that have followed the most courses with our academy, and we were able to recognize their participation, thank them for their trust, and invite them to become ambassadors of our academy. Getting feedback from the persons that know us most is a very important achievement.

VII. Future Development of the Platform

At the time of writing of this paper we are working on integrating new social web technologies in our platform. We want to implement the OpenID [4] standard to allow students to register for classes using their social network or web accounts (Facebook, Google, Yahoo or any OpenID provider) in order to ease students' interaction with our platform.

We are also testing different wiki solutions in order to implement a collaborative knowledge base that would add value both to our students and to our instructor all over the country.

We intend to take one additional step in the blending of Cisco Press books with the study materials. In the present, we are showing relevant Cisco Press titles in the study materials area. Students are directed to the website pages describing the book and they can place orders.

However this is not an actual online shop, but only an online order placement system. We intend to fully integrate an online book shop on our platform.

In this respect, we are close to finishing the process of adding the possibility for students to register and pay online for our courses, in a desire to constantly keep improving our overall offer, aiming to provide a fully satisfactory learning experience.

VIII. Conclusions

Although this is an academic and not-for profit environment we found that our processes need to be designed the business way. That is why we have developed business specific tools and have customized them for our academic activity.

Direct results are a significant improvement of work relationships, due to the fact that everyone's responsibilities are clearly defined together with the expectations for each employee.

Furthermore, productivity tools enable us and our instructors to concentrate on increasing the quality of our students' learning experience in order to obtain even higher degrees of student satisfaction, retention and graduation as a direct measure of teaching/learning success.

We aim that our students will speak about the Credis learning experience as the 'best learning experience' ,which is why we need to be constantly developing and testing new ideas and technologies, especially more collaborative ones using Web 2.0 tools like Facebook, Twitter, blogs, and wiki's.

References

[1] http://en.wikipedia.org/wiki/Enterprise_resource_planning [retrieved March 14, 2010]
[2] http://en.wikipedia.org/wiki/Customer_relationship_management [retrieved March 15, 2010]
[3] http://www.ciscopress.com/index.asp [retrieved March 16, 2010]
[4] http://openid.net/[retrieved March 15, 2010]

 ISBN: 978-1-61208-133-5

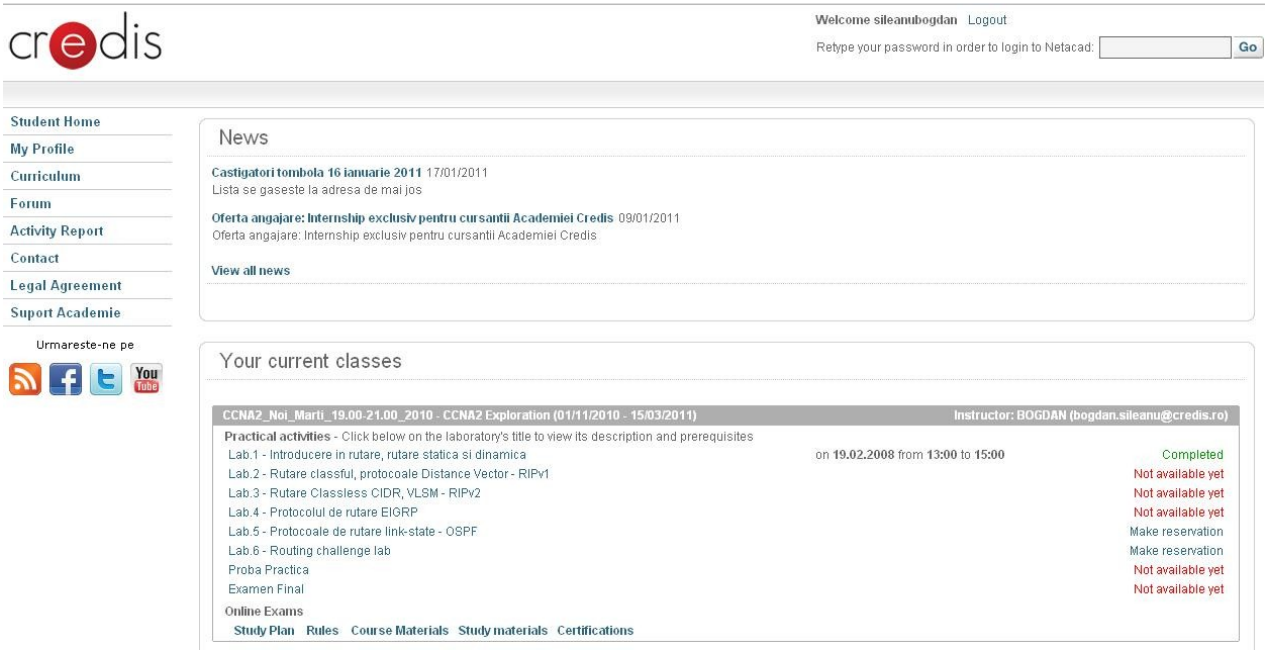


Figure 1. Student Experience Home Page.

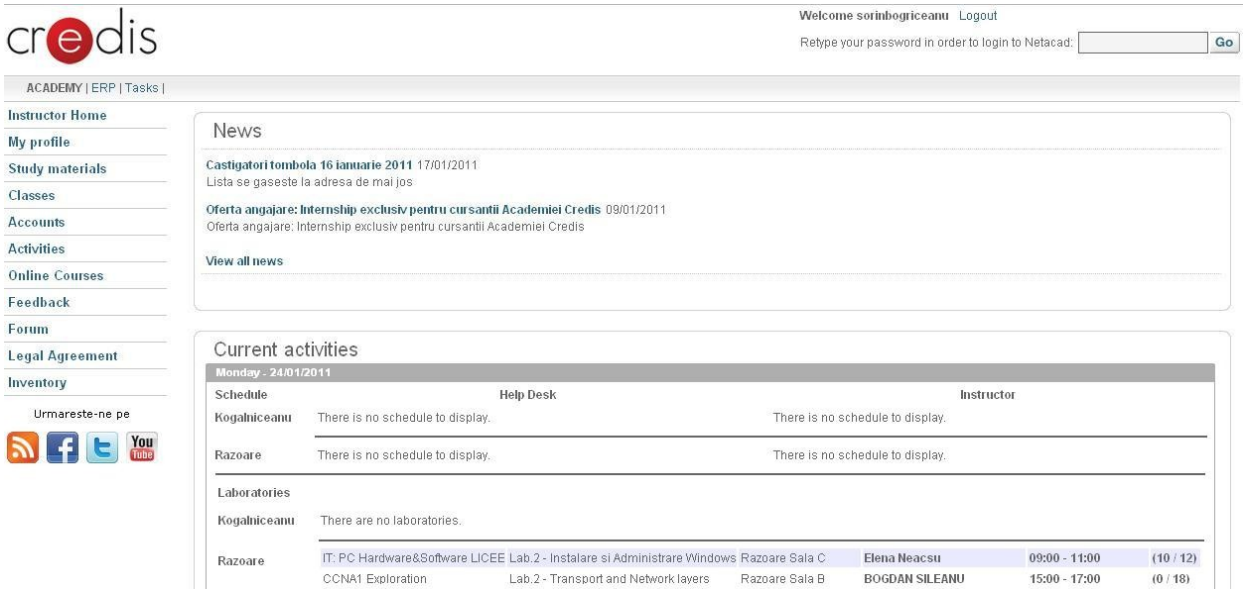


Figure 2. Instructor Experience – Home Page

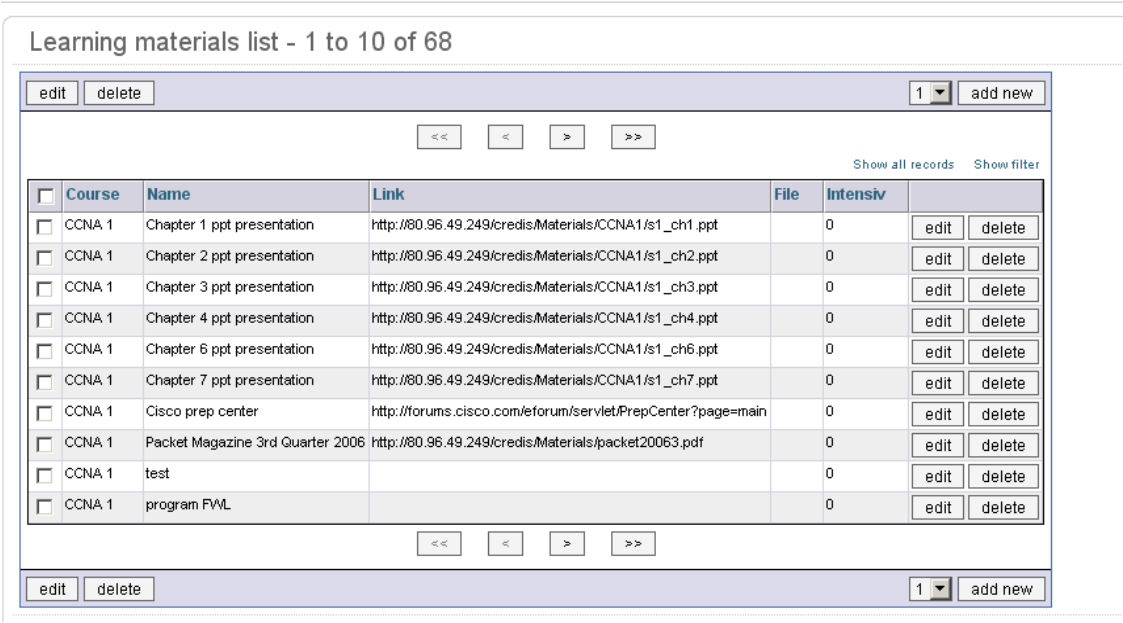


Figure 3. Learning materials added for a CCNA 1 class.

Building Interactive Multi-User In-Class Learning Modules For Computer Networking

Oleg Sotnikov, Ammar Musheer and Shahram Shah Heydari
University of Ontario Institute of Technology, Oshawa, Canada
{oleg.sot@gmail.com, musheer.ammar@gmail.com, shahram.heydari@uoit.ca }

***Abstract*—The new multiuser capability in Cisco Packet Tracer tool has provided great potential for developing interactive multiuser activities that can be completed by Cisco Academy students in the classroom. This approach will not only greatly enhance students' interest in the technical topics, but will also allow the instructor to create group activities where student progress can be monitored and tracked easily on the instructor side. In this research we focus on primary design principles for creating such interactive activities. We present a general architecture for a library of interactive modules we developed for CCNA exploration 1-4 topics. We provide solutions for a number of technical problems that must be resolved for multiuser access, such as scalable addressing, VLAN handling, individual student monitoring, and offline exporting. We also provide testing results to evaluate the performance of packet trace in an interactive multiuser environment.**

Keywords-component; blended learning; packet tracer; multiuser; simulation; teaching; ccna; cisco; networking; academy;

I. INTRODUCTION

A. Background

Introductory computer networking education (e.g. Cisco CCNA level) focuses on teaching students both the theory and the practical knowledge about networking technology that has rarely been covered at their prior educational levels (high school). Students often have to bring themselves up to speed quickly on a large amount of content that includes the entire OSI model and TCP/IP protocol suite, several routing and switching protocols, local and wide area networks etc. The sheer volume of technical details in this case can quickly result in a dry and non-interactive environment that can impede student learning process. Lab components often improve the dryness of the material by giving students practical examples to reinforce theory. However, more can be gained by increasing the level of interaction during lecture times and in the classroom. This can be achieved by implementing short in-class activities that encourage students to collaborate with their peers throughout the learning process. Interactive learning is becoming more prominent and show great potential in teaching IT concepts.[1,2] Likewise, e-learning tools are used with great benefit to teach CCNA Discovery and Exploration curriculum of the Cisco Networking Academy.[3] One such tool is Packet Tracer (PT), which provides a network simulation environment with sufficient details of IOS for CCNA levels [4].

The objective of this work is to use the recently added multiuser feature in Packet Tracer to create in-class interactive learning activities that would enhance students' understanding of complex networking concepts. The multiuser feature in Packet Tracer allows students to work in an environment that is affected by their peers and is under control of the instructor. Despite some technological problems, PT multi-user activities can make networking more interesting to learn and lead to greater student engagement.

B. Related Work

The Multiuser functionality was added to Packet Tracer in August 2008. However, at the time of writing this paper Cisco has not yet developed any curriculum activities that feature multiuser operation, leaving it up to individual academies. In 2010 the Open University of UK reported on implementation of PT's multiuser functionality into their blended distance learning CCNA courses.[5] Their results offer an extensive guide to the multiuser architecture as well as the implementation of multiuser over the WAN, and the inherent problems. We build on their work by implementing our multiuser architecture in a traditional classroom through a LAN. This method bypasses the majority of the technical limitations in [5] and gives students more interaction with the class.

Basic technical specifications of the Packet Tracer Messaging Protocol (PTMP) and Inter-Process Communication is available in [5,6], and while many proprietary details were not available or could not be made public, the available information helped understanding the communication between two hosts running a Packet Tracer multiuser connection.

The rest of this paper is organized as follows. In Section II we describe our objectives and design specifications for interactive learning modules. In Section III we discuss some technical challenges in creating interactive learning modules and present our solutions to overcome these challenges. This section also includes some performance results. Some performance results are presented in Section IV. Section V includes our conclusions as well as future plans.

II. LEARNING MODULE SPECIFICATIONS

The problem of engaging students in the process of learning networking is normally solved by using a blended learning approach that Packet Tracer is already

 ISBN: 978-1-61208-133-5

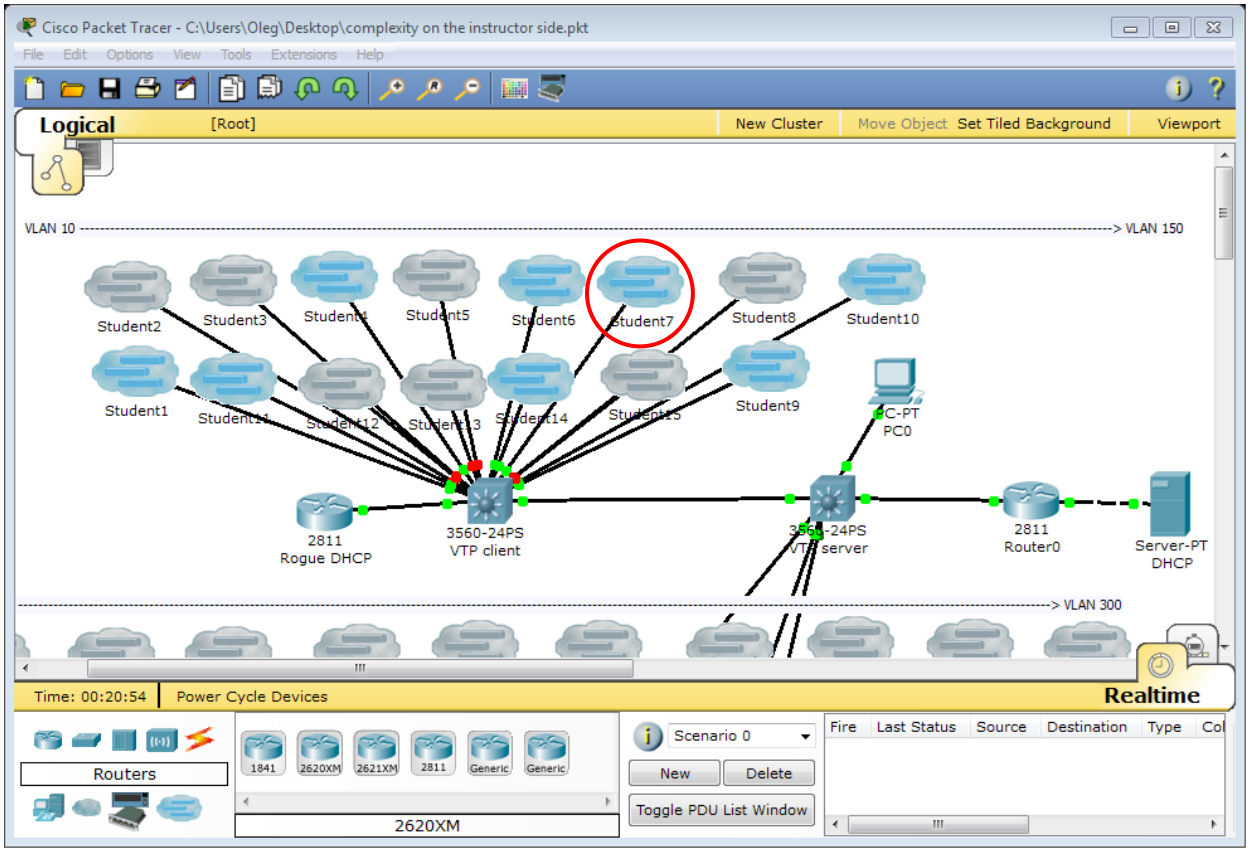


Figure 1: Example of an Instructor-Side Module

part of. However, current Packet Tracer activities are highly scripted, having little interactivity and class participation. Multiuser activities allow Packet Tracer to be used in a more dynamic way, allowing the instructor to affect student's Packet Tracer environment in real-time. Having multiuser activities as part of networking course would fill a gap of short, interactive and extensible activities that can be used to promote student participation in lectures. The activities must be simple and require minimal configuration by the students, and they should be used in-class and be marked through student participation. The result of this project included 40 multiuser activities and two games covering the entire CCNA curriculum. The design of the activities closely followed the framework described in this paper. The activities are currently being integrated into the two introductory networking courses at UOIT that cover the CCNA curriculum.

A. Architechture

Significant differences between standard and multiuser CCNA activities create a whole different set of factors that have to be taken into account in the latter case. The main task is minimizing the disadvantages and limitations of the multiuser feature and maximizing the advantages it offers in interactivity and real-time communication. We achieve this objective by creating a thin client-server environment in which most of the complexity and configurations are on the instructor side. The fact that UOIT is a laptop-based university in which all students use pre-configured university-issued laptop in the classroom, greatly facilitated the implementation and use of this model.

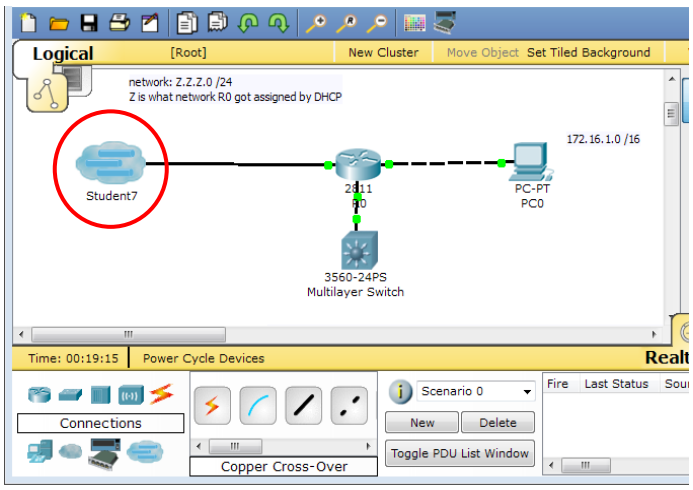


Figure 2: Example of a Student-Side Module

The instructor-side module is responsible for providing

the hub part of a hub-and-spoke topology. The students can then start a multi-user PT activity on their respective laptops and establish connections to the instructor side of the activity. This usually involves two PT activity files: the instructor file, and the student file. Student-specific configurations or modifications must be kept to minimum. Figures 1 and 2 display examples of the instructor and student side modules, respectively.

The instructor should also be able to save the student's progress at any point and inspect it on his/her own time. This approach is ideal for completing an activity in the classroom and marking it later. Depending on who the instructor file is released to, the activities are open for modification and can feature any new content that gets added to PT in the future. In general, a laptop per student is required. The requirements for the instructor computer depends on the number of students that are expected to connect and will be discussed later in this paper.

B. *Multi-user activities specifications*

To retain student attention, multi-user activities were designed to be completed in 10-20 minutes. The activities would be presented in the middle of the lecture and allow the instructor to demonstrate a particular networking concept interactively with the students. The activities assume that the students have limited hands-on experience and are mainly used to demonstrate the networking concept. In general, the activities follow these requirements:

1) Minimal configuration or pre-configured for the student
2) Student tracking through offline saving
3) Allow group work
4) Task variety between activities

The major limitation to this architecture is that PT's simulation mode cannot be used as easily during a multiuser connection. Therefore this architecture pushes the focus of the activities to the real-time interaction between the instructor and student networks.

C. *Modifying and extending activities*

Creating and editing router and switch configurations in text files is considerably easier and more manageable than working within PT. To create or extend the functionality of a certain activity, it is easiest to first export all configurations of each device into text, and edit the configurations using a text editor. Finally you can either erase the configuration in the old topology and load the new configurations, or simply load the new configuration overtop of the old one. One must be aware that certain configuration options (VTP) are not saved in the run configuration and must be configured each time. Knowing how to effectively use the "find and replace" function featuring RegExp (or similar extended options) is key to creating new activities in a timely manner. Figure 3 shows an example.

The intent of multiuser is not to replace hands-on portions of networking. Instead it is meant to increase student engagement during lectures. As such in multiuser activities the learning curve to get started and the amount of tedious configuration should be minimized. This can be done by putting commonly accessed devices, such as servers and core devices, on the instructor side. To minimize configuration, the student side has devices that are preconfigured, allowing the instructor to limit the focus to a particular topic discussed in class. Minimal configuration from the student is necessary to differentiate the students and allow the networks to communicate.

As we shall explain in the next section, student devices can be assigned addresses in a unique network depending on the how VLANs are configured. This feature allows students to configure routing protocols between each other. Alternatively, students can be split up into groups on different VLANs, and assigning them addresses accordingly.

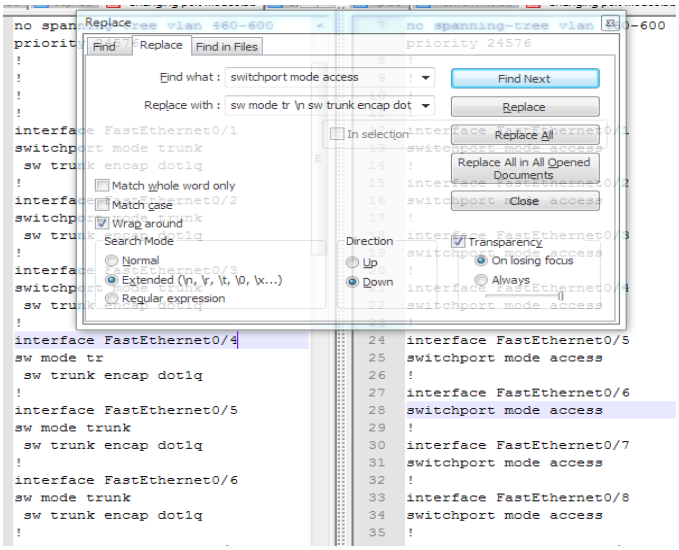


Figure 3: Editing Activities

D. *Student Evaluation and Monitoring*

PT's activity wizard offers an extensible script-based evaluation system that examines the parameters of each device in the student network. In our work students were evaluated based on their participation in the multiuser activities. The TCP traffic between PT instances is unencrypted and information such as hostnames can be found by capturing that traffic. Additionally, the instructor side can poll the student network and generate an offline copy of the whole network. This offline saving capability allows he instructor to the save the state of the network at that particular time and view the commands that each student entered on their respective activity.

III. COMMON ROADBLOCKS

In designing multiuser PT activities several good practices were developed. The main principle is maximizing the advantages gained by having interactive activities. Configuring multi-user activities can be split into three layers. Layer 1 is the physical and data connection between hosts running the PT instances. Layer 2 is the data connections between switches and

 ISBN: 978-1-61208-133-5

relates to MAC addressing and VLAN assignment within the PT environment. Layer 3 is responsible for the rest of the connectivity between PT environment such as IP addressing and routing.

A. Layer 1 Issues

Documentation on how to setup multi-user connections between packet tracer instances is readily available [5]. Connections between instances are done by matching IP address, port, cloud name, and password parameters between instances. PT offline saving accounts for the most tolling activity on computer and network resources. Remote student computers are polled at the same time resulting in bursts of traffic that is well handled over the LAN but may result in problems in a WAN implementations like that of Open University of UK. Traffic encryption is optional and useful information can be gathered by collecting the traffic data in Wireshark.

B. Layer 2 Issues

As mentioned in the previous section, in most activities students start out with an identical student-side file of the network. This presents a problem because all students now share the same MAC addresses within the PT environment. One way to work around duplicate MAC addresses is by using NAT (Network address translation), but this effectively cuts off inter-student communication. Multiple VLAN's on the other hand maintain inter-student communication when configuring routing protocols. Duplicate MAC addresses can still interfere with various protocols on the instructor-side. STP (Spanning tree protocol) must be properly configured on the switches within PT. The switchport connections in Figure 1 can be inadvertently blocked off by STP due to the adjacent port MAC addresses being identical. The switches on the instructor side should be configured with a very low STP priority to guarantee that they become the root and no ports gets blocked.

When designing activities that teach VLAN concepts. Problems can be encountered on the instructor side if past configurations with different VLAN schemes are reintroduced to the instructor network. It is recommended to save the start configuration, erase the vlan.dat and restart all the routers simultaneously to avoid VLAN/VTP problem when reconfiguring VLANs. Despite the additional complexity on the instructor side, in most activities no other changes have to be made on the student side.

C. Layer 3 Issues

DHCP is primarily responsible for achieving layer 3 connectivity with minimal configuration from the students. Prior to developing addressing schemes, layer 2 should be problem free. While DHCP servers can be set up on routers, for classless networks the DHCP has to be setup on a separate server as shown in Fig. 1. Although DHCP is not intended to provide IP addresses to devices that with identical MAC address, if layer 2 is problem free, the only DHCP problems encountered can be solved by issuing DHCP release/renew commands.

D. Scalability, Security and other Limitations

The scalability of the hub and spoke topology in use is limited. During 60 user activities, the hub computer will be responsible for any bottleneck encountered. It may be necessary to split the class into two 30 student groups that will connect to different hub computers (instructor laptops). The hub computers can also optionally be connected to each other, providing connectivity in a more scalable manner than a single hub computer.

Accountability between authenticated users is very limited and productive use of PT would require mutual trust between the users. Although, offline saving and capturing a Wireshark trace provide limited logging capability. While specific task evaluations can be automated using Packet Tracer's Activity Wizard, we were not able to find a practical method to account for every detailed student action in the multiuser network unless the PT traffic captured by Wireshark can be reverse-engineered.

Denial of service is possible because students can configure a Layer 2 loop between switches in PT which can affect the instructor side. Although technological solutions may be written and added onto the PT platform by using the included extensions interface. Due to lack of documentation and added complexity these resources were not considered when implementing module activities. However, if the activities are only used for participation and have a very small academic weight assigned to them, the risk for abuse is much lower than if the activities were worth a significant portion of the course grade.

Using multiuser mode restricts access to the simulation mode, which is available when PT is used offline. This problem restricts students' access to a very useful tool that visualizes many theoretical components of networking. In simulation mode the student can choose to view the level of information that makes them comfortable in understanding a topic. Although it may be possible to use simulation mode on a multiuser network, its integration may not be practical for application in a classroom. Moreover, since students learn at different paces, as activities get longer the gap between the fast and slow learners widens. As such, in our implementation we have limited the activities to 10-20min in length.

 ISBN: 978-1-61208-133-5

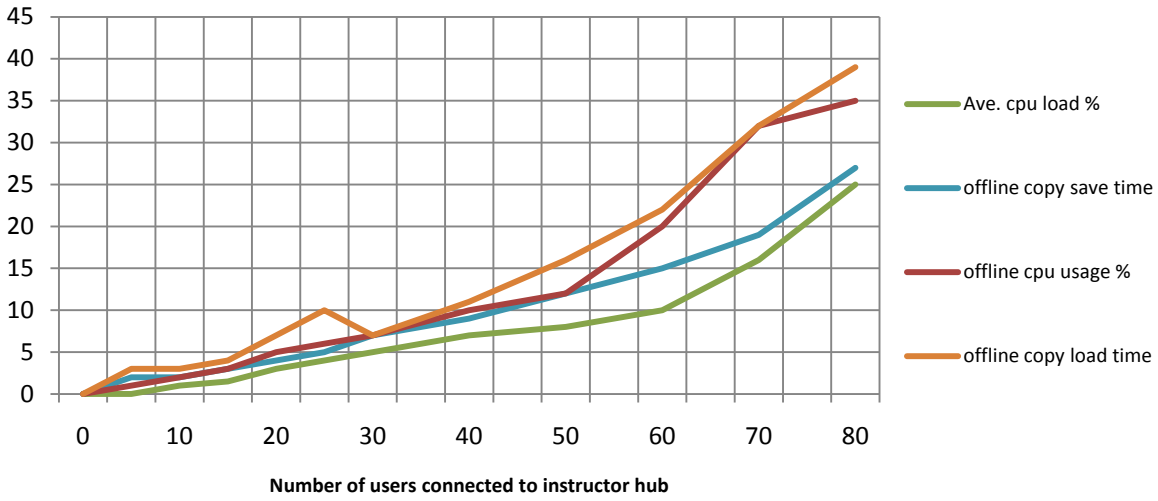


Figure 4: Packet Tracer scalability with 1 hub

Lack of commands is probably the most common limitation that will be encountered in designing multiuser activities. The level of commands is acceptable for a CCNA level at the student side. However to facilitate the large interconnected network, it often requires a CCNP level of commands on the instructor side. The level of complexity on the instructor side is also much higher than the student side, making troubleshooting more difficult. For example for a 60-student class activity: the instructor network makes use of about 60 VLANs that are used to assign different network addresses to each student. VTP will propagate VLANs when they don't have to be, and PT is limited in that you cannot turn on VTP pruning, nor turn off VTP. This often results in needless complexity in the way STP operates (PT by default uses PV-STP), and has on occasion unnecessarily blocked ports and caused loss of connectivity.

E. Exporting

PT is limited to interact only with devices that are simulated in the PT environment; external syslog servers cannot receive logs. Text, such as router configurations is the most easily exportable information, but not all devices have text interfaces. For example, DHCP servers that address all the VLANs can only be configured through a GUI. As such it can be more practical to build on prior versions of the instructor file rather than create a new one. Unlike GNS/Dynagen and other network simulators, PT cannot generate a network from a text file, nor export a network into any other format. The simplest way to save the work of a student is by using the offline saving feature on the instructor side.

IV. SCALABILITY TESTING

A. Methodology

A single prototype activity file was used for all the student connections, and a single instructor file was used that would accepts all the connections. Statistics about resource consumption were recorded by observing CPU, memory, and network usage as more and more students were connected to the instructor's simulated network. The hub was based on a school Lenovo T61p laptop (Core 2 Duo T8300 2.4GHz, 2GB RAM). The method of gathering the system information has its limitations in precision but portrays an accurate picture of the increase in system resources as the number of student rises.

B. Results

Several times the instructor file became unresponsive and testing had to be restarted. The growth for memory resource was very scalable to the amount of students. However, CPU usage and loading times grew at a greater rate

C. Discussion

In most cases for the instructor using more than 1 hub would greatly reduce the practicality of multiuser activities. As such, multiuser activities remain most practical in classes that have less than 60 students. Likewise, if classes are small, the requirements for reliability can be lower, as technological problems can be easily managed by the instructor.

 ISBN: 978-1-61208-133-5

V. CONCLUSIONS AND FUTURE WORK

Networking courses that cover CCNA material suffer from having to cover a large amount of theory to students that may have little networking experience. In class lectures can be improved by using interactive activities that foster student participation. Multiuser achieves greater class interaction by allowing students to form what is essentially one large network supervised by the instructor. Unfortunately there is no library of multiuser activities provided by cisco, and the development rests with individual academies. UOIT research has generated 40 activities and 2 educational game activities to cover the CCNA material and provide greater class participation. The activities are implemented as short in-class exercises maximizing the benefit of real-time communication between student networks and working around multiuser limitations. Multiuser does not replace normal PT activities or lab time, and should not be used to significantly evaluate students.

Although PT multi-user activities experience some limitation when used over a WAN[5], as the tool matures multi-user activities may provide an additional level of interactivity for distance-learning. The multiuser architecture does not take account the capabilities of simulation mode and focuses on the real-time communication advantages offered. Simulation mode can offer greater understanding for students but is not fully integrated into multiuser.

Future work on the multiuser feature would have the most impact by improving the quality and user experience offered by multiuser activities. Other academies may seek to modify the instructions in order to change the difficulty of some activities. Student feedback is critical to effectively improving activities and bringing them up to the standard of activities offered by Cisco. In the meantime it is up to the instructor judgment to adjust the difficulty or instructions on the spot.

Various multiuser architectures may be attempted. In contrast to hub and spoke topology in use, ad-hoc PT connections are low resource and can make use of the simulation mode of PT. The problem with ad-hoc connection is that the instructor is not present to assist students and it is difficult to account for the activities completed. This may offer a different learning experience to the student. However user input would be needed for a successful implementation. The activity wizard provided by PT can be used to create more locked down activities for the student and may address certain security limitations. The challenge with using the activity wizard lies with the increased development time and the requirement to account and restrict/facilitate all the student actions that may occur in the activity.

CCNP level multiuser activities are possible as long as the activities are limited in the depth of the topic they attempt to cover. These activities would increase the complexity of the instructor network and be highly dependent on PT limited amount of commands.

Compared to regular PT activities, by being connected to the same network, multiuser allows students to collaborate and work towards a common goal. This allows the creation of activities that have before been impractical to implements, such as group troubleshooting, capture the flag, and relay race games. Under the right conditions PT multiuser activities can be used as another to aid in teaching networking and maintain student interest.

ACKNOWLEDGMENT

This research was supported through a 2010 Teaching Innovation Funding (TIF) grant from the office of the Associate Provost, Academic of the University of Ontario Institute of Technology. Packet Tracer is a product of Cisco Networks and is provided free of charge to Cisco Networking Academy students and instructors.

REFERENCES

[1] D.D. Burdescu, M.C. Mihaescu, C.M. Ionascu and B. Logofatu, "Support system for e-Learning environment based on learning activities and processes," in Research Challenges in Information Science (RCIS), 2010 Fourth International Conference on, pp. 37-42, 2010.

[2] Maiga Chang and Kinshuk, "Web-Based Multiplayer Online Role Playing Game (MORPG) for Assessing Students' Java Programming Knowledge and Skills," in Digital Game and Intelligent Toy Enhanced Learning (DIGITEL), 2010 Third IEEE International Conference on, pp. 103-107, 2010.

[3] F. Jakab, M. Bucko, I. Sivy, L. Madarasz and P. Cicak, "The system of career promotion of networking professionals based on industrial certificates," in Intelligent Engineering Systems, 2009. INES 2009. International Conference on, pp. 221-226, 2009.

[4] Packet Tracer Reference Guide and Tutorials, Cisco Networking Academy, 2010.

[5] A. Smith and C. Bluck, "Multiuser Collaborative Practical Learning Using Packet Tracer," in Networking and Services (ICNS), 2010 Sixth International Conference on, pp. 356-362, 2010.

[6] Packet Tracer Messaging Protocol (PTMP) Specification Document, Cisco Networks, 2008.

 ISBN: 978-1-61208-133-5

Network Simulation and Remote Laboratory Systems for Students with Vision Impairment

Iain Murray
Department of Electrical & Computer Engineering
Curtin University
Perth, Australia
i.murray@curtin.edu.au

Alan Ng
Department of Electrical & Computer Engineering
Curtin University
Perth, Australia
alan.ng@student.curtin.edu.au

Abstract— **This paper describes further developments in laboratory and simulation tools for use in the Cisco Academy for the Vision Impaired (CAVI) program. Due to the geographic distribution of vision impaired students it is necessary to offer the normally "hands on" aspects of the Cisco Academy Program in a remote, virtual or simulated manner. Two major issues currently faced in delivering this program is that of the inaccessibility of Packet Tracer and scheduling equipment access to the Remote Bundles.**

Keywords-component; Remote laboratories, network simulaltion, assistive technology, blind, vision impairment

I. INTRODUCTION

The Cisco Academy for the Vision Impaired (CAVI) is a learning environment that modifies the "standard" Cisco Networking Program in a manner that is suitable for people with significant vision impairment. Students at CAVI acquire skills in computer networking design and administration through a variety of tailored facilities, such as a laboratory of network equipment that can be accessed remotely via the Internet.

Disabilities inhibit many from obtaining gainful employment and the building of skills in IT enables those with a vision disability to be better equipped to take on employment in the IT industry. With an unemployment rate of more than 60% in Australia, the vision impaired battle to compete for employment with those who are sighted [2]. Technology has become an integral part of the business world with little being achieved without some use of information and communication technologies. In many cases organizations use technology to gain a competitive advantage, hence the need for IT skills is paramount for employment in today's business world. Education courses in introductory skills in computing are offered by many disability support organizations to their members, however very few advance beyond the basics of how to use email, Word, and how to search the Internet. A scan of the literature on accessible e-learning environments results in little in the way of new developments.

People who are blind rely on computers and the Internet for information essential to their lives, availability of goods and services, directions to get from one location to another, pay bills and do their banking, communications with other people, transferring information and documents to businesses, family and friends. The ability to set up their own home computers and networks is a great advantage to them. With advanced IT skills they can also troubleshoot problems and fix their own systems when they fail. They can also do this for their employer, making IT help desk and network administration a relevant employment role for people with vision impairments.

E-learning environments suitable for remote delivery of accessible IT curriculum have been a long time emerging. The needs of those with vision disabilities are specialized, requiring assistive technologies to access the information technology and curriculum. Research and reporting of advances in this area have not been forthcoming, with pockets of work on accessible teaching methods and curriculum, but little spans the digital divide between sighted and vision impaired students in the area of information technology and network administration education. Employing an e-learning environment consisting of a virtual classroom, remote laboratories and curriculum fully accessible to the vision impaired, the Cisco Academy for the Vision Impaired (CAVI) delivers advanced IT network education to students situated around the world. Students access the learning system via the Internet, enabling those in remote locations to work through the courses at their own pace and within their own time zones. This paper describes the remote e-learning environment designed for the delivery of advanced IT networking to vision impaired students so that others may benefit from the low-cost and effective design.

The paper first describes the development of an accessible external application to connect with Packet Tracer.

Secondly, users of the remote laboratory face the problem of not knowing ahead of time whether the equipment is going to be occupied or not. Therefore, a mechanism for scheduling laboratory usage is required. This mechanism must also be accessible and usable by vision impaired students. This paper presents a prototype booking system that fulfills these requirements. This booking system was designed with assistive technologies in mind, and implemented as a web-based application

 ISBN: 978-1-61208-133-5

II. THE LEARNING ENVIRONMENT

The CAVI accessible e-learning environment was established to deliver advanced IT networking courses to vision impaired students situated in different parts of the world. The presentation of Cisco curricular were also redesigned to ensure the teaching and learning materials were accessible, however the environment structure can be applied to any learning environment that has accessible educational materials. The curriculum presented is comprised of the Discovery, Exploration (previously CCNA courses) and IT Essentials that form part of the Cisco Networking Academy Program. "Networking Academy utilizes a blended learning model that combines face-to-face teaching with engaging online content and hands-on learning activities to help students prepare for industry-standard certifications, entry-level and advanced careers, and higher education in engineering, computer science, information systems, and related fields."[4]. Details on the modifications to the various curricular may be found in [5].

The CAVI e-learning environment is illustrated in Figure 1 and comprises:

1. a local classroom where local vision impaired students can attend and where lectures and tutorials are broadcast for the virtual classroom,
2. support academies situated at other remote locations providing local vision impaired tutors,
3. direct access via the Internet for remote learners at home,
4. a curriculum server housing the Cisco e-learning materials and a file and applications server housing the course management system plus additional accessible teaching materials and applications,
5. a local laboratory where local students can carry out their laboratory exercises,
6. a virtual classroom consisting of a voice server, webserver, and podcast server,
7. a remote laboratory where students can configure the network equipment and test their configurations.

Vision impaired instructors deliver the lectures and tutorials in the local classroom and these are broadcast to the support academies and remote students. Vision impaired tutors provide a one-on-one capacity with students to further explain and walk through concepts and exercises. The lectures and tutorials are recorded and the audio files are stored for later access by the students on demand. The virtual classroom enables the students to interact with the instructors and other students during the broadcasts. The virtual classroom also provides the opportunity for lectures and tutorials to be delivered by lecturers in remote locations, as the physical location of the teaching staff is flexible. Assessing assignments can also take place electronically. Students can complete and submit their work electronically, and lecturers in any location can grade the students' work. One-on-one communication between instructors, tutors and remote students is carried out via Skype.

The above components allow the vision impaired students to login and hear broadcast lectures and tutorials, and then work through the accessible curriculum at their own pace. The curriculum and applications enable the students to design network architectures, then implement and test these in the remote laboratory which is a real networked environment.

The Cisco courses require students to design and build IT networks, and trouble-shoot these networks. This means the vision impaired students must have access to network equipment and/or simulation software so that they may configure and test configurations, hence the need for an accessible remote laboratory facility and simulation software.

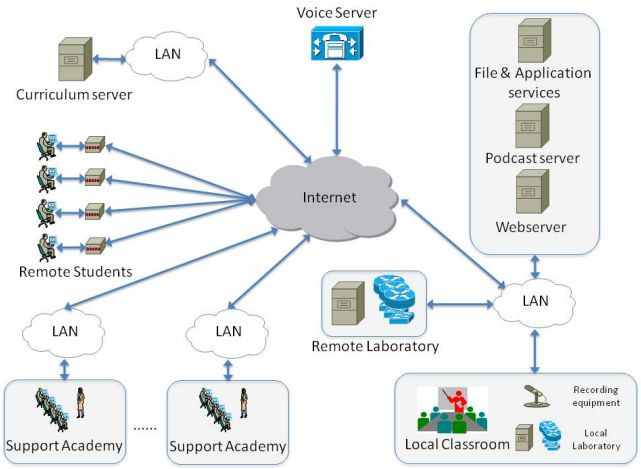


Figure 1. CAVI Teaching environment

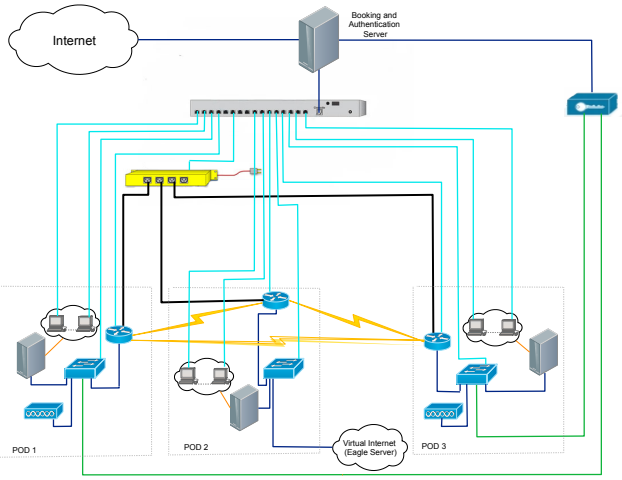


Figure 2. Remote Laboratory "bundles" utilised in the CAVI program.

III. AN ACCESSIBLE INTERFACE TO PACKET TRACER

iNetSim [3] was an early prototype application that allowed vision impaired students to run basic network configuration simulations. Whist this prototype was successful in its aim of illustrating that network simulators may be made fully accessible if accessibility is built into the design stage of application development, maintaining and building a separate simulator for use by the vision impaired is not feasible. Any such application would lag development

 ISBN: 978-1-61208-133-5

and features of commercial, well resourced projects such as Packet Tracer. It was therefore decided to develop an external application to connect to Packet Tracer utilizing the newly released APIs and multi user features. Initial development has aimed at examining the flexibility and ability of the framework underlying the Packet Tracer application development environment. The information obtained was then used to decide on the feasibility of an external application for vision impaired people, allowing them to use and manipulate the Packet Tracer software package to carry on networking simulation, particularly where the CCNA curriculum is concerned.

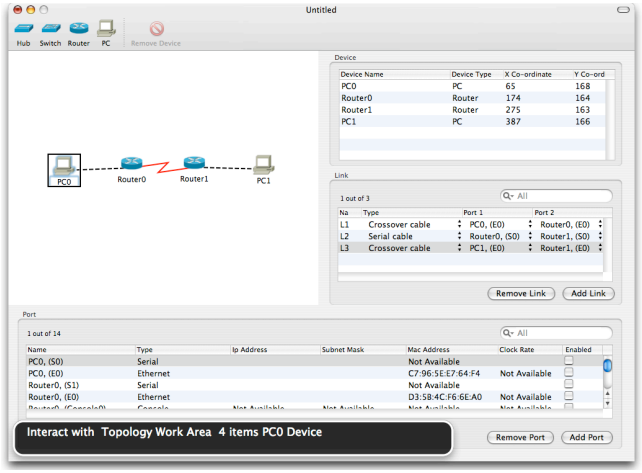
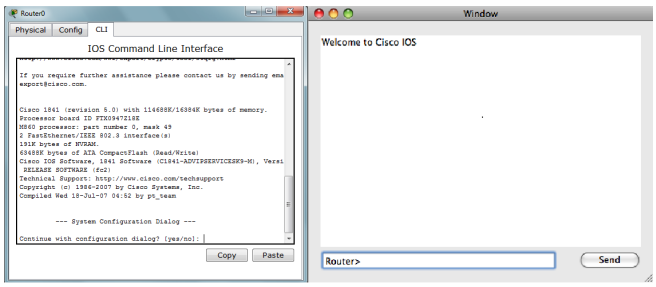


Figure 3. Example iNetSim Scenario

As may be seen when comparing the iNetSim interface (Figure 3) and that of the new prototype, some modification to the layout have been made in particular the reduction in the number of tables used in the connection and devices configuration areas. This is intended to increase the navigation speed when using a screen reader.



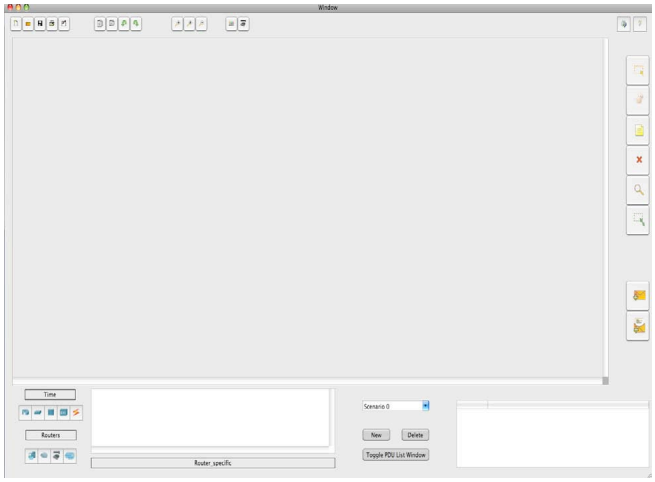


Figure 4. Example screen captures of the external application communication with a running instance of Packet Tracer

IV. REMOTE LABORATORY BOOKING SYSTEM

A. Available Booking Systems

As stated earlier, users of the remote laboratory face the problem of not knowing ahead of time whether the equipment is going to be occupied or not and therefore an accessible booking system is required.

One possible booking system is contained within the Netlab appliance by Network Development Group (NDG). While it is tailored towards remote laboratory setups such as the one within CAVI, there are areas where this particular system falls short of CAVI's needs. Firstly, NDG prices the initial purchase of its Netlab Academy Edition appliance at $6795 USD [6]. Armstrong and Murray [7] also note how the cost of the appliance prevented its use within the CAVI remote laboratory, but perhaps more importantly, they describe how the Java based applications in Netlab, including the booking system, are not accessible by screen reader software. Figure 5 below shows a screenshot of Netlab's booking system interface.

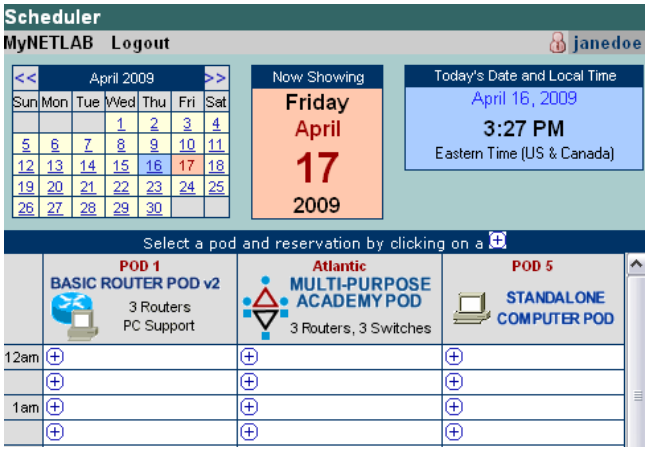


Figure 5. Netlab scheduler interface (NDG 2010)

Even if a screen reader was able to interpret this Netlab Scheduler interface, it would still be of low usability to the vision impaired user. The screen reader would attempt to read every number within the monthly calendar aloud. The links within the monthly calendar are not meaningful on their own, and it is likely that a screen reader would ignore the colours associated with the selected date and the current date. Finally, while the interface may look relatively simple to a sighted user, this may not be the case for their vision impaired counterparts.

The Meeting Room Booking System (MRBS) and WebCalendar are two open source booking systems, and are geared towards more general purpose applications. They are free to use, but they also suffer from accessibility problems, similar to the ones that affected Netlab. Both systems are built using PHP, meaning that these systems generate HTML pages that are translatable by a screen reader. Like the Netlab scheduler, these open source systems also display monthly calendars which add unnecessary verbosity, and use links that cannot stand alone. Figure 6 below shows a screenshot of MRBS.

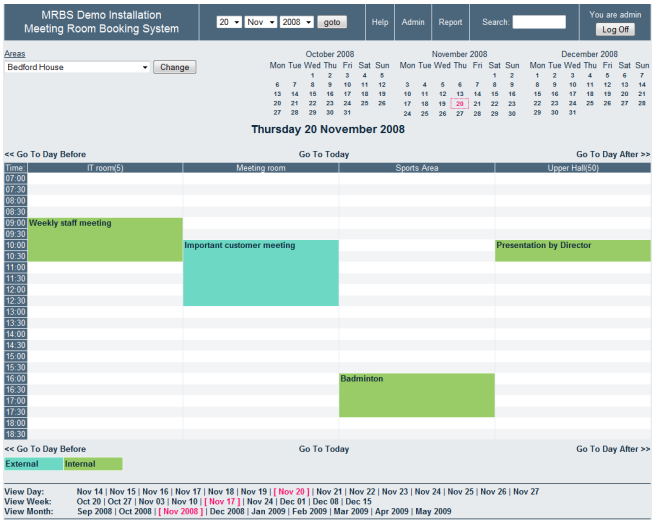


Figure 6. Screenshot of MRBS (MRBS Development Team 2010)

It can be seen that MRBS employs colours in an entirely inaccessible fashion. From the home page of this system, a user of a screen reader would not be able to easily distinguish when a particular reservation starts or finishes, since this information is solely conveyed through the use of colour. The times listed to the left of the daily schedule, and the listings to the bottom of the home page would also be read aloud by the screen reader, which further reduces the simplicity and accessibility of this system.

B. *Implementation*

All the interfaces of the booking system (see Figure 7) use no more than these three colours: black, white, and red. The use of red is exclusively for the indication of error conditions, such as a missing username or missing password for the user login page. Users who are completely blind are still notified of any errors since their screen reader should be able to translate the textual error messages.

The titles of each page are designed to fully identify the purpose of each interface, and unambiguously distinguish these pages from any other websites the user may be browsing simultaneously. If the login interface here was simply entitled “Login”, then a vision impaired user may not be able to tell if this was a login page for the CAVI booking system, or, for example, a login page for their online discussion board. The link to the administration interface does not appear to users who do not have the necessary privileges. This link is close to the top of the page in order to enable quick identification by vision impaired instructors, and hence save them from having to listen to the rest of the page in the instance where they only want to manage the booking system.

The two message lines that subsequently follow are placeholders for any notifications or error messages that are to be output to the user. When a user first logs in, these message lines are initially empty. In a manner similar to the user login interface, any error messages here are also printed in red. Furthermore, their location near the top of the page also allows a vision impaired user to speedily ascertain the results of their actions.

The values of the user bookings table are populated dynamically from the database. If a user has no bookings for the future, then the table does not appear. Information regarding a user future bookings and the selected day’s open timeslots was not arbitrarily placed above the drop boxes for a new booking. As a screen reader goes through the page, it will read out the open timeslots and the future bookings first, which will in turn give the vision impaired user an idea of when to (or when not to) book before they encounter and use the drop boxes for their next booking. This can increase the interface’s usability. If the arrangement was reversed, then a user could hastily select drop box options to start with, only to change their choices soon after when they later discover through their screen reader that their timeslot was actually not available.

 ISBN: 978-1-61208-133-5

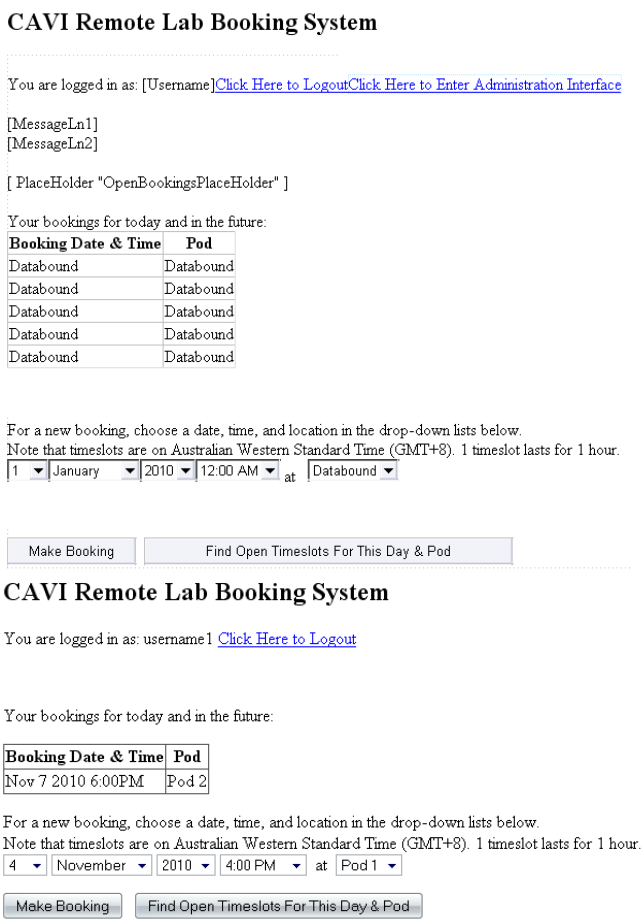


Figure 7. The booking interface

V. CONCLUSIONS

A. Packet Tracer Accessibility

Although the developments described are in their early stages of implementation several significant milestones have been achieved. A functional and fully accessible external application to communicate with Packet Tracer has been implemented. At this point only basic functionality has been included it is only a matter of coding additional calls to the Packet Tracer APIs to further increase usability.

Ideally, resources would be allocated to the developers of Packet Tracer to modify the user interface so that there is no longer the need for an external application. This project has shown that network simulators may be designed to be accessible to people with vision impairment without reducing the appeal and usability for sighted students.

B. Remote Laboratory Booking System

The accessible booking system has met its design specifications but still requires further development and refining. The achievements so far include:

- The design of an equipment booking system for the CAVI remote laboratory. This design encompassed aspects of its user interfaces, system architecture, and user processing.
- The development and implementation of the booking system as an accessible application that can be hosted on a web server.
- Research into the methods of integrating the booking system into the CAVI remote laboratory environment.
- The design, development, and implementation of this prototype booking system can serve to prove the feasibility of a complete laboratory front-end system. The work also demonstrates that such a system can be accessible to the vision impaired, and can be realised at an economical cost.

The developed booking system only constitutes a core component of the CAVI remote laboratory's front-end system. In order to refine and improve the current system, several courses of action can be pursued in the future.

Firstly, the booking system should be ported to a free and open-source implementation. Although the development tools employed during the project are available without monetary cost, the current implementation does bias itself towards a Windows-based web server. This is due to the use of ASP.NET and C#, which are both Microsoft technologies [8][9]. This can be attained by converting the interface pages' markup and programming logic into PHP. MySQL can also be used to host the booking system's database in the future. Alternatively, a Linux-based web server can execute ASP.NET pages and C# code using the Mono platform and the mod_mono Apache module [10].

The user authorisation mechanisms should also be incorporated into the booking system, but a method of relaying generated passwords to the booking system server will have to be devised. User accounting functions may be of use. To be more precise, student users may be tracked to see whether they actually use the laboratory equipment at their chosen booking time.

If the management console switch was to be retained when integrating the booking system server, then a customised Telnet client could be written and embedded into one of the system's web interfaces. This can free the user from having to Telnet into the booking system server, and could in turn obviate the need to run a Telnet daemon on the server.

ACKNOWLEDGMENT

The authors would like to acknowledge the support of the Association for the Blind (WA) and Fundi Software whose financial and in kind support make the Cisco Academy for the Vision Impaired possible.

REFERENCES

[1] Murray, I and Armstrong H, Remote Laboratory Access for Students with Vision Impairment, The International Journal on Advances in Life Sciences, Vol 1, no 2-3, 2009 pp77-89

[2] Vision Australia, 2007. Employment Report 2007. Vision Australia, Available http://www.visionaustralia.org/info.aspx?page=1651, retrieved 5 May 2011

[3] J. Hope, B. vonKonsky, I. Murray, L. C. Chew and B. Farrugia, A Cisco Education Tool Accessible to the Vision Impaired, ASSETS06, Portland, Oregon USA, October 23-25, 2006, pp235-236

[4] Cisco. 2010. Cisco Systems, available http://www.cisco.com/web/learning/netacad/get_involved/BecomeAnAcademy.html, retrieved 2nd July 2010

[5] Armstrong, H. and Murray, I., Remote and Local Delivery of Cisco Education for the vision Impaired, Proceedings of ITICSE 2007, June 23–27, Dundee, Scotland, United Kingdom.

[6] NDG. 2010. *NDG NETLAB+ Remote Lab Solutions Pricing Summary*. http://www.netdevgroup.com/ordering/pricing.html retrieved 5 May 2011.

[7] Armstrong, H., and I. Murray. Spanning the Digital Divide: A Remote IT Learning Environment for the Vision Impaired. Freiburg, Germany. IADIS Multi Conference on Computer Science and Information Systems, July 26-31, 2010:

[8] Hanselman, S., What is ASP.NET? Streaming video. What is ASP.NET?: The Official Microsoft ASP.NET Site. http://www.asp.net/general/videos/what-is-asp-net retrieved 5 May 2011

[9] Burnt Jet. 2010. *Burnt Jet - Glossary of Terms for Programming*. http://burntjet.co.uk/programs/help/glossary.php, retrieved 5 May 2011

[10] Mono. 2010. *Main Page – Mono*. http://www.mono-project.com/Main_Page, retrieved 5 May 2011

 ISBN: 978-1-61208-133-5

An Evaluation of Blended Learning Components of the Cisco Network Academy Using a Rasch Model

Kevin Sealey *(Author)*
Curtin University
Perth, Australia
kevin.sealey@ozemail.com.au

***Abstract*—The blended model of e-learning espoused by the Network Academy curriculum involves the interaction of students with its various components - online content, laboratory exercises, simulations, online assessment, texts and most importantly instructors. These components were evaluated from the student viewpoint using two surveys and data from the final online examination. The analysis used a Rasch probabilistic model. This was done in the West Australian, and in the case of the online examination, an Australia-wide context.**

Keywords-curriculum evaluation; blended learning; Rasch model, student

I INTRODUCTION

This is part of a wider study to evaluate the Network Academy curriculum in secondary and post-secondary academies in Western Australia, looking at the intended, implemented, achieved and perceived curriculum [1][2] from the point of view of various stakeholders - students, instructors, employers and Cisco.

Considerable research has been carried out into the value of the Network Academy in various educational settings [3][4]. In this study students were asked to respond to two online surveys [5], one at the beginning of their course and another later in their course. For the first survey, the 175 respondents comprised 12 secondary students (7%), 109 Technical and Further Education (TAFE) students (62%) and 54 university students (31%). 135 of these students (76%) studied full time and 42 part time (24%). 20% of these students (36) were in full time employment, 38% in part time employment (66) and 42% not employed (73). In the second survey, the 93 respondents were comprised of 13 secondary (14%), 56 TAFE (60%) and 24 university (26%) students. The responses to the final online examination for the Exploration curriculum for the first semester, 2010, were also analysed. This involved 750 students drawn from academies across Australia.

Analyses of surveys and examination were performed using a Rasch probabilistic model, which yields precise, interval-level measures of both person (ability) and item (difficulty). This model was used to investigate if it could give additional insight beyond that derived from standard analysis.

II WHY DO STUDENTS STUDY THE CURRICULUM? THE INTENDED CURRICULUM

It was anticipated that students would take the Network Academy course for a wide variety of reasons, chiefly centred on their current and future study goals and their employment considerations

In order to gain further insight into the survey responses, the data was analysed using RUMM2030[6]. The summary of the analysis indicates that the data fits the Rasch model [7]. Both the reliability indices and the power of test-of-fit indicate that the data was suitable for analysis using this approach.

TABLE 1. FIRST SURVEY RESULTS

Survey Item	Description	Location (Difficulty to Endorse)	Fit Residual (Rasch Model Fit)	Standard Error of Estimate
7	Graduation	-0.28	0.24	0.087
8	Further Education	0.008	-0.204	0.086
9	Job	-0.662	0.24	0.098
10	Current Job	0.55	2.211	0.074
11	Interest	0.714	1.057	0.081
12	Peers	1.529	2.294	0.082
13	Practical	-0.798	-0.206	0.107
14	Theory	-0.852	-0.913	0.112
15	Certification	-0.209	0.389	0.082

 ISBN: 978-1-61208-133-5

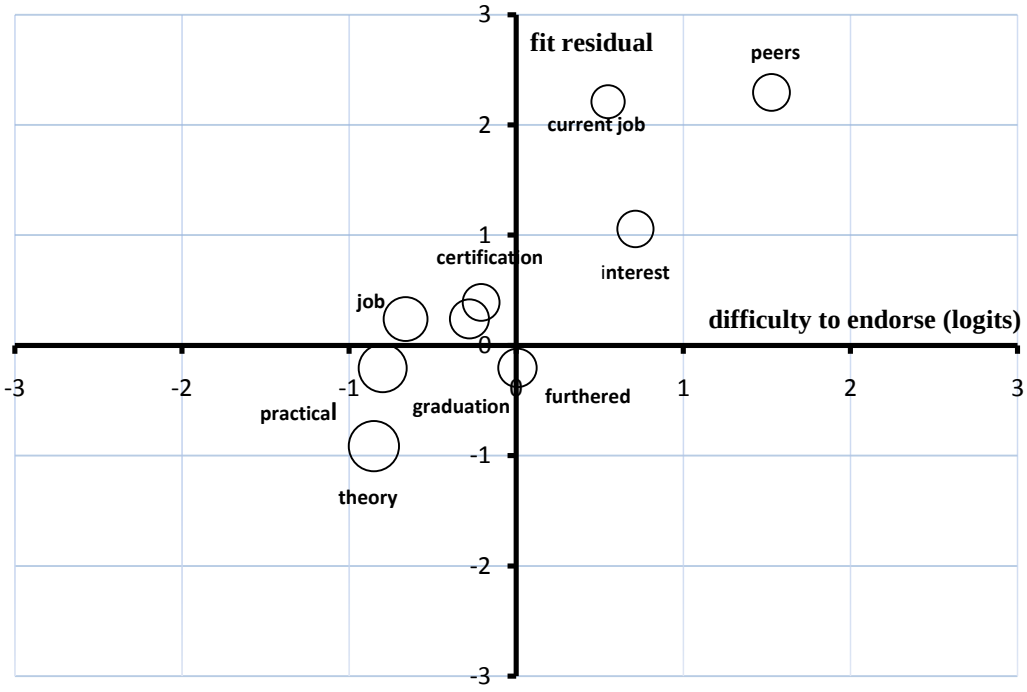


Figure 1. Intended curriculum

Table 1 shows the item details for the first survey. This displays the difficulty to endorse (Location), the test of fit (Fit Residual) and the standard error of the measurement of the Location. The unit of measurement for Location is logit (log odds). In this context, the more negative the Location for an item the easier it is for the respondent to endorse the body of the item. That is, the more negative the item's Location the stronger is the agreement of the respondent to the item's statement. The Fit Residual, as listed here, measures the correspondence of the data with the model. It represents the standardized difference between the actual data and the values calculated using the item estimates and the mathematical model. It is considered acceptable fit if the Fit Residual lies between +2 and -2. Values between 2 and 2.5 and between -2 and -2.5 are considered marginal. Values above 2.5 and below -2.5 are considered to represent poor fit to the model. The standard error estimates are meant to indicate the degree of uncertainty in the Location values for a particular item. It is considered a bonus for this method of analysis that each item has its own estimate of error. With classical test theory the standard error is calculated to be the same for all items.

It can be seen from Table 1 that the items vary significantly in difficulty to endorse, and that most items fit the model, with the exception of items 4 and 6, which have only marginal fit. If the data is presented as a bubble plot (Fig. 1), the differences between the items become more evident.

In this plot difficulty to endorse is plotted horizontally and fit residual vertically. The width of the bubbles represents the standard error (each bubble's width is approximately shown as double the value of the standard error).

It is evident that the strongest intentions of the students were to gain theoretical and practical knowledge and to enhance their employment prospects. To a lesser extent, the opportunity to progress towards graduation and certification and also to prepare for further education were seen to be important. Pursuing the course as a matter of interest was not well endorsed. The fact that three quarters of the cohort were not employed means that the responses to the item asking whether they thought that the course would help in their current employment was less reliable. This item did not have a good fit to the model. The influence of peers on their decision to take the course was not endorsed, and the relevant item was not a good fit to the model. Perhaps this reflects the large age range and variability of maturity and experience of the cohort. Table 2 shows the results for the analysis of the second survey. These results are used to investigate the implemented and perceived curriculum.

III HOW DO STUDENTS STUDY THE IMPLEMENTED CURRICULUM?

Subjecting the data from the second student survey to analysis using RUMM2030 indicated that the data had an excellent fit to the model. Once again, graphical representation of the data indicates significant differences in endorsement of the items by students. Although the item data is shown in the one table, different groups of items represented particular features of the analysis.

 ISBN: 978-1-61208-133-5

TABLE 2. SECOND SURVEY RESULTS

Item	Description	Location (Difficulty to endorse)	Fit Residual	Standard error	Item	Description	Location (Difficulty to endorse)	Fit Residual	Standard error
5	Online	-1.513	0.057	0.137	37	Explain Solutions	-1.141	0.399	0.146
6	Texts	0.264	0.194	0.114	38	Teamwork	-1.933	1.15	0.127
7	Lecture	-0.19	-0.345	0.125	39	Work/Others	-1.479	3.446	0.125
8	Labs	-1.594	-0.783	0.13	40	Learn/Others	-1.791	2.655	0.127
9	Packet Tracer	-0.27	5	0.121	41	Self Paced	-2.235	1.156	0.134
10	Tests	-0.336	-0.022	0.124	42	Relate/Contexts	0.977	0.092	0.112
11	Cases	0.509	0.087	0.109	43	Interest	0.692	-0.199	0.126
12	Others	0.318	0.089	0.121	44	Ability Match	0.292	-0.308	0.135
13	Explanation	-0.298	-0.156	0.125	45	Control/Learning	0.206	-0.212	0.124
14	Icg	0.72	0.128	0.114	46	Use Feedback	0.036	-0.064	0.127
15	Other Networks	0.567	0.082	0.108	47	Understanding	0.071	-0.222	0.127
16	Online	-0.042	0.001	0.165	48	Relevant	-0.351	-0.116	0.134
17	Packet Tracer	-0.837	0.234	0.199	49	Important	-0.029	-0.252	0.122
18	Tests	-0.309	-0.002	0.175	50	Weaknesses	0.005	0.044	0.119
19	Discussions	-0.614	-0.214	0.19	51	Explain	0.198	-0.152	0.104
20	Labs	-0.776	-0.285	0.199	52	Pace	0.408	-0.311	0.122
21	Texts	-0.001	-0.058	0.188	53	Examples	0.488	-0.37	0.124
22	Lectures	-0.447	-0.495	0.179	54	Involvement	0.524	-0.221	0.134
23	Quest/Contrib	0.298	0.397	0.144	55	Texts	0.88	0.213	0.116
24	Presentation	1.863	0.126	0.135	56	Effective	0.09	-0.395	0.107
25	Presentation	1.815	0.142	0.13	57	Prepared	-0.581	-0.35	0.141
26	Unprepared	1.198	0.741	0.132	58	Clear Course	0.524	-0.161	0.125
27	Worked/ Others	0.181	0.614	0.132	59	Good Answers	-0.602	-0.73	0.131
28	Mentor	1.543	0.155	0.149	60	Difficulty	0.281	-0.213	0.129
29	Extra-Curricular	1.88	0.134	0.141	61	Activities	-1.395	-1.619	0.147
30	Challenged	-0.99	1.534	0.225	62	Cases	0.413	-0.106	0.145
31	Discuss Ideas	0.872	-0.117	0.126	63	Working/ Others	0.187	0.104	0.124
32	Give Opinions	1.104	0.054	0.124	64	Well Matched	-0.025	-0.118	0.147
33	Inst Questions	0.823	0.164	0.125	65	Understanding	-0.277	-0.12	0.143
34	Ideas Used	1.894	-0.003	0.124	66	Available	-1.682	0.428	0.137
35	Quest/Contrib	0.246	0.276	0.118	67	Attitude	0.369	-0.172	0.111
36	Problem Solving	-0.999	0.664	0.166					

 ISBN: 978-1-61208-133-5

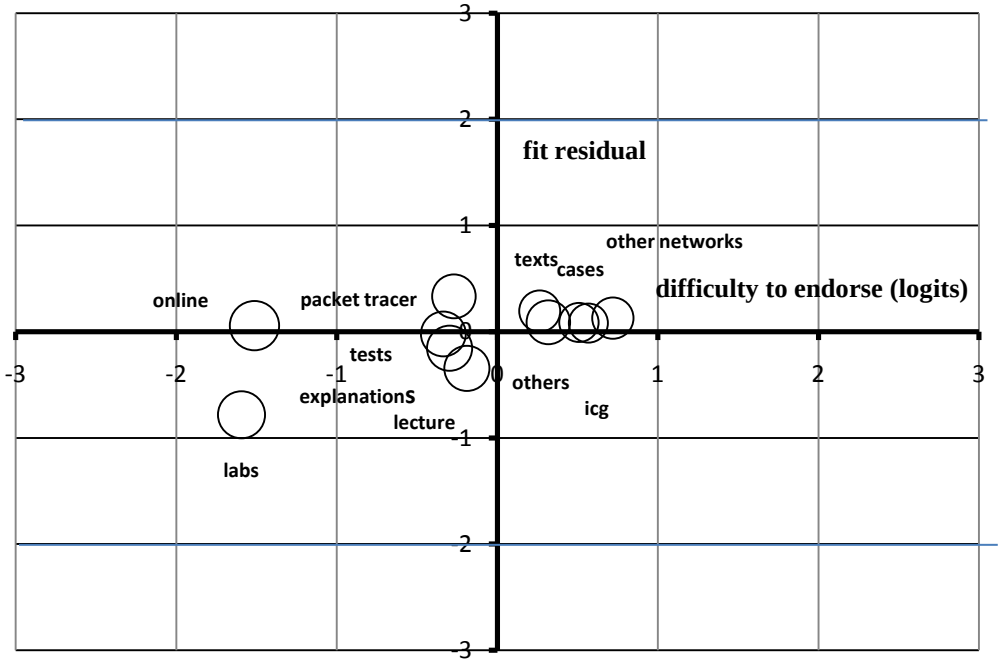


Figure 2. Implemented Curriculum

The items relating to student implementation of the curriculum (items 5 to 15) were concerned with the strategies found most effective by the students to proceed through the curriculum. The results of this analysis are shown in Fig 2. The fit residuals and standard errors indicate that all items have a good fit and that there are broadly three groups of items when comparing aspects relating to students' difficulty in endorsement. Clearly the online curriculum and the laboratory exercises were the most important resources for the students. The second most important group of resources included quizzes and chapter tests, simulations (Packet Tracer), explanations by the instructor and the lecture presentations by the instructor. The least important resources included texts, working with other students, case studies, working with other networks outside the laboratory and interactive course guides. This rating of importance of resources is confirmed by class observation. It is interesting to note that the online assessment resources are rated as highly as the interaction with the instructor. Interactive Course Guides were not commonly used by students as they were a recent innovation and targeted chiefly at the instructors.

Another group of items (items 23 to 37) were related to the students' engagement in the class. This could be considered to be another aspect of how the students implemented the curriculum. The most strongly endorsed items (most negative Location) relate to the degree of academic challenge of the curriculum. Students were challenged by the course and frequently engaged in problem solving and explained solutions to their classmates. To a lesser extent students worked with others, asked questions and made contributions to the class. It is notable that students considered that they always prepared well for the classes. The other group of items could indicate that students rarely made unsolicited contributions to the class and did not work with other members of the class outside of scheduled class times.

A third group of items (items 38 to 42) indicated the learning style adopted by the students. These items suggest that students embraced the self-paced aspect of the curriculum and that when working in groups they felt that teamwork was important.. However, the material learnt in class was not well related to the non-class situation. Items relating to working with and learning from others were poorly fitting items and thus no conclusions may be drawn from their values.

IV. How Do Students Feel About the Curriculum? The Perceived Curriculum

When asked to rate the effectiveness of the components of the curriculum (items 16 to 22), the students' responses indicated that they perceived all of the components as effective. Once again the fit residuals indicate a good fit to the model. The fact that all items have a negative Location shows that all components were strongly endorsed as being effective. The spread of the Location of the items is smaller than for those items discussed previously. It may be inferred that the laboratory exercises and the simulations with Packet Tracer and classroom discussions were perceived as the most effective components of the course. The lectures, online assessment and online curriculum were in turn more effective than the texts.

On the other hand, some aspects of the curriculum were not endorsed as strongly. These items (items 43 to 45)

 ISBN: 978-1-61208-133-5

appear to indicate that the students' perception of the curriculum materials was that they did not match their ability and interests, and that very little was left to the students' initiative as far as progression through the course was concerned. In effect, the Cisco course materials and the order in which they are presented is very much controlled by the instructors and course developers.

As part of the online aspect of the curriculum, the online assessment materials form an integral part of the package. The responses to items 46 to 50 indicated that the assessments are viewed as well matched to the curriculum, and an important resource in the overall package. The feedback delivered with the assessment results and the use of assessment to discover weaknesses and generally help with the understanding of course material are not always perceived as helpful.

Instructors are an integral part of this blended e-learning initiative. The role and effectiveness of instructors, as perceived by the students, is an important aspect of the curriculum evaluation. The items relating to students' perception of the instructors (items 51 to 62) fit the model well and there is some significant variation in difficulty to endorse the statement in the item. The class activities chosen by the instructor were strongly endorsed. The preparation of the instructor and his/her response to in-class questions were well regarded. The pace at which the course was conducted, the student involvement in the class, the number and quality of the examples chosen by the instructor, the teaching effectiveness, the clarity of the scope of the course, the difficulty level of the course materials and the case studies used by the instructor were not endorsed. The printed material of the course was least strongly endorsed.

The final group of questions (items 63 to 67) relate to the overall perception of the curriculum. The most strongly endorsed perception was that the instructor was freely available for consultation. The course content was well matched to the abilities and expectations of the students and led to increase understanding of course content. The value of working with other students was less strongly endorsed. The overall attitude to coming to class probably reflects the challenging nature of the curriculum materials.

V. WHAT DO STUDENTS GET OUT OF THE CURRICULUM? THE ACHIEVED CURRICULUM

An analysis of the response from all of the students in Australia for the final Exploration examination (Semester one, 2010) were analysed using RUMM2030. All except two of the 120 items in the item bank returned an acceptable level of fit to the model. An analysis such as this might lead to a modification or removal of the poorly fitting items, which in the current situation merely add 'noise' to the assessment of student ability.

The person fit as estimated in the first analysis indicated that students' responses conformed well to the model. Those students with perfect scores (ten students) were classified as having a poor fit to the model, since the ability estimate could not be made. Only one other student in the group of 750 had a poor fit to the model.

The ability (Location) frequency graph (Fig 3) shows abrupt changes at 0, 0.4, 1.1 and 1.5. If these are used to signify boundaries of different ability groups, then those students above 1.5 (raw score 95) could be awarded an A, those between 1.1 (raw score 90) and 1.5 a B, those between .4 (raw score 70) and 1.1 a C, those between 0 (raw score 50) and .4 a D and those below 0 an E. Examination of the frequency distributions for each of the module tests suggests a very similar distribution of abilities and this suggests that this could be a feasible approach to awarding grades.

When managing the assessment for the class, the instructor can stipulate the form with which the students are confronted when they access the test. The data analysed for the Australian cohort consists of responses to three forms, each comprising a subset of the 120 items in the item pool. Rasch analysis of the data provides a means of comparing the ability estimates made with the different forms. The 'equating tests' option of RUMM2030 provides a graphical indication of the relative difficulty of different tests which are made up of different selections of items from a pool of calibrated items. The resulting graph for the three forms of the final module examination is displayed in Figure 4 and this shows that the ability measure (Location) over the whole range of scores is essentially independent of the form used. Students are not disadvantaged by being assigned one form rather than another.

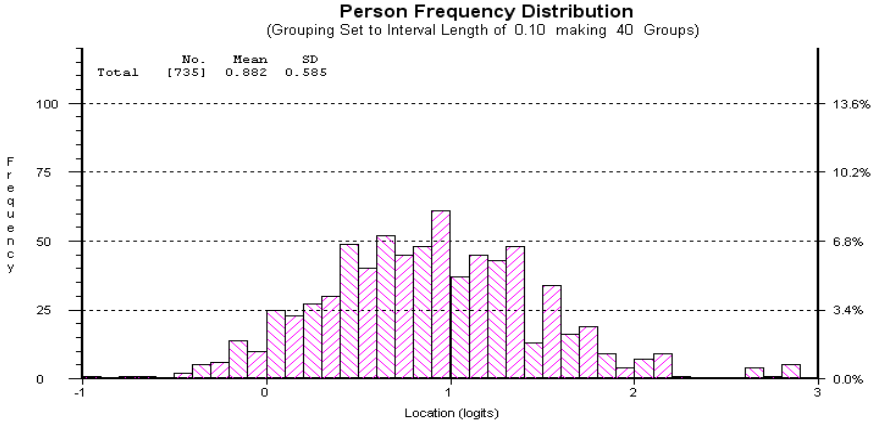


Figure 3. Ability Frequency Distribution

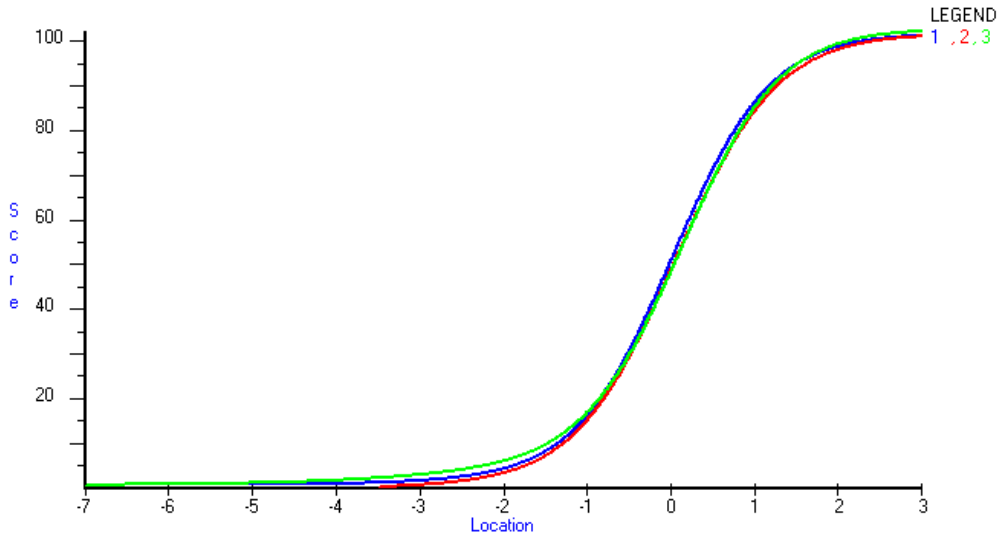


Figure 4. Comparison of Different Forms

VI. CONCLUSION

The study indicates that Rasch analysis is appropriate for an in-depth evaluation of the components comprising the blended model of e-learning used in the Network Academy curriculum. Such an evaluation may be useful in refining the various components, guiding the relative importance given to particular components by the different stakeholders and modifying student recruitment strategies of educational institutions.

The chief expectations for students were that the curriculum would enhance their practical skills and theoretical knowledge as well as improve their prospects for employment. The online content and laboratory exercises, online assessment and instructor centred activities provided the means by which the students progressed through the program. The importance of the instructor's role is evident in the students' perception of this style of e-learning. Analysis of the online examinations shows that they are a fair and reliable tool for assessing student ability. It is clear, however, that a Rasch analysis of the data adds value to the procedure.

REFERENCES

[1] Hartley, M. S., Treagust, D. F., & Ogunniyi, M. B. (2008). The application of a CAL strategy in science and mathematics for disadvantaged Grade 12 learners in South Africa. International Journal of Educational Development, 28, 596-611.

[2] Van den Akker, J. (1998). The science curriculum: Between ideals and outcomes. In B.J. Fraser & K. G. Tobin (Eds.), International handbook of science education. (pp. 421-427). Dordrecht, The Netherlands: Kluwer Academic Press.

[3] Cakir, H., Bichelmeyer, B., Dennis, A., Bunnage, J. C., Duffy, T., Oncu, S., et al. (2006). Value of the CCNA program: perspectives on satisfaction and applicability from CCNA and comparison group students Cisco Networking Academy Evaluation Project. Bloomington, IN: Indiana University

[4] Randall, M. H., & Zirkle, C. J. (2005). Information technology student-based certification in formal education settings: who benefits and what is needed. Journal of Information Technology Education, 4, 287-306.

[5] The first student survey is at www.tinyurl.com/ccna-int-out and the second survey for students is at www.tinyurl.com/ccna-final-out

[6] Sheridan, B. (2009). Rasch Unidemensional Measurement Model (Version 2030). Duncraig, Western Australia: Rumm Laboratory Pty Ltd. Retrieved from www.rummlab.com

[7] Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model: fundamental measurement in the human sciences* (second ed.). London: Lawrence Erlbaum Associates, Publishers.

 ISBN: 978-1-61208-133-5

The Effectiveness of Blended Distance Learning

A Multi-Dimensional Analysis within ICT Learning

Ron Kovac
Ball State University
Muncie, USA
e-mail: rkovac@bsu.edu

Kristen E. DiCerbo
Independent Researcher
Phoenix, USA
e-mail: kdicerbo@cisco.com

***Abstract*—eLearning as an educational mode of delivery presents many new challenges to educators and students. The change in modality of the teaching/learning environment begs many questions in regard to effectiveness and efficiency of student learning (the main goal of our educational system). The focus of this research and paper is the effectiveness of Blended Distance Learning for training instructors. Blended Distance Learning, growing in popularity recently, has learners doing part of their studies in person, and part of their studies remotely (either synchronously, asynchronously, or a combination). Effectiveness of Blended Distance Learning will be measured in two dimensions: instructor outcomes and then subsequent outcomes of students. The study looks at both attitudinal and academic outcomes of students and instructors by instructor training modality. The results of this study hold implications for education in general, but specifically for those making decisions about learning methodologies and platforms to use when training teachers.**

Keywords-Blended distance learning; e-learning; teacher training

I. INTRODUCTION

With the current attention paid to eLearning, we become aware that this modality of education is growing and here to stay. Although eLearning is not a panacea for all of education problems, it does present another tool, with benefits and detriments for the student and the instructor, in terms of both attitudes and resulting learning (the ultimate goal of our educational system). With this new modality, a series of questions arise, centering on educational effectiveness (learning), the efficiency of this new modality, and the ways to use different instructional methodologies within the framework of eLearning.

The focus of this research and paper is on the effectiveness of Blended Distance Learning (BDL), a specific form of eLearning, for training teachers. BDL has learners doing part of their studies in person and part remotely (synchronously, asynchronously, or a combination). BDL has the attraction of not fully abstracting the course or the people involved, while also allowing the freedom and flexibility that remote learning offers. Many major institutions and entities have settled on this as a key learning methodology in their business plan. However, others have expressed concern that BDL classes can be as challenging and effective as in-person classes. While there has been a fairly large body of research on student BDL training, less is known about training teachers via BDL methodologies. This is the question this paper will explore.

Effectiveness of BDL will be measured in two dimensions. The first dimension will compare the final exam and satisfaction survey scores of instructors who took a course via BDL with instructors who took the course in a purely face-to-face traditional manner. Although the instructors were not randomly assigned to the two test groups, similarity was sought in age, gender, and background of the instructors.

The second dimension of this study will look at the students of these two instructor groups. Each of these instructors, after training in a specific ICT cognate area, went to their classrooms to instruct their students in the same ICT cognate area. Effectiveness was measured in their students by their final exam scores and their course satisfaction survey. The instructors and the students were provided the same final exam score and satisfaction survey.

The results of this study hold implications for education in general, but specifically within the BDL environment. The results of this study also hold merit for studying the effects of BDL instructor training on the instructors directly and their students that they teach. In the paper below we will first review relevant literature, then describe the analysis of instructor and student outcomes based on instructor training type, and finally end with a discussion of implications.

II. LITERATURE REVIEW

The National Center for Education Statistics [1] reports that the K-12 public school enrollment in distance learning classes in the U.S. grew 65% in the years from 2002 to 2005. A more recent study by Picciano and Seaman [2] finds that more than a million students were educated via DL methodologies in the academic year of 2007-2008. Caution should be used when looking at the phenomenal growth of DL first because DL courses, and especially BDL courses, and include everything from correspondence courses to course with minimal remote use [3], but it is clear that the practice of taking courses remotely is increasing.

Analysis and study in the field of eLearning has been going on since eLearnings' inception. The results have been varied but meta-studies conducted recently point to some consistencies and trends in evolution and results [3]. Because this DL and BDL field has existed long enough and

 ISBN: 978-1-61208-133-5

has high appeal for researchers and policy makers, enough studies of BDL have been done to warrant a meta-analysis of results. Meta-analysis is a technique for combining the results of many research studies to obtain a composite estimate of effect. It is essentially research on research, combining all the results of similar studies. Of over 90 studies reviewed by this meta-analysis only ½ provided sufficient statistical data and methodologies required to fit the rigor of this analysis. Most of these studies were conducted in higher education and/or specific cognate areas (Military, Training, ICT area). The studies ranged from 1994 – 2008.

The main finding of the study was that those in distance education classes had slightly better outcomes than those in face to face classes [3]. However, on further examination, it became clear that in fact the outcomes from face to face instruction and pure online instruction were approximately equal (all other elements being equal), but that BDL offers an advantage as seen by differences in exam scores across the studies in the analysis. The authors note that students in BDL classes often have both additional learning time and more instructional elements than those in face to face classes. Therefore, differences in outcomes may be due to these factors rather than any media or delivery method per se.

There has also been some work done on student satisfaction in BDL courses as applied to traditional face to face courses [4]. Besides the obvious benefits of DL and BDL, comfort and convenience [5], other items came up in research related to satisfaction in BDL courses. Five themes seem to emerge as constant and consistent; 1) classroom climate, 2) learning needs, 3) learner efficacy, 4) interaction and 5) appropriate format for the content [6]. At first blush one would not think of these are benefits of BDL course but the clientele in the study. The implications are many but seem to center around the recognition of different formats require different methodologies, Responsiveness to the variety of learning styles in the "classroom", empowering students and quality of material and format used in distance learning.

In regard to the degree of blendedness (mixture of face to face and remote), Voos [7] suggested it is unlikely that the proportion makes the difference in the course but that reconsideration of course design, new instructional media choices, and learning strengths and weaknesses make the difference. As Privateer [8] states so eloquently, "Opportunities for real change lie in creating new types of professors, new uses of instructional technology and new kinds of institutions whose continual intellectual self-capitalization continually assures their sites as learning organizations" (p. 72). Interestingly most of these studies have focused on the effects of BDL on students [9]. This study examines the effect of BDL learning during instructor training.

III. STUDY 1

The first study looks at the effects of instructor training method on instructor outcomes.

A. Participants

The context of this study is the Cisco Networking Academies. This study analyzed existing data from two groups of instructor trainees in the Academy: one trained in a BDL class and one trained via in-person classes. Instructors in the Network Academies are required to complete training courses in each of the classes they are going to teach. These courses have traditionally been five day, eight hour per day, in-person classes. However, in more cases, these classes have been distributed over time and place. Instructor trainees who completed instructor training in one of the four Exploration courses in the 2009 calendar year were included in the study. The BDL sample was also limited to trainees with completed course feedback forms and final exam scores who were trained in a class with more than one student were included in this study. The determination of whether a class was offered in the BDL format was made based on the instructor trainers' indication of method of offering classes indicated in the online class management system.

There were 364 instructors trained via BDL in the year. There were 10,412 instructors trained via in-person classes. In an attempt to get a better-matched sample, in-person trainees taught by the same instructors who taught BDL students were selected. We then randomly selected participants to get a similar sized sample with equivalent geographic and education level characteristics, resulting in a sample of 400 instructors trained via In-person classes. The groups are distributed as shown in Table 1

TABLE 1. DISTRIBUTION OF SAMPLE BY CLASS MODE

	BDL		In-Person	
	n	%	n	%
Network Fundamentals	135	37.5	150	37.5
Routing Protocols and Concepts	95	26.4	105	26.25
LAN Switching and Wireless	68	18.9	75	18.75
Accessing the WAN	62	17.2	70	17.5
Total	364	100.0	400	100.0

 ISBN: 978-1-61208-133-5

The participants were distributed by geographic theater as follows: Asia Pacific: 3.0%, Western Europe: 20.9%, Emerging Markets (Latin America, Middle East, Africa): 49.7%, and United States and Canada: 26.4%. There were relatively few participants from Asia as the BDL approach has been less adopted there in the Networking Academy.

B. Measures

Four measures were used to assess outcomes for both the instructor trainees: Satisfaction, Confidence, and Instructor rating subscales from the Course Feedback form and Final Exam scores. The Satisfaction, Confidence, and Instructor rating scores are each the means of a set of questions on the Course Feedback Form that the instructor trainees complete after a class. The Satisfaction scale asks students to rate their overall satisfaction with items such as labs, assessments, and course materials. Ratings are made on a five point scale (1 = Very Dissatisfied; 5 = Very Satisfied). The Confidence scale asks students to rate their confidence in performing various networking-related tasks taught in the course. Ratings are again completed on a 5 point scale (1 = Not at all confident; 5 = Very confident). The Instructor scale asks students to rate their instructor on things such as preparedness and approachability. These are rated on a 5 point agreement scale (1 = Strongly Disagree; 5 = Strongly Agree). The final exam is taken by each student at the end of every class. It is a 50 question multiple choice exam. Requirements from the Networking Academy require that the exam be proctored, whether the class is BDL or in-person.

C. Results

There are significant mean differences between the BDL and In-person groups for all four measures, as seen in Table 2. The Confidence, Satisfaction, and Instructor subscales have significantly higher means for the BDL group, while the Final Exam scores are significantly higher for the In-person group. However, it should be noted that the effect sizes are extremely small. This means that although there were statistically significant differences between the two groups, for most practical purposes, their ratings were very similar.

TABLE 2. INSTRUCTOR RESULTS BY CLASS MODE

	BDL	Mean	Effect Size
Final Exam*	BDL	90.63	-0.15
	In-Person	92.84	
Confidence*	BDL	4.19	0.14
	In-Person	3.98	
Instructor*	BDL	4.67	0.07
	In-Person	4.58	
Satisfaction*	BDL	4.33	0.10
	In-Person	4.21	

Analysis by Curriculum

Comparisons were also conducted by curriculum subgroups: Network Fundamentals, Routing, Switching, and WAN. Tables 3 and 4 show these results. Effect sizes are provided for statistically significant differences

In the Network Fundamentals subgroup, the only significant mean difference between the BDL and In-person groups is for the Final Exam. The mean Final exam score for the In-person group is significantly higher than the BDL group with a small effect size. Again, this means that the difference between the groups is small, however, it may be that taking the first class in the Networking Fundamentals curriculum is slightly more difficult with a remote component.

In the Routing subgroup, the means for the BDL group are significantly higher than the means for the In-person group for the Confidence and the Instructor Subscales. In the Switching subgroup, the mean scores for the BDL group were significantly higher than the In-person group for the Confidence and the Instructor Subscales. In the WAN subgroup, the mean scores for the BDL group were significantly higher than the mean scores for the In-person group on the Confidence subscale.

TABLE 3. DIFFERENCES BY CLASS-NF AND ROUTING

		Network Fundamentals		Routing	
	BDL	Mean	ES	Mean	ES
Final Exam	BDL	88.76		91.44	
	In-Person	93.55	-0.27	92.14	
Confidence	BDL	4.24		4.26	
	In-Person	4.14		4.03	0.15
Instructor	BDL	4.60		4.76	
	In-Person	4.62		4.59	0.16
Satisfaction	BDL	4.31		4.34	
	In-Person	4.22		4.22	

 ISBN: 978-1-61208-133-5

TABLE 4. DIFFERENCES BY CLASS – SWITCHING AND LAN

		Switching		WAN	
	BDL	Mean	ES	Mean	ES
Final Exam	BDL	91.84		91.87	
	In-Person	92.58		93.34	
Confidence	BDL	4.14		4.04	
	In-Person	3.93	0.14	3.74	0.19
Instructor	BDL	4.76		4.59	
	In-Person	4.59	0.15	4.53	
Satisfaction	BDL	4.36		4.34	
	In-Person	4.21		4.19	

IV. STUDY 2

Study 2 examines the outcomes of students based on the training modality of their instructor.

A. Participants

This study examined existing data of students who took classes from instructors who were examined in Study 1. This resulted in overall samples of 3514 students of BDL-trained instructors and 3421 students of In-person trained instructors.

B. Measures

As with the instructors, four measures were used to assess outcomes for both the instructor trainees: Satisfaction, Confidence, and Instructor rating subscales from the Course Feedback form and Final Exam scores.

C. Results

When the data for all four courses is combined, the means for the Confidence, Instruction, and Satisfaction subscales and the Final Exam scores are significantly higher for students enrolled in classes taught by BDL-trained instructors (see Table 5). However, it should be noted that the effect sizes were very small.

TABLE 5. OUTCOMES FOR STUDENTS BY INSTRUCTOR TRAINING MODE

	Group	N	Mean	ES
Confidence*	BDL	3514	3.61	
	In-person	3410	3.51	0.06
Instruction*	BDL	3470	4.30	
	In-person	3328	4.23	0.05
Satisfaction*	BDL	2477	3.80	
	In-person	1723	3.72	0.06
Final Exam*	BDL	2430	80.53	
	In-person	2476	79.59	0.03

For Network Fundamentals, the Confidence subscale, Instruction subscale, and Satisfaction subscale means were significantly higher for students enrolled in classes taught by BDL instructors. For Routing, the Confidence subscale mean is significantly higher for students enrolled in classes taught by BDL-trained instructors, although the effect size was small. There were no significant differences in the other measures.

For Switching, the means for the Confidence, Instruction, and Satisfaction subscales and the Final Exam scores are significantly higher for students enrolled in classes taught by BDL-trained instructors. The difference in final exams is the only one that approaches even a small effect.

For WAN, the means for the Confidence, Instruction, and Satisfaction subscales and the Final Exam scores are significantly higher for students enrolled in classes taught by BDL-trained instructors. Both satisfaction and final exam approach a small effect.

V. CONCLUSIONS

This section discusses the results of Study 1 and 2, the limitations of the studies, and conclusions we might reach.

A. Discussion

The purpose of this study was to determine whether there were differences in outcomes for instructors trained via BDL and those trained via in-person classes. In addition, it explored potential differences in their students' outcomes. As the results are examined, it is important to keep in mind the general rule of thumb that effect sizes less than .20 are negligible and likely not clinically important (i.e., there will be little noticeable difference in the individuals). Effect sizes from .20 to .40 are generally considered small [9].

When looking at differences in instructors, the difference in final exam scores between BDL and In-person trainees in the Network Fundamentals course is significant and falls in the range of a small effect. This suggests that for the first course, students may perform slightly better when trained in In-person classes. Given that this is the first class in the sequence and for many students may be their first exposure

 ISBN: 978-1-61208-133-5

to the content, it is possible that the camaraderie and support available during In-person classes may be particularly helpful. In addition, it may be that access to real equipment is more important in this class. Although there are other significant differences in the opinion survey questions favoring BDL classes, these are of negligible effect size.

When examining differences in student outcomes, effect sizes of differences are even smaller. The results overall indicate that there are not meaningful differences in student outcomes dependent on mode of instructor training.

B. Limitations

It should be noted that the instructors who participated in the BDL model of instructor training were self selected. Therefore, a causal link cannot be established because this is not an experimental study. In addition, the instructor trainers for the two groups were not identical. Results therefore cannot absolutely be attributed to class format. We also do not have any visibility into the details of the BDL offering (e.g., how many days/weeks long the course is, what proportion and activities are offered remotely vs. in-person etc.) This analysis relies heavily on student survey responses. Although we have removed students who were clearly not taking this seriously (e.g., those with the same response to each question), the heavy reliance this potentially unreliable source should be considered. Finally, we rely in instructor trainers to accurately report whether their class is offered in a BDL format. It is unknown the extent to which trainers may mis-label their classes.

C. Conclusion and Future Work

Future research should look more closely at the variables that are associated with successful BDL offerings for teachers. There are likely both characteristics of instructor trainees and course practices that are related to positive outcomes for both instructors and their subsequent students. In addition, research might explore differences between initial teacher training and ongoing professional development.

There are continued questions about the impact of training instructors via blended distance learning. This study examined this question with a global sample of instructors and revealed there are very few differences in instructor outcomes or the outcomes of the students they teach. There was some suggestion that instructors may have slightly lower exam scores in the first course in the sequence when taken via BDL. Other than this, differences were negligible and, if anything, favored the BDL solutions. These findings align to a growing body of literature that suggests that BDL solutions produce similar results to In-person learning.

REFERENCES

[1] I. Zandberg and L. Lewis, Technology-based distance education courses for public elementary and secondary school students: 2002-03 and 2004-05. Washington, D.C.: National Center for Educational Statistics, 2008.

[2] A. G. Picciano and J. Seaman, K-12 Online learning: A Follow-up of the Survey of U.S. School District Administrators. Needham, MA: The Sloan Consortium, 2009.

[3] B. Means, Y. Toyama, R. Murphy, M. Bakia, and K. Jones, Evaluation of Evidence-Based Practices in Online learning: A Meta-Analysis and Review of Online Learning Studies. Washington, D. C.: U.S. Department of Education, 2010.

[4] P. Gerbic, E. Stacey, B. Anderson, M. Simpson, J. Mackey, C. Gunn, and G. Samarawickerma, "Blended learning: Is there evidence for its effectiveness?" Proceedings ascilite Auckland, 2009, pp. 1214-1216.

[5] D. Parkinson, W. Greene, Y. Kim, and J. Marioni, "Emerging themes of student satisfaction in a traditional course and a blended distance course," TechTrends, vol. 47, Jul-Aug. 2003, pp. 22-28.

[6] F. Spooner, L. Jordan, B. Algozine, and M. Spooner, "Evaluating instruction in distance learning classes," J. Educ. Res., vol. 92, 1999, pp. 132-140.

[7] R. Voos, "Blended learning: What is it and where might it take us?" Sloan-C View, vol. 2, 2003, pp. 2-5.

[8] P. M. Privateer, "Academic technology and the future of higher education: Strategic paths taken and not taken," J. Higher Educ., vol. 70, pp. 60-79.

[9] A. Smith and N. Moss, "Large scale delivery of Cisco networking Academy program by blended distance learning," Sixth International Conference on Networking and Services, March 2010, Cancun, MX.

[10] J. Cohen, Statistical Power for the Behavioral Sciences. Hillsdale, NJ: Erlbaum.

 ISBN: 978-1-61208-133-5