IARIA

ICN 2013

The Twelfth International Conference on Networks

ISBN: 978-1-61208-245-5

January 27 - February 1, 2013

Seville, Spain

ICN 2013 Editors

Pascal Lorenz, University of Haute Alsace, France

Tibor Gyires, Illinois State University, USA

Iwona Pozniak-Koszalka, Wroclaw University of Technology, Poland

ICN 2013

Foreword

The Twelfth International Conference on Networks [ICN 2013], held between January 27th- February 1st, 2013 in Seville, Spain, continued a series of events focusing on the advances in the field of networks.

ICN 2013 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 fora or in industry consortia, survey papers addressing the key problems and solutions, short papers on work in progress, and panel proposals.

We take here the opportunity to warmly thank all the members of the ICN 2013 Technical Program Committee, as well as the numerous reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and efforts to contribute to ICN 2013. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

Also, this event could not have been a reality without the support of many individuals, organizations, and sponsors. We are grateful to the members of the ICN 2013 organizing committee for their help in handling the logistics and for their work to make this professional meeting a success.

We hope that ICN 2013 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress in the field of networks.

We are convinced that the participants found the event useful and communications very open. We also hope the attendees enjoyed the charm of Seville, Spain.

ICN Chairs:

Pascal Lorenz, University of Haute Alsace, France
Tibor Gyires, Illinois State University, USA
Eva Hladká, Masaryk University - Brno / CESNET, Czech Republic
Iwona Pozniak-Koszalka, Wroclaw University of Technology, Poland

ICN 2013

Committee

ICN General Chair

Pascal Lorenz, University of Haute Alsace, France

ICN Advisory Chairs

Tibor Gyires, Illinois State University, USA
Eva Hladká, Masaryk University - Brno / CESNET, Czech Republic
Iwona Pozniak-Koszalka, Wroclaw University of Technology, Poland

ICN 2013 Technical Program Committee

Pascal Anelli, University of Reunion, France
Jalel Ben-Othman, Université de Versailles, France
João Afonso, FCCN - Fundação para a Computação Científica Nacional - Lisboa, Portugal
Max Agueh, LACSC - ECE Paris, France
Kari Aho, University of Jyväskylä, Finland
Pascal Anelli, Université de la Réunion, France
Cristian Anghel, Politehnica University of Bucharest, Romania
Harald Baier, Hochschule Darmstadt, Germany
Alvaro Barradas, University of Algarve, Portugal
Carlos Becker Westphall, Federal University of Santa Catarina, Brazil
Zdenek Becvar, Czech Technical University in Prague, Czech Republic
Djamel Benferhat, University of South Brittany, France
Ilham Benyahia, Université du Québec en Outaouais - Gatineau, Canada
Robert Bestak, Czech Technical University in Prague, Czech Republic
Jun Bi, Tsinghua University, China
Fernando Boronat Seguí, Universidad Politécnica de Valencia, Spain
Agnieszka Brachman, Silesian University of Technology - Gliwice, Poland
Arslan Brömme, Vattenfall Europe AG | Corporate Security | Security Centre Germany - Berlin, Germany
Matthias R. Brust, Technological Institute of Aeronautics, Brazil
Jorge Luis Castro e Silva, UECE - Universidade Estadual do Ceará, Brazil
Joaquim Celestino Júnior, Universidade Estadual do Ceará (UECE), Brazil
Eduardo Cerqueira, Federal University of Para, Brazil
Marc Cheboldaeff, T-Systems International GmbH, Germany
Buseung Cho, KREONET Center, KISTI - Daejeon, Republic of Korea
Andrzej Chydzinski, Silesian University of Technology - Gliwice, Poland
Nathan Clarke, Plymouth University, UK
Guilherme da Cunha Rodrigues, Federal Institute of Education, Science and Technology Sul -Rio Grandense (IFSUL) - Brasil
Javier Del Ser Lorente, TECNALIA-TELECOM, Spain

Lars Dittman, Technical University of Denmark, Denmark
Daniela Dragomirescu, LAAS/CNRS, Toulouse, France
Matthew Dunlop, Virginia Tech, USA
Sylvain Durand, LIRMM - Montpellier, France
Inès El Korbi, High Institute of Computer Science and Management of Kairouan, Tunisia
Emad Abd Elrahman, TELECOM & Management SudParis - Evry, France
Jose Oscar Fajardo, University of the Basque Country, Spain
Weiwei Fang, Beijing Jiaotong University, China
Mário F. S. Ferreira, University of Aveiro, Portugal
Mário Freire, University of Beira Interior, Portugal
Wolfgang Fritz, Leibniz Supercomputing Centre - Garching b. München, Germany
Holger Fröning, University of Heidelberg, Germany
Laurent George, University of Paris-Est Creteil Val de Marne, France
Eva Gescheidtova, Brno University of Technology, Czech Republic
S.P. Ghrera, Jaypee University of Information Technology - Waknaghat, India
Markus Goldstein, German Research Center for Artificial Intelligence (DFKI), Germany
Anahita Gouya, AFD Technologies, France
Vic Grout, Glyndwr University - Wrexham, UK
Mina S. Guirguis, Texas State University - San Marcos, USA
Huaqun Guo, Institute for Infocomm Research, A*STAR, Singapore
Tibor Gyires, Illinois State University, USA
Keijo Haataja, University of Eastern Finland- Kuopio / Unicta Oy, Finland
Jiri Hajek, FEE-CTU - Prague, Czeck Republic
Mohammad Hammoudeh, Manchester Metropolitan University, UK
Fanilo Harivelo, Université de la Réunion, France
Hiroyuki Hatano, Shizuoka University, Japan
Luiz Henrique Andrade Correia, University of Lavras, Brazil
Eva Hladká, Masaryk University - Brno / CESNET, Czech Republic
Osamu Honda, Onomichi City University, Japan
Xin Huang, Deutsche Telekom, Inc. - Mountain View, USA
Florian Huc, EPFL - Lausanne, Switzerland
Jin-Ok Hwang. Korea University - Seoul, Korea
Muhammad Ali Imran, University of Surrey - Guildford, UK
Raj Jain, Washington University in St. Louis, USA
Borka Jerman-Blažič, Jozef Stefan Institute, Slovenia
Aravind Kailas, UNC - Charlotte, USA
Omid Kashefi, Iran University of Science and Technology-Tehran Iran
Andrzej Kasprzak, Wroclaw University of Technology, Poland
Sokratis K. Katsikas, University of Piraeus, Greece
Abdelmajid Khelil, TU Darmstadt, Germany
Sun-il Kim, University of St. Thomas, USA
Wojciech Kmiecik Wroclaw University of Technology, Poland
Christian Köbel, Technische Hochschule Mittelhessen - Raum, Germany
Leszek Koszalka, Wroclaw University of Technology, Poland
Tomas Koutny, University of West Bohemi-Pilsen, Czech Republic
Polychronis Koutsakis, Technical University of Crete, Greece
Francine Krief, University of Bordeaux, France
Michał Kucharzak, Wroclaw University of Technology, Poland

Radek Kuchta, Brno University of Technology, Czech Republic
Hadi Larijani, Glasgow Caledonian University, UK
Angelos Lazaris, University of Southern California, USA
Steven S. W. Lee, National Chung Cheng University, Taiwan R.O.C.
Jun Li, Qinghua University, China
Yan Li, Conviva, Inc. - San Mateo, USA
Feng Lin, Tennessee Tech University, USA
Diogo Lobato Acatauassú Nunes, Federal University of Para - Belem, Brazil
Andreas Löffler, Friedrich-Alexander-University of Erlangen-Nuremberg, Germany
Pascal Lorenz, University of Haute Alsace, France
Richard Lorion Université de la Réunion, France
Pavel Mach, Czech Technical University in Prague, Czech Republic
Damien Magoni, University of Bordeaux, France
Ahmed Mahdy, Texas A&M University - Corpus Christi, USA
Zoubir Mammeri, IRIT - Paul Sabatier University - Toulouse, France
Anna Manolova Fagertun, Technical University of Denmark, Denmark
Gustavo Marfia, University of Bologna, Italy
Rui Marinheiro, ISCTE - Lisbon University Institute, Portugal
Boris M. Miller, Monash University, Australia
Pascale Minet, INRIA - Rocquencourt, France
Mario Montagud Climent, Universidad Politécnica de Valencia, Spain
Katsuhiro Naito, Mie University - Tsu City, Japan
Frank Oldewurtel, Audi AG - Ingolstadt, Germany
Go-Hasegawa, Osaka University, Japan
Constantin Paleologu, University Politehnica of Bucharest, Romania
Konstantinos Patsakis, University of Piraeus, Greece
João Paulo Pereira, Polytechnic Institute of Bragança, Portugal
Kun Peng, Institute for Infocomm Research, Singapore
Ionut Pirnog, "Politehnica" University of Bucharest, Romania
Marcial Porto Fernandez, Universidade Estadual do Ceara (UECE), Brazil
Iwona Pozniak-Koszalka, Wroclaw University of Technology, Poland
Jani Puttonen, Magister Solutions Ltd., Finland
Shankar Raman, Indian Institute of Technology - Madras, India
Victor Ramos, UAM-Iztapalapa, Mexico
Priyanka Rawat, INRIA Lille - Nord Europe, France
Shukor Razak, Universiti Teknologi Malaysia (UTM), Malaysia
Yenumula B. Reddy, Grambling State University, USA
Krisakorn Rerkrai, RWTH Aachen University, Germany
Karim Mohammed Rezaul, Glyndwr University - Wrexham, UK
Wouter Rogiest, Ghent University, Belgium
Simon Pietro Romano, University of Napoli Federico II, Italy
Jorge Sá Silva, University of Coimbra, Portugal
Teerapat Sanguankotchakorn, Asian Institute of Technology - Klong Luang, Thailand
Susana Sargento, University of Aveiro, Portugal
Panagiotis Sarigiannidis , University of Western Macedonia - Kozani, Greece
Masahiro Sasabe, Osaka University, Japan
Thomas C. Schmidt, HAW Hamburg, Germany
Hans Scholten, University of Twente- Enschede, The Netherlands

Copyright Information

Table of Contents

MAC Protocol for Ad Hoc Network Using Smart Antenna with Pulse/Tone Exchange

Jing Ma, Hiroo Sekiya, Nobuyoshi Komuro, Shiro Sakata
Graduate School of Advanced Integration Science, Chiba University
1-33, Yayoi-cho, Inage-ku, Chiba, 263-8522 Japan
Email: maggie@graduate.chiba-u.jp, sekiya@faculty.chiba-u.jp, kmr@faculty.chiba-u.jp, sakata@faculty.chiba-u.jp

***Abstract*—This paper proposes a MAC protocol for smart antenna used networks. The basic idea is that the Pulse/Tone exchange, instead of the RTS/CTS handshake, is applied prior to the data transmission. By only changing the transmission initiation method, we can obtain four advantages. First, collisions due to general hidden-node problem can be reduced. Second, the collision between the control frame and the DATA frame due to the directional hidden-node problem can be mitigated. Third, the transmission delay can be reduced by using the fixed contention window value. Finally, the overhead can be reduced because the duration of the Pulse/Tone exchange is much shorter than that of the RTS/CTS handshake. As a result, the higher network throughput can be achieved compared with previous protocols. Simulation results demonstrate the validity and effectiveness of the proposed protocol.**

***Keywords*-Smart antennas; ad hoc networks; Pulse/Tone.**

I. Introduction

Recently, wireless communication systems using a beamforming of the smart antenna have attracted many researchers‘ attention [2–6]. Smart antennas provide two separate modes. One is the omni-mode, where the antenna radiates in omni-directions. The other is the directional mode, where the antenna can point its main lobe towards any specified direction. A MAC protocol for smart antenna used networks was proposed in [2], in which IEEE 802.11 with RTS/CTS is applied to smart antenna used networks. Because the spatial-reusability efficiency is enhanced by using smart antennas, the network throughput can be improved. However, there are two dominant factors for degrading the network throughput. One is the collision occurrence. Collisions often occur between two control frames due to general hidden-node problem. Additionally, collisions between the control frame and the DATA frame due to the directional hidden-node problem newly occur in smart antenna used networks. The other factor is the transmission delay due to the deafness problem. The deafness problem is a typical drawback for smart antenna used networks. When the transmission fails, the transmitter cannot understand the cause for transmission failure. The transmitter initiate the retransmission with doubled contention window value(CW) in all cases, because the effect of the transmission failure caused by the collision is larger than that due to the effect of the transmission failure caused by the deafness problem. The binary exponential increase in the CW value causes the transmission delay.

On the other hand, narrow-band and short-time in-band signals, called Pulse and Tone hereafter, are proposed in [7]. According to [7], it is sufficient for nodes to detect the Pulse or Tone signal in 5 μs, which is much shorter than the control frame length, e.g. RTS and CTS frames. Therefore, the simultaneous transmission of Pulse or Tone signals from multiple nodes rarely occurs. Additionally, the Pulse or Tone signal does not interfere with frames. The characteristics of Pulse and Tone signals are suitable to avoid collisions which occur between the control frame and the DATA frame due to the directional hidden-node problem.

This paper proposes a MAC protocol for smart antenna used networks. Each node has only one transceiver in the proposed system. The basic idea is that the Pulse/Tone exchange, instead of the RTS/CTS handshake, is applied to smart antenna used networks. By only changing the transmission initiation method, we can obtain four advantages. First, collisions due to general hidden-node problem can be reduced, because the simultaneous transmission of Pulse or Tone signals from multiple nodes rarely occurs. Second, the collision between the control frame and the DATA frame due to the directional hidden-node problem can be mitigated, because the Pulse or Tone signal does not interfere with DATA frames. Third, the transmission delay can be reduced by using the fixed contention window value. Because collisions due to the general or directional hidden-node problem rarely occur by using Pulse/Tone exchange, it can be stated that the reason for transmission initiation failure is the deafness problem. Therefore, the retransmission is conducted using the fixed CW, which can achieve the transmission delay reduction. Finally, the overhead can be reduced because the duration of the Pulse/Tone exchange is much shorter than that of the RTS/CTS handshake. As a result, the higher network throughput can be achieved compared with previous protocols. Simulation results demonstrate the validity and effectiveness of the proposed protocol.

The rest of the paper is organized as follows: The related work is discussed in Section 2. In Section 3, the details of the proposed protocol is introduced. Performance evaluations are presented in Section 4. Section 5 concludes.

ISBN: 978-1-61208-245-5

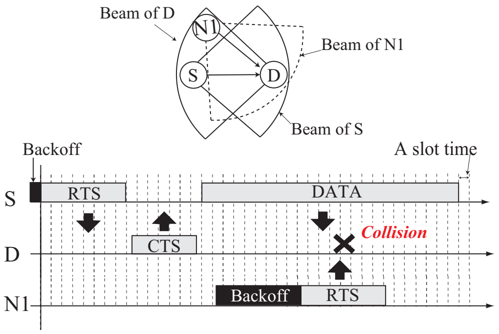


Fig. 1. An example scenario of the collision due to the directional hidden-node problem in the DMAC protocol.

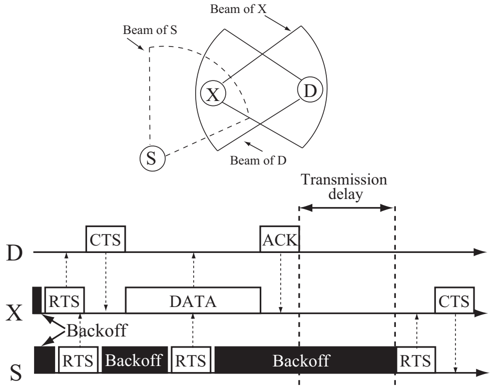


Fig. 2. The transmission failure example due to the deafness problem in the DMAC protocol.

II. Related Works

Wireless communications in smart antenna used networks can enhance the spatial reusability of the network [2–6]. The DMAC (Directional Medium Access Control) [2] protocol is a basic MAC protocol for smart antenna used networks. In the DMAC protocol, a channel is reserved by using RTS/CTS handshakes. Because all frames are transmitted in the directional mode, the network spatial-reusability efficiency is high. Therefore, the throughput can be improved compared with the omni-directional antenna networks.

However, the network throughput is degraded because of two dominant factors in the DMAC protocol. One is the control frame collisions due to general hidden-node problem. This kind of collision often occurs when control frames are transmitted by multiple nodes simultaneously when the offered load is heavy. Additionally, collisions between the control frame and the DATA frame due to the directional hidden-node problem newly appear in the smart antenna used networks. Fig. 1 shows an example scenario of a collision due to the directional hidden-node problem. In Fig. 1, we consider the case that the node N1 communicates with a certain node, which is in the opposite direction of the node S. In this case, the node S cannot hear the RTS/CTS handshake between the nodes S and D. There is a possibility that the node N1 transmits an RTS-frame to the node D after the previous communication. Therefore, the RTS-frame transmission of the node N1 interferes with the DATA-frame transmission of the node S. In this case, the frame transmissions from both the nodes S and N1 are in failure. In Fig. 1, the node N1 is a hidden node of the node S due to the smart-antenna usage. Therefore, this collision problem is called ”directional hidden-node problem”.

The other factor is the deafness problem, which causes the transmission delay according to [2]. The deafness problem is a typical drawback for smart antenna used networks. When a transmitter transmits an RTS frame to a receiver, which transmits or receives a frame to or from another node, the deafness problem occurs. Since the receiver is beamformed toward the direction away from the transmitter, the receiver is unable to hear the RTS frame transmission. Fig. 2 shows an example of transmission failure due to the deafness problem in the DMAC protocol. Since node X is beamforming toward node D, node X cannot comprehend the RTS frame transmission from node S. Accordingly, node S cannot receive the CTS frame for response. Then, node S retransmits the RTS frame after the BT decreases to 0. The initial BT value is set randomly in the range of 0 to CW. When the RTS frame transmission fails, the CW is doubled and the BT is reset. This kind of binary exponential increase in the CW can reduce the RTS frame collision probability. However, if the transmission failure is caused by the deafness problem, then the binary exponential increase in the CW causes wastage of channel resources, as shown in Fig. 2. In Fig. 2, node S fails in the RTS/CTS handshake process due to the deafness problem. It is not necessary to double the CW when the transmitter retransmits the RTS frame. In the DMAC protocol, the transmitter cannot understand the reason for the transmission failures. When an RTS frame retransmission is needed, the CW is doubled in all cases. This is because the throughput decrease due to the effect of the transmission failure caused by the hidden-node problem is larger than that due to the effect of the transmission failure caused by the deafness problem.

III. Proposed MAC protocol

In this paper, a MAC protocol for ad hoc networks with smart antennas is proposed. The basic idea is that the Pulse/Tone exchange, instead of the RTS/CTS handshake, is applied to smart antenna used networks. In the proposed protocol, we only focus on the MAC protocol design. It is assumed that each node knows all the neighbor nodes and their directions, which is the same assumption as the smart-antenna systems [2], [5]. There are some techniques for identifying the node positions. GPS technique [3] is one of the methods which determine the location of a node in the network. Fig. 3 shows a flowchart of the proposed protocol for the transmitter. Compared with the DMAC protocol, the short-duration Pulse/Tone exchange is conducted prior

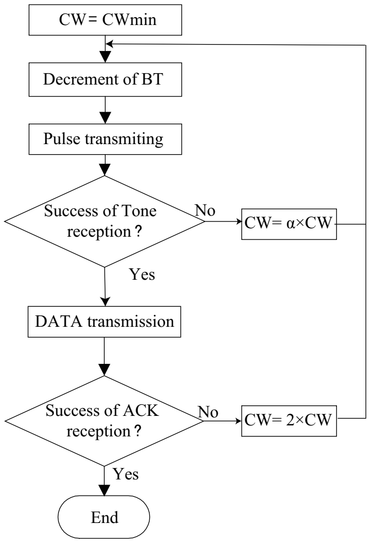


Fig. 3. Flowchart of the proposed protocol.

to the DATA-frame transmission instead of the RTS/CTS handshakes in the proposed protocol.

Pulse and Tone are narrow-band and short-time in-band signals [7]. According to [7], it is sufficient for nodes to detect the Pulse or Tone signal in 5 μs, which is much shorter than the control frame length, e.g. RTS and CTS frames. Therefore, the simultaneous transmission of Pulse or Tone signals from multiple nodes rarely occurs. Additionally, the Pulse or Tone signal does not interfere with frames. Because of the characteristic of Pulse and Tone signals, Pulse and Tone signals can be exchanged within only one time slot [8]. In the proposed protocol, Pulse/Tone exchange, instead of RTS/CTS handshake, is conducted at the final count of the backoff timer (BT).

A. The proposed protocol design

A node, which has no transmission frame, is in the idle state with omni-mode. When a node has a transmission frame, it sets the BT and senses the channel in omni-mode. If the transmitter confirms that the channel is idle, it transfers to directional mode, requests the physical layer to beamform toward the receiver, and sends a Pulse signal. After that, the transmitter sets a Tone-wait timer and waits for the Tone signal in the directional mode. If the transmitter detects the Tone signal, it starts transmitting a DATA frame. Inversely, if the transmitter cannot detect the Tone signal during the Tone-wait timer duration, it transfers to the omni-mode and sets the BT again with the fixed CW, namely α=1 in Fig. 3. After transmitting the DATA-frame, the transmitter sets an ACK-frame-wait timer. If the transmitter receives an ACK-frame from the receiver successfully, the transmission process is finished successfully. If the transmitter cannot receive the ACK-frame from the receiver, it sets the BT again after doubling the CW value as shown in Fig. 3.

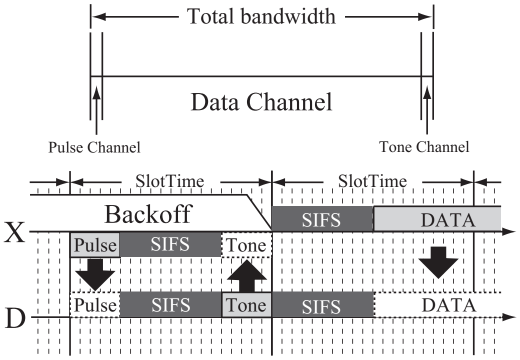


Fig. 4. Pulse/Tone exchange process.

Nodes, which detect the Pulse signal, reply a Tone signal and prepare to receive a DATA frame in directional mode. Nodes, which only detect the Tone signal, will freeze their transmissions of the Tone detected direction.

B. Features of the proposed protocol

In the proposed protocol, each node is only required to have one transceiver. The Pulse/Tone exchange, instead of the RTS/CTS handshake, is applied prior to the DATA frame transmission in smart antenna used networks. By only changing the transmission initiation method, we can obtain four advantages. First, collisions due to general hidden-node problem can be reduced. Second, the collision between the control frame and DATA frame due to the directional hidden-node problem can be mitigated. Third, the transmission delay can be reduced by using the fixed CW. Finally, the overhead can be largely reduced. As a result, the network throughput can be effectively improved.

1) Reduction of frame collisions due to the general hidden-node problem.: Fig. 4 shows the Pulse/Tone exchange process in the proposed protocol. In Fig. 4, the transmitter X sends a Pulse to the receiver D prior to the DATA frame transmission. The duration of the Pulse/Tone exchange is only one slot time, which is 20 μs in IEEE 802.11b, as shown in Fig. 4. The probability that multiple Pulses are sent simultaneously is much lower than the probability of DATA frame collisions. As a result, DATA frame collisions due to the general hidden-node problem are effectively inhibited.

2) Mitigation of directional hidden-node problem: By using Pulse/Tone exchanges, collisions between the control frame and the DATA frame due to the directional hidden-node problem can be mitigated. Fig. 5 shows an example for avoiding the collision due to the directional hidden-node in the proposed protocol. As shown in Fig. 5, the Pulse/Tone exchanges are carried out only one time slot at the final count of the BT. Therefore, the probability of the concurrent transmission of the Pulse signals from multiple nodes is very low. In the shown scenario in 5, when the node N1 finishes the previous communication and wants to

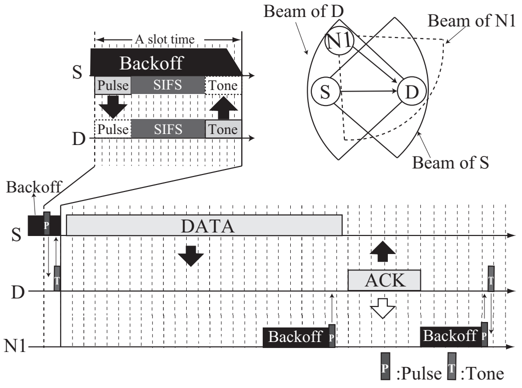


Fig. 5. An example of mitigating the directional hidden-node problem in the proposed protocol.

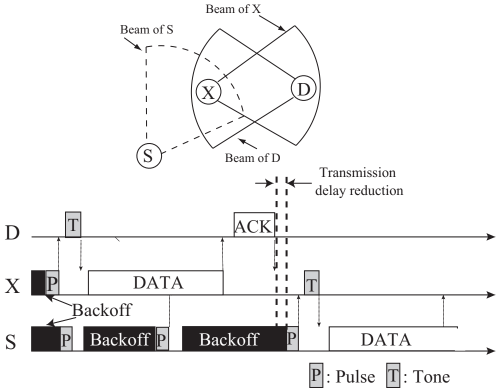


Fig. 6. An example of reducing the transmission delay due to the deafness problem in the proposed protocol.

transmit a new frame to the node D, the node N1 is unaware of the communication between the nodes S and D. In this case, the node N1 sends a Pulse signal as shown in Fig. 5. Because the Pulse signal does not interfere with the frame, the node D can receive the DATA-frame from the node S successfully. This means that the directional hidden-node problem is solved by using Pulse/Tone exchanges.

3) The transmission delay reduction: Fig. 6 shows an example of reducing the transmission delay due to the deafness problem in the proposed protocol. Node S transmits a Pulse to node X, as shown in Fig. 6. Since node X communicates with node D, node S cannot detect a Tone for response. Thus, node S recognizes that node X is busy communicating with another node in another direction. This is because the probability of Pulse-transmission overlapping due to the hidden-node problem is very low. Therefore, node S retransmits a Pulse with the fixed CW, namely $\alpha = 1$ in Fig. 3. As a result, the network throughput can be improved by applying the Pulse/Tone exchange due to the transmission delay reduction.

4) The overhead reduction: In the proposed protocol, the Pulse/Tone exchange is conducted prior to the DATA frame transmission instead of RTS/CTS frame handshake. Because the duration of the Pulse/Tone exchange is much shorter than that of the RTS/CTS handshake [1], [7], the overhead can be largely reduced by using Pulse/Tone exchange instead of RTS/CTS handshake.

By the way, because Pulse and Tone signals do not contain any information, the all nodes, which only detect the Tone signal, freeze their transmission processes in the proposed protocol. It seems that many exposed node will appear in this proposed protocol. Smart antennas are, however, applied in the proposed system. By using smart antennas, the transmission range can be squeezed. Therefore, exposed nodes appearance is limited. The exposed nodes appearance due to the Pulse/Tone exchange is not a serious problem in the proposed protocol.

IV. Performance Evaluations

In order to evaluate the performance of the proposed protocol, we have simulated ad hoc networks implementing the proposed protocol and other conventional protocols using a simulation program written in C. In order to confirm the credibility of our simulator, the throughputs of the IEEE 802.11 DCF obtained using our simulator were verified to be the same as those obtained using the NS-2 simulator. The effects of PHY and upper layer are not included in the results of this paper. Additionally, it is assumed that the bandwidth consumption of the in-band Pulse and Tone signals is negligible compared to the bandwidth of the data channel. This assumption is the same as assumptions in [7], [8]. Each node has both the omni mode and the directional mode with an adaptive array antenna. Generally, directional transmissions have larger transmission range than omni-directional transmissions. Therefore, the directional beamforming may potentially interfere with communications taking place far away. In this paper, however, we would like to focus on the gains from spatial reuse exclusively. Therefore, it is assumed that the transmission range of the directional antenna is the same as that of the omni-directional antenna. Each node can know all neighbor nodes and their directions. Receivers can know the transmitter direction by receiving frames and the sensing Pulse or Tone signals in the omni-mode. It is possible for the nodes to transmit only one frame or one signal at a time.

A. Simulation parameters and results

The simulation parameters are given in Table I, which basically follow those in IEEE 802.11b standard [1]. Data-channel and control-channel rates are 11 Mbps and 1 Mbps, respectively. Both the Pulse and Tone signals are sent for 5 μs duration [7]. Nodes are placed in the 300 m × 300 m square area at random. Each node randomly selects one of the neighbor nodes as a receiver. The traffic model follows the Poisson arrival. The node mobility is not considered in this paper. The angle of the antenna beam is set to $\pi/2$. In this paper, MAC protocol using smart antennas

Table I
SIMULATION PARAMETERS.

Antenna type	Adaptive antenna array antenna
Angle of antenna beam	$\pi/2$
Node density	9.11×10^{-4} nodes/ m^2
Transmission range	135 m
PHY layer	IEEE 802.11b
Data channel rate	11 Mbps
Control channel rate	1 Mbps
Slot time	20 μs
DIFS time	50 μs
SIFS time	10 μs
Minimum CW size	31 slot
Max CW size	1023 slot
Frame payload	1024 bytes
RTS-frame length	20 bytes
CTS/ACK-frame length	14 bytes
Pulse/Tone tx time	5 μs
Simulation area	300 m × 300 m
Simulation time	20 s

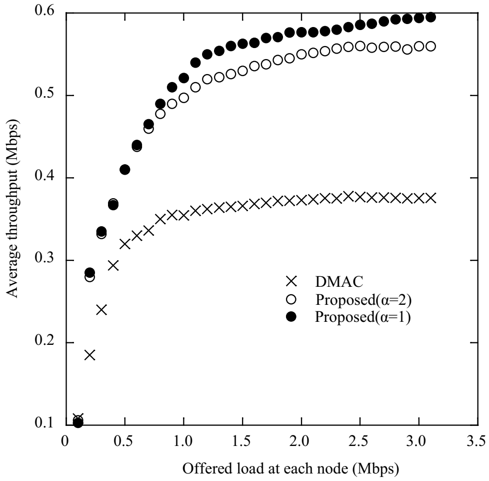


Fig. 7. Average throughput as a function of the offered load at each node.

(DMAC) [2] and the proposed protocol (Proposed) are investigated. Additionally, the proposed protocol is evaluated for α=1 and 2, where α is defined as shown in Fig. 3.

Fig. 7 shows the average throughput as a function of offered load at each node for 9.11×10^{-4} nodes/m^2 of node density. Additionally, Figs. 8 and 9 show the average of backoff time (Aver_backoff) and overhead time (Aver_overhead) per one DATA-frame transmission success as functions of offered load at each node. Aver_backoff, and Aver_overhead are defined as ratio of the total backoff time to the number of the DATA-frame transmission success and ratio of the total control-frame-transmission duration to the number of the DATA-frame transmission successes, respectively. Here, the total control-frame-transmission duration includes RTS-, CTS-, and ACK-frame transmission periods. Pulse and Tone signal durations are not included in the overhead time since Pulse/Tone exchanges are conducted in the final time slot in the backoff stage.

It is seen from Fig. 7 that the proposed protocol provides the higher throughput than DMAC. This is because the collisions due to the general and directional hidden-node problem are reduced by applying the Pulse/Tone exchanges, as well as the overhead is largely reduced in the proposed protocol. DMAC suffers from frame collisions and the deafness problem. It can be confirmed from Fig. 8 that DMAC achieves the highest Aver_backoff. This is because collisions cause many retransmission, as well as the deafness problem causes the transmission delay. In the proposed protocol, by using the Pulse/Tone exchange, collisions can be reduced. Therefore, it can be confirmed from Fig. 8 that Aver_backoff of the proposed protocol is much lower than that of DMAC. Additionally, the overhead is reduced compared with DMAC. It can be confirmed in Fig. 9 that Aver_overhead of the proposed protocol is lower than that of DMAC. As a result, the proposed protocol achieves the higher throughput compared with DMAC.

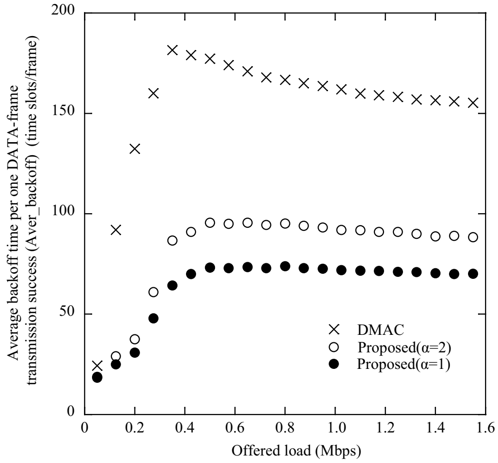


Fig. 8. The average backoff periods per successful frame transmission at each node.

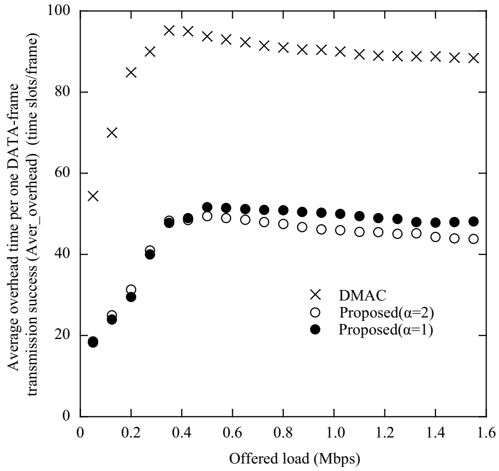


Fig. 9. The average overhead per successful frame transmission at each node.

In addition, it is seen from Fig. 7 that the throughput

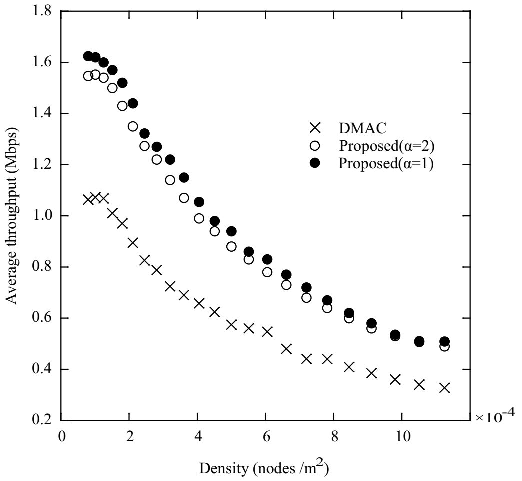


Fig. 10. Average throughput as a function of the node density

of the proposed protocol for α=1 is higher than that for α=2. This is because transmission delay induced by the deafness problem is reduced by retransmitting with the fixed CW in the proposed protocol for α=1. It can be confirmed from Fig. 8 that the proposed protocol for α=1 shows lower Aver_backoff than that for α=2. Therefore, the transmission delay reduction enhances the network throughput by using the fixed CW value. By the way, it is seen from Fig. 9 that the Aver_overhead of the proposed protocol for α=1 is a little higher than that of the proposed protocol for α=2. This is because fixing the CW increases the retransmission probability in the proposed protocol for α=1. Note that the positive factor of the transmission delay reduction can overcome the negative factor of the increase in retransmissions, which can be confirmed in Figs. 8 and 9.

Fig. 10 shows the average throughput as a function of the node density for 2.5 Mbps of offered load. It is seen from Fig. 10 that the throughput decreases as the node density increases for all the protocols. When the node density is high, it is inevitable that the network throughput is degraded due to collisions. However, it is seen from Fig. 10 that the proposed protocol for α=1 achieves the highest throughput regardless of the node density. In the proposed protocol, the positive factor of the collision reduction, the transmission delay reduction, and the overhead reduction improves the network throughput even if the node density is high.

Additionally, it is seen from Fig. 10 that the throughput difference between the proposed protocol for α=1 and that for α=2 becomes small as the node density increases. As the node density increases, the possibility that the transmitter detects the unexpected Tone signals becomes high in spite of the smart antenna used networks. Therefore, most of the Pulse/Tone exchanges are in success. Therefore, the behavior of the proposed protocol for α=1 is almost the same as that for α=2 as the node density increases. In this case, the DATA-frame collisions due to the directional-hidden node problem occur.

V. Conclusions

This paper has proposed a MAC protocol for smart antenna used networks. The basic idea is that the Pulse and Tone signals are exchanged, instead of the RTS/CTS handshake, prior to the data transmission. In the proposed protocol, first, collisions due to general hidden-node problem can be reduced. Second, the collision between the control frame and the DATA frame due to the directional hidden-node problem can be mitigated. Third, the transmission delay can be reduced by using the fixed contention window value. Finally, the overhead can be reduced because the duration of the Pulse/Tone exchange is much shorter than that of the RTS/CTS handshake. As a result, the higher network throughput can be achieved compared with previous protocols. Simulation results demonstrate the validity and effectiveness of the proposed protocol.

References

[1] *IEEE 802.11 Standard: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification*, IEEE, 1999.

[2] R. R. Choudhury, X. Yang, R. Ramanathan, and N. H. Vaidya, “Using Directional Antennas for Medium Access Control in Ad Hoc Networks,” *Proc. ACM MobiCom*, Atlanta, Georgia, USA, pp.59-70, Sept. 2002.

[3] S. Motegi, H. Sekiya, J. Ma, K. Sanada, and S. Sakata, “A directional MAC protocol with the DATA-frame fragmentation and short busy advertisement signal for mitigating the directional hidden node problem, “ *Proc. PIMRC‘ 12*, Sydney, Australia, Sept. 2012.

[4] J. L. Bordim, K. Nakano, ”Deafness Resilient MAC Protocol for Directional Communications,” *IEICE TRANSACTIONS on Information and Systems*, Vol.E93-D, No.12, pp.3243-3250, 2010.

[5] K. Sundaresan, R. Sivakumar, “Cooperating with smartness: using heterogeneous smart antennas in multihop wireless networks,” *Mobile Computing, IEEE Transactions on*, vol.10, no.12, pp.1666-1680, Dec. 2011.

[6] K. Takano, S. Shiokawa, ”Fairness aware back-off control scheme in wireless network using directional MAC protocol”, IEICE Tech. Rep., vol.110, no.449, IN2010-187, pp.259-264, March 2011. (in Japanese)

[7] Fouad A. Tobagi and L. Kleinrock, ”Packet switching in radio channels: part II- the hidden terminal problem in carrier sense multiple-access and the busy-tone solution, “ *IEEE Transactions on Communications*, COM-23(12), Dec. 1975.

[8] K. P. Shih, W. H. Liao, H. C. Chen, and C. M. Chou, “On avoiding RTS collisions for IEEE 802.11-based wireless ad hoc networks,” *Computer Communications*, vol.32, no.1, pp.69-77, Jan. 2009.

Achieving connectivity in an Unstructured Wireless Sensor Network using Optimal Assignment of Mobile Nodes

Paritosh Ramanan, Prathamesh Gaikwad, Sreejith Vidyadharan
Department of Computer Science and Information Systems
BITS-Pilani, K.K. Birla Goa Campus
Email: *paritosh.ramanan@gmail.com*, *prathameshgaikwad09@gmail.com*, *srev@ymail.com*

***Abstract*—Wireless Sensor Networks (WSNs) have been proposed as a solution to problems in a variety of monitoring applications. Most of the random deployments in sensor networks lack proper connectivity. In this paper, an approach is presented to achieve connectivity in a disconnected random deployment of sensor networks. This can be accomplished by using mobile nodes for establishing connectivity in the network as well as for information gathering. The concept of Steiner point has been used to find the co- ordinates for placing the mobile nodes. The paper proposes a variant of Hungarian algorithm for placing mobile nodes in the field in minimum time. The simulation results indicate that the proposed approach can be effectively used in an unstructured deployment of WSN.**

Keywords-WSN, Mobility, Assignment Problem

I. Introduction

The field of WSNs has emerged over recent years with a wide variety of applications. WSNs have been applied in scenarios such as health-monitoring, industrial and consumer applications. Another application of WSNs is that of intruder detection in remote areas, where sensor nodes continuously monitor any movement in their vicinity [1].

A WSN, typically, consists of a *sensor node*-a PCB (Printed Circuit Board) having a micro-controller, source of power supply, sensors for measuring the physical parameters and a radio transceiver for sending and receiving data. Nodes are deployed over a wide area forming a multi hop network in order to relay the data to a central node commonly referred to as a base station. The base station in a WSN has a dedicated power supply and a higher processing capability. The base station receives the information from the static nodes relayed over the network and processes the data.

There are many constraints associated with the maintainability of WSNs; one of the main constraints is connectivity. In case the node runs out of power it will disconnect from the network, which might have serious consequence on the connectivity of the network. In most of the applications the node once deployed cannot be accessed manually thereby highlighting the importance of maintaining connectivity.

In this paper, the issue of connectivity in the network has been addressed in order to ensure a seamless communication between each node and the base station. In case of a random deployment of sensor nodes, clusters of nodes may be formed which are disconnected from the base station. Such networks consist of clusters of nodes which are connected to each other but may not be connected to the base station. This may lead to a breach of communication between the individual nodes which are part of such clusters and the base station. A mechanism is hence required to ensure that the network remains connected.

The paper aims for an efficient placement of mobile nodes to connect the network. Since the problem calls for a technique of graph augmentation, the concept of Steiner point [2] has been used to determine positions of the mobile nodes with respect to the individual clusters. As mobile nodes themselves are scattered across the area, they must be assigned to points within the field such that the total time for the entire assignment is minimized.

The technique illustrated in this paper can find applications in unstructured deployments where clusters of nodes may be formed leading to a disconnected network. It will be useful in scenarios such as deploying nodes for volcanic activity monitoring, military purposes like object tracking and monitoring environmental conditions in remote areas. The approach also could be adapted with little modification in cases where a deployed network might get disconnected due to malfunctioning of sensor nodes.

This paper is structured as follows: Section II talks about the problem of connectivity in the network and an overview of the techniques to solve them. Section III describes the network architecture explaining the various phases like deployment, discovery, calculation of mobile nodes' positions, optimal assignment of mobile nodes and finally communication to the base station. Section IV illustrates the sequence of events during the simulation along with the results. Section V concludes the paper and highlights the future plans.

II. Background

Achieving connectivity is a significant challenge in WSNs. Connectivity can be achieved by calculating the exact density of nodes required for covering an area. The greater the density, the higher is the chance of attaining connectivity. However in the case of random deployment, covering the entire area becomes a problem due to an uneven distribution of nodes.

This paper proposes the use of mobile nodes to address the problem of connectivity in the case of random deployment. Mobile nodes have been used for data collection as well as for node discovery. Various techniques can be employed for data collection as illustrated in [3]. The concept of mobility has already been employed through Data MULEs (Mobile Ubiquitous LAN Extension) and mobile gateways in this regard.

Data collection and transfer to the base station could be accomplished by using a Data MULE which are based on the concept of DTNs (Delay Tolerant Networks). These devices are nothing but miniature mobile computers which roam around the whole field and gather data and remain in motion during most of the lifetime of the network. Data MULES transmit the data to the base station wirelessly through ZigBee, Wi-Fi or Bluetooth, by physically travelling to the base station to complete the data transfer. In [4], an approach to optimally patrol disconnected clusters of static nodes using Data MULES is presented. However, the Data MULES need to be aware of a certain optimal path which they must follow to reach the destination. An attempt has been made in [5] to address this issue.

In cases where the sensor node is at a considerable distance from the base station, a lot of energy is used for multi-hop routing. To avoid such energy losses, a device known as a mobile gateway is used in the field. The primary aim of this device is to act as a mobile base station. In [6], an attempt has been made to reduce the energy consumption due to excessive communication traffic between the static nodes and the mobile gateways.

Many problems may arise during the deployment of an unstructured sensor network. Also the sensor nodes do not have a reliable collision detection mechanism and they do not have knowledge about the network topology. Various techniques to solve the above problems are mentioned in [7].

In this approach, the base station is required to communicate with all the mobile nodes in the network not only for instructing them to occupy particular positions in the network, but also for collecting data about all the clusters and their respective centroids. There is, therefore, a requirement for a technique which enables such a communication over a long range. This can be achieved by using cellular technology. Since cellular technology consumes more power, it is not advisable to use the same for a long duration communication.

In order to save energy, the bulk of data transfer in the network is done through ZigBee [8] although cellular technology is used for communicating with mobile nodes. ZigBee is an IEEE specification which is used extensively in sensor networks due to its low power consumption and longer battery life compared to specifications like Bluetooth or Wi-Fi (Wireless Fidelity). Cellular technology is used mainly during the initial deployment to gather data and to instruct mobile nodes. Cellular technology consists of a variety of standards developed over the years, including GPRS (General Packet Radio Service), EDGE (Enhanced Data Rates for Global Evolution), and so on, which are used for data transfer.

The mobile node communicates on two fronts, with sensor nodes through ZigBee and with the base station through cellular radio. An implementation similar to [9] and [10] could be used in this regard which has both ZigBee and cellular module (GPRS) on the same node.

The concept of Data MULES could be used instead of cellular technology. However, in such a case the mobile node will have to travel quite a distance to transmit data to the base station, which leads to an increase in battery consumption and introduce a considerable delay in the assignment of mobile nodes.

Though the use of cellular technology itself to transfer sensor data from static nodes to base station appears more convenient, it is however power expensive. Cellular technology will put a massive strain on the limited power source of a sensor node for it to be used effectively. Therefore in this paper it is proposed that cellular technology be used only in the initial stages. Once mobile nodes are assigned to positions in the field, the cellular radio can be put to sleep.

Further, for assigning the mobile nodes to the respective positions, the Hungarian method [11] is used. It assigns jobs by a one-is-to-one matching to identify the lowest-cost solution. Each job must be assigned to only one machine. It is assumed that every machine is capable of handling every job, and that the costs or values associated with each assignment combination are known and fixed. The above mentioned algorithm after modification could be used in optimal placement of mobile nodes. The detailed algorithm and the steps involved are explained in the following work [11]. Although [3] mentions some techniques for data collection using mobile elements, the focus of our paper is network connectivity.

The routing algorithm for communication between the nodes and the base station is explained in Section IV.

Assumptions

In this work, the following assumptions have been made:

- The static nodes in the network have already been localized, and have prior knowledge of their global positions. The localization of the nodes can be done using range free techniques such as the APIT (Approximate Point-In-Triangulation Test) algorithm [12].
- The mobile nodes are equipped with a GPS (Global Positioning System) module for localization.
- The mobile node also has a GPRS module with which it can communicate with the base station. The base station using GPRS, can instruct these mobile nodes to occupy certain positions in the network.

III. Network Architecture

The network architecture will be described in 5 phases. The phases are divided according to the sequence of operations which take place in the network right from deployment of the nodes to the final seamless communication establishment between each node and the base station. Subsection A elaborates on the modules that are used for communication and basic assumptions about static and mobile nodes. Subsection B outlines the various steps involved in data collection by mobile nodes. In Subsection C, the exact process of calculating the positions of mobile nodes using the Steiner point algorithm is described. Subsection D illustrates the requirement for optimal assignment of mobile nodes and the Modified Hungarian algorithm which is used to accomplish the same. Finally, in Subsection E, the routing algorithm which can be used to send data from the nodes to the base station is presented. It is to be noted that all the phases henceforth are executed only during the initial stage.

A. Deployment Phase

In the field of WSNs, nodes are often deployed without an organized structure. In such cases the nodes are randomly deployed over the area of interest and often lead to cost effective deployment scheme. However, such a deployment can give rise to issues in network connectivity, wherein the static nodes form disconnected clusters amongst themselves.

B. Discovery Phase

The discovery phase involves data collection, i.e., the positions of each individual node along with the centroid of its cluster. Each static node in the network maintains a $discovery$ bit, which is initially set to zero. This work assumes that there are enough mobile nodes to carry out this step initially. The mobile nodes take a random walk and establish contact with any static nodes in the vicinity. Such static nodes are referred to as gateway nodes and facilitate the gathering of information by the mobile node. The $discovery$ bit is required by each static node in the network to determine whether its position has been relayed to the base station or not. This is done to avoid repeating the same sequence of steps if any another static node is contacted by a mobile node in the future. There are three different types of messages which are exchanged, they are:

- $co-ordinate\ request$: It is sent by the gateway to all the nodes in the cluster to gather their co-ordinates.
- $co-ordinate\ response$: It is sent by the node to the gateway containing its co-ordinates.
- $discovered$: It is sent by the gateway to all the nodes in the network to change the $discovery$ bit from default zero to one.

The following are the sequence of events for node discovery, from the individual cluster to the base station:

STEP 1: The mobile node initially establishes contact with the nearest node which is part of a cluster. The node which is contacted by the mobile node now serves as the $gateway$ which employs a one way broadcasting [13] to gather co-ordinates of each node. It broadcasts a $co-ordinate\ request$ packet into the network.

STEP 2: On receiving the $co-ordinate\ request$ packet, a node responds back to the gateway with a $co-ordinate\ response$, and in turn forwards this $co-ordinate\ request$ packet to its own neighbours. This process repeats until all nodes in the cluster have been covered.

STEP 3: On receiving the responses from nodes, the gateway forwards individual positions of the nodes to the mobile node.

STEP 4: On receiving the co-ordinates from all nodes, the mobile node calculates the centroid of the entire cluster. The mobile node in turn relays the data obtained to the base station using cellular network.

STEP 5: The gateway now sends a $discovered$ message back to all the nodes. On receiving this message all nodes will set their $discovery$ bit to one.

STEP 6: The base station now has the information of all nodes in the form of tuples <position of node, centroid of cluster>. The centroid can be used to uniquely identify a given node as part of a particular cluster.

STEP 7: In case another mobile node approaches the above cluster, it will find the $discovery$ bit of the node already set to one, and will move ahead in search of an undiscovered cluster.

The centroid obtained from each cluster is used as an identifier to the cluster.

In [14], the time distribution required for the mobile nodes to visit all the sensor nodes is given.

C. Calculation of the positions of the mobile nodes

On the successful completion of the discovery phase, the base station has obtained the information regarding all the static nodes and the position of the centroid of their clusters. It now seeks to calculate positions for the assignment of the mobile nodes for obtaining connectivity using the minimal number of mobile nodes. To achieve this the concept of Steiner point in geometry is used.

1) Steiner Point: Steiner Point [2] in a graph is an extra vertex, not originally part of the vertex set of the graph, which is introduced into the graph with the aim of making the graph connected in an efficient way. The Steiner Point has the property that the sum of the distances from itself to all the other vertices is minimum. When the Steiner Point is determined with respect to three other vertices, it is also known as the Fermat's point.

Geometrically the construction of the Steiner point is as follows.

- Let the triangle whose Steiner Point is to be calculated be ABC.
- Initially, construct two equilateral triangles on any two out of the three sides of the triangle, wherein the equilateral triangles so constructed have one side common with the original triangle ABC.
- For each of the equilateral triangles, draw a line from its non-common vertex, i.e., the vertex which does not lie on the shared side with ABC, to the vertex of ABC which lies opposite to the shared side.
- The Steiner Point is obtained from the intersection of the two lines.

The Steiner Point with respect to three vertices is calculated in the following way.

- Calculate the barycentrics of the Steiner Points

  f(a,b,c) : f(b,c,a) : f(c,a,b), where

$$f(a,b,c) = a^4-2(b^2-c^2)^2+a^2(b^2+c^2+4(\sqrt{3})Area(ABC)) \quad (1)$$

where a,b,c are the edge lengths of the triangles and f(a,b,c),f(b,c,a),f(c,a,b) are the barycentric co-ordinates.
- Let A,B,C be the cartesian co-ordinates and let the barycentrics be p, q, r then the cartesian co-ordinates of the Steiner point can be obtained by

$$\frac{pA+qB+rC}{p+q+r} \tag{2}$$

2) Iterative Procedure for connecting clusters: The concept of the Steiner point can now be employed by the base station to calculate positions in the graph which the mobile nodes will eventually occupy. As mentioned earlier, the base station has the knowledge of the centroid of each cluster. The centroid positions can now be used along with the Steiner point to connect the graph in an iterative procedure.

$centroid[i] \leftarrow centroid\ of\ i^{th}\ cluster$

Algorithm 1 Steiner Point algorithm

```
$c_1 \leftarrow centroid[j],\ where\ j\ is\ index\ of\ base\ cluster$
$n_1 \leftarrow number\ of\ nodes\ in\ base\ cluster$
$S = set\ of\ centroids$
$S = S - c_1$
while S $\neq \phi$ do
    $find\ clusters\ 2\ and\ 3\ such\ that\ their\ centroids\ c_2, c_3\ are\ nearest\ to\ c_1$
    $n_2 \leftarrow number\ of\ nodes\ in\ cluster 2$
    $n_3 \leftarrow number\ of\ nodes\ in\ cluster 3$
    $calculate\ Steiner\ Point\ w.r.t\ c_1, c_2, c_3$
    $c_1 \leftarrow \frac{n_1*c_1+n_2*c_2+n_3*c_3}{n_1+n_2+n_3}$
    $S = S - c_2 - c_3$
end while
```

The algorithm initially isolates the centroid of the base cluster and stores it in c_1. By comparing the distance of c_1 with all other centroids, it obtains the centroids of two clusters c_2 and c_3 nearest to c_1. It calculates the Steiner point with respect to c_1,c_2 and $c3$. It now sets c_1 to the centroid of all three clusters.

After having obtained the Steiner Point with respect to the clusters, optimally connecting it to the cluster is also important. Fig. 1 illustrates the connection between the Steiner Point and the cluster. In order to connect to the cluster, the base station selects a point on the periphery of the cluster such that it is at an optimal distance from the Steiner point itself.

The cluster is initially divided into two regions (region I and region II) based on the line L which passes through the centroid C and L ⊥ AC. Let the nodes falling in the region (region I) on whose side the Steiner point lies, be part of a set Q. Now, only nodes belonging to set Q are considered.

The Steiner point is located at A(x_2, y_2), the centroid is located at point C(x_3, y_3)

Let D=$max\ distance\ of\ i^{th}\ node \forall\ i\epsilon\ Q$

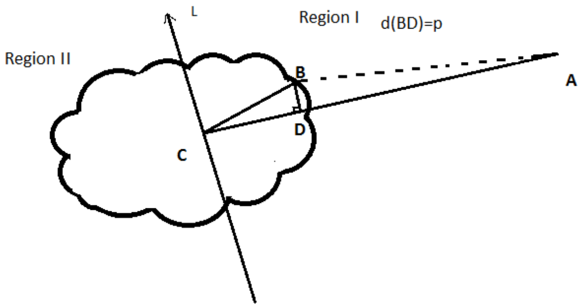


Fig. 1: Cluster Head

let p_i=perpendicular distance of the i^{th} node in cluster to the line joining AC.

Select point B(x_1, y_1) such that,
if, E(i)=$D - d_i + p_i$

then, E(B)=$min\ E(i) \forall\ i\epsilon\ Q$

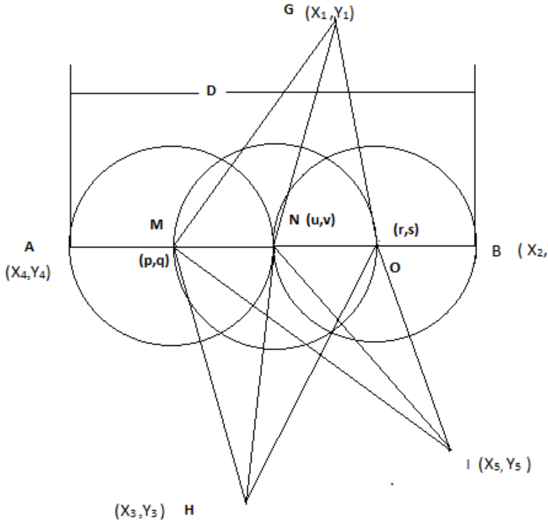


Fig. 2: Bisection of line AB

After the clusterhead is determined mobile node positions are calculated by dividing the line joining the Steiner point to the clusterhead.

In Fig. 2, points A(X_4,Y_4) and B(X_2,Y_2) are the clusterhead and Steiner point respectively. The paper discusses the bisection algorithm to divide line AB. The distance between A and B is D. The slope of line AB is given by:

$$k = \frac{Y_4 - Y_2}{X_4 - X_2} \tag{3}$$

Assuming a spherical range pattern of all nodes, Fig. 2 illustrates the positions obtained after applying the bisection algorithm on line AB.

The mobile nodes are to be placed in such a way that the successor mobile nodes are just within the range of their predecessor. In Fig. 2, the successor N of mobile node M is placed on line AB and in such a way that it is just within the range of mobile node N. However, the predecessor of node M is the clusterhead itself. The given equations express this relation,where d is the range of a mobile node and points (p,q) are the co-ordinates of the final mobile node position.

$$(X_4 - p)^2 + (Y_4 - q)^2 = d^2 \tag{4}$$

$$\frac{Y_4 - q}{X_4 - p} = k \tag{5}$$

On solving equations 4,5 :

$$p = \frac{\pm d}{\sqrt{k^2 + 1}} + X_4 \tag{6}$$

$$q = \frac{\pm kd}{\sqrt{k^2 + 1}} + Y_4 \tag{7}$$

In the case wherein the x co-ordinates of points A and B are equal, the slope of line AB is undefined. In this case, the x co-ordinate of the final mobile node position will be equal to that of A and B, and the y co-ordinate can be obtained by adding the range directly to the y co-ordinate of A.

Algorithm 2 calculates the co-ordinates of the mobile nodes on a line AB with slope k. Using the maximum range of the mobile nodes, it determines the number of mobile nodes required to connect AB.

Algorithm 2 Bisection Algorithm

```
$dist = distance(A, B)$
$ratio = dist/d$
$n = \lceil ratio \rceil - 1$
for $i = 0 \to n$ do
    $range = d * (i + 1)$
    if $A.x! = B.x$ then
        $find\ p_1\ p_2\ using\ 6$
        $find\ q_1\ q_2\ using\ 7$
    else
        if $A.y < B.y$ then
            $q_1 = A.y + range$
            $q_2 = A.y + range$
        else
            $q_1 = A.y - range$
            $q_2 = A.y - range$
        end if
    end if
    $calculate\ d11 = distance\ of\ (p_1, q_1)\ from\ B$
    $calculate\ d12 = distance\ of\ (p_2, q_2)\ from\ B$
    $choose\ the\ lesser\ one\ as\ final\ position$
end for
```

Equations 6 and 7 give two sets of co-ordinates (p_1, q_1) and (p_2, q_2) and one of these is chosen as the correct point on the basis of its distance from B. In order to get the next mobile node position, iteratively increase the $range$ in multiples of d.

D. Optimal Assignment of mobile nodes

Since the mobile nodes are scattered all over the network, the base station needs to assign a particular mobile node to one of the previously calculated positions. The mobile nodes use GPS to navigate to their calculated positions and will cease to use GPS on occupying the positions and hence save power. There are implementations like [15] which provide a low power GPS solutions which is useful for saving energy of the mobile nodes. Since it is an assignment problem, a modified form of the Hungarian algorithm is used to achieve this.

Modified Hungarian Algorithm: This approach however demands an algorithm which can minimize the maximum time taken for the simultaneous movement of all mobile nodes from their current positions to the calculated positions. This problem is therefore a BAP (Bottleneck Assignment Problem). The paper uses the technique illustrated in [16] to solve the BAP. This modified hungarian algorithm takes a two dimensional array as input with each element pq having the value of the distance of the current position of the p^{th} mobile node from the q^{th} calculated position. It gives as output n assignments such that each node is assigned to a particular calculated position and also such that maximum time taken for the entire process is minimized.

The algorithm illustrated in [16] follows the following steps

STEP 1: Given a $n \times n$ matrix

$$A = \begin{pmatrix} a_{00} & a_{01} & \ldots\ldots & & a_{0n} \\ a_{10} & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ a_{n0} & a_{n1} & \ldots\ldots & & a_{nn} \end{pmatrix}$$

arrange the elements in non descending order and assign ranks to each element according to its position. Equal elements will have the same rank. Calculate another two dimensional array of size $n \times n$ B with $O(n^2 \log n)$
b_{ij}=rank of element a_{ij}
The above can be done by quicksort algorithm with an input of n^2 elements in .
STEP 2:
Choose α in such a way that
$b_{ij}^{\alpha} \geq n * (b_{ij} - 1)^{\alpha}\ \forall\ b_{ij}$

Set,
$b_{ij} = b_{ij}^{\alpha}\ \forall\ elements\ b_{ij}\ \epsilon\ B$

This can be done in $O(n^2)$
STEP 3: Apply Hungarian algorithm as illustrated in [16] to get the assignment. This is done in $O(n^3)$.

Hence the total complexity for the entire process comes out to be $O(n^3)$, with $O(n^2 \log n)$ complexity for arranging the elements and $O(n^2)$ for setting the values in matrix B and

$O(n^3)$ for the Hungarian algorithm itself. Since the algorithm works in polynomial time, its time complexity is better than a non polynomial time one.

E. Communication

The assignment phase must be followed by the setting up of a communication protocol in the network, so that the nodes can communicate with the base station seamlessly. At this stage, the mobile nodes will occupy the positions as calculated in the previous section. The mobile nodes augment the network by establishing connectivity with respect to cluster nodes and the base station. Hereafter there is a requirement for a routing mechanism to route the data generated from the sensor nodes to the base station. The Ad-hoc On-demand Distance Vector (AODV) is used as a routing protocol in this regard. The advantages of AODV are that it is an *on demand* protocol which creates the route only when required by a particular node and also scales well with increase in number of nodes in the network. An application of AODV to a network consisting of ZigBee (IEEE 802.15.4) along with minimum battery consumption is provided in [13].

In the case presented herein, the base station propagates a one way broadcast message in the network. Each successive node on receiving this message updates the address of its parent node and in turn broadcasts another such messsage with its own address as the source address to its neighbours. This process continues until all nodes in the network know their parent node. If a sensor node is required to send data, it will forward it to its parent node.

IV. Result

The testing and simulation of the proposed algorithm as discussed in Section III is done using Omnet++ simulator [17]. The MiXiM framework of Omnet++ is used in this regard. The modules and flow is as shown in Fig. 3. Once the base station node has the cluster information, it will calculate the positions of the mobile nodes. The algorithm begins initially by choosing the cluster which contains the base station node and it is designated as the base cluster. The nearest two clusters with respect to the base cluster are determined. The Steiner point is calculated with respect to three clusters. The centroids of the three clusters are taken as the three vertices of a triangle. If the Steiner point does not lie within any of the three cluster, the cluster head is calculated with respect to the Steiner point. Otherwise, if the Steiner point is within the range of the three clusters, the cluster head is calculated by choosing a pair of nodes belonging to two different clusters with the least distance. Using the bisection algorithm the mobile node positions are calculated. As all the three clusters are now connected, a bigger cluster comprising of all the three clusters is now formed. This bigger cluster is now designated as the base cluster for the next iteration. This process repeats until all clusters are exhausted.

After all the mobile node positions are determined, the modified Hungarian algorithm as illustrated in Section III(D) will be applied to the given scenario to assign the mobile nodes to their respective positions.

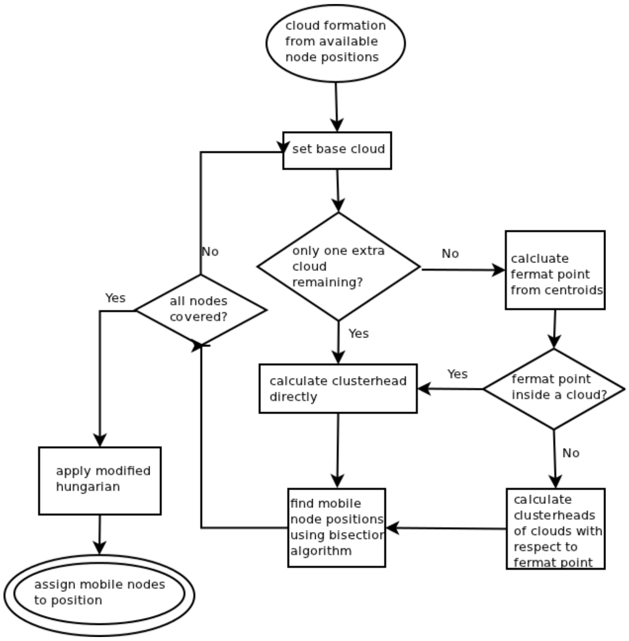

cloud formation from available node positions
set base cloud
only one extra cloud remaining?
No
calcluate fermat point from centroids
Yes
calculate clusterhead directly
Yes
fermat point inside a cloud?
No
calculate clusterheads of clouds with respect to fermat point
find mobile node positions using bisection algorithm
No
all nodes covered?
Yes
apply modified hungarian
assign mobile nodes to position

Fig. 3: Flowchart

Once the network is connected, AODV routing technique is used to form the path from every sensor node to the base station. Sensor nodes will send data using multihop communication to the base station using the route obtained using AODV.

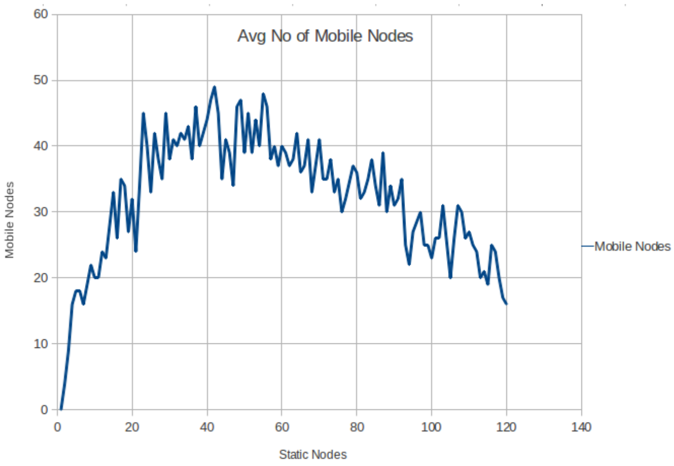


Fig. 4: Mobile nodes vs Static nodes

In order to determine the relation between the number of static nodes to the number of mobile nodes, a graph between the average number of mobile nodes required for a particular number of static nodes has been illustrated in Fig. 4.

In Fig. 4, the graph was generated by simulating the average number of mobile nodes required for a particular number of static nodes spread across an area of 1000m X 1000m with

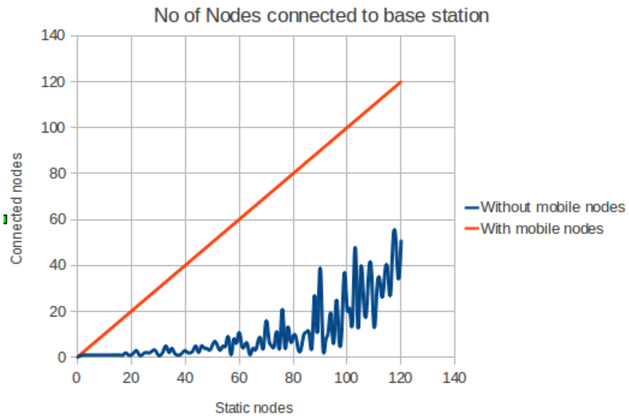


Fig. 5: Connected nodes vs Static nodes

each node having a range of 103.7m.

It can be inferred from Fig. 4 that the number of mobile nodes required to connect the network keeps increasing. Initially as the network is sparse, with an increase in the number of static nodes, the number of mobile nodes also increases. Gradually as the density of static nodes increases, the number of mobile nodes stabilizes. With further increase in density of static nodes the number of mobile nodes starts to decline. This is due to the fact that with an increase in density of static nodes there are lesser number of disconnected clusters formed and hence a lesser number of mobile nodes required.

Fig. 5 plots the number of static nodes deployed in the network to the number of static nodes connected to the base station. The straight line denotes the number of static nodes connected to the base station after mobile nodes have been deployed.

It can be seen that with the use of mobile nodes all the static nodes are connected to the base station. Whereas, without the use of mobile nodes there are static nodes which are unconnected to the base station. The number of nodes connected to base station keep increasing as the network becomes more dense.

Scenario

The algorithm has been simulated for a number of combinations of area and network size. One such scenario is illustrated in Fig. 6., which depicts a network of 36 nodes randomly deployed over an area of $700m \times 700m$ with the range of a node being 103.7m. The node 35 is designated as the base station.

Fig. 6 shows the screenshot of Omnet++ simulation with 36 nodes. As shown in the figure, clusters of nodes are formed which are disconnected from the base station. From Fig. 6 it is clear that none of the clusters are connected to the base station and hence the packet delivery is null.

Fig. 7 shows the screenshot of the simulation after applying the algorithm. The resulting network is connected to the base station with ten mobile nodes.

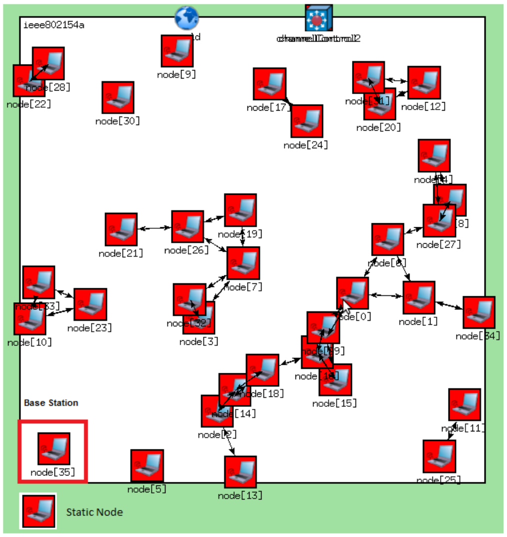


Fig. 6: Before assignment of mobile nodes

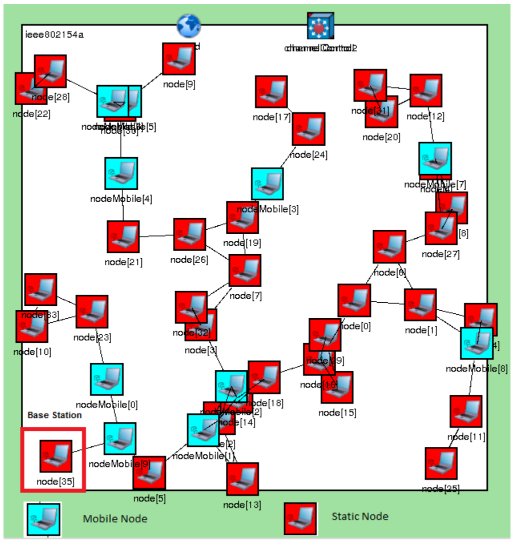


Fig. 7: After assignment of mobile nodes

Once the mobile nodes are placed, the connectivity is established between all the nodes and the base station.The simulation results show that the base station received packets from all the nodes in the network, hence it can be inferred that all the nodes in the network are connected with the base station.

After the assignment of mobile nodes, the throughput has been measured as indicated in Table 1. The data in Table 1 is obtained by varying the delay time. Delay time refers to

TABLE1 : Throughput

Delay Time	Sending Rate(B/s)	Receiving Rate(B/s)	Throughput
0.5s	700	415.5	59.3
1s	350	230.5	65.8
2s	175	147.5	84.2

the time interval between two successive transmissions of a sensor node.

ZigBee uses Carrier Sense Multiple Access (CSMA) to send a packet, so that it doesn't interfere with any of the transmission occurring in the vicinity. This causes hold up of some packets in the network thereby affecting the throughput. The throughput before deploying the mobile nodes is zero. The results in Table 1 show the throughput for different delay times after deploying the mobile nodes.

From the simulation results it is clear that the connectivity is established by optimally assigning mobile nodes in the network.

V. Conclusion and Future Work

The paper addressed the problem of connectivity in unstructured WSNs by using mobile nodes. The network architecture suggested in the paper is suitable for connectivity issues in large area WSNs. The paper uses ZigBee for majority of the communication leading to low power consumption and hence a low cost solution. Using this approach the number of mobile nodes required for connectivity are found to be minimal. The placement of mobile nodes has been optimised using a variant of the Hungarian algorithm and hence saves time. The throughput can be further improved by solving the problem of bottlenecks which arise at some points in the network. The plan for future work includes solving the problem of bottlenecks so as to improve the performance. optimised using a variant of the Hungarian algorithm and hence saves time. Considering large scale WSNs the approach suggested here can find many applications in sensor networks.

The future extension of this work consists of performance improvement in the network by removing bottlenecks and addressing the localisation issues. The throughput can be further improved by solving the problem of bottlenecks which arise at some points in the network. The plan for future work includes solving the problem of bottlenecks so as to improve the performance.

Acknowledgement

We would like to thank Dr Shalu M.A, IIITDM, Kancheepuram, Prof. Abhiram G. Ranade, Department of Computer Science and Engineering, Indian Institute of Technology Bombay and Dr. Saby John, BITS-Pilani Goa Campus for their guidance and suggestions.

References

[1] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," *Computer Networks*, vol. 52, no. 12, pp. 2292 – 2330, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1389128608001254

[2] Encyclopedia of triangle centres. [Online]. Available: http://faculty.evansville.edu/ck6/encyclopedia/ETC.html

[3] M. Di Francesco, S. K. Das, and G. Anastasi, "Data collection in wireless sensor networks with mobile elements: A survey," *ACM Trans. Sen. Netw.*, vol. 8, no. 1, pp. 7:1–7:31, Aug. 2011. [Online]. Available: http://doi.acm.org/10.1145/1993042.1993049

[4] C.-Y. H. Chih-Yung Chang, Chih-Yu Lin and Y.-J. Ho, "Patrolling mechanisms for disconnected targets in wireless mobile data mules networks," in *Parallel Processing (ICPP), 2011 International Conference on*, October 2011, pp. 30 – 41.

[5] R. Sugihara and R. Gupta, "Improving the data delivery latency in sensor networks with controlled mobility," in *Distributed Computing in Sensor Systems*, ser. Lecture Notes in Computer Science, S. Nikoletseas, B. Chlebus, D. Johnson, and B. Krishnamachari, Eds. Springer Berlin Heidelberg, 2008, vol. 5067, pp. 386–399.

[6] W. Seino, T. Yoshihisa, T. Hara, and S. Nishio, "A sensor data collection method with a mobile sink for communication traffic reduction by delivering predicted values," in *Proceedings of the 2012 26th International Conference on Advanced Information Networking and Applications Workshops*, ser. WAINA '12. Washington, DC, USA: IEEE Computer Society, 2012, pp. 613–618. [Online]. Available: http://dx.doi.org/10.1109/WAINA.2012.133

[7] F. Kuhn, T. Moscibroda, and R. Wattenhofer, "Initializing newly deployed ad hoc and sensor networks," in *Proceedings of the 10th annual international conference on Mobile computing and networking*, ser. MobiCom '04. New York, NY, USA: ACM, 2004, pp. 260–274. [Online]. Available: http://doi.acm.org/10.1145/1023720.1023746

[8] Zigbee alliance. [Online]. Available: http://www.zigbee.org

[9] S. H. Yujie Zhang, "The design of network coordinator based on zigbee and gprs technology," in *Proceedings of the 2012 International Conference on Computer Science and Electronics Engineering*, 2012.

[10] W. Z. Hujing, "Design of remote intelligent home system based on zigbee and gprs technology," in *Consumer Electronics, Communications and Networks (CECNet), 2012 2nd International Conference on*, 2012.

[11] H. W. Kuhn, "The hungarian method for the assignment problem," *Naval Research Logistics Quarterly.*, 1955.

[12] T. He, C. Huang, B. M. Blum, J. A. Stankovic, and T. Abdelzaher, "Range-free localization schemes for large scale sensor networks," in *Proceedings of the 9th annual international conference on Mobile computing and networking*, ser. MobiCom '03. New York, NY, USA: ACM, 2003, pp. 81–95. [Online]. Available: http://doi.acm.org/10.1145/938985.938995

[13] A. Bhatia and P. Kaushik, "A cluster based minimum battery cost aodv routing using multipath route for zigbee," pp. 1 –7, dec. 2008.

[14] H. Z. Qinghai Gao, "Improving probabilistic coverage and connectivity in wireless sensor networks: Cooperation and mobility," 2010.

[15] T. Horng-Wen Lee, Meng, "A low power gps receiver architecture," in *Global Telecommunications Conference, 1999. GLOBECOM '99*, 1999.

[16] E. S. Page, "A note on assignment problems," *Computer Journal 6*, p. 241243, 1963.

[17] Omnet++, discrete event simulation system. [Online]. Available: http://www.omnetpp.org

ISBN: 978-1-61208-245-5

A Spectrum Sharing Method Considering Users' Behavior for Uncooperative WiFi/WiMAX Providers

Hiroaki Takemoto*, Keita Kawano†, Kazuhiko Kinoshita‡, Koso Murakami‡
*Graduate School of Information Science, Nara Institute of Science and Technology
8916–5 Takayama, Ikoma, Nara 630–0192, Japan
†Center for Information Technology and Management, Okayama University
3–1–1 Tsushimanaka, Okayama, Okayama 700–8530, Japan
‡Graduate School of Information Science and Technology, Osaka University
1–5 Yamadaoka, Suita, Osaka 565–0871, Japan
E-mail: *hiroaki-ta@is.naist.jp,† keita@cc.okayama-u.ac.jp, ‡{kazuhiko,murakami}@ist.osaka-u.ac.jp

***Abstract*—The number of wireless network users has remarkably grown by recent advances in wireless communication technologies such as WiFi and WiMAX. This has led to a lack of spectrum resources, which has therefore become an important issue. To overcome this problem, spectrum sharing technology, whereby a WiFi system temporarily uses a spectrum band of a WiMAX system, is receiving much attention. Although existing work assumes that the WiMAX and WiFi providers are cooperative, it may not be realistic. In addition, user behavior model is too simple. In this paper, we propose a spectrum sharing method that behaves properly even if the WiMAX provider, WiFi providers, and users are mutually uncooperative. Finally, we confirm the effectiveness of the proposed method by simulation experiments.**

***Keywords*-spectrum sharing; uncooperative providers; satisfaction; WiFi; WiMAX; user model**

I. Introduction

In recent years, users have had access to various wireless systems such as cellular, WiFi, and WiMAX. With increased bandwidth becoming available, multimedia services via wireless networks are now widely used and the traffic demand is increasing.

However, with available spectrum resources being finite, techniques that use wireless resources more effectively should be considered. As one approach to this problem, cognitive radio [1], [2] is receiving attention. Cognitive radio is a technology whereby wireless devices can select between a number of independent wireless systems according to the condition of each system. The frequency spectrum can then be used more efficiently than by using these systems independently.

For cognitive radio, a spectrum-sharing method has been proposed [3], in which a WiFi system temporarily uses a spectrum band of a WiMAX system. In this method, the WiFi access points (APs) that are to use the additional WiMAX channel are decided according to the "load", where the "load" is defined as the total number of users connecting to the WiFi AP. It was confirmed that this method could improve the overall average throughput for the network compared to a method without spectrum sharing.

However, the proposed method assumed that the WiMAX and WiFi providers cooperated to improve the overall average throughput. Therefore, if the WiMAX provider and WiFi providers do not cooperate and pursue only their own interests, this method might not work.

An alternative auction-based method has been proposed [4], which adopts a pricing model for lending channels. This method can behave appropriately even if the WiMAX provider and WiFi providers do not cooperate. However, because this method focuses on the WiMAX provider's profit, the improvement in the average throughput may not be optimal.

To address these problems, we propose a spectrum-sharing method that behaves properly even if the WiMAX provider, WiFi providers, and users are mutually uncooperative. This is an extended version of[5].

The rest of this paper is organized as follows. In Section II, we introduce some existing methods and point out their problems. In Section III, we elaborate our proposed method. Section IV shows the excellent performance of the proposed method by simulation experiments. Finally, Section V makes some conclusions and indicates future work.

II. Related Work

A. Integrated Wireless Networks

Although several wireless systems such as cellular, WiFi, and WiMAX have been developed, each system is used independently. However, if these were to be integrated, users could access services seamlessly. Therefore, WiFi/WiMAX integrated networks [6], [7] have been investigated recently, aiming to improve quality of service (QoS) and load balancing between the WiFi and WiMAX systems by using each system selectively according to the condition of the systems and the demands of user applications. The coverage area for a WiFi AP is about 100 meters, whereas that for a WiMAX base station (BS) is a few kilometers. As shown in Fig. 1,

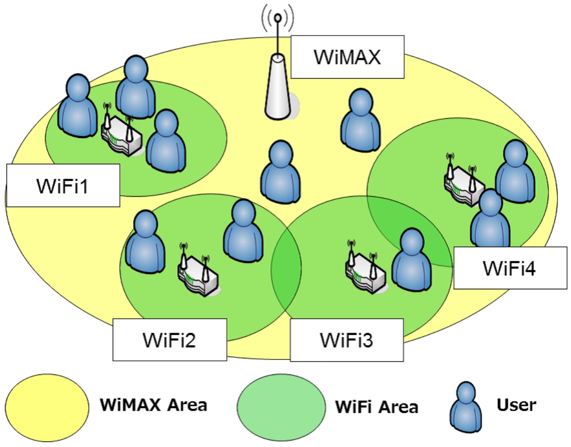


Figure 1. Network Model

this means that two or more WiFi APs may exist inside a WiMAX service area.

As an alternative, a spectrum-sharing method that temporarily assigns a spectrum band of a WiMAX system to WiFi system APs has been proposed. Here, the same spectrum can be used repeatedly without causing interference between adjacent WiFi APs. This enables more efficient utilization of the spectrum, thereby providing users with higher-throughput services.

B. Spectrum-sharing Method to Improve Overall Throughput

In [3], a spectrum-assignment method for improving the overall average throughput in the network was proposed. In this method, the assignment of a WiFi AP to an additional channel in the WiMAX system is decided by using a genetic algorithm (GA). The number of users who connect to the assigned target WiFi AP is used as the evaluation value and the channel assignment is carried out under the constraint that adjacent WiFi APs cannot be assigned the same channel simultaneously. It was confirmed that this method improved the overall average throughput for the network. Moreover, in [8], a spectrum-assignment method that also minimized the difference in throughput between WiFi and WiMAX users was able to provide higher throughput. This method not only improved the overall average throughput but also reduced the coefficient variance.

However, for these methods, the WiMAX system has to lend channels to WiFi APs without itself receiving any direct reward. This implies that either there is one provider of the WiMAX and WiFi services or that the separate providers are prepared to cooperate. In reality, providers do not always cooperate, preferring to pursue their own profit. Therefore, an effective spectrum-sharing method that behaves properly even if the WiFi and WiMAX providers do not cooperate is needed.

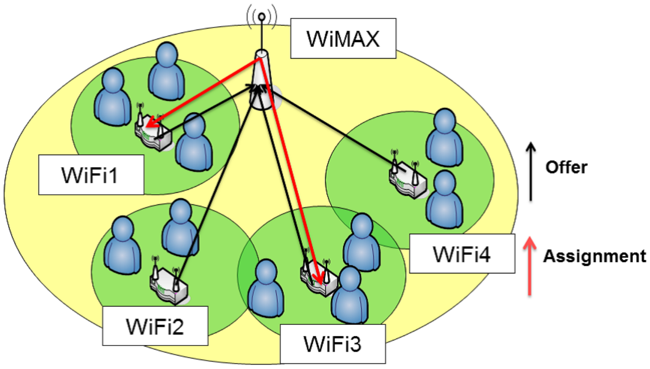


Figure 2. Auction Method

C. Spectrum-sharing Method Based on an Auction

There is a spectrum-sharing method that uses the model of an auction [4]. In this method, if WiFi providers receive an additional channel from the WiMAX provider, they pay for that channel. As shown in Fig. 2, each WiFi provider can make an offer for a channel. By considering these offers, the WiMAX provider selects an assignment pattern that maximizes the WiMAX provider's revenue.

This enables the WiMAX provider to obtain additional profit by lending channels and the WiFi providers to increase their effective bandwidth and user throughput. Furthermore, this method can behave appropriately even if the WiMAX provider and WiFi providers are uncooperative.

However, because this method focuses only on increasing the WiMAX provider's profit, more effective assignment patterns may be overlooked, which implies that the improvement in average throughput is not optimal. Moreover, for some of these proposed methods, the offered prices for an additional channel are decided randomly, which is unrealistic.

III. Proposed Method

A. Summary of the Proposed Method

To overcome the problems described above, we propose a spectrum-sharing method that behaves properly even if the WiMAX provider, WiFi providers, and users are mutually uncooperative. To achieve this, we introduce *satisfaction* as an indicator of users' behavior. In this method, we assume that the user arrival rate at WiFi APs and the WiMAX system varies according to the user satisfaction. Furthermore, given that the WiMAX provider and WiFi providers are not expected to cooperate, the WiFi providers must pay for receiving an additional channel from the WiMAX provider, and must decide for themselves how much they are prepared to pay. The WiMAX provider will select the assignment pattern that maximizes its own profit via a GA method, using the WiFi APs' offered prices and the constraint that

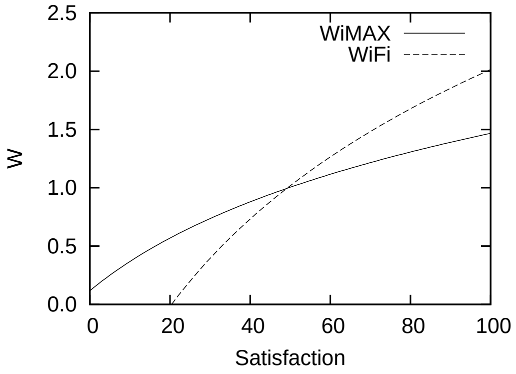


Figure 3. Relation between Satisfaction and Magnification W

adjacent WiFi APs cannot be assigned the same channel simultaneously.

In the following subsections, we first introduce the concept of satisfaction. We then describe how the WiFi APs decide about payments and the algorithm for spectrum assignment.

B. Satisfaction

Satisfaction is associated with each WiMAX and WiFi AP according to the throughput, and may vary from 0 to 100. APs have increased satisfaction as the user throughput increases.

In general, users hope to connect to an AP with as high a throughput as possible. On the other hand, WiMAX and WiFi providers hope to increase the number of connecting users, which increases their profit. Therefore, WiMAX and WiFi providers aiming to increase their profit will improve satisfaction.

The details are as follows. We assume that the user-arrival rate changes as a function of the satisfaction at each AP. Fig. 3 shows the magnification W of the arrival rate as a function of the satisfaction. The arrival rate of APs increases according to the magnification W as the satisfaction increases.

C. Calculation of the WiFi Providers' Offer Price

This subsection explains how WiFi providers calculate the payment for being assigned a channel from the WiMAX provider. As explained in Section III-A, the WiFi providers decide for themselves how much they are prepared to pay. For the WiFi providers, increasing the number of connected users leads to increased profit. Therefore, WiFi providers need to raise their user satisfaction. If WiFi providers receive an additional channel from the WiMAX system, the throughput and the satisfaction will both increase, and an increase in revenue would therefore be expected.

However, WiFi providers must pay the WiMAX provider for the assigned channels. Therefore, WiFi providers must consider both revenue and payment in deciding whether to borrow a channel. If a WiFi provider wants to borrow a single channel, they can calculate the appropriate payment using Eq. (1). For multiple channels, the same calculation will be repeated.

$$P_F * E_i * \alpha, \tag{1}$$

where P_F, E_i and α refer to the price that WiFi providers impose on users, the estimated number of increased users, and the expected profit to the WiFi provider, respectively. E_i is calculated by Eq. (2), assuming that the satisfaction changes from s to s'.

$$E_i = \lambda_F * T * (W_F(s') - W_F(s)), \tag{2}$$

where λ_F and T are the arrival rate at the WiFi AP, and the interval times for spectrum assignment, respectively. $W_F(s)$ is the magnification of the arrival rate, as shown in Fig. 3. Therefore, $W_F(s') - W_F(s)$ indicates by how much the number of users connecting to the WiFi AP will increase by borrowing additional channels. Whenever a WiFi provider judges, using this equation, that it can increase its revenue, it offers the calculated payment for borrowing additional channels. Otherwise, it does not seek to borrow any additional channels.

D. Procedure for Channel Assignment

We now explain the procedure for channel assignment. The WiMAX provider decides on the number of assignment channels and the APs of the assignment targets according to the WiFi APs' offer prices described in Section III-C. However, the assignment of channels to WiFi providers will cause the throughput of the WiMAX provider to decrease. Therefore, the WiMAX provider's satisfaction will change from s to s', which causes a decrease in both arrival rate and revenue. Because of this, the WiMAX provider should perform channel assignment by considering the difference between the revenue decline and the payment from the WiFi APs in terms of Eq. (3).

$$\textit{Estimated Decreased Revenue} = P_M * E_d, \tag{3}$$

where E_d and P_M refer to the estimated number of decreased users and the price that the WiMAX provider imposes on users, respectively. E_d is calculated by Eq. (4).

$$E_d = \lambda_M * T * (W_M(s) - W_M(s')), \tag{4}$$

where λ_M is the arrival rate for the WiMAX system. $W_M(s) - W_M(s')$ indicates the decrease in the number of users connecting to the WiMAX system because of the decrease in available channels.

The APs of the assignment targets and the number of assignment channels are decided according to the following steps.

1) WiFi APs calculate the payment required to borrow channels.
2) WiMAX provider selects the assignment pattern that maximizes the sum of payments offered by WiFi APs.
3) WiMAX provider calculates the estimated revenue decrease from lending channels.
4) If the revenue from lending channels exceeds the estimated revenue decrease, then perform the channel assignment.
5) Repeat Steps 1 to 4 until all channels are lent or the estimated revenue decrease exceeds the revenue from the target WiFi APs.

Note that, in Step 2, we use the algorithm proposed in [8].

IV. Performance Evaluation

A. Simulation Model

In this subsection, we evaluate the performance of the proposed method by simulation experiments. The network model assumed that WiFi and WiMAX were able to carry out spectrum sharing. There was one WiMAX BS, whose area was divided into $10 \times 10 = 100$ small areas. 50 small areas were selected at random, each having a WiFi AP. The spectrum bandwidth for the WiMAX system was set to 100[MHz] and divided into channels of 20[MHz] each. Each WiFi AP can use one or more additional channels assigned from the WiMAX system. The WiMAX system was assumed to provide 40[Mbps] per channel in accordance with the WiMAX Forum [9]. The WiFi systems were assumed to provide 17.5[Mbps] per channel according to preliminary experiments that used the ns2 discrete-event simulator [10].

In addition, the interval time T for spectrum assignment was set to 300[sec]. The price that the WiMAX provider imposes on its users was 100, and the price that WiFi providers impose on their users was set to 80. The WiFi APs' price was lower because the coverage for WiFi is narrower than that for WiMAX. We considered the example of a user downloading a 10 MByte file.

In this simulation, calls occured according to a Poisson arrival process, with each area having its own arrival rate. Because WiFi APs tend to be set up in places where people gather, such as offices, rail stations, and cafes, the call arrival rate for a WiFi AP was assumed to be x times that for WiMAX. We defined the arrival rate for the whole network as λ_{all}. The initial arrival rate per WiFi AP area (λ'_F) and that for WiMAX (λ'_M) were then calculated by Eq. 5 and Eq. 6.

$$\lambda'_F = \lambda_{all} * \frac{x}{50(x+1)} \tag{5}$$

$$\lambda'_M = \lambda_{all} * \frac{1}{100(x+1)} \tag{6}$$

In this simulation, we set $x = 3$.

Now, as described in Section III-B, the arrival rate would change according to the magnification W of the initial arrival rate. Therefore, the actual arrival rates for WiMAX (λ_M) and per WiFi AP area (λ_F) satisfied the following equations.

$$\lambda_M = \lambda'_M * W_M \tag{7}$$

$$\lambda_F = \lambda'_F * W_F \tag{8}$$

It followed that the arrival rates λ_1 (without WiFi AP) and λ_2 (with WiFi AP) were calculated by the following equations.

$$\lambda_1 = \lambda_M \tag{9}$$

$$\lambda_2 = \lambda_F + \lambda_M \tag{10}$$

If a new user arrived in an area with a WiFi AP, the system with the higher throughput was used. Otherwise, the WiMAX BS was used. In addition, users would stay in the arrival area until the end of their download.

We chose two other methods for comparison. One method was the spectrum-sharing method described in Section II-B, which improved the overall average throughput. In this paper, we call this method the “existing method”. The other method did not share any spectrum. As performance measures, we observed the average time to complete the download (download time) and the revenue for the WiMAX and WiFi providers.

Under the same conditions as described above, we ran five simulations with various initial overall arrival rates λ_{all}. The simulation ended after 250,000 calls were completed. We set the parameter α to 0.1. $W_F(s)$ and $W_M(s)$ are logarithmic functions of the satisfaction s. As explained above, with channel assignment, the throughput changes and the satisfaction of the WiMAX and WiFi providers changes from s to s', where s' is calculated from the following equations.

$$s'(WiFi) = s + 190/s \tag{11}$$

$$s'(WiMAX) = s - 25/s \tag{12}$$

Note that, we assume that the satisfaction s is also a logarithmic function of throughput (Tp), as defined in Eq. 13.

$$s = 70 * log((Tp + 5)/5) \tag{13}$$

B. Simulation Results

Fig. 4 shows the average download times as a function of the initial overall arrival rate.

This indicates that both of the existing method and the proposed method improve the average overall throughput, since the capacity of the system increased by assigning one or more channels from the WiMAX BS to several WiFi APs.

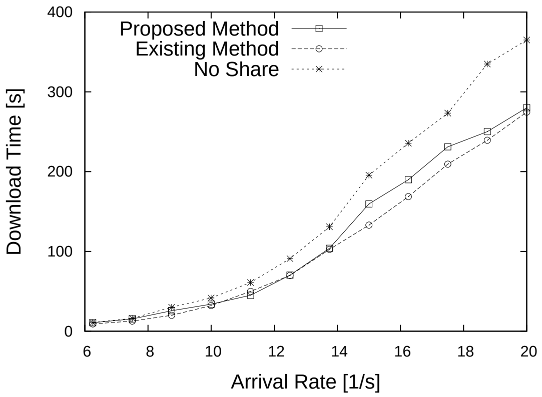


Figure 4. Mean Download Time with Variable Arrival Rate

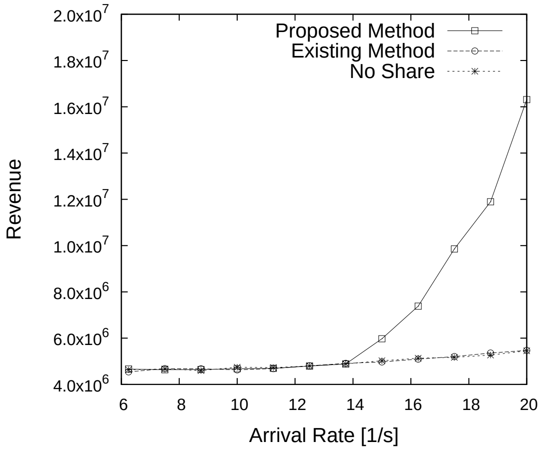


Figure 5. Overall Revenue with Variable Arrival Rate

The average throughput of the proposed method is almost equal to that of the existing method. Therefore, this means that our proposed method behaves properly even if the WiMAX provider, WiFi providers and users are mutually uncooperative.

Figs. 5-7 show the sum of the revenues of the WiMAX provider and WiFi providers, the WiMAX provider's revenue, and the WiFi providers' revenue as a function of the initial overall arrival rate, respectively. There is not so much difference between the overall revenue of the existing method and that of the proposed method, when λ_{all} is up to 13.75. However, the overall revenue in the proposed method becomes much higher when λ_{all} is bigger than 15.0. Moreover, Fig. 5 and Fig. 6 indicate that the revenues of the WiMAX provider and that of the WiFi providers are increasing. This is because the proposed method considers the change of the revenue adequately to assign an additional channel assignment based on satisfaction.

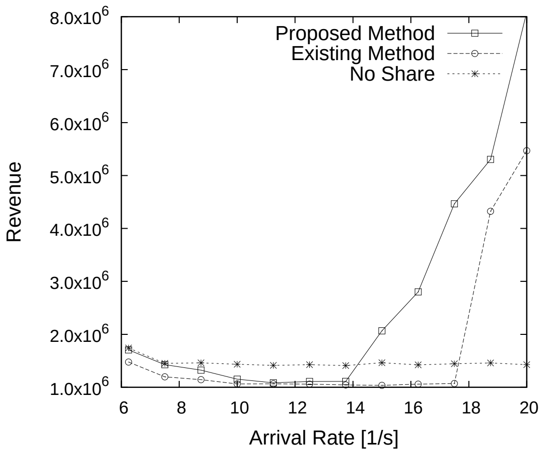


Figure 6. WiMAX Provider's Revenue with Variable Arrival Rate

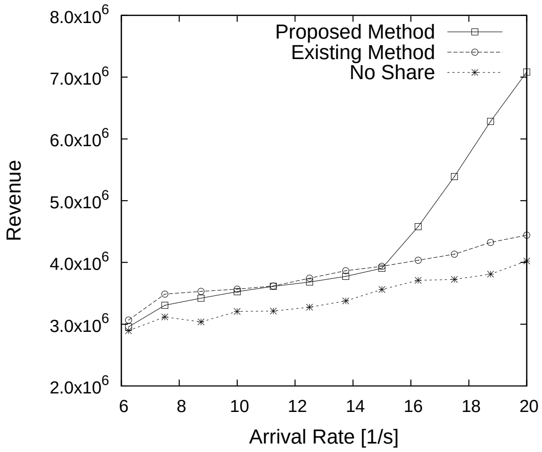


Figure 7. WiFi Providers' Revenue with Variable Arrival Rate

However, from Fig. 6, we can see that the revenue of the WiMAX provider is smaller than that of the non-sharing method when the arrival rate is small. In this simulation, when a new user arrives at an area with WiFi AP, he/she chooses a system with higher throughput. In the methods with spectrum sharing, the throughput of the WiMAX users decreases because of the decrease in its available channels. This might cause the situation that more users connect to a WiFi AP rather than WiMAX BS. The proposed method avoids this situation by consideration of the satisfaction. Consequently, when λ_{all} is large, the revenue for the WiMAX provider in the proposed method is higher than that in the existing method.

These results indicate that the proposed method achieves well spectrum sharing even if the WiMAX provider, WiFi providers, and users are mutually uncooperative.

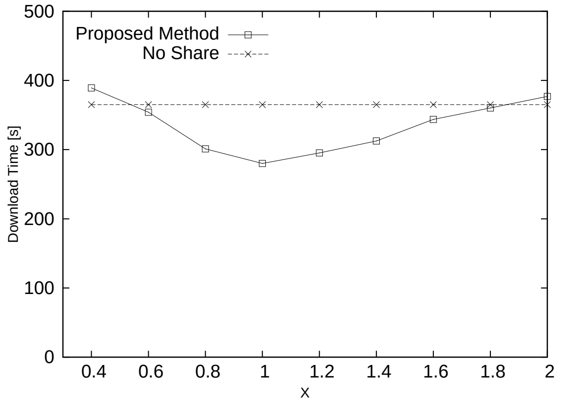


Figure 8. WiMAX: Mean Download Time with Variable X ($\lambda = 20$)

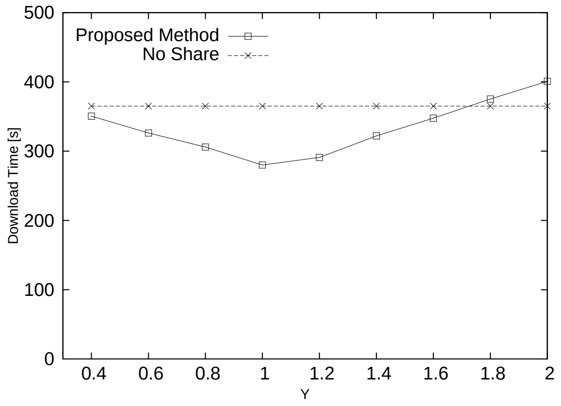


Figure 9. WiFi: Mean Download Time with Variable Y ($\lambda = 20$)

C. Robustness

In this simulation, we assumed that WiMAX and WiFi providers exactly knew the relationship between satisfaction and the arrival rate. Therefore they can expect the change of the number of uses properly. However, this is difficult in fact.

To confirm this robustness, we verified the case where WiMAX (WiFi) provider misestimated the change of users X (Y) times more than the actual value. In other words, if X is 1.0, WiMAX provider can expect the change properly.

Figs. 8 and 9 show the average download times as a function of the parameter X (or Y) compared with the no spectrum assignment method. We set $\lambda_{all} = 20$. Fig. 8 indicates that even if WiMAX provider expects about 50% more or less, the proposed method can improve the average throughput. Similarly, when the differece between what WiFi providers expect and the actual change is less than 60%, the improvemnet of the average throughput is achieved.

V. Conclusion and Future Works

In this paper, we proposed a new spectrum sharing method that works well even if WiMAX provider, WiFi providers, and users are mutually uncooperative. It introduces the satisfaction as an indicator of users' behavior. Furthermore, it was confirmed that the proposed method could keep the average throughput compared with the existing method and improve the revenues.

In a future work, we evaluate the sensitivity of each parameter.

Acknowledgement

This work is partly supported by JSPS KAKENHI Grant Number 21760290.

References

[1] J. Mitola III and G. Q. Maguire, "Cognitive Radio for Flexible Mobile Multimedia Communications," IEEE International Workshop on Mobile Multimedia Communications, pp. 3–10, Nov. 1999.

[2] I. Akyildiz, W. Lee, M. Vuran, and S. Mohanty "Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey," Computer Networks, vol. 50, pp. 2127–2159, Sep. 2006.

[3] M. Nakagawa, K. Kawano, K. Kinoshita, and K. Murakami, "A Spectrum Assignment Method Based on Genetic Algorithm in WiFi/WiMAX Integrated Networks," 5th ACM International Conference on Emerging Networking EXperiments and Technologies (CoNEXT2009), Dec. 2009.

[4] G. Losifidis and L. Koutsopoulos, "Challenges in Auction Theory Driven Spectrum Management," IEEE Communications Magazine, Vol. 49, No. 8, pp. 128–135, Aug. 2011.

[5] H. Takemoto, K. Kawano, K. Kinoshita, and K. Murakami, "A Study on Spectrum Sharing between Unncoperative WiFi/WiMAX Providers," IEICE Tehnical Report, NS2012-100, pp. 113–118, Oct. 2012. (in Japanese)

[6] Y. Chen, J. Hsia, and Y. Liao, "Advanced Seamless Vertical Handoff Architecture for WiMAX and WiFi Heterogeneous Networks with QoS Guarantees," Computer Communications. Vol. 32, Issue 2, pp. 281–293, Feb. 2009.

[7] L. Berlemann, C. Hoymann, G. R. Hiertz, S. Mangold, "Coexistence and Interworking of IEEE 802.16 and IEEE 802.11(e)," Vehicular Technology Conference, vol. 1, pp. 27–31, May. 2006.

[8] K. Kinoshita, M. Nakagawa, K. Kawano, and K. Murakami, "A Fair and Efficient Spectrum Assignment for WiFi/WIMAX Integrated Networks," 6th International Conference on Systems and Networks Communications (ICSNC2011), pp. 117–121, Oct. 2011.

[9] WiMAX Forum, http://www.wimaxforum.org/. (retrieved: Dec. 2012)

[10] ns-2, The Network Simulator, http://www.isi.edu/nsnam/ns/. (retrieved: Dec. 2012)

Optimization of Wireless Sensor Network Using Network Coding Algorithm

Jéssica Bartholdy Sanson
GPSCOM-Department of Electronics and Computer Science
Federal University of Santa Maria
Santa Maria, Brasil
jessikbs.37@hotmail.com

Natanael R. Gomes, Renato Machado
Andrei P. Legg and Osmar M. dos Santos
GPSCOM-Department of Electronics and Computer Science
Federal University of Santa Maria
Santa Maria, Brasil
natanael.rgomes@gmail.com, renatomachado@ufsm.br
andrei.legg@gmail.com and osmar@inf.ufsm.br

***Abstract*—A wireless sensor network is a kind of *ad hoc* network that can be used to monitor a variety of environmental characteristics aiming, for instance, military purposes, environmental purposes, etc. A typical application of these networks is to collect and send historical information from all network sensor nodes to the base station. In this work we propose a new network coding technique and evaluate its performance in terms of time and energy saving. We show that the proposed technique has considerably improvements on sensor networks with small number of nodes.**

Keywords- ***Computer Modeling, Mobile Communication Systems, Channel of Communication.***

I. Introduction

Wireless Sensor Networks (WSN) consist of sensor nodes responsible for sensing tasks in a distributed way, according to the client applications. These networks act as data acquisition systems environment, allowing monitor physical or environmental phenomena, such as temperature, sound, pressure and vibration. The sensor nodes collect data and forward them to an exit point of the network, called the sink node, destination node, or base station (depending on the application), to be analyzed and processed.

In conventional sensor networks, the propagation of information is traditionally performed by a method called routing, where the information is stored by the intermediary nodes and then forwarded until it reaches its destination. It was believed, until a few years ago, that the information processing in the intermediary nodes do not bring any benefits in the replication and dissemination of data. However, in [4], Ahlswede, Cai, Li & Yeung demonstrated that applying such a processing it is possible to achieve higher data throughput. The processing of information in intermediary nodes is denominated network coding.

The topology of wireless sensor networks consists of multiple source nodes and a sink node – base station. This originates problems such as data congestion and limited resources. Consequently, it is important to apply techniques for data reduction so that fewer bits are transmitted into the wireless channel [1]. Network coding is a technique that can be used for this. The technique combines algebraically by using "exclusive or" operation (denoted by the symbol " $\oplus$" or by the word "XOR"), over the received packets [2]. This strategy reduces the traffic of packets in the network communication channels and, consequently, the data congestion. Also, the capacity and transmission speed of the network are increased without need of complex routing algorithms.

The research which is considered the work demonstrates that the problem of transmission of wireless sensor networks, described above can be softed through network coding, remains to know whether in practice this technique has the same efficiency that achieved by simulation

This paper is organized as follows. The next section provides a brief discussion of the network coding technique. Section III presents the methodology used, defines and evaluates the technique proposed in this paper. For comparison purposes, Section IV describes a previous technique found in the literature. The technique proposed in this paper is shown in Section V. Results are presented in Section VI, which is followed by the conclusions.

II. Network Coding

Some of the advantages of network coding were introduced in terms of flow in a butterfly network [1]. This kind of network represents a communication network as a directed graph in which the vertices correspond to the network nodes (terminals) and the edges represent the channels as shown in Fig. 1. The network is composed of two source nodes (A and B) and two destination nodes (R_1 and R_2). It is assumed that the sources A and B can only send one bit at each time interval. Hence, it would take more than one time interval to transmit a bit from "A" and "B" to nodes R_1 e R_2. In contrast, by using network coding, there is the possibility of processing the bits at the intermediary node "X". Such processing enables the reception of bits from nodes "A" and "B" in one time interval. The node R_1, which receives the bits from node A and $A \oplus B$, obtains the bit from node B by calculating $A \oplus (A \oplus B) = B$. Similarly, the node R_2 is also able to decode the bit information from nodes A and B [3]. Hence, there is a benefit in terms of throughput when the processing of information is allowed at the intermediary network nodes, therefore, justifying the use of network coding.

ISBN: 978-1-61208-245-5

Practical implementation of network coding is fully described in [6], [7].

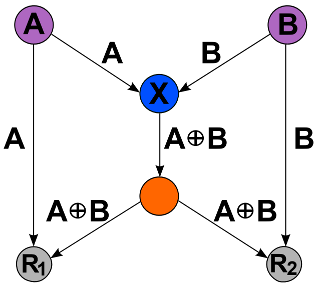


Fig. 1. Representation of the Butterfly Network

III. Methodology

The encoding process described in the previous section is employed in this work to obtain a higher transmission data rate in wireless sensor networks. The direction of transmission is from sensors towards the base station (receiver). The nodes of the network are deployed in a systematic way aiming an efficient coding of information by the intermediate nodes called encoders. We have considered two different network sizes: (A) the first one, denominated network type A, has a smaller number of nodes (around 20 nodes); (B) the second one, denominated network type B, has a large number of nodes (40 nodes). These two different sizes of network allow us to evaluate, by means of time delay comparison, the throughput of the network.

In this work, we assume that, in addition to sensor nodes, the networks are formed by intermediary nodes called “relays”, which relay the information from other nodes toward the base station. Some “relays” perform the “XOR” operation on received data and, therefore, are called encoders.

Simulations were implemented and performed on a Mathematical Software by using normalized transmission rates and data frames. A representation of the network is illustrated in Fig. 2.

In this work we simulated two coding techniques. Technique 1 was previously proposed by [1]. It employs network coding and is efficient for sensor networks with a large number of nodes. Nevertheless, its performance is similar to conventional wireless sensor networks (no network coding) for a network with a smaller number of nodes. This technique is described in the next section. In order to improve the performance of this technique, specially for smaller networks, this paper proposes Technique 2, which is presented in Section V.

IV. Technique 1

A. System Model

The sensor nodes collect data from the environment and transmit them to relay nodes. At the relay nodes the data is evaluated in order to verify the need to perform network coding. The data is then forwarded to the receiving node.

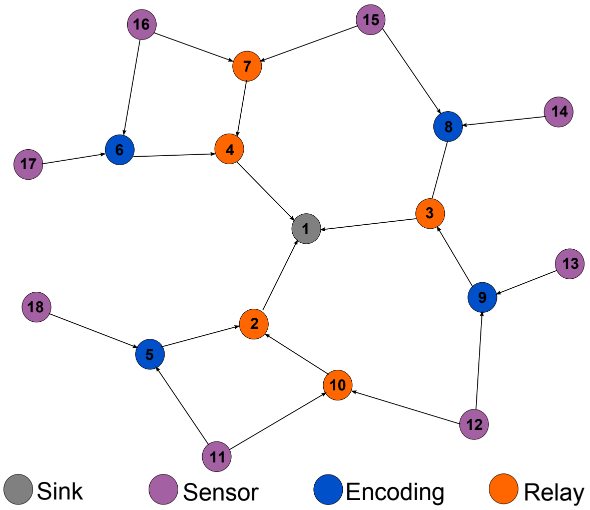


Fig. 2. Representation of Sensor Network

The necessity to perform network coding is evaluated by using function f, which is defined as follow: Let p and q be the two data received by a encoder node and f a binomial function that calculates the significant difference between two data packets and returns true or false. If p and q do not differ more than a threshold γ, then the value of the function f is false (0), otherwise it returns true (1). The absolute value of difference is denoted by [5]:

$$d = p - q; \qquad f : \begin{cases} 0, & \text{d} < \gamma; \\ 1, & \text{otherwise.} \end{cases} \tag{1}$$

B. Description

Assume that sensor nodes 17 and 16 have some information to send to the network, as shown in Fig. 2. Node 16 sends information to node 7 and to the node 6 (relay encoder). Sensor node 17 transmits its information only to node 6. Node 7 only transmits the information towards the receiving node. However, encoder node 6 has two packets to transmit, then it employs function f. If function f returns true, node 6 encodes the packets using the XOR operation and forwards the result towards node 4. Now node 4 has two packets: one is a data packet transmitted by node 7 and the other one is a coded packet transmitted by node 6. Node 4 transmits a packet at each time interval, and the same procedure is adopted to the others non-coding relay nodes when there is more than one packet to retransmit. When these packets reach the receive node, they decode a packet at a time using XOR operation. In this study it was considered that the difference between two data packets is always greater than the threshold γ. In other words, a pair of packets is always encoded by an encoder node.

C. Algorithm

When an encoder node has to encode two packets, for instance, pkt1 and pkt2, the following procedure is adopted: XOR operation is performed on these two packets and the result is encapsulated in a new packet, which is routed forward to the receiver. The receiver decodes the data by using XOR operation on the appropriate packets. When an encoder node receives data from sensor nodes, it calculates the difference between the data using function f. If the difference is less than γ, then there is no need for coding and a randomly selected data is routed towards the receiving node. If the difference is greater than γ then the encoder node encodes and forwards the data to the receiver. A bit is used to confirm if encoding will be performed on the data. The algorithm that was implemented in the coding nodes in [4] is described in Algorithm 1.

```
Algorithm 1: Technique 1(packet pkt1, packet pkt2)
//pkt1i and pkt2i is ith packet sent by leaf node 1 and
leaf node 2 respectively.
{
If f(data(pkt1i), data(pkt2i)) == 0
   {
     Select either packet and transmit
   }
Else
   {
     Perform network coding on pkt1 and pkt2
     Transmit data obtained by encoding in previous
step
   }
}
//End of Algorithm
```

V. Technique 2

A. System Model

In our proposed technique (Technique 2), each encoder node receives and processes data from a pair of sensor nodes. Moreover, only a single sensor node from that pair sends information to a relay node, as shown in 3. The information of this sensor (marked in Fig.3) is used to decode the data across the network. This technique does not use function f, and thus the nodes encoders always perform encoding.

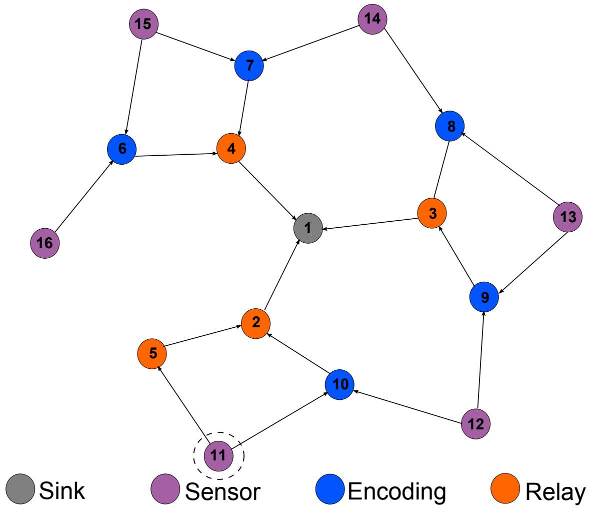


Fig. 3. Representation of Sensor Networks using Technique 2

B. Description

Observing Fig. 3, it is assumed that sensor nodes 16 and 15 have some information to send through the network. Node 16 transmits the information to the encoder node 6, while sensor node 15 transmits its information to the encoders nodes 6 and 7. Encoder node 6 performs XOR operation on data received from sensors 16 and 15. Node 7 will encode the data received from sensors 15 and 14. The other subsequent encoder nodes execute similar procedure until sensor 11 is reached, which sends its data to the encoder node 10 and to relay node 5. The information sent to relay node 5 will be used to decode the entire network, by using exclusive OR operation among successive pairs of data in the receiver.

C. Algorithm

An encoding node in Technique 2 (proposed in this paper), as well as in Technique 1, has to encode a pair of packets, namely, pkt1 and pkt2. This is done as following: XOR operation is executed on pkt1 and pkt2, the result is encapsulated in a new packet that is routed forward. The decoding procedure is accomplished by using XOR operation on the appropriate packet.

It is relevant to notice that in this technique, every encoder node always applies XOR operation on two received packets. The encoder node does not evaluate the difference between a pair of packets by employing function f. Due to this fact, the algorithm becomes simpler, consisting only of a XOR operation on a pair of data packets. Algorithm 2 presents the proposed algorithm for the encoding procedure.

```
Algorithm 2: Technical 2(packet pkt1, packet pkt2)
//pkt1i and pkt2i is sent by leaf node 1 and leaf node
2 respectively.
  {
    Perform network coding on pkt1 and pkt2
    Transmit data obtained by encoding in previous step
  }
//End of Algorithm
```

VI. Results

Table 1 shows the results for three types of wireless sensor networks: (1) network with no encoding, (2) considering the Technique 1 and (3) considering the proposed technique (Technique 2). For data analysis, the following definitions are used: *cycle* is the time interval in which the information from all sensors arrives at the receiver node; *transmission delay* is the time required for data being exchanged between two nodes; NN is the number of network nodes; B is the amount of data arriving at the receiver in a time interval necessary for the network with no encoding to complete a cycle; TD represents the transmission delay needed to complete one network cycle; NT is the number of transmissions required to complete a cycle; R is the transmission rate given in bits/nodes/time.

Technique 2, which is proposed in this work, is more effective for a network size of type A. In other words, it has a higher transmission rate, requires a smaller number

TABLE I
COMPARISON RESULTS

Network Type A	NN	B	TD	NT	R
Without coding	18	8	5	24	0.08889
Technique 1	18	8	5	24	0.08889
Technique 2	16	9	4	18	0.11250
Network Type B	NN	B	AT	NT	R
Without coding	38	16	8	64	0.05263
Technique 1	38	19	9	64	0.0625
Technique 2	34	15	7	48	0.05514

of transmissions, and presents a smaller number of nodes in the network. Hence, Technique 2 leads to a saving of time and energy in the transmission of information. Technique 1 does not provide advantages when a network of type A is considered. Considering a network of type B, we see that the Technique 2 has a loss of performance, but still performs better than sensor networks with no coding. However, Technique 1 begins to have a better performance then the others techniques. In this scenario, Technique 1 presents a higher transmission rate and a decreasing in the number of network transmissions related to the number of data information arriving at the base station. Its performance increases as the network size increases, being a good technique for large networks.

VII. CONCLUSION AND FUTURE WORK

In this paper, we compared and analyzed two robust implementation of network coding for transmission in sensor networks, denominated Technique 1 and Technique 2. Technique 1 was already discussed in [4]. Technique 2 is proposed in this work, which allowed a better network performance and a reduction in the number of messages transmitted on the network. It was observed that the network coding applied to wireless network, considering both methods, has advantages over the conventional transmission techniques (no encoding techniques).

Technique 2 presented a performance improvement on small sensor networks. The proposed technique allows data compression and concatenation of the data path to the base station in some aggregation points. However, as the network size increases, Technique 2 becomes less efficient. In contrast, Technique 1, which in small sensor networks has no advantages, has its performance improved as the network size increases. This is due to the number of collisions on the network, that is, in a network with no encoding there is a higher number of collisions when compared to an encoded network.

For the next steps of the research is necessary to create a protocol that can be implemented in practice in wireless sensor networks, in order to have a more precise calculation of the gain technique, and variables to calculate performance as due to processing delays obtained protocol, bits of overhead, packet loss and others.

ACKNOWLEDGEMENT

This work was supported, in part, by the Brazilian Research Support Foundation of the State of Rio Grande do Sul (FAPERGS)

REFERENCES

[1] N. Jain, S. Sharma, and S. Sahuv, “Efficient flooding for a large sensor networks using network coding,” *International Journal of Computer Applications*, v. 30, no. 9, pp. 1-4, September 2011.

[2] J. L. Rebelatto, B. F. Uchôa-Filho, Y. Li, and B. Vucetic, “Adaptive distributed network-channel coding,” *IEEE Transactions on Wireless Communications*, vol. 10, no. 9, pp. 2818-2822, September 2011.

[3] R. W. Nóbrega, B. F. Uchôa-Filho, “Multishot Codes for Network Coding: Bounds and a Multilevel Construction,” *in Proc. IEEE International Symposium on Information Theory (ISIT’09)*, Seul, South Corea, June 28-July 3, 2009, pp. 428-432.

[4] R. Ahlswede, S. Y. R. Li, and R.W. Yeung, “Network information flow,” *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1204-1216, July, 2000.

[5] R. R. Rout, S.K. Ghosh, and S. Chakrabarti, “Network Coding-aware data aggregation for a distributed wireless sensor network,” *in. Proc. IEEE International Conference on Industrial and Information Systems (ICIIS)*, Sri Lanka, Dec. 28-31, 2009, pp. 32-36.

[6] P. A. Chou, Y. Wu, and K. Jain, “Practical network coding,” *in Proc. Conf. Comm., Control and Comp.*, Oct. 2003.

[7] Qunfeng Dong, Jianming Wu, Wenjun Hu, and Jon Crowcroft, “Practical Network Coding in Wireless Networks,” *in Proceedings of the 13th Annual ACM International Conference on Mobile Computing and Networking*, Montréal, QC, Canada, Sept. 9-14, 2007, pp. 306-309.

Blind Estimation of Frame Synchronization Patterns in Telemetry Signal

Byunghoon Oh, Jinwoo Jeong, Yeonsoo Jang, and Dongweon Yoon
Department of Electronic Engineering
Hanyang University
Seoul, Korea
elfinston5@hanyang.ac.kr, jhjeong@hanyang.ac.kr, ysjang83@hanyang.ac.kr and dwyoon@hanyang.ac.kr

***Abstract*— Telemetry signals are widely used in order to control electric equipment, collect information for launching satellites and for testing aircraft and guided weapons. In non-cooperative contexts, a blind estimation algorithm is essential to acquire data from telemetry signals. Generally, telemetry signals have a frame structure which consists of data words and a unique frame synchronization pattern. To obtain the telemetry signals blindly, we have to estimate the synchronization patterns in the frame structure first. In this paper, we propose a new algorithm for the estimation of frame synchronization patterns for the blind detection of telemetry signals and verify the proposed algorithm through computer simulations. By exploiting this algorithm, structures of unknown telemetry signals can be reconstructed.**

Keywords- detection and estimation; telemetry signal; frame synchronization.

I. INTRODUCTION

Telemetry signals containing data measurements made at distance [1] are commonly used in vehicles such as aircraft, missiles, automobiles, and satellites, and applied in various fields including space exploration, flight control, and traffic control systems [2]. For example, National Aeronautics and Space Administration (NASA), European Space Agency (ESA) and other space agencies have used telemetry systems for the collection of data from spacecraft and satellites since telemetry signals can contain various measurements such as temperature, pressure and a radian level. Telemetry signals have a frame structure for transmitting each measurement as a single stream. Generally, the frame structure consists of data words and a unique synchronization pattern. Data words are composed of measurements, counters, commands, or other information. The synchronization pattern is a unique sequence for identifying each minor frame. To acquire the telemetry signals blindly, we first have to estimate the frame synchronization patterns in the frame structure.

In this paper, we propose an algorithm for the estimation of frame synchronization patterns for the blind detection of telemetry signals, and verify the proposed algorithm through computer simulations. With this algorithm, unknown telemetry signals can be reconstructed.

II. FRAME STRUCTURE OF THE TELEMETRY SYSTEM

There are various telemetry standards, such as IRIG 106, CCSDS 102.0-B-4, and PSS-04-106 [3-5]. Among these the standards, IRIG 106 is widely used in both military and commercial fields because it is made to ensure the interoperability in aeronautical telemetry application from Range Commanders Council [6]. In this paper we consider IRIG 106 as a telemetry standard in this paper.

Telemetry signals can be classified by the method of modulation: Pulse Code Modulation (PCM), Frequency Modulation (FM), and Pulse Amplitude Modulation (PAM). Among these schemes, PCM is widely used in telemetry systems because it has the advantages of better accuracy, greater dynamic range, and less noise than other schemes. In Chapter 4 of IRIG 106 standard [3], the PCM formats are divided into two classes, Class 1 and Class 2. Table 1 shows the specifications of the two classes, and Figure 1 depicts the frame structure of Class 1 [3].

TABLE I. SPECIFICATION OF CLASS 1 AND CLASS 2 IN IRIG 106

	Class 1	Class 2
Format change	X	O
Word length	4~32 bits	4~64 bits
Max. minor frame length	8192bits & 1024 words	16384 bits & 1024 words
Max. major frame length	256 minor frames	
Min. frame synchronization	16~33 consecutive bits	

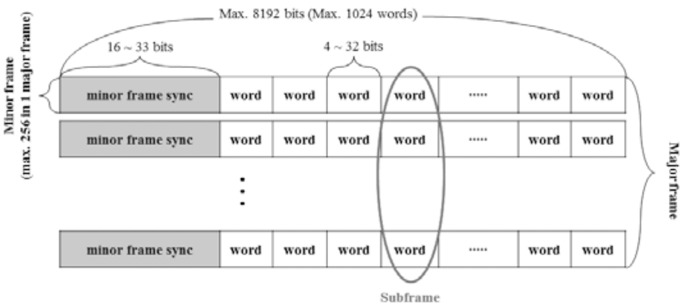

Max. 8192 bits (Max. 1024 words)
16 ~ 33 bits
4 ~ 32 bits
Minor frame (max. 256 in 1 major frame)
minor frame sync
word
·····
Major frame
Subframe

Figure 1. Frame structure of IRIG Class 1

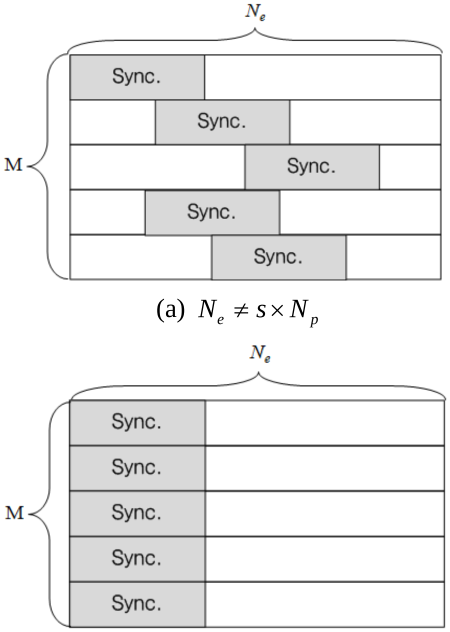
As shown in Figure 1, the major frame structure of IRIG 106 Class 1 consists of up to 256 minor frames. Each minor frame has the maximum length of 8192 bits and 1024 words, and each word consists of 4 to 32 bits. Note that there is a minor frame synchronization(sync) pattern having 16 to 33 bits at the head or tail of each minor frame, and each minor frame sync pattern is also treated as a word in a minor frame. Table 2 shows the minor frame sync patterns of IRIG 106 Class 1. In a major frame, each minor frame has the same sync pattern.

TABLE II. MINOR FRAME SYNC PATTERNS OF IRIG 106 CLASS 1

Length	Sync Patterns
16	111 010 111 001 000 0
17	111 100 110 101 000 00
18	111 100 110 101 000 000
19	111 110 011 001 010 000 0
20	111 011 011 110 001 000 00
21	111 011 101 001 011 000 000
22	111 100 110 110 101 000 000 0
23	111 101 011 100 110 100 000 00
24	111 110 101 111 001 100 100 000
25	111 110 010 110 111 000 100 000 0
26	111 110 100 110 101 100 110 000 00
27	111 110 101 101 001 100 110 000 000
28	111 101 011 110 010 110 011 000 000 0
29	111 101 011 110 011 001 101 000 000 00
30	111 110 101 111 001 100 110 100 000 000
31	111 111 100 110 111 110 101 000 010 000 0
32	111 111 100 110 101 100 101 000 010 000 00
33	111 110 111 010 011 101 001 010 010 011 000

III. BLIND ESTIMATION OF FRAME SYNCHRONIZATION PATTERNS IN TELEMETRY SIGNALS

The length of minor frame in IRIG 106 Class 1 is up to 8192 bits, and one minor frame can have up to 1024 words with the same sync pattern at the same position. If we divide the unknown telemetry data sequence into data blocks having the estimated minor frame length of N_e and load the data blocks to a matrix $H(M,N_e)$ row by row, then we can make the matrix for the analysis of minor frame sync pattern as shown in Figure 2, where N_e is an arbitrarily chosen minor frame length, N_p is the original minor frame length, M is the numbers of rows in the matrix $H(M,N_e)$, and s is an integer.



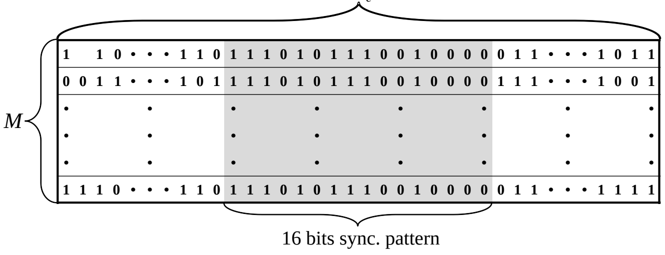
(a) $N_e \neq s \times N_p$

(b) $N_e = s \times N_p$

Figure 2. Matrix $H(M,N_e)$ for the estimation of minor frame sync pattern

In Figure 2, each matrix consists of sync patterns and words, where shaded areas represent sync patterns, and plain areas represent words. As shown in Figure 2 (a), when the arbitrarily chosen minor frame length N_e is not an integer multiple of the original minor frame length N_p, the sync patterns are not aligned in the same column in the matrix $H(M,N_e)$. On the other hand, if N_e is an integer multiple of N_p, as shown in Figure 2 (b), the sync patterns are aligned in the same column and the correlation between the rows can be seen in the matrix $H(M,N_e)$. If there is an equal bit sequence of the length over 16 bits in every row, we can find a sync pattern in the unknown telemetry data sequence.

Figure 3 depicts the matrix $H(M,N_e)$ constructed from an unknown telemetry data sequence with 16 bits minor frame sync pattern. As shown in Figure 3, if N_e is estimated to be the original minor frame length N_p, the 16 bits sync pattern is aligned in the same position in the matrix.



1   1 0 • • • 1 1 0 1 1 1 0 1 0 1 1 1 0 0 1 0 0 0 0 0 1 1 • • • 1 0 1 1

0 0 1 1 • • • 1 0 1 1 1 1 0 1 0 1 1 1 0 0 1 0 0 0 0 1 1 1 • • • 1 0 0 1

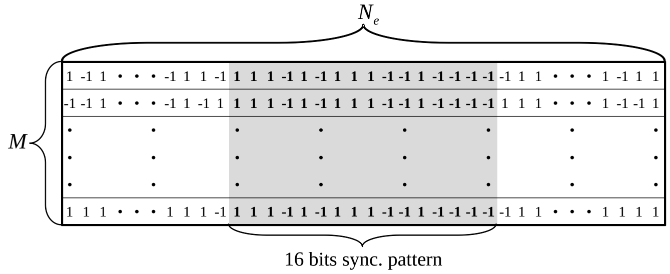
1 1 1 0 • • • 1 1 0 1 1 1 0 1 0 1 1 1 0 0 1 0 0 0 0 0 1 1 • • • 1 1 1 1

16 bits sync. pattern

Figure 3. Matrix $H(M,N_e)$ including 16 bits sync Pattern

ISBN: 978-1-61208-245-5

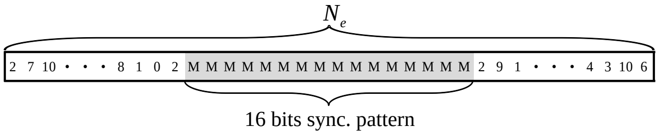
To verify the equality of column bits in the 16 bits sync pattern, we change the data bit representations: zeros to -1 and one to 1 in Figure 3, and show the results in Figure 4.



1 -1 1 • • • -1 1 1 -1 **1 1 1 -1 1 -1 1 1 1 -1 -1 1 -1 -1 -1 -1** -1 1 1 • • • 1 -1 1 1
-1 -1 1 • • • -1 1 -1 1 **1 1 1 -1 1 -1 1 1 1 -1 -1 1 -1 -1 -1 -1** 1 1 1 • • • 1 -1 -1 1
1 1 1 • • • 1 1 1 -1 **1 1 1 -1 1 -1 1 1 1 -1 -1 1 -1 -1 -1 -1** -1 1 1 • • • 1 1 1 1

Figure 4. Matrix $H(M, N_e)$ after bit mapping

In Figure 4, if we estimate the sync pattern correctly, the column-wise sum of all data bits in the 16 bits sync pattern will be M or –M. In this case, if we take an absolute value for the sum, there are continuous 16 values equal to minor frame number M when the estimated frame length N_e is equal to the original frame length N_p as shown in Figure 5.



2 7 10 • • • 8 1 0 2 M M M M M M M M M M M M M M M M 2 9 1 • • • 4 3 10 6

Figure 5. Absolute value of summation result when $N_e = N_p$

We summarize the overall process of blind estimation of minor frame length and sync pattern in Figure 6.

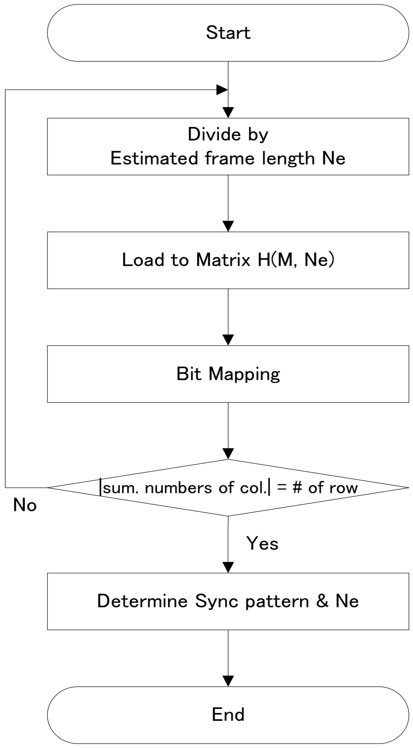

Start
Divide by
Estimated frame length Ne
Load to Matrix H(M, Ne)
Bit Mapping
|sum. numbers of col.| = # of row
No
Yes
Determine Sync pattern & Ne
End

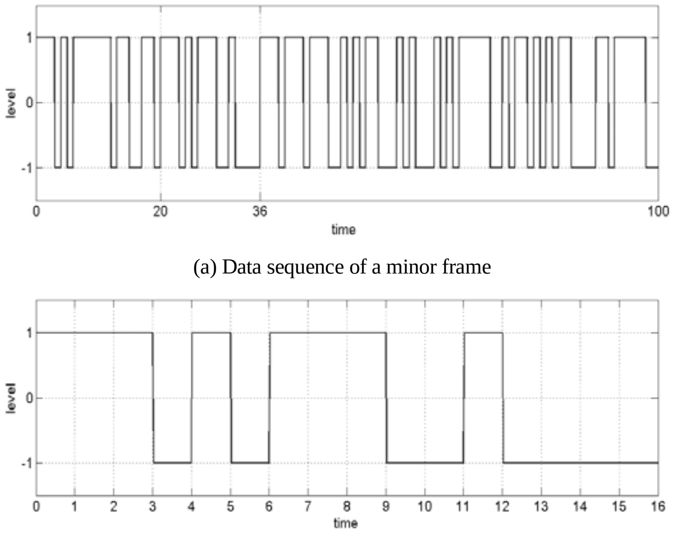
Figure 6. Blind estimation process of the minor frame length and sync pattern

As shown in Figure 6, we first divide the unknown telemetry data sequence into data blocks of the estimated minor frame length N_e and load the data blocks to the matrix $H(M, N_e)$ row by row. Then we map the data bits onto 1 or -1. If the absolute values of the column-wise sum of the matrix are equal to the number of divided data blocks through 16 consecutive columns, we can determine that the sequential bits in every row as a minor frame sync pattern and the estimated minor frame length N_e as the original minor frame length.

IV. SIMULATION RESULTS

For simulation of the algorithm, two telemetry data sequences depending on IRIG 106 Class1 are made. First telemetry data sequence has one major frame which consists of 128 minor frames. Each minor frame has the length of 1024 bits, and minor frame sync pattern of 16 bits. Second telemetry data sequence has one major frame which consists of 128 minor frames. Here, each minor frame has the length of 1024 bits, and minor frame sync pattern of 24 bits.

Figure 7 depicts a minor frame of the telemetry data sequence which has minor frame sync pattern of the length 16 bits and the estimated minor frame sync pattern.



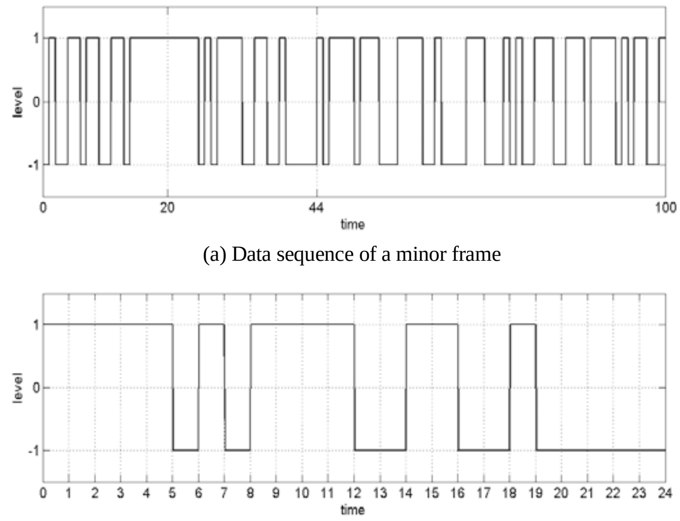
(a) Data sequence of a minor frame

(b) Estimated minor frame sync pattern of 16 bits

Figure 7. Determination of the minor frame sync pattern

As shown in Figure 7 (b), the estimated minor frame sync pattern is identical to 16 bits from 20th bit to 36th bit of Figure 7 (a). Figure 8 depicts a minor frame of the telemetry data sequence which has minor frame sync pattern of the length 24 bits and the estimated minor frame sync pattern.

ISBN: 978-1-61208-245-5



(a) Data sequence of a minor frame

(b) Estimated minor frame sync pattern of 24 bits

Figure 8. Determination of the minor frame sync pattern

As shown in Figure 8 (b), the estimated minor frame sync pattern is identical to 24 bits from 20th bit to 44th bit of Figure 8 (a).

If there is an error in the minor frame sync pattern, the proposed algorithm can be modified to estimate the original frame sync pattern, by introducing the threshold coefficient that has a real value of 0 to 1.

V. Conclusions

In this paper, we have presented a new blind estimation algorithm to find frame synchronization patterns in telemetry systems based on IRIG 106, using the matrix constructed using estimated minor frame length. By computer simulations, we have confirmed performance of the proposed algorithm. Simulations are carried out for a major frame consisting of 128 minor frames with 1024 bits. By exploiting the proposed algorithm, we estimated 16 bits long and 24 bits long sync patterns. In error environments, we can estimate the frame sync pattern also by using the threshold coefficient α . Our results can be applied to an unknown telemetry signal reconstruction for various practical cases in non-cooperative contexts such as spectrum surveillance systems and guided weapon systems.

Acknowledgment

This research was supported by Agency for Defense Development of Korea (2011-00000002370).

References

[1] Telemetry: Summary of concept and rationale, NASA. NASA Technical Reports Server. Retrieved 26 September 2011

[2] F. Carden, R. Jedlicka, R. Henry, “Telemetry Systems Engineering”, Artech House, 2002

[3] Telemetry standards, IRIG Standsard 106-11 Part 1, Secretatiat, Range Commander Council, U.S. Army White Sands Missile Range, New Mexico, 2011

[4] Packet Telemetry, CCSDS 102.0.B.5, Blue Book, Consultative Committee for Space Data Systems, Washington D.C, November, 2000

[5] Packet Telemetry Standard, PSS-04-106, ESA – European Space Agency, January 1988

[6] O.J.Strock, “Introduction to Telemetry”, Instrument Society of america, 1987

[7] Peter Fortescue, John Stark, Graham Swinerd, “Spacecraft systems engineering -3rd ed.”, Wiley, 2003

[8] G. Scot, S. Houcke, “Blind detection of interleaver paremeters”, IEEE International Conference on Acoustics, Speech and Signal Processing, Proceedings, Philadelphia, USA, vol. 3 pp.829-832, March 2005

[9] G. Burel, R. Gautier, “Blind estimation of encoder and interleaver characteristics in a non cooperative context”, IASTED International Conference on Communications, Internet and Information Technology, pp. 275–280, Scottsdale, Arizona, USA, November 2003

ISBN: 978-1-61208-245-5

Performance Analysis of Channel Switching with Various Bandwidths in Cognitive Radio

Po-Hao Chang, Keng-Fu Chang, Yu-Chen Chen, and Li-Kai Ye
Department of Electrical Engineering, National Dong Hwa University, 1,Sec.2, Da-Hsueh Rd., Shou-Feng,
Hualien, Taiwan, R.O.C.
po@mail.ndhu.edu.tw, m9823023@ems.ndhu.edu.tw, m9723017@ems.ndhu.edu.tw, d9623008@ems.ndhu.edu.tw

Abstract—**Dynamic spectrum access is a key technique in cognitive radio. Whenever primary users appear, secondary users must evacuate the primary channel rapidly, and switch the appropriate channel to primary users. There are two types of channel access methods, namely reactive and proactive channel access. In reactive method, cognitive radio does not need to switch channel until primary users appear, while in proactive method, secondary users predict future channel traffic by using channel history and switch channel before primary users appear. Most of the previous researches assume that all primary users possess the same channel bandwidth. In this paper, we take various channel bandwidths into consideration to make spectrum handoff decision with the method proposed before under real cognitive radio environment, and use this method to analyze the performance. Channel utilization rate in both methods is enhanced while considering various bandwidths.**

Keywords- ***reactive channel access; proactive channel access; Bandwidth.***

I. Introduction

Spectrum is a precious and limited natural resource, and most portions of it have been authorized. With the development of network techniques, the demand for spectrum is increasing. “The biggest problem of spectrum application is not scarcity but ineffective,” according to [1]. Not only do we have to solve the ineffective problem, but we also need to make good use of the spectrum. With "spectrum sensing capability”, cognitive radio technology has been proposed [2], hoping that through time and space configuration, full use of those unused spectrum resources can effectively solve the problem of the spectrum congestion or the unequal distribution.

The first step of developing cognitive radio is dynamic spectrum access [3]. Cognitive radio users must monitor the idle spectrum periodically, analyze the surrounding wireless messages at the same time to adapt themselves to the environment, and use these messages to learn and adjust the transmission state and parameters in radio. As soon as the primary users appear, cognitive radio users can seamlessly switch to other idle channels, making transmission continue in spite of the appearance of the primary users.

Liu, *et al.* [4] discusses the concept of spectrum mobility, and develops the probability models of spectrum holes and spectrum handoff according to the characteristics when primary users appear. The probability of handoff is an important indicator to affect spectrum mobility, and it is also an important part of spectrum management.

There are two types of channel access methods, namely proactive channel access and reactive channel access. It is still unclear under what condition we shall use which of these handoff methods. Wang, *et al.* [5] introduces these two methods, and uses a PRP M/G/1 model to derive the formula to analyze which of the two spectrum handoff methods can achieve the best efficiency with the variation of spectrum sensing time.

Without interference to primary users, secondary users are allowed to temporarily use the channel for transmission. Aravinda, *et al.* [6] uses proactive channel access method without interference under TV broadcast condition. Secondary users can use the information of channel history to build a prediction model, and use it to make the spectrum handoff decision. Not only can it enhance the channel utilization, but it can also reduce the interfering time produced by primary users.

Höyhtyä, *et al.* [7] proposes a simple method to classify the channels to either periodic or stochastic patterns, which may in turn help the secondary users to schedule the channel. Yang, *et al.* [8] sets the period when primary users appear as an alternative exponential ON/OFF model, and predicts future primary channel state according to previous primary users’ traffic. Secondary users can try not to interfere with the primary users by quickly switching to another unused primary channel to continue its transmission. Xue, *et al.* [9] emphasizes the importance of handoff delay in real life so that the secondary users can take it into consideration to make the spectrum handoff decision.

Most of the previous researches assume that all primary users possess the same channel bandwidth. In reality, however, cognitive radio needs to adapt to the heterogeneous network architecture in which not all primary users possess the same channel bandwidths. According to IEEE 802.22 standard [10], cognitive radio is applied for 54 ~ 862 MHz UHF/VHF TV bands and must adapt to 6MHz, 7MHz, and 8MHz TV bands. This will lead us to make different spectrum handoff decision. Hence, we take various channel bandwidths into consideration to make spectrum handoff decision with the method proposed before under real cognitive radio environment, and use this method to analyze the performance.

The rest of the paper is organized as follows: In Section II, we introduce the system model. In Section III, we discuss the strategy of handoff decision. In Section IV, we show our simulation results. In Section V, we provide our conclusion.

II. System model

A. System prediction model

At the beginning of channel prediction model shown in Fig. 1, secondary users will perform spectrum sensing, and then save the sensing information into channel history database. Primary users channel's traffic information can be obtained from the historical database. Cognitive radio can use spectrum sensing results and channel history to predict channels idle time in the next slot. If it needs spectrum handoff, the channels idle time in the next slot subtracts channel switch cost and spectrum sensing time to get the channels remaining time. The packet capacity can be derived from the product of channels remaining time and the channel bandwidth. Finally, in the handoff decision section, cognitive radio will look for the channel that has the maximum packet capacity to be the handoff channel, and use it to transmit data.

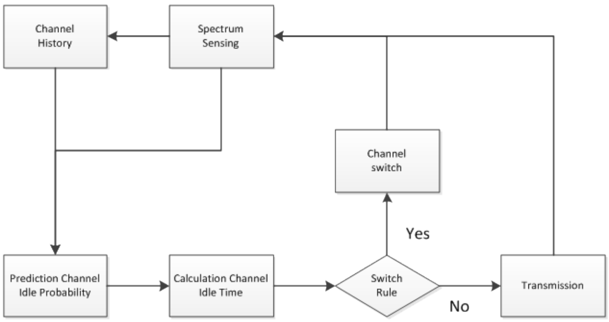


Figure 1. Flowchart of proactive spectrum access

B. Channel model

We use the common alternative exponential ON-OFF model to be our channel model as shown in Fig. 2. The difference with previous research is that our bandwidth is variable. At first, all primary channels will be random in ON or OFF state. When the channel is in the "ON" state, which means that the channel is "BUSY", the secondary users cannot access the channel; when the channel is in the "OFF" state, which means that the channel is "IDLE", the secondary users can access the channel for transmission. The durations that the primary user passage shows ON (BUSY) and OFF (IDLE) are independently exponentially distributed. For channel n, the period of ON, B_n, follows an exponential distribution with mean $1/\lambda_{B_n}$. On the other hand, the period of OFF, I_n, also follows an exponential distribution with mean $1/\lambda_{I_n}$.

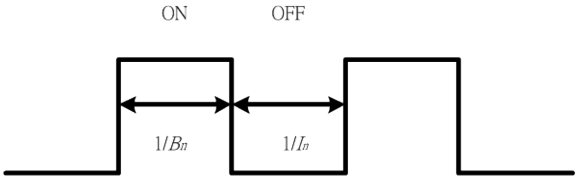


Figure 2. The alternative exponential ON/OFF channel model

$$f(B_n)=\begin{cases}\lambda_{B_n}e^{-B_n\lambda_{B_n}}, B_n \geq 0\\ 0 \qquad\qquad\quad , B_n < 0\end{cases} \tag{1}$$

In our simulation environment, we assume that cognitive radio is operated in a slotted model, and has N primary channels for access as shown in Fig. 3. At the period of spectrum sensing, cognitive radio will use the spectrum sensing result and the information from channel history database to predict the probability of channel idle time in next slot.

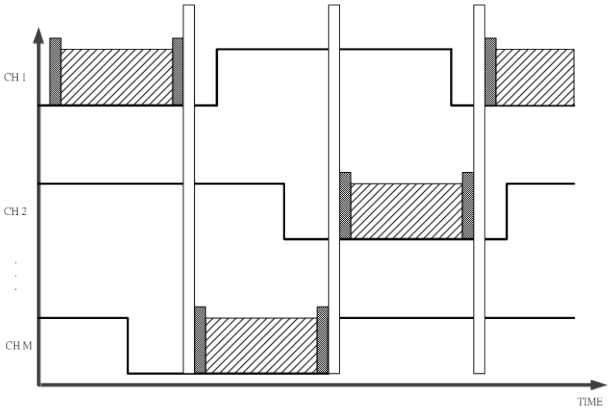


Figure 3. The slotted structure of the secondary user

III. The strategy of handoff decision

A. Reactive v.s. Proactive channel access

Spectrum handoff occurs when the primary user appears in the licensed channel that is temporarily used by the secondary users. The main significance of the spectrum handoff is to help the secondary users switch to the suitable idle channels to resume transmissions. The types of spectrum handoff are divided into two types:

Reactive channel access: Whenever a primary user occurs, secondary users have to handoff by following steps: first, after detecting any primary user, secondary users interrupt transmission and do spectrum sensing. Second, according to the real-time spectrum sensing, find idle channels and switch to the idle channel to resume transmissions. Not only does it have to spend real time spectrum sensing, some real time applications (for example, watching movies online) will also cause great harm. This is shown in Fig. 4.

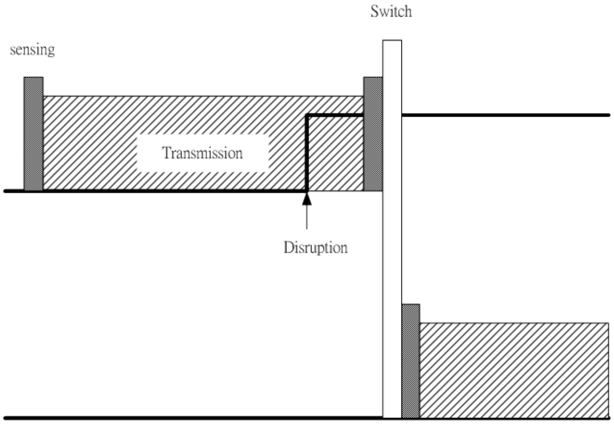


Figure 4. The reactive channel access model

Proactive channel access: Secondary users use channel history to predict future channel traffic, and schedule the channel access intelligently. They may use the predicable method with the result of spectrum sensing to predict future channel traffic before primary users appear in order to avoid disruption by primary users, and to maintain reliable communication. The advantage of using proactive channel access is the reduction of communication disruption by primary user, for secondary users can switch to other channels before primary users appear and resume transmission faster. This is shown in Fig. 5.

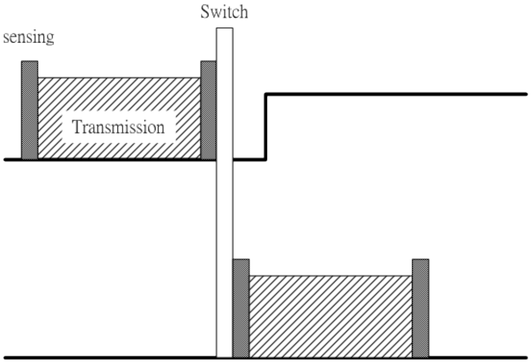


Figure 5. The proactive channel access model

B. Channel prediction

We assume that the primary channels are a series of ON/OFF model; when the channel is in ON state, the channel is being used; when the channel is in OFF state, the channel is IDLE and secondary users can access this channel. The frequency of primary users' appearance can be built in an alternative exponent ON/OFF model. The average channel IDLE time can be set as $\mathrm{E}[\mathrm{T}^{\mathrm{n}}_{\mathrm{IDLE}}]={}^{1}/_{\lambda_{I_n}}$, and the average channel IDLE time can be set as $\mathrm{E}[\mathrm{T}^{\mathrm{n}}_{\mathrm{ON}}]={}^{1}/_{\lambda_{B_n}}$. We can use the built model to predict future channel traffic. We assume that the idle probability of channel N in next slot is P_n which can be derived based on renewal theory [11]:

$$Pn=\begin{cases}\frac{\lambda_{B_n}}{\lambda_{B_n}+\lambda_{I_n}}+\frac{\lambda_{I_n}}{\lambda_{B_n}+\lambda_{I_n}}\mathrm{e}^{-(\lambda_{B_n}+\lambda_{I_n})\Delta\mathrm{t}} & ,\mathrm{s}=0\\ \frac{\lambda_{B_n}}{\lambda_{B_n}+\lambda_{I_n}}-\frac{\lambda_{B_n}}{\lambda_{B_n}+\lambda_{I_n}}\mathrm{e}^{-(\lambda_{B_n}+\lambda_{I_n})\Delta\mathrm{t}} & ,\mathrm{s}=1\end{cases} \qquad (2)$$

We can use the following formula to predict the channel idle time in next slot:

$$\mathrm{T}^{\mathrm{n}}_{\mathrm{IDLE}}={}^{Pn}/_{\lambda_{I_n}} \qquad (3)$$

C. Intelligence channel switching

When we make the spectrum switch channel decision, we consider [9] which mentioned switching cost to achieve more realistic simulation. In real life, spectrum switch decision making must consider more situations such as packet-loss-ratio, synchronization and delay. It must produce non-negligible handoff delay. The handoff delay needs to be taken into consideration, and it must affect the handoff decision as follows:

$$\mathrm{CH}=\arg\max[\,(\mathrm{T}_{\mathrm{IDLE}}-\Gamma\mathrm{T}_{\mathrm{switch}})] \qquad (4)$$

$$\Gamma=\begin{cases}1 & ,\text{if CH(Prev)}\cong\text{CH(Current)}\\ 0 & ,\text{else}\end{cases}$$

D. Bandwidth consideration

When we make a switch decision in the process of proactive channel access, we take the channel bandwidth into consideration to achieve the best QoS.

$$\mathrm{PACKET}^{\mathrm{n}}=\mathrm{T}^{\mathrm{n}}_{\mathrm{remain}}*\ \mathrm{BW}^{\mathrm{n}} \qquad (5)$$

$$\mathrm{T}^{\mathrm{n}}_{\mathrm{remain}}=\mathrm{T}^{\mathrm{n}}_{\mathrm{IDLE}}-\alpha\,(\,\mathrm{Tswitch}-\mathrm{Tsense}\,) \qquad (6)$$

$$\alpha=\begin{cases}1 & ,\text{if CH(Prev)}\cong\text{CH(Current)}\\ 0 & ,\text{else}\end{cases}$$

$$\mathrm{CH}=\arg\max_{\mathrm{n}}(\,\mathrm{PACKET}^{\mathrm{n}}) \qquad (7)$$

$\mathrm{PACKET}^{\mathrm{n}}$ means the packet capacity of the channel which performs transmission once and is equal to the product of channel remaining idle time and channel bandwidth. $\mathrm{T}^{\mathrm{n}}_{\mathrm{remain}}$ means the channel remaining idle time, and BW^{n} means channel bandwidth. If it needs to do channel switch, channel idle time has to subtract switch cost and spectrum sensing to calculate the channel remaining idle time. Finally, we will find the channel which has the maximum packet capacity to be our spectrum switch channel. The flowchart of this proactive channel access model is shown in Fig. 6.

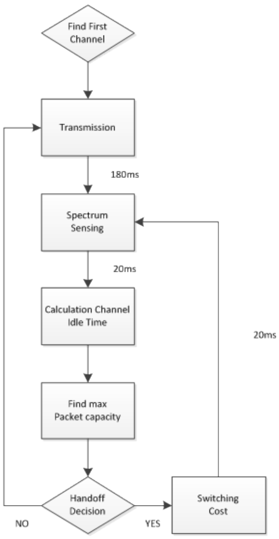


Figure 6. The flowchart of proactive channel access model

The available amount of the bandwidth is the key to determine the channel switch. When the remaining channel idle time is the same, the channel with bigger bandwidth possesses higher data transmission rate so that secondary users can finish their transmission faster.

Example: In Fig. 7, the channel remaining idle time of CH Y is larger than that of CH X. It is the best choice to switch to CH X according to conventional channel switch method which does not consider the channel bandwidth. After taking it into consideration, even if the channel remaining idle time of CH Y is larger than that of CH X, CH X can achieve faster data transmission since the channel bandwidth of CH X is twice that of CH Y. Selecting CH X not only reduces the cost of channel switch and the slots needed for further spectrum sensing so that more slots can be used for data transmission, but also requires less time to finish transmission.

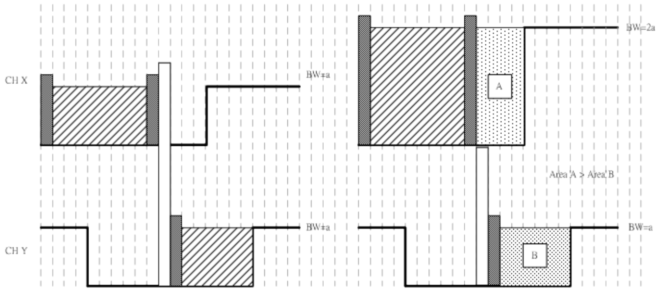


Figure 7. Channel selection with bandwidth consideration

As for the process of reactive channel access shown in Fig. 8, cognitive radio will first select a channel which is in IDLE state with maximum bandwidth to be “the First channel”, and start data transmission for 180ms. After finishing data transmission for 180ms, cognitive radio will perform spectrum sensing for 20ms periodically. Once primary user appears, cognitive radio will switch to the channel which has maximum bandwidth to resume communication if idle channels exist; otherwise, cognitive radio will keep spectrum sensing until the first idle channel appears.

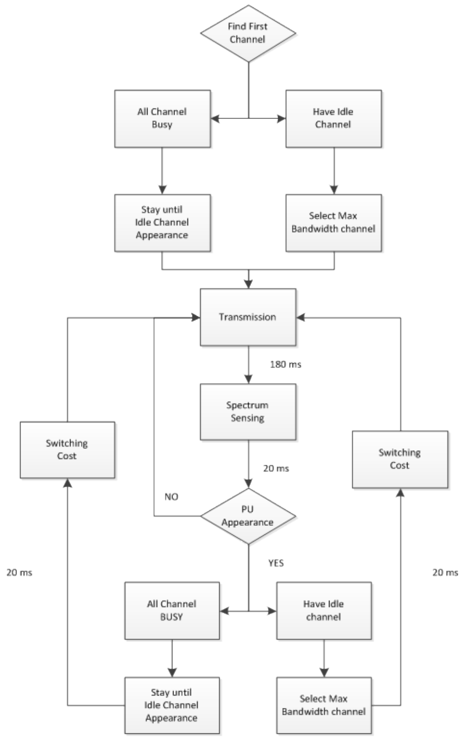


Figure 8. The flowchart of reactive channel access model

IV. SIMULATION RESULTS

We take various channel bandwidths into consideration to make spectrum handoff decision with the method proposed before under real cognitive radio environment, and use this method to analyze the performance. In our simulation environment, there are a secondary user and ten primary users. The bandwidths of ten primary channels are uniformly distributed from 1 to 10 MHz and the mean idle and busy periods are also uniformly distributed in min=0.5 and max=0.6~2.0. Sensing period and switch cost is 20ms. The period of data transmission is 180ms. Simulation time is set to 10000 sec.

Fig. 9 shows the number of channel switch in proactive and reactive methods. Note that proactive channel access method will process channel switch before primary users appear in order not to interfere with the primary users. On the other hand, reactive method just switch after primary users appear. Therefore, the number of channel switch in proactive method will be larger than that in reactive method.

Fig. 10 illustrates the number of interruption by primary users in these two methods. We find that the number of interruption by primary users in proactive method is always smaller than that in reactive method since proactive method predicts future primary channel state according to previous primary users’ traffic.

Channel utilization rate in both methods is shown in Fig.11. Instead of choosing the longest transmission time, proactive method always chooses the channel with maximum data transmission rate to switch. Comparing with reactive method which switches to the first channel with the idle state, the channel utilization rate of proactive method is lower than that of reactive method.

Proactive method can derive channel remaining time according to channel prediction model, and use the product of channel remaining idle time and the channel bandwidth to switch to channel with maximum data transmission rate, while in reactive method, cognitive radio will switch to the channel which has the maximum bandwidth without considering the probability of appearance of primary users, as shown in Fig. 12.

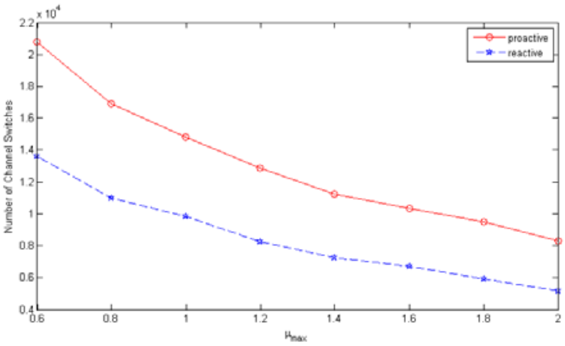


Figure 9. Number of channel switches

ISBN: 978-1-61208-245-5

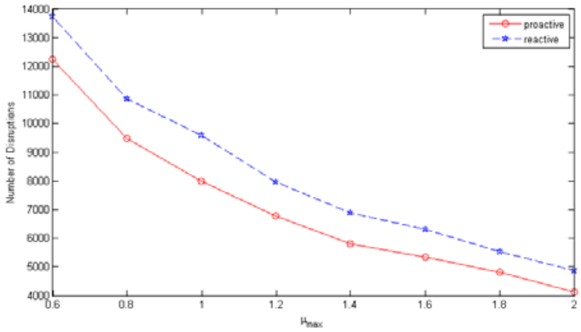


Figure 10. Number of disruptions by primary users

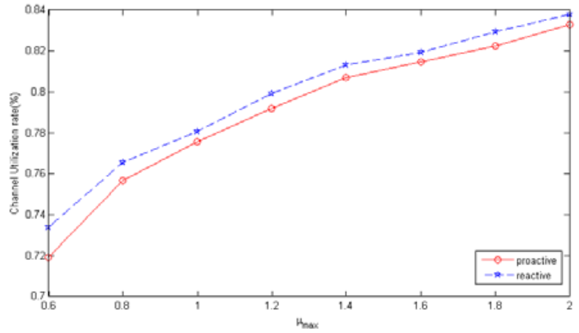


Figure 11. Channel utilization rate

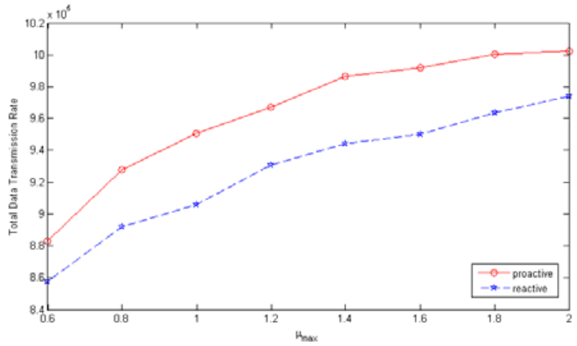


Figure 12. Total data transmission rate

V. CONCLUSION

Previous methods of channel switch just apply to channels with the same bandwidth. In this paper, we take different bandwidths into consideration for that purpose. Under these circumstances, the rule of channel switch should be changed in order to find the channel which meets the requirements of the users. By comparing the Proactive with the Reactive methods, we conclude that the former has better performance in terms of total transmission data rate and the number of disruptions by primary users, while the latter enjoys less channel switches and higher channel utilization rate.

REFERENCES

[1] Federal Communications Commission, “Spectrum policy task force report, FCC 02-155,” Nov. 2002.

[2] Simon Haykin, “Cognitive Radio: Brain-Empowered Wireless Communications” *IEEE Journal on Selected Areas in Communications*, vol.23, no.2, pp. 201- 220, Feb. 2005.

[3] Z. Tabakovic, S. Grgic, and M. Grgic, “Dynamic spectrum access in cognitive radio,” *ELMAR, 2009. ELMAR '09. International Symposium* , pp.245-248, Sept. 28-30, 2009.

[4] Hong-jie Liu, Zhong-xu Wang, Shu-fang Li, and Min Yi, “Study on the Performance of Spectrum Mobility in Cognitive Wireless Network” *IEEE Singapore International Conference on Communication Systems*, Nov. 19-21, 2008.

[5] Li-Chun Wang and Chung-Wei Wang, “Spectrum Handoff for Cognitive Radio Networks: Reactive-Sensing or Proactive-Sensing?” *Performance, Computing and Communications Conference, 2008. IPCCC 2008. IEEE International* , vol., no., pp.343-348, 7-9 Dec. 2008.

[6] Prashanth Aravinda, Kumar Acharya, Sumit Singh, and Haitao Zheng, “Reliable Open Spectrum Communications Through Proactive Spectrum Access” *First International Workshop on Technology and Policy for Accessing Spectrum*, August 5, 2006, Boston, MA, United States.

[7] Marko Höyhtyä, Sofie Pollin, and Aarne Mämmelä, “Performance Improvement with Predictive Channel Selection for Cognitive Radios” *First International Workshop on Cognitive Radio and Advanced Spectrum Management, 2008 (CogART 2008)*, pp.1-5, Feb. 14, 2008.

[8] Lei Yang, Lili Cao, and Haitao Zheng, “Proactive Cannel Access in Dynamic Spectrum Networks,” *2nd International Conference on Cognitive Radio Oriented Wireless Networks and Communications, 2007 (CrownCom 2007)* , pp.487-491, Aug. 1-3, 2007.

[9] Xue Feng, Qu Daiming, Zhu Guangxi, and Li Yanchun, “Smart Channel Switching in Cognitive Radio Networks” *2nd International Congress on Image and Signal Processing, 2009 (CISP '09)*, pp.1-5, Oct. 17-19, 2009.

[10] C. Cordeiro, K. Challapali, D. Birru, and N. Sai Shankar, “IEEE 802.22: the first worldwide wireless standard based on cognitive radios,” *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 2005 (DySPAN 2005)*, pp.328-337, Nov. 8-11, 2005.

[11] Cox, D. *Renewal Theory*, Spottiswoode Ballantyne and Co. Ltd, 1962.

A Fuzzy Inference System for Increasing of Survivability and Efficiency in Wireless Sensor Networks

Jose V. V. Sobral*, Aldir S. Sousa*, Harilton S. Araujo*, Rodrigo A. R. S. Baluz†, Raimir Holanda Filho†,
Marcus V. S. Lemos‡ and Ricardo A. L. Rabelo‡
*Computer Science Coordination
Unified Center of Teresina (CEUT), Teresina, Piaui, Brazil
Email: victormld@gmail.com, aldirss@yahoo.com.br, HariltonAraujo@ceut.com.br
†University of Fortaleza (UNIFOR - PPGIA)
Fortaleza, Ceara, Brazil
Email: rodrigobaluz@gmail.com, raimir@unifor.br
‡LABoratory of Intelligent Robotics, Automation and Systems (LABIRAS)
State University of Piaui (UESPI), Teresina, Piaui, Brazil
Email: marvinlemos@gmail.com, ricardor_usp@ieee.org, labiras.ctu@gmail.com

***Abstract*—The nodes of a Wireless Sensors Network (WSNs) are composed of small devices capable of sensing and transmitting data related to some phenomenon in the environment. These devices, named sensor nodes, have severe constraints, such as lower processing and storage capacity, and, mainly, they have severe constraints related to battery energy. Therefore, the developing of strategies to reduce the power consumption is one of the main challenges in WSNs, and thereby helping to increase the survivability and efficiency of these networks. This paper proposes a new approach to help multi-path routing protocols to choose the best route based on Fuzzy Inference Systems and ant colony Optimization (ACO). The Fuzzy System is used to estimate the degree of the route quality, based on the number of hops and the lowest energy level among the nodes that form the route. The Ant Colony Optimization (ACO) algorithm is used to adjust the rule base of the fuzzy system in order to improve the classification strategy of the route, and hence increasing the energy efficiency and the survivability of the network. The simulations showed that the proposal is effective from the point of view of the energy, the number of received messages, and the cost of received messages when compared against other approaches.**

***Keywords-WSN*; *Energy*; *Routing*; *Fuzzy Inference Systems*; *Ant Colony Optimization*; *Sink Nodes*.**

I. Introduction

Wireless sensor networks (WSNs) consist of a large number of sensor nodes distributed over a geographic area. Each node belonging to the network has the property to sense and transmit events, such as: luminosity, humidity, atmospheric pressure, pollution levels, temperature, among other. It is import to stress that each node has one or more sensors, processing capacity, storing and communication.

The WSN has motivated the interest of the research community because of its applicability in many areas, for instance, household applications [1], medical [2], military, environmental, farming [3] and vehicular environments. The WSNs differ from the traditional networks in many aspects, such as: the WSN are composed of a large number of sensor nodes; sensor nodes have a limited power supply, lower processing capacity and memory. In addition, some WSN applications require self-organization, in which the nodes must adjust themselves in an autonomous way, responding to structural changes, due to failure in some equipment, battery depletion, or by an external user request.

According to Araujo et al. [4], the main goal of a WSN is to collect data from the environment and transmit this data to a special node, called sink-node. In addition, the WSNs must provide some interface to permit the extraction of this information by external entities (i.e., users, or others information systems). Because of the sensor nodes limited power supply, the energy consumption is a relevant factor at all the stages of the life cycle in WSN applications [5].

Communication in WSNs consumes more energy than processing and sensing performed by the network nodes. An important challenge in WSN field is the energy consumption reduction, since in most cases these nodes are deployed in a harsh environment, making hard the battery replacement. The importance of energy consumption in WSNs is also depicted by Pinz et al. [6]. The authors have showed that transmission is the main cause of energy consumption.

This feature requires the implementation of routing policies that enable the sensor nodes to communicate efficiently and effectively with minimum power consumption. For this reason, the routing protocols must work with information based on the quality of routes, related to relevant network metrics, such as the energy level of the network sensor nodes.

Thus, the routing protocols for WSNs must have self-configuration properties, enabling to find out which is the best way to transfer information, considering the guaranteed delivery and energy level among the nodes of the network. If a sensor node will fail due to lack of energy, routes must be

ISBN: 978-1-61208-245-5

recalculated, so that the information collected can reach the destination node. The communication between the sensor nodes must optimize the energy consumption in order to increase the lifetime of the network.

This work proposes a Fuzzy Inference System [7], [8] to help the Directed Diffusion routing protocol [9] to choose a route for the communication between any nodes in the network. The Directed Diffusion was chosen because of its wide acceptance in current works and the multi-path properties. The proposed fuzzy system uses as input to the inference process, the number of hops and the lowest energy level among the nodes that comprise the route. From the quantitative values of the inputs, the system estimates a quantitative value associated with the quality of each route, in order to assist the routing protocol in the selection of several feasible routes. Therefore, based on the quality of the route, the routing protocol should define which route to be used for sending the data collected with the aim of increasing the network lifetime.

The design of a fuzzy inference system can be seen as a search/optimization problem in a search space of high dimensionality (multidimensional). Each point of the search space represents a particular fuzzy knowledge base (fuzzy database + fuzzy production rules base). Therefore, finding the best design of a fuzzy inference system means to obtain an optimal point on the search space. However, this search space is characterized as infinitely large, non-differentiable, complex, noise, multimodal and deceptive [10]. Thus, obtaining a fuzzy inference system optimized for a particular application can be a very complex task. In the proposed approach, the adjustment of the Fuzzy Inference System for classification of routes is performed automatically by the Ant Colony Optimization (ACO) algorithm [11]. ACO is a computational model for optimization inspired by the foraging behavior of real ants [12]. Specifically, ACO algorithms are based on the ability of real ants for finding the shortest path from their nest to the food source by exploiting pheromone information. Pheromone is a chemical substance layed by real ants while walking. Paths that have higher concentrations of pheromone have more chances to be followed by the other ants. Therefore, the pheromone plays an important role by biasing the decision mechanism of the real ants. By following a specific path, the ants lay its own pheromone in order to reinforce the pheromone concentration. Based on trail-laying and trail-following mechanism, the real ants can find the shortest path connecting the nest and the food source.

The main benefits of the proposed approach are:

- Application of fuzzy inference systems for inferring the quality degree of routes. Therefore, the process for calculating the quality degree of each route can be accompanied and tracked from the linguistic viewpoint. Additionally, fuzzy inference systems are able to express and manipulate qualitative information, which can help the domain experts in understainding the results produced;
- As sensor nodes have limited resources, the sink nodes are responsible for calculating the quality degree of each route. This means that the fuzzy inference system is executed inside the sink nodes instead of inside sensor nodes. The sink nodes are not strongly limited as the sensor nodes;
- Application of an ant colony optimization algorithm for adjusting the rule base of the fuzzy inference system. The ACO algorithm implemented is the Ant System [13]. As a heuristic algorithm, ant colony optimization does not require special properties of the search space such as convexity, smoothness, existence of derivatives. Additionally, ACO algorithm is a population-based technique and includes stochastic components to update the solutions which results in lower chances of the optimization process to get trapped in local minima. It is noteworthy that the rule base stores the strategy of action/control implemented on the fuzzy system. Therefore, an optimal adjustment of the rule base must result in an efficient strategy for dealing with the limited resources in a WSN;
- Besides the use of pheromone information by artificial ants, a heuristic function is used for enhancing the process of constructing solutions to the problem. Although ACO algorithms are able for solving problems without using a heuristic function [14], the incorporation of a heuristic information normally results in better solutions [15]. The use of a heuristic function requires specialized information related to the problem being solved. For this purpose, the experience of the expertise in WSNs is used as heuristic information for guiding the construction of solutions by artificial ants;
- In order to represent realistically the behavior of a Wireless Sensor Network, an energy dissipation (consumption) model is applied in the proposed approach. Most of the works in WSNs are focused on the routing protocols itself, without taking into account the energy consumption that happens in the sensor nodes [16]. The sensor nodes are highly dependent on the limited battery source for its communication (data collection and transmission) and computation operations [17]. Since energy is drawn for both operations, it is important to consider the rate at which the energy is consumed for both operations. As the energy consumption in transmission is greater than the consumption in computation operations, the majority of the works only handles the consumption of energy during the communication. Therefore a energy consumption is included for avoiding misleading calculation to the overall energy consumption of the WSN.

The article is organized as follows: Section 2 presents a description of the problem to be solved. Section 3 describes the proposed approach. The evaluation of results is discussed in Section 4, followed by conclusions in Section 5.

II. Routing in Wireless Sensor Networks

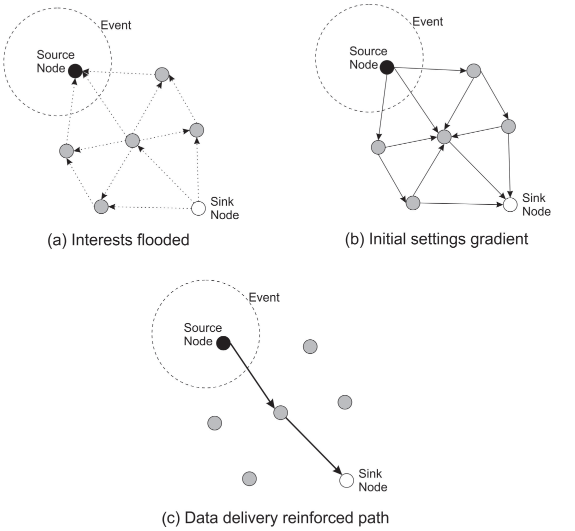
A. Routing Protocol

The Directed Diffusion [9] routing protocol was used for the communication between the sensor nodes and the sink nodes, which is designed for Wireless Sensor Networks where the network designer is responsible for defining the type of event that must be observed by sensors and monitored area [5], [18], [19], [20].

Directed Diffusion protocol has its operation based on the following elements: named data, interest, gradient, and reinforcement. Data are named by using a pair (attribute, value) and represent an event detected by the sensor nodes. The interest is the phenomenon that represents the search attributed to the network. The gradient is the pointer that represents the reverse path addressed to the sink node. The task to be sensed is diffused by the sensors network through a interest message sent by the sink node through periodic broadcasts. The interest messages can be originated by one or more sink nodes, according to the network design. Because of this feature, interest messages, when disseminated by the network, create gradients, which are states stored by the sensor nodes that have received interest messages, identifying the nodes that sent interest. Thus, the gradients define the nodes that should receive data related to the disclosed interests. Finally, there is the reinforcement, where the sink node receives messages from events occurring at a low rate of transmission through various available paths, and then it chooses one of these paths and reinforces the transmission rate for the event to be informed through this path at a higher transmission rate. The sink performs reinforcement by resending the original interest to the selected path, forcing the data source node (node that detected the event) to increase its transmission rate of data collected through this path.

Figure 1 illustrates some aspects of Directed Diffusion. The propagation of interest is shown in Figure 1 (a), at this initial moment, the sink node broadcasts over the network a interest message containing the named data (attribute, value) from which it wants to receive information. The interest message is periodically updated by simply changing the time stamp of interest. This is necessary because the sensor network is not reliable in transmitting packets.

After broadcasting the interest message, sensor nodes in the network associate gradients to each interest message received, creating routes between the sink and the data source node, as shown in 1 (b). Through the use of interests and gradients, several paths are established between the sink node and source node, but only one of these paths is selected by the reinforcement mechanism, as illustrated in Figure 1 (c).



(a) Interests flooded

(b) Initial settings gradient

(c) Data delivery reinforced path

Figure 1. Directed Diffusion

The Directed Diffusion protocol meets dynamic networks by a trading scheme with the spread of interest and reinforcements paths, allowing the network to converge before any topological change. Moreover, the Directed Diffusion uses delay as a metric for route choosing, which usually is the better choice.

A disadvantage of this approach is the high cost of communication for route repair when failures occur due to the need of periodically broadcast in network in order to reinforce others routes.

B. Related Works

A proposal for optimization of energy consumption is treated in Chan et al. [21]. The idea of this proposal is to put some sensor nodes in sleep mode (off) to conserve energy while maintaining connectivity between the nodes that comprise the network. This strategy has the benefit of energy savings due to switching execution between sensor nodes.

Shah and Rabaey [22] describe the protocol EAR (Energy Aware Routing). The basic operation is the occasional use of a set of paths chosen by a probability function (which is dependent on the power consumption of each path). This prevents that the best route has its energy exhausted. Therefore, the network lifetime is increased. It is assumed that each node has an address and information about your location. This proposal has advantages due to which route selection is performed by probability function allowing the selected route is not used until the death of the nodes that make up said route.

A fuzzy expert system for clustering of sensor nodes [23] is presented with the target of conserving energy. Three linguistic variables are employed in the design of fuzzy expert system, including the selection probability, the distance from the base station, and the sum of the distances between the selected node and the other nodes with lower energy than the average energy.

Most of the energy consumed in WSNs is the transmission of data to the sink node. Singh et al. [24] propose mobile sink to minimize power consumption. The movement of the sink is determined by the mechanism based on fuzzy logic. The authors state that the base station can only move in a predefined circular path in accordance with input variables, such as the node residual energy, and distance to the base station. The simulation results are compared with the methods that have a stationary base station.

A hybrid approach involving a Mamdani fuzzy system optimized with Genetic Algorithms is presented in [25]. The hybrid intelligent system is employed to assist the routing protocol Directed Diffusion [9]. The sensor nodes have a fuzzy inference system implemented, which is tuned by a Genetic Algorithms, resulting in a fuzzy-genetic system [26]. The fuzzy system set is used to estimate the quality of each route associated with the sensor node. The results show that the use of genetic fuzzy system in conjunction with the protocol Directed Diffusion increases the lifetime of the network protocol when compared to Directed Diffusion without the use of fuzzy systems. However, the proposed approach does not consider a dissipation energy models in the sensor nodes, i.e, the approach does not account consumption (dissipation) of energy during the calculation of the degree of quality of the route through fuzzy inference system (fuzzification, inference procedure and defuzzification). The adoption of a dissipation energy model would make the processing more realistic, since the fuzzy system is implemented in sensor nodes, reducing the level of battery power consumption.

III. Proposed Approach

A. Fuzzy Inference System

Fuzzy inference systems are capable of dealing with highly complex processes, which are represented by qualitative information. Normally, fuzzy inference systems are based on linguistic rules of the type “if condition then action”, in which the fuzzy set theory [27] and fuzzy logic [28] provide the necessary mathematical basis to deal with qualitative information and with the linguistic rules.

A fuzzy inference system, generally, is composed by four components (Figure 2):

- Fuzzification Interface: responsible for mapping the quantitative input variable to the fuzzy domain, representing the assignment of linguistic values (primary terms), defined by membership functions, to the input variables;
- Knowledge Base: is formed by two components, the Data Base and the Rule Base. The data base contains the primary terms for each variable considered in the linguistic rules and the membership functions associated to each primary terms. The rule base is comprised of linguistic rules that determine the policies of control-action strategy and decision-making. The rule base realizes the mapping from the input domain to the output domain, and this way, plays an important role to generate the results produced by the fuzzy inference system;

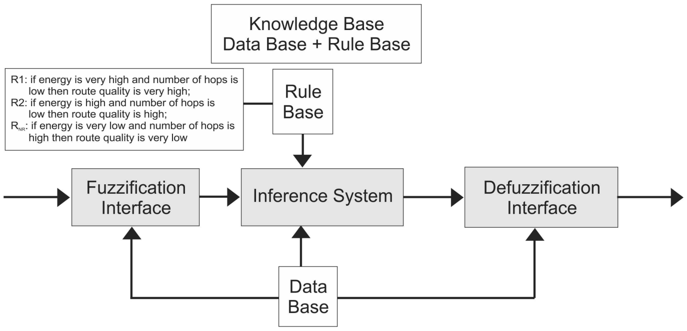

Knowledge Base
Data Base + Rule Base
R1: if energy is very high and number of hops is low then route quality is very high;
R2: if energy is high and number of hops is low then route quality is high;
R_{NR}: if energy is very low and number of hops is high then route quality is very low
Rule Base
Fuzzification Interface
Inference System
Defuzzification Interface
Data Base

Figure 2. Fuzzy Inference System

- Inference System: responsible for evaluating the primary terms of the input variables, by applying linguistic production rules contained in rule base, in order to obtain the fuzzy output value of the inference system. Therefore, the fuzzy output value is function of the rule base specified;
- Defuzzification Interface: responsible to assign a numerical value to the output fuzzy value. Thus, defuzzification can be considered a kind of synthesis of the final fuzzy output set by means of a numerical value.

B. Ant Colony Optimization

Ant Colony Optimization (ACO) algorithms constitute a relevant subset of ant algorithms. ACO is composed of algorithms for optimization/search problems and is inspired on observations of how some ant species forage for food. Therefore the ACO meta-heuristic concerns about developing algorithmic models of the foraging behavior of real ants. Besides of the complex behavior for foraging, other collective behaviors of real ants that have been proposed and applied include the division of labour, cemetery organization, brood care and construction of nests.

The emergence of shortest path selection in foraging behavior is explained by the differential path length effect and autocatalysis (positive feedback, reinforcement learning through pheromone deposit) [11], [15].

The ACO algorithm involves two basic procedures:

- Procedure for building a solution, in which Na (number of ants) ants build in parallel way Na solutions to the problem.
- Procedure for updating the pheromone concentration. The built solutions by the ants are evaluated through

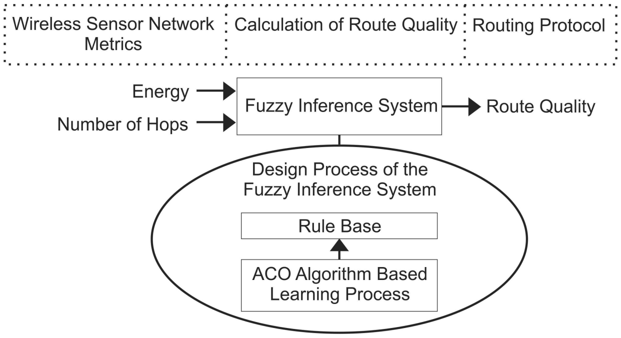

Wireless Sensor Network Metrics
Calculation of Route Quality
Routing Protocol
Energy
Number of Hops
Fuzzy Inference System
Route Quality
Design Process of the Fuzzy Inference System
Rule Base
ACO Algorithm Based Learning Process

Figure 3. Design Process of the Fuzzy Inference System

an evaluation function in order to measure the quality of the solutions produced. The update of the pheromone concentration is based on the evaluation function, this way, better solutions result in more pheromone deposited in its parts.

C. ACO Algorithm for Adjusting the Rule Base of the Fuzzy Inference System

In our approach, an Ant Colony Optimization algorithm has been used for adjusting in an optimal way the rule base of the fuzzy inference system. This way, the tour of an artificial ant is regarded as a combination of primary terms to the output linguistic variable (route quality) from every rule of the rule base. Therefore, for each rule, N_{pt} linguistic values are available to be selected, where N_{pt} is the number of primary terms for the output linguistic variable. During its tour, the ant has to choose one primary term for each rule from a total of N_{pt} options. This way, the complete specification of the rule base of the whole fuzzy inference system is given by the tour of an ant. Suppose that N_r is the number of linguistic rules present on the rule base, there are $N_{pt}^{N_r}$ combinations associated to the output linguistic variable. As the rule base relates the mapping of input values to the output value, an optimal adjustment of the rule base enhances the results produced by the fuzzy inference system. For our purpose, the result produced by the fuzzy inference system is the quality degree of the routes (route quality). As better the result produced by the fuzzy system, higher is the lifetime of the WSN. Our target is to find the best combination that maximizes the performance of the fuzzy inference system to classify the routes at the WSN. Therefore, after the training phase (learning process) via ACO algorithm, the fuzzy inference system is optimally adjusted and is ready to be incorporated in a sink node of a real Wireless Sensor Networks for classifying the routes associated to itself (Figure 3). The route quality is used by the routing protocol for selecting a specified path to send a message.

The fuzzy inference system proposed has two input variables: the lower energy level associated to some sensor node of a determined route and the number of hops necessary for sending the message to the sink node. The definition of the partition fuzzy for each input variable has been made in advance, based on the knowledge of the expertise. 5 primary terms were defined for the variable related to the lower energy level, and 3 primary terms were defined for the variable associated to the number of hops. This way, the rule base contains 15 rules. The output variable which determines the quality degree of the route has 5 primary terms. Therefore, for each rule, 5 options are available for the linguistic value. The artificial ants have to find an optimal setting for the linguistic rules from a total of 5^{15} (30517578125) combinations.

Besides of using artificial pheromone to help the choice of a specified path by the ants, the Ant System algorithm incorporates a heuristic function. The inclusion of an heuristic information normally results in better solutions but requires specialized information related to the problem being solved. The problem of designing the heuristic information is solved by using the expertise knowledge. Therefore, the accumulated experience of the expertises is used for helping the decision-making process by the ants.

The main aspects involved with the optimization of the fuzzy rule base are:

- Initialization of the parameters: at this step, the parameters of the ACO algorithm are initialized. The number of ants, the evaporation rate, the parameters that control the relative importance of pheromone information versus heuristic information.
- Initial placement of the ants: all the ants are placed on the start node which can resemble the nest.
- Selection of the primary term for each rule: the ants execute a probabilistic decision-making concerning what node should be visited. The decision-making process is based on the pheromone information and heuristic function. The primary term represented through the selected node by the ant is inserted in the linguistic value of the associated rule. This way, the selection of the primary term represents the process of building a solution which is equivalent to determine the primary term for each rule.
- Evaluation of the built solutions (produced tours): after the ants finish the solution construction process, it is necessary to measure the obtained solutions. The evaluation of the produced solutions is used for determining the quality of the solutions with respect to the problem being optimized. This way, it is possible to indicate what ant adjusted better the rule base. For evaluating the solution produced by a specific ant, the rule base obtained is inserted in the fuzzy inference system and a simulation of the WSN is realized. The lifetime of the WSN is used as the value for measuring the quality of the fuzzy inference system because this value represents the energy level of the sensor nodes. Therefore, as

higher the lifetime of the WSN, better is the rule base obtained.

- Updating of the pheromone concentration: in the last stage, the ants deposit their own pheromone. The pheromone deposited is proportional to the lifetime of WSN. Therefore, the highest value of pheromone deposited is obtained by the fuzzy inference system that classified better the routes and this way extends the lifetime of the WSN.

IV. Results and Discussions

In order to verify the applicability of the proposed approach, the Sinalgo simulator [29] has been used. Sinalgo is a framework developed in Java that allows the simulation of wireless network, abstracting the lowest layers of the protocol stack. The proposed solution is compared to the Directed Diffusion routing protocol (DD) and the Directed Diffusion routing protocol with a fuzzy inference system (DDF) incorporated, but designed in a manual way. The simulator scenario was designed to allow a didactic comprehension of the proposed algorithm, with a topology containing 10 X 10 sensor nodes. The initial energy stored at the battery was considered equal to 1,0 Joule. The simulations took into account the energy consumption model proposed in [30] and used in [17]. It is noteworthy that the fuzzy inference procedures for obtaining the route quality are realized at sink nodes, therefore the cost of the energy consumption associated to fuzzy operations (fuzzification, inference and defuzzification) is not considered because the sink nodes are not limited for energy. However, the cost for receiving, processing and sending messages is taken into account at all sensor nodes.

We choose the following four metrics in order to evaluate the proposed approach (DD-ACO-Fuzzy), against the Directed Diffusion (DD) and Directed Diffusion with fuzzy system manually adjusted: number of received messages, residual power, number of received messages versus time simulation, and the cost of incoming messages. The number of received messages corresponds to the amount of incoming messages versus the time simulation measured on rounds (time scale of the simulator). The residual power measures the amount of remaining power in the sensor node after sending messages to the sink node. The number of received messages versus the time simulation corresponds to the amount of messages that the sink node received during the network simulation. The cost of incoming messages is the ratio between the consumed power and the amount of received messages into the sink node. The network lifetime is the elapsed time between the start of the simulation till the moment that the sink node is unable to receive messagens collected and sent by the network.

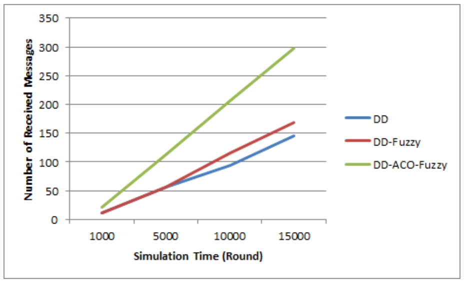


Figure 4. Number of received messages into the sink node versus the time

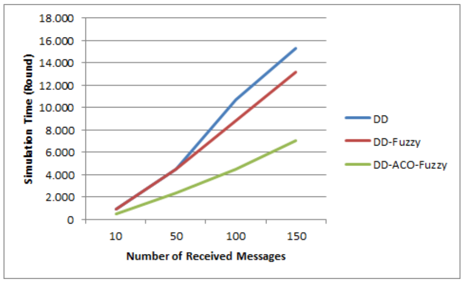


Figure 5. Simulation time necessary for a specified number of received messages

Figure 4 shows that using the proposed approach in this work, the number of received messages into the sink node versus the time is greater than in the DD and DDF approaches. This way, the proposed approach is able to increase the number of messages for the same simulation time. This means that a highest number of messages is capable of be transmitted, which maximizes the benefits of the limited resources of the sensor nodes. The Figure 5 complements the Figure 4 by showing the necessary time for the other approaches be able to receive the same number of messages received by the DD-ACO-Fuzzy. The proposed approach reaches a higher amount of delivered messages to the sink node during the simulation and requires a lowest quantity of time for receiving a specified number of messages. The difference was about 8,000 rounds.

The remainder amount of power in the sensor nodes after the receiving of messages into the sink node is greater with the DD-ACO-Fuzzy approach, as showed in the Figure 6. This is reached because the proposed approach guarantees a higher delivery rate by selecting smartly the route with the lower number of hops and the greater residual power of the sensor nodes along the route. Therefore, the routing protocol uses an information, the route quality, which is inferred from the number of hops and the residual power of the sensor nodes. The route quality is adjusted whenever the network conditions (metrics) are modified. This adjustment increases the reliability of the route quality because it takes into account instantaneously any modifications in energy level of the sensor nodes, for instance.

The Figure 7 shows that the cost of incoming messages, using the DD-ACO-Fuzzy approach, is decreased of approximately 60% if compared against the Directed Diffuzion and Directed Diffusion Fuzzy approaches. Therefore, a lowest

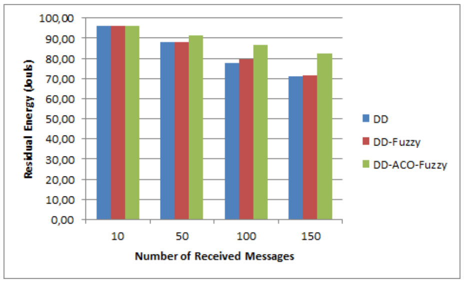


Figure 6. Residual power (energy) in the sensor nodes for a specific number of received messages

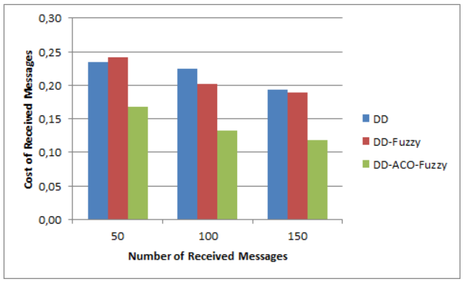


Figure 7. The cost of incoming messages

cost for the message communication is realized by using the proposed approach. This implies that a highest level of energy will be available which results in a highest life time for the wireless sensor network.

V. Conclusion and Future Work

This work proposes a Fuzzy Inference System to help the Directed Diffusion routing protocol to choose a route for the communication between any nodes in the network. The proposed fuzzy system estimates a quantitative value associated with the quality of each route, in order to assist the routing protocol in the selection of several feasible routes. Therefore, based on the quality of the route, the routing protocol should define which route to be used for sending the data collected with the aim of optimizing the network lifetime, the number of received messages, the necessary time to send a specified number of messages and the cost of received messages. As a hard task, the adjustment of the Fuzzy Inference System for classification of routes is performed automatically by the Ant Colony Optimization (ACO) algorithm. The ACO algorithm is used for adjusting the rule base of the fuzzy inference system. The rule base stores the strategy of action/control implemented on the fuzzy system, and therefore, an optimal adjustment of the rule base must result in an efficient strategy for dealing with the limited resources in a WSN.

The results showed that the Directed Diffuzion with Fuzzy approach using the ACO algorithm, for all metrics, is more efficient than the others, showing positive results with relation to the amount of received messages, residual power, number of received messages verus time simulation and the cost of incoming messages. Therefore the inclusion of a fuzzy inference system is capable of improving the use of limited computational resources associated to a wireless sensor netowrk. Although the use of a fuzzy inference system adjusted by trail and error approach (DD-Fuzzy) makes a better use of informations to classify the routes, this kind of adjustment is not so powerful as the automatic adjustment by ACO. The ACO algorithm is able to explore the search space and identify good regions to be exploited, in order to optimize the benefits of using a fuzzy inference system for helping a routing protocol. As future works, the authors are applying ACO algorithms for a simultaneous adjustment on fuzzy data base and fuzzy rule base.

References

[1] S. Hussain, S. Z. Erdogen, and J. H. Park, “Monitoring user activities in smart home environments,” *Information Systems Frontiers.*, vol. 11, no. 5, pp. 539–549, 2009.

[2] W. Chen, S. Hu, J.and Bouwstra, S. B. Oetomo, and L. Feijs, “Sensor integration for perinatology research,” *International Journal of Sensor Networks*, vol. 9, pp. 38–39, 2011.

[3] M. E. Martnez-Rosas, H. C. vila, J. I. N. Hiplito, and J. R. G. . Lpez, *Wireless Sensor Networks (WSN) Applied in Agriculture*. IGI Global, 2011, pp. 115–135.

[4] H. S. Arajo, W. L. T. Castro, and R. Holanda Filho, “Wsn routing: An geocast approach for reducing consumption energy.” in *IEEE Wireless Communications & Networking Conference*, 2010.

[5] ——, “A proposal of self-configuration in wsn for recovery of broken paths.” in *IEEE Sensors Applications Symposium - SAS 2010*, 2010.

[6] A. Pinz, M. Prantl, H. Ganster, and H. K. Borotschnig, “Active fusion a new method applied to remote sensing image interpretation,” *Pattern Recognition Letters*, vol. 17, pp. 1349–1359, 1996.

[7] L. A. Zadeh, “The Concept of a Linguistic Variable and its Application to Approximate Reasoning -I,” *Information Sciences*, vol. 8, no. 3, pp. 199–249, 1975.

[8] W. Pedrycz and F. Gomide, *Fuzzy Systems Engineering: Toward Human-Centric Computing*. John Wiley & Sons, 2007.

[9] C. Intanagonwiwat, R. Govindan, D. Estrin, J. Heidemann, and F. Silva, “Directed diffusion for wireless sensor networking,” *IEEE/ACM Transactions on Networking*, vol. 11, no. 1, pp. 2–16, 2003.

[10] Y. Shi, R. Eberhart, and Y. Chen, “Implementation of Evolutionary Fuzzy Systems,” *IEEE Transactions on Fuzzy Systems*, vol. 7, no. 2, pp. 109–119, 1999.

[11] M. Dorigo and T. Stutzle, “The ant colony optimization metaheuristic: Algorithms, applications, and advances,” *Handbook of metaheuristics*, pp. 250–285, 2003.

[12] A. P. Engelbrecht, *Computational Intelligence: An Introduction*. John Wiley & Sons, Ltd, 2007.

[13] M. Dorigo, V. Maniezzo, and A. Colorni, "Ant system: Optimization by a colony of cooperating agents," *IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics*, vol. 26, pp. 29–41, 1996.

[14] M. Dorigo and T. Stutzle, "An experimental study of the simple ant colony optimization algorithm," in *WSES International Conference on Evolutionary Computation*, 2001.

[15] M. Dorigo, E. Bonabeau, and G. Theraulaz, "Ant algorithms and stigmergy," *Future Generation Computer Systems*, vol. 16, no. 8, pp. 851–871, 2000.

[16] K. Yang and H. Wu, Y. Zhou, "Research of optimal energy consumption model in wireless sensor network," in *2nd International Conference on Computer Engineering and Technology*, 2010, pp. 421–424.

[17] D. P. Dahnil, Y. P. Singh, and H. C. Kuan, "Minimum energy dissipation protocol in degree-based clustering in wireless sensor networks," in *International Conference on Computer and Communication Engineering (ICCCE 2012)*, 2012, pp. 582–586,.

[18] A. N. Eghbali and M. Dehghan, "Load-balancing using multi-path directed diffusion in wireless sensor networks," in *Proceedings of the 3rd international conference on Mobile ad-hoc and sensor networks (MSN'07)*, 2007.

[19] M. Chen, T. Kwon, Y. Yuan, Y. Choi, and V. C. M. Leung, "Mobile agent based directed diffusion in wireless sensor networks," *EURASIP Journal of Applied Signal Processing*, 2007.

[20] X. Zhu, "Pheromone based energy aware directed diffusion algorithm for wireless sensor network." *Lecture Notes in Computer Science*, vol. 4681, pp. 283–291, 2007.

[21] B. Chan, K. Janieson, H. Balakrishnan, and R. Morris, "Span: An energy-efficient coordination algorithm for topology maintenance in ad hoc networks," *ACM Wireless Networks Journal*, vol. 8, no. 5, pp. 481–494, 2002.

[22] R. Shah and J. Rabaey, "Energy aware routing for low energy ad hoc sensor networks," in *Wireless Communications and Networking Conference (WCNC2002)*, 2002, pp. 350–355.

[23] Y. Shen and H. Ju, "Energy-efficient cluster-head selection based on a fuzzy expert system in wireless sensor networks," in *2011 IEEE/ACM International Conference on Green Computing and Communications (GreenCom)*, 2011.

[24] A. Singh, A. Alkesh, and N. Purohit, "Minimization of energy consumption of wireless sensor networks using fuzzy logic," in *2011 International Conference on Computational Intelligence and Communication Networks (CICN)*, 2011.

[25] R. A. L. Rabelo, M. V. S. Lemos, L. B. Leal, R. Holanda, and F. A. S. Borges, "An integration of fuzzy inference systems and genetic algorithms for wireless sensor networks," *International Journal of Hybrid Intelligent Systems*, vol. 9, pp. 61–74, 2012.

[26] F. Herrera, "Genetic Fuzzy Systems: Taxonomy, Current Research Trends and Prospects," *Evolutionary Intelligence*, vol. 1, no. 1, pp. 27–46, 2008.

[27] L. Zadeh, "Fuzzy Sets*," *Information and control*, vol. 8, no. 3, pp. 338–353, 1965.

[28] L. A. Zadeh, "Fuzzy logic = computing with words," *IEEE Transactions on Fuzzy Systems*, vol. 4, no. 2, 1996.

[29] Sinalgo. Last Access: 21/02/2012. [Online]. Available: http://dcg.ethz.ch/projects/sinalgo/

[30] W. R. Heinzelman, A. Chandrakasan, and H. Balakrisnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Transactions on Wireless Communications*, vol. 1, pp. 660–670, 2002.

Peak-to-average power ratio reduction scheme in impulse postfix OFDM system

Byung Moo Lee
Infra Laboratory
KT
Seoul, Korea
Email: blee@kt.com

Youngok Kim
Department of Electronic Engineering
Kwangwoon University
Seoul, Korea
Email: kimyoungok@kw.ac.kr

***Abstract*—A recently introduced impulse postfix OFDM (IP-OFDM) system achieves the enhanced bit error rate (BER) performance compared to that of conventional OFDM systems, but there is an important peak-to-average power ratio (PAPR) issue of using impulse postfix (IP) that needs to be resolved. This paper proposes a combined IP-OFDM scheme with the selected mapping technique and the optimum power boosting factor (PBF) determination method to resolve the PAPR issue while achieving the enhanced BER performance. In this paper, the effectiveness of proposed scheme is analyzed in terms of the BER performance as well as the input back-off (IBO) to high power amplifier. The analytic results show that the proposed scheme provides the remarkable BER performance enhancement with relatively low IBO (or PBF) rather than with high IBO (or PBF).**

***Index Terms*—Impulse Postfix OFDM (IP-OFDM), Peak-to-Average Power Ratio (PAPR), SLM, Power boosting factor.**

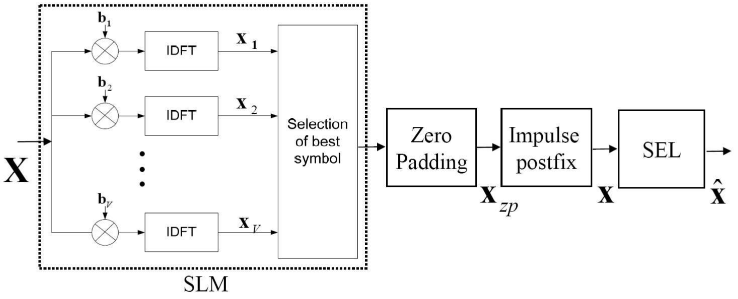


Fig. 1. Simplified block diagram for the proposed scheme

I. Introduction

A novel channel estimation technique for OFDM systems, which is called as impulse postfix OFDM (IP-OFDM), has recently been introduced [1], [2]. The IP-OFDM system exploits the IP, which consists of a high power impulse sample and several zero samples at the end of a zero padded-OFDM symbol block, to estimate channel impulse responses in time-domain. As shown in [1], [2], the IP-OFDM system achieves the enhanced bit error rate (BER) performance compared to that of conventional OFDM systems. However, there is an important peak-to-average power ratio (PAPR) issue, which can degrade the BER performance of IP-OFDM systems due to the nonlinear distortion of the IP. For this reason, an optimum power boosting factor (PBF) determination method for IP was proposed to avoid nonlinear distortion of IP [3]. The PBF of IP should be decided using the CCDF and the IBO, which are determined by HPA characteristics. It is also shown that boosting the IP without considering these factors may degrade the BER performance significantly.

In this paper, we propose a combined IP-OFDM scheme with a selected mapping (SLM) technique and the optimum PBF determination method to resolve the PAPR issue, while achieving the enhanced BER performance. In the scheme, the signal with minimum PAPR is selected from the generated multiple signals with different PAPRs and then, the optimum PBF of IP is determined to enhance the BER performance of the system. According to the analytic results, the proposed scheme provides the remarkable BER performance enhancement with relatively low input back-off (IBO) rather than with high IBO.

II. System description

The time-domain representation of OFDM symbol with N subcarriers can be represented as follows:

$$x(t) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X[k] e^{j2\pi f_k t},\ 0 \le t \le T_s, \tag{1}$$

where T_s is the duration of the OFDM symbol and $f_k = \frac{k}{T_s}$.

The IP-OFDM proposed in [1] is a modified OFDM system, which adds an impulse sample and several zero samples at the end of a zero padded-OFDM symbol [4] for time-domain channel estimation. If we assume the length of the guard interval is L, we can represent the time-domain IP-OFDM symbol vector $\mathbf{u}$ as follows:

$$\begin{aligned} \mathbf{u} &= [u(0), u(1), \cdots, u(N+2L)]^T \\ &= [\mathbf{x}, \mathbf{g}, \mathbf{p}]^T, \end{aligned} \tag{2}$$

where $\mathbf{x}$ is the $1 \times N$ OFDM data vector, $\mathbf{g}$ is the $1 \times L$ zero vector, $\mathbf{p}$ is the $1 \times (L+1)$ IP for channel estimation, and T represents transpose operation. The IP, $\mathbf{p}$, is composed of an impulse sample, s, and L zeros, and can be expressed as follows:

$$\mathbf{p} = [s, \mathbf{0}_{(1\times L)}], \tag{3}$$

where $\mathbf{0}_{(1\times L)}$ is the $(1 \times L)$ zero vector.

ISBN: 978-1-61208-245-5

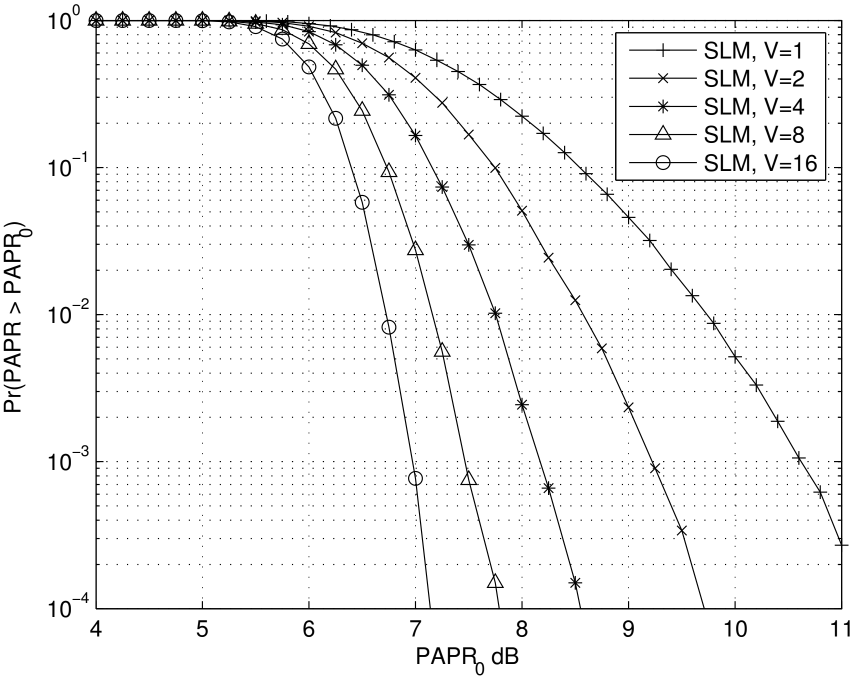


Fig. 2. PAPR performance of SLM with various V.

III. Proposed scheme

A. System Framework

The simplified block diagram for a transmitter part of the proposed scheme is shown in Fig. 1. A SLM block and a soft envelope limiter (SEL) are employed to obtain the signal with the reduced PAPR and to represent the nonlinearity of high power amplifier (HPA), respectively. In the figure, $\mathbf{X}$, the input symbol of SLM, represents a QPSK/QAM modulated baseband signal and $\mathbf{b_i}$, $i = 1, 2, \cdots, V$, represents V different phase sequences to be multiplied with $\mathbf{X}$. Each phase sequence consists of N phases, which are selected from $\{\pm 1\}$ for simplicity. Note that the length of both $\mathbf{X}$ and $\mathbf{b_i}$ are same with the number of subcarriers N. If a signal block $\mathbf{X}$ is multiplied with V different phase sequences and the inverse discrete Fourier transform (IDFT) is applied to each block, then V different OFDM signal blocks are generated from one original signal block. Since the PAPR of OFDM signal is very sensitive to phase variation, the V different signal blocks have different PAPRs. Therefore, we can select the signal with minimum PAPR from the generated multiple signals with different PAPRs [5]. After that, zero samples and IP are added to constitute IP-OFDM signal [1].

Generally, the PAPR in the digital domain is not necessarily the same as the PAPR in the analog domain, where nonlinear distortion due to high PAPR occurs. However, it is shown that the PAPR in the analog domain can be closely approximated by oversampling the signal in the digital domain [6]. Usually, an oversampling factor $D = 4$ is sufficient to approximate the PAPR in the analog domain. Therefore, we can express the PAPR of the OFDM signal as follows:

$$PAPR = \frac{\max_{0 \le n \le DN-1} |x(n)|^2}{E(|x(n)|^2)}, \tag{4}$$

where $E(\cdot)$ denotes the expectation operator.

B. Analysis of CCDF

It is known that the CCDF(Complementary Cumulative Distribution Function) of the SLM-OFDM signal can be represented as follows [7]:

$$\begin{aligned} CCDF &= P(PAPR \ge PAPR_0) \\ &\approx (1 - (1 - e^{-PAPR_0})^{\alpha N})^V, \end{aligned} \tag{5}$$

where α is an adjustment factor for the close approximation of the CCDF of OFDM signal, N and V are the numbers of subcarriers and different phase sets, respectively. As we can see from (5), if the number of subcarriers, N, is increased, the PAPR of OFDM signals is increased. On the other hand, if the number of phase set, V, is increased, the PAPR of OFDM signals is reduced as shown in Fig. 2

Fig. 3 shows the CCDF simulation results of the signal of proposed scheme, when the randomly generated four different phase sets $V = 4$ and various PBFs, p = 0, 2, 5, 6, 7, 8 dB, are assumed. In the simulations, we use QPSK modulation with $N = 64$ subcarriers and $D = 4$ oversampling factor. As shown in the figure, the PAPR of the signal of the proposed scheme is remarkably reduced compared to that of the original signal and a HPA that is linear up to around 8 $\sim$ 9dB is good enough if allowable nonlinear distortion probability is 10^{-4}.

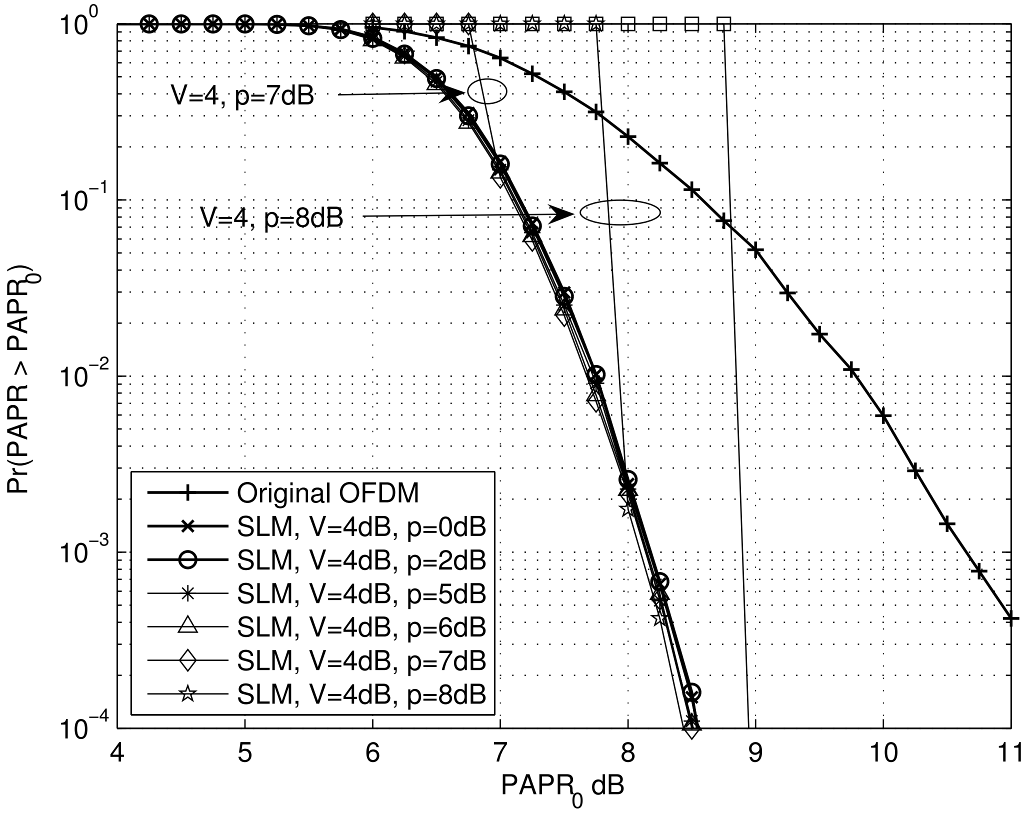


Fig. 3. CCDF of the signal of proposed scheme, when p=0, 2, 5, 6, 7, 8 dB.

Note that the actual PBF, p, needs not to be reduced because the maximum allowable amplitude is given with the HPA in real systems, when the PAPR of OFDM signals is reduced.

To represent the nonlinearity of HPA, the SEL, which is equivalent to Rapp's SSPA model with an infinity smoothness factor [8], is employed. With the input signal $x(n)$, the output signal $\hat{x}(n)$ of the SEL can be represented as follows [6]:

$$\hat{x}(n) = \begin{cases} x(n), & |x(n)| \leq A_{\max} \\ A_{max}e^{j\phi(n)}, & |x(n)| > A_{\max} \end{cases} \tag{6}$$

where A_{max} is the maximum allowable amplitude without distortion and $\phi(n)$ is the phase of the input signal.

In [3], it is shown that the PBF of IP should be carefully determined with two criteria, the CCDF of signal and the value of IBO defined as follows [9]:

$$IBO(dB) = 10\log_{10}\left(\frac{A_0^2}{P_{in}}\right), \tag{7}$$

where A_0 is the maximum allowable input amplitude and P_{in} is the average input power of the OFDM signal before going through the HPA. In the SEL, A_{max} is equivalent to A_0, which is the maximum allowable amplitude in the HPA, and is fixed regardless of the PAPR of the input signals. As shown in [3], therefore, the optimum PBF of IP can be determined as the same value of IBO, where the amplitude of the impulse sample is the same with $A_{max}(= A_0)$. That is, even though the PAPR of OFDM signals is reduced by the SLM technique, the optimum PBF is same with IBO, as is without the SLM technique.

IV. Numerical results

Fig. 4 shows the BER performance of the proposed scheme over the frequency selective Rayleigh fading channel with 8-taps exponential power delay profile, where the parameters are set as N=64 subcarriers, $L = 16$ guard interval, 16QAM, four different values of IBO=2, 3, 4, 5dB with the optimum PBF, $p(= IBO)$, and V=1(w/o SLM), 4, 8. We assume perfect SLM side information at the receiver. As shown in the figure, we can enhance the BER performance of IP-OFDM by increasing V in the SLM technique at the low IBO. Note that the BER performance is remarkably enhanced by increasing V 1 to 8 at IBO=2, 3. However, the effect of SLM is significantly reduced at the high IBO=4, 5.

This is because the effect of OFDM PAPR reduction for HPA is reduced as increasing IBO. Note that the power efficiency of HPA, which is another system constraint, is degraded as increasing the IBO to the HPA.

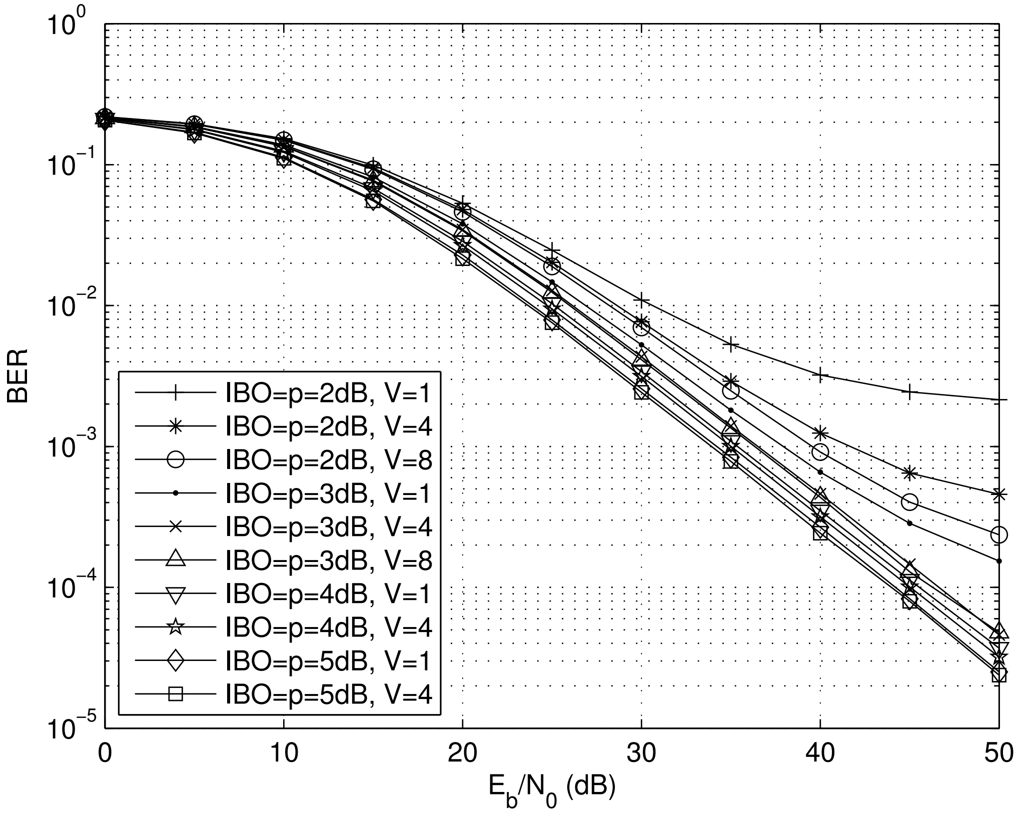


Fig. 4. BER performance of the proposed scheme.

V. Conclusion

A combined IP-OFDM scheme with the SLM technique and the optimum PBF determination method is proposed to resolve the PAPR issue, while achieving the enhanced BER performance. The analytic results show that the proposed scheme can provides the remarkable BER performance enhancement with relatively low IBO. On the other hand, the BER performance can be enhanced by increasing the IBO at the cost of degradation of the power efficiency of HPA. As a future work, a practical IP-OFDM system with a nonlinear HPA will be considered and the impact of nonlinear HPA on the determination of the PBF of IP as well as the BER performance will be analyzed.

Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (No.2012-0001840).

References

[1] N. Chang and J. Kang, "Impulse Symbol Based Channel Estimation in OFDM Systems," *IEEE Internation Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 1-5, Sep. 2006.

[2] N. Chang, N.Y. Kim, J. Kang, Y. Kim, and H. Lee, "Statistic-Based magnitude determination of impulse sample in impulse postfix OFDM Systems," *IEICE Transactions on Communications*, vol.E90-B, no.12, Dec. 2007.

[3] B. Lee and Y. Kim, "Practical determination of impulse sample power boosting factor in impulse postfix OFDM systems," *IEEE Communications Letters*, pp. 187-189, Mar. 2009.

[4] B. Muquet, Z. Wang, G.B. Giannakis, M. de Courville, P. Duhamel, "Cyclic prefixing or zero padding for wireless multicarrier transmissions?," *IEEE Transactions on Communications*, vol. 50, Issue 12, pp. 2136-2148, Dec. 2002.

[5] R.W. Bauml, R.F.H. Fischer, and J.B. Huber, "Reducing the peak-to-average power ratio of multicarrier modulation by selected mapping," *Electronics Letters*, vol.32, pp. 2056-2057, Oct. 1996.

[6] L. Wang and C. Tellambura, "A Simplified Clipping and Filtering Technique for PAR Reduction in OFDM Systems," *IEEE Signal Processing Letters*, vol.12, no.6, Jun. 2005.

[7] H. Ochiai and H. Imai, "Performance of the deliberate clipping with adaptive symbol selection for strictly band-limited OFDM systems," *IEEE Journal on Selected Areas in Communications*, vol.18, pp.2270-2277, Nov. 2000.

[8] C. Rapp, "Effect of HPA-nonlinearity on 4-DPSK/OFDM-signal for a digital sound broadcasting system," *Proc. the Second European Conference on Satellite Communications*, Liege, Belgium, pp.179-184, Oct. 1991.

[9] B. Lee and R.J.P. de Figueiredo, "Adaptive Pre-Distorters for Linearization of High Power Amplifiers in OFDM Wireless Communications," *Circuits, Systems and Signal Processing*, Birkhauser Boston, vol.25, pp. 59-80, 2006.

An Approach for Reduction of the Security Overhead in Smart Grid Communication Infrastructure Employing Dedicated Encryption

Miodrag J. Mihaljević
Mathematical Institute, Serbian Academy
of Sciences and Arts, Belgrade, Serbia, and
RISEC, National Institute AIST, Tsukuba, Japan
Email: miodragm@turing.mi.sanu.ac.rs

Aleksandar Kavičić
Department of Electrical Engineering
University of Hawaii
Honolulu, USA
Email: kavcic@hawaii.edu

***Abstract*—This paper considers an approach for partial reduction of the overhead implied by request for data security within the Smart Grid communications infrastructure. A significant part of the implementation and processing overhead appears as a consequence of the data confidentiality request and the related requirement for massive data encryption. Accordingly, this paper points out to the related requirements and employment of certain light-weight and highly secure encryption dedicated to the noisy communication channels.**

***Keywords*—Smart Grid; information-communication infrastructure; overheads; data confidentiality; light-weight encryption; randomness; coding.**

I. Introduction

Generally, a Smart Grid is an autonomous system consisting of an information collection network, a data management center, a power grid control center and power generation and transmission infrastructures (see [16], for example). The information collection network is a complex network involving a multi-hop ad hoc network, WiFi, cellular network, and Internet. The sensor nodes, such as Phasor Measuring Units (PMUs) and Smart Meters (SMs), are deployed over the power grid to monitor the states of the system.

A communication infrastructure is an essential part in the Smart Grid and a scalable and pervasive communication infrastructure is crucial for the operation of a Smart Grid. To ensure the correct functionality of a Smart Grid, it is essential that communications are secured, devices are protected, and privacy is respected. Two main requirements regarding communications are data authentication and confidentiality, and the information-communication infrastructure as a whole must be robust. Note that confidentiality of communications also support the privacy of Smart Grid customers.

In certain domains of information-communications infrastructure of the Smart Grid, the communications channels suffer from unavoidable and high noise (see [2][7], for example). A particular example are the floating wind turbines where only wireless and power line cable (PLC) are available for communications and control purposes, and both of these channels are extremely noisy.

On the other hand, it is interesting to address the following issues: utilization of the inherent noise for design dedicated cryptographic algorithms (based on which the security mechanism are built), when appropriate, to employ the available error-correction coding within a cryptographic algorithm.

Extensive employment of cryptography as ”a must” implies request for light-weight cryptographic primitives in order to minimize the overheads. Overheads as a consequence of information security requests can be listed as follows: (i) implementation overhead; (ii) computational overhead; (iii) communications overhead.

Motivation for the Work. In order to reduce the overheads in a lot of scenarios the requirement is employment of the lightweight cryptographic techniques On the other hand, also in a lot of scenarios, the requirement is high (provable, preferably) security of the employed cryptographic primitives; particularly the above requirements appear regarding a number cryptographic techniques required in the Smart Grid. As illustrations of the previous statements, note the following: (i) a lot of IT components requires power supply from batteries, because it is too expensive to equip each tiny IT device with AC to DC convertor; (ii) the smart-meters are two-way communication-control devices which could remotely control the power availability at a home, and in order to avoid potential disastrous impacts of malicious control the employed cryptographic techniques should be highly secure and preferably provably secure ones.

Organization of the Paper. A background on Smart Grid relevant for this paper is given in Section II, and Section III discusses corresponding data security issues and implications. Section IV points out to the noisy nature of the main communication channels within Smart Grid. An approach for light-weight and dedicated encryption which fits into the security and implementation requirements is given in Section IV, and its implementation complexity is considered in Section V. Concluding discussion is given in Section VI.

II. A Background on Smart Grid

In the Smart Grid, data collection is performed by measuring devices, including PMUs and SMs. The data manage-

ISBN: 978-1-61208-245-5

ment center communicates with the sensors and the control centers through the network. It analyzes the information of the power grid and makes corresponding decisions. The power grid control centers receive instructions from the data management center and rule the power system according to the received instructions. The whole system works in a real-time manner, which implies real-time situation awareness, real-time response, and real-time control.

According to the above discussion, Fig. 1 points out the important two-way nature of the information transmission within the information-communications sub-infrastructure related to SMs (advanced metering infrastructure - AMI), and the required basic cryptographic techniques for providing security of this infrastructure.

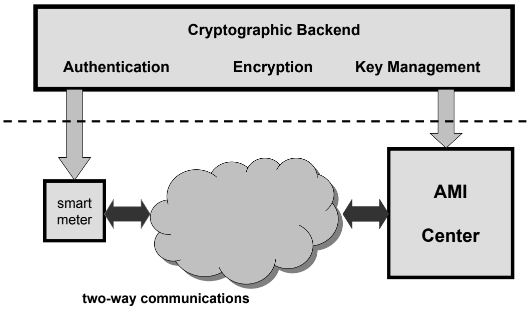


Figure 1. Illustrative model of security issues regarding communications within advanced metering infrastructure (AMI).

In addition to the above, particularly note the following. The measuring nodes are usually implemented as embedded systems to perform data processing and two-way communications. Due to the vast amount of deployment, each node is relatively cheap and simple. Thus, they have very limited on-board resources including power availability. For example, they usually have one simple low speed Micro Control Unit (MCU) as the processor, very limited memory space and very tight power budget due to their deploying areas, cost and physical sizes. The MCU is a small computer integrated on a single chip. It runs programs to support computation and control tasks in an embedded system. The MCUs are usually much simpler than the CPUs for the general purpose computers. Therefore, only some basic tasks, like simple computation and communication, can be implemented on these nodes.

Also, basically, most of the PMUs and SMs are deployed ”in the wild”. The collected data are usually transmitted through wireless links that rely on the open media (i.e., wireless channel). The adversaries can easily physically destroy or replicate these devices by capturing them. The attackers can also launch attacks by setting up some hacker equipment.

III. Data Security Issues and Implications

A. Preliminaries

The network based monitoring system provides efficient remote monitoring and control, but also exposes the Smart Grid to potential cyber attacks. On the other hand, misleading information (or stale information) can cause severe (even disastrous) consequences to the system, as well as to the customers. Illustrations of the previous claim are given as follows. For example malicious modification of phasor information can cause wrong management operations. Malicious analysis against the smart meter data can reveal the living schedule of the householders or production activities of a plant. In an attacking scenario, which should not be excluded due to the interconnection of the system, the terrorists could collect (for example) 80% of the sensitive information that can be used to mount further attacks on the whole Smart Grid. Accordingly, communication security which provides integrity, authenticity, availability and confidentiality over the whole system has to be enforced by appropriate cryptographic algorithms (as illustrated in Fig. 1).

B. On the Need for Advanced Cryptographic Techniques for Information Security within the Smart Grid

A consideration of suitable cryptographic components could begin from the following simple question: Do we face any specific security requirement regarding the Smart Grid information-communication infrastructure, in comparison with traditional security requirements for wireless networks, for example. If we claim ”Yes”, a natural follow-up question is: ”Why”. On the other hand, if we need advanced cryptographic techniques a natural question is: ”What are the related requirements”. This section addresses the previously mentioned issues.

First, we stress the following:

- Security requirements should be related to the impacts of potential security problems. We can identify a number of very different impacts of the security problems in the Smart Grid information-communication infrastructure in comparison with a mobile telephony network, for example.
- Compromising the security of the information-communication infrastructure could imply a collapse of the Smart Grid, and recovering the Smart Grid is much more complex then recovering a wireless network. Accordingly, the Smart Grid information-communication infrastructure requires stronger security in comparison with a wireless network because of possible more severe impacts of compromising the security.

Regarding the question ”Do we need cryptographic techniques different than currently standardized ones”, we emphasis the following:

- For example, if AES (Advanced Encryption Standard) is enough, why we do not employ it in mobile telephony and in a number of other possible applications;

- If the existing cryptographic techniques are enough than we already have the best ones, and this is very unlikely ...

C. Requirements on Cryptographic Techniques for the Smart Grid

In a number of application scenarios within the Smart Grid, an important request is employment of light-weight cryptographic algorithms in order to reduce the overhead to the system implied by involved cryptographic mechanisms (see, for example, discussions in [5][12][16]). At the same time, beside the light-weightiness, the employed cryptographic algorithms should be highly secure. These issues are elaborated in more details as follows.

1) Light-Weight and Highly Secure: Employed cryptographic techniques:

- should provide low overhead implied by implementation of cryptographic elements - massive employment of cryptographic techniques if they are not light-weight, cumulatively could imply high overhead, and particularly regarding extensive device-to-device (M2M) communications; particularly, note that complex encryption algorithms also imply heavy power consumption and when the power is obtained from batteries a consequence is the excessive drain on batteries.
- should not be a weak point because impacts of compromised security could be much more stronger in comparison with other systems - a Smart Grid requires employment of strong cryptographic primitives because impacts of compromising a cryptographic primitive could be disastrous.

2) Dedicated: In order to achieve the above implementation and security requirements, we need advanced and dedicated cryptographic primitives which are light-weight and provably secure ones in order to provide/support efficient and effective Smart Grid security and privacy; taking into account the entire overhead which security requirements imply, dedicated cryptographic techniques that meet security requirements and minimize the overhead are very welcome.

IV. Noisy Communication Channels in the Smart Grid

In certain domains of information-communications infrastructure of the Smart Grid, the communications channel suffer from unavoidable and high noise. A particular example are the floating wind turbines where only wireless and power line cable (PLC) are available for communications and control purposes, and both of these channels appear as very noisy. Accordingly, for reliable communications we need an adequate error-correction coding scheme. On the other hand it is interesting to address the issues of employment the inherent noise for design dedicated cryptographic algorithms based on which the security mechanism are built, as well as, when appropriate, to employ the available error-correction coding within a cryptographic algorithm.

Particularly, note the following communications problems regarding floating wind turbines.

- Highly reliable (low noise) communication channels are not available, and standard Internet-like communication channels are not available.
- Basically, there are only two communication options: Wireless or via Power Line Cable (PLC). Both options suffer from high noise (as discussed in [2][7], for example).
  - Wireless Channel: Mostly assumed AWGN, presence of impulsive noise in certain environments
  - PLC Channel: More complicated noise structure: colored background noise, narrow band noise and impulsive noise.

V. An Approach for Dedicated Encryption

As discussed in the previous sections, in certain Smart Grid scenarios we need light-weight and highly secure cryptographic components and at the same instance we face a noisy implementation environment. This and the next section address design and analysis of a cryptographic algorithm for encryption which employ the channel noise for the cryptographic security enhancement. Light-weight cryptographic algorithms are very important but currently available ones suffer from security weaknesses (as an illustration, see [13] and [14]). On the other hand, it has been shown that physical noise could play a supporting role in cryptographic security enhancement (see [8][11], for example). This section points out to an encryption scheme suitable for the scenarios where (i) the requirement is employment of lightweight (in order to reduce the overweight implied by the employed cryptography) and highly secure algorithms; (ii) physical noise is available for enhancing the security. It is assumed that the symmetric key management is based on a pre-distribution paradigm.

A. Encryption and Decryption Algorithms

The approach [15] pointed out in this section yields a framework for achieving light-weight implementation and processing complexity and a high cryptographic security level implied by employment of the randomness which appears as a supporting element for enhancing the security implied by hardness of the so called LPN problem (Learning Parity with Noise).

We assume the following notation:

- $\mathbf{a} = [a_i]_{i=1}^{\ell}$: ℓ-dimensional binary vector of message/plaintext data;

- $\mathbf{r} = [r_i]_{i=1}^{m-\ell}$: $(m-\ell)$-dimensional binary vector of random data where each r_i is a realization of the binary random i.i.d. variable R_i such that $\Pr(R_i = 1) = \Pr(R_i = 0) = 1/2$, $i = 1, 2, ..., n$;

- $\mathbf{u} = [u_i]_{i=1}^{k}$: k-dimensional binary vector of random data

where each u_i is a realization of the binary random i.i.d. variable U_i such that $\Pr(U_i = 1) = \Pr(U_i = 0) = 1/2$, $i = 1, 2, ..., k$;
- $\mathbf{S} = [s_{i,j}]_{i=1}^{k}{}_{j=1}^{n}$: $k \times n$-dimensional binary matrix of the secret key
- $\mathbf{v} = [v_i]_{i=1}^{n}$: n-dimensional binary vector of random data where each v_i is a realization of the binary random i.i.d. variable V_i such that $\Pr(V_i = 1) = p$ and $\Pr(V_i = 0) = 1 - p$, $i = 1, 2, ..., n$;
- $C_H(\cdot)$ and $C_H^{-1}(\cdot)$: operators of the homophonic encoding and decoding, respectively; $C_H(\cdot)$ denotes a mapping $\{0,1\}^m \rightarrow \{0,1\}^m$;
- $C_{ECC}(\cdot)$ and $C_{ECC}^{-1}(\cdot)$: operator of the error-correction encoding and decoding, respectively; $C_{ECC}(\cdot)$ denotes a mapping $\{0,1\}^m \rightarrow \{0,1\}^n$.

This section points out to a symmetric key encryption scheme [15], where the encryption and decryption operations are specified by the following.

- *Encryption*
  1) Employing $\mathbf{r}$, perform the homophonic encoding of $\mathbf{a}$, and the error-correction encoding of the resulting vector as follows: $C_{ECC}(C_H(\mathbf{a}||\mathbf{r}))$ where $||$ denotes concatenation.
  2) Generate the ciphertext in form of n an dimensional binary ciphertext vector $\mathbf{z}$ as follows:

$$\mathbf{z} = C_{ECC}(C_H(\mathbf{a}||\mathbf{r})) \oplus \mathbf{u} \cdot \mathbf{S} \oplus \mathbf{v} \ . \qquad (1)$$

- *Decryption*
  Assuming availability of the pair $(\mathbf{u}, \mathbf{z})$ decrypt the ciphertext as follows:

$$\mathbf{a} = tcat_\ell(C_H^{-1}(C_{ECC}^{-1}(\mathbf{z} \oplus \mathbf{u} \cdot \mathbf{S}))) \ , \qquad (2)$$

  where $tcat_\ell(\cdot)$ denotes truncation of the argument vector to the first ℓ bits and the assumption is that the employed code which corresponds to $C_{ECC}(\cdot)$ and $C_{ECC}^{-1}(\cdot)$ can correct the errors introduced by a binary symmetric channel with the crossover probability p.

Note that the random vector $\mathbf{u}$ is a public one, and the decryption part assumes availability of the pair $(\mathbf{u},\ \mathbf{z})$. Also note that the decryption does not require knowledge of $\mathbf{r}$.

The proposed encryption scheme is displayed in Fig. 2.

B. Algebraic Structure Assuming Employment of Linear Codes

Encoding Issues. When the employed homophonic and error-correcting codes are linear, the encoding operations in both cases are vector-matrix multiplications. Accordingly, the encoded version of $\mathbf{a}||\mathbf{r}$ is given by the following:

$$C_H(\mathbf{a}||\mathbf{r}) = [\mathbf{a}||\mathbf{r}]\mathbf{G}_H, \qquad (3)$$

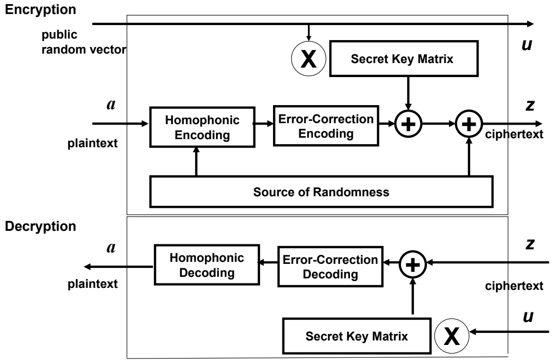


Figure 2. Encryption scheme [15] which involves randomness and dedicated coding.

and $\mathbf{G}_H$ is an $m \times m$ matrix, and thus

$$\begin{aligned} C_{ECC}(C_H(\mathbf{a}||\mathbf{r})) &= C_{ECC}([\mathbf{a}||\mathbf{r}]\mathbf{G}_H) \\ &= [\mathbf{a}||\mathbf{r}]\mathbf{G}_H\mathbf{G}_{ECC} \\ &= [\mathbf{a}||\mathbf{r}]\mathbf{G} \qquad (4) \end{aligned}$$

where $\mathbf{G}_{ECC}$ is an $m \times n$ binary generator matrix corresponding to $C_{ECC}(\cdot)$, and $\mathbf{G} = \mathbf{G}_H\mathbf{G}_{ECC}$ is an $m \times n$ binary matrix summarizing the two successive encodings at the encryption side, implying that

$$\mathbf{z} = [\mathbf{a}||\mathbf{r}]\mathbf{G} \oplus \mathbf{u} \cdot \mathbf{S} \oplus \mathbf{v} \ . \qquad (5)$$

Decoding Issues. Assuming that the employed error-correction code can correct all the errors introduced by the vector $\mathbf{v}$ we have

$$C_{ECC}^{-1}(\mathbf{z} \oplus \mathbf{u} \cdot \mathbf{S}) = C_{ECC}^{-1}(C_{ECC}(C_H(\mathbf{a}||\mathbf{r})) \oplus \mathbf{v}) = [\mathbf{a}||\mathbf{r}]\mathbf{G}_H \qquad (6)$$

and accordingly,

$$C_H^{-1}(C_{ECC}^{-1}(\mathbf{z} \oplus \mathbf{u} \cdot \mathbf{S})) = [\mathbf{a}||\mathbf{r}]\mathbf{G}_H\mathbf{G}_H^{-1} = [\mathbf{a}||\mathbf{r}] \ , \qquad (7)$$

implying that

$$tcat_\ell(C_H^{-1}(C_{ECC}^{-1}(\mathbf{z} \oplus \mathbf{u} \cdot \mathbf{S}))) = \mathbf{a} \ . \qquad (8)$$

C. A Summary of Security Issues

The encryption technique [15] is based on a framework for enhancing security of light-weight stream ciphers employing randomness and dedicated coding. Security evaluation of the considered framework has been discussed from information-theoretic and computational-complexity points of view in [15]. Regarding the information-theoretic approach, the equivocation of the secret key is analyzed. The computational-complexity evaluation approach shows that recovering of the secret key appears as hard as decoding of certain general linear block codes, i.e. certain problems of learning the parity in noise (the LPN problem) assuming appropriate design a linear block codes, which provide joint error-correction and homophonic coding.

LPN based schemes offer a very strong security guarantee. The LPN$_{k,\epsilon}$ problem is equivalent to the problem of decoding certain random linear code (k, n) after the binary symmetric channel with the crossover probability (”noise level”) equal to ϵ, a problem that has been extensively studied in the last half century and which is provably NP-complete in the worst case scenario as shown in [3]. On the other hand hardness of the LPN problem in the average case has been studied and still, the fastest known algorithms run in exponential time.

VI. Implementation Issues

This section discusses the implementation resources required and complexity of the considered encryption/decryption which shows that the entire implementation is light-weight. Particularly, we point out to a technique for the inner product evaluation which provides an opportunity for a trade-off between time and space implementation complexity. Finally, a brief summary of the Advanced Encryption Standard (AES), relevant for comparison with the approach given in this paper regarding the implementation, is given.

A. Implementation Requirements

The implementation of the considered encryption requires the following: (i) a source of randomness, (ii) suitable error correction code; (iii) suitable homophonic code, and (iv) resources for the computations over GF(2). The following discussion show that the above requirements fit into a framework of the light-weight encryption.

Regarding requirement for a source of randomness, note that it is the same as requirement discussed regarding HB-authentication protocols (see [9], for example) designed for resource limited implementation scenarios like RFID ones, and accordingly the assumption on availability of a ”light-weight” source of randomness is justified, particularly noting that the generation of randomness could be supported by the assumed environmental noise.

Following [1], the ECC reported in [17] can be employed in the proposed cipher as well. Particularly note that the codes reported in [17], have the property that the encoding can be computed via a circuit of size $O(n)$ and the decoding can be decoded by a circuit of size $O(n\log_2 n)$ making them the candidate codes.

Regarding the required homophonic coding we point out to the following. Homophonic coding or ”multiple substitution” (see [10], for example) is a technique for mapping source data employing certain random bits into the encoded data which are the randomized form of the source ones so that the source data can be recovered from the noise-free encoded ones without knowledge of the random bits. Homophonic encoding provides that many particular outputs of encoding become possible substitutes (or ”homophones”) of the source data based on employment of different random sequences. Perfect homophonic code provides that the encoded data appear as truly random ones.

A particular class of homophonic codes are the universal ones reported in [10]: These codes provide the randomization without knowledge of the source data statistics which is a request for some homophonic coding schemes. The source data can be recovered from the homophonic encoder output without knowledge of the randomizing data by passing the encoded data through the decoder and then discarding the randomizer bits.

Finally, note that the Wire-tap channel coding [18] is based on assigning multiple codewords to the same information vector and from that point of view, particularly when the main channel is noise-free, it shares the same underlying idea employed in the homophonic coding.

Next subsection discusses some of the issues regarding implementation of the operations over GF(2).

B. Discussion of the Encryption and Decryption Operations

We assume that implementation of the proposed scheme is based on employment of light-weight source of randomness, based on the existing channel noise, and error-correction coding. Particularly, we assume that a low-implementation complexity linear error correcting code is employed: such coding scheme has encoding and decoding complexities linearly proportional to the codeword length n. Also note that the proposed scheme requires only one additional (homophonic) encoding at the encryption side and one additional (homophonic) decoding at the decryption side

Accordingly, the algebraic representation of encryption and decryption, when liner codes are employed, implies the following: (i) Encryption requires $mn + kn$ binary multiplications and $mn + kn + n$ mod 2 additions (see (5)); (ii) Decryption requires $kn+m^2$ binary multiplications, $kn + n + m^2$ mod 2 additions, and $O(n)$ operations for decoding of the employed linear error correcting code.

Also, note the following: When the linear codes are employed, the algebraic representation of encryption and decryption directly shows that the implementation complexity is dominated by the required number of the inner products between binary vectors. The next section points out that we can employ dedicated look-up tables instead of binary multiplications and additions for obtaining (if appropriate) certain trade-offs between the time and space implementation complexity.

C. An Implementation of the Inner Products Evaluation

Here is given an approach for time-memory trade-off based on a read-only memoriy (which play role of a number of look-up tables) In a look-up table implementation, all the possible outputs of the function are pre-calculated and stored in the memory. Each time the output of an input is looked up in the memory instead of being calculated calculated. The look-up table can save the computation operations at cost of storage space.

Let $\mathcal{A}$ be a set of binary vectors $\mathbf{a} = [a_i]_{i=1}^{n}$, and $\mathcal{B}$ be a set of binary vectors $\mathbf{b} = [b_i]_{i=1}^{n}$ with the cardinalities $|\mathcal{A}| >> |\mathcal{B}|$.

Any inner product

$$\mathbf{a}\cdot\mathbf{b} = \bigoplus_{i=1}^{n} a_i b_i \quad , \tag{9}$$

can be considered as

$$\mathbf{a}\cdot\mathbf{b} = \bigoplus_{j=0}^{\frac{n}{\Delta}-1} (\bigoplus_{i=1}^{\Delta} b_{j\Delta+i}\ a_{j\Delta+i}) \quad , \tag{10}$$

assuming that $\frac{n}{\Delta}$ is an integer. On the other hand, each sub-sum $\bigoplus_{i=1}^{\Delta} b_{j\Delta+i}\ a_{j\Delta+i}$ can be considered as a liner Boolean function of Δ arguments. Accordingly, the inner product (9) can be considered as the modulo 2 sum of of the outputs of Δ linear Boolean functions. Taking into account that $|\mathcal{A}| >> |\mathcal{B}|$, it is suitable to consider that a segment of $\mathbf{b}$ specifies a linear Boolean function $f_j(\cdot)$, and a segment of $\mathbf{a}$ the arguments of this Boolean function. Accordingly, we have:

$$f_{j,b}([a_{j\Delta+i}]_{i=1}^{\Delta}) = \bigoplus_{i=1}^{\Delta} b_{j\Delta+i}\ a_{j\Delta+i} \ . \tag{11}$$

Consequently, the inner product (10) appears as

$$\mathbf{a}\cdot\mathbf{b} = \bigoplus_{j=0}^{\frac{n}{\Delta}-1} f_{j,b}([a_{j\Delta+i}]_{i=1}^{\Delta}) \ . \tag{12}$$

It is well known that any Boolean function of Δ arguments (see [4], for example) can be implemented employing a look-up table of dimension 2^{Δ}. Also, the cumulative inner product (12) can be evaluated employing a binary look-up table of dimension $2^{n/\Delta}$. (Of course, instead of one-step look-up table evaluation of (12), a multiple step approach could be considered employing a cascade of look-up tables, but this approach is out of the scope of the current consideration.)

According to the above discussion, any of the considered inner products can be evaluated with time complexity $O(1)$ employing $\frac{n}{\Delta}$ binary look-up tables each of dimension 2^{Δ}, and an additional look-up table of dimension $2^{n/\Delta}$. So, the total space complexity C_S of the above approach for the inner products evaluation is upper-bounded as follows:

$$C_S \le 2^{n/\Delta} + |\mathcal{B}|\frac{n}{\Delta}2^{\Delta} \tag{13}$$

where m is a parameter. In order to have a balance between the space complexity required for the evaluation of partial inner products (11) and the cumulative one (12), we have the following requirement:

$$2^{n/\Delta} \le |\mathcal{B}|\frac{n}{\Delta}2^{\Delta} \tag{14}$$

which for the given parameters $|\mathcal{B}|$ and n yields the maximal $\frac{n}{\Delta}$ such that (14) is fulfilled.

D. Elements for an Illustrative Comparison

A number of different encryption schemes are employed in Smart Grid: Some are proprietary algorithms (not disclosed) and some are the standardized/recommended ones like Advanced Encryption Standard (AES).

For the illustrative preliminary comparison of the proposed encryption technique and the currently recommended ones, we point out to the recommendation from [6], where AES is pointed out as an encryption algorithm and discussions of its energy consumption from [16].

AES is an encryption algorithms which generates the ciphertext through a number of iterative recalculations. AES with 128-bit secret key consists of the following operations: (i) $Key\_Expansion$ round keys are derived from the secret key; (ii) Initial Round each byte of the state is combined with the round key using bitwise xor (iii) 10 Rounds each consisting of the following four procedures: (a) Sub_Bytes - a non-linear substitution step where each byte is replaced with another according to a lookup table; (b) $Shift\_Rows$ - a transposition step where each row of the state is shifted cyclically a certain number of steps. (c) $Mix\_Columns$ - a mixing operation which operates on the columns of the state, combining the four bytes in each column. (d) Add_Round_Key; (iv) Final Round (no $Mix\_Columns$): Sub_Bytes, $Shift\_Rows$, Add_Round_Key.

The above description illustrates that AES has significantly higher implementation complexity, particularly because AES operates through 10 main rounds plus the initial and final ones, and the employed operations are more complex in comparison with mod 2 additions. This higher implementation complexity, in a number of scenarios, implies a heavy overhead at least regarding the power consumption.

VII. Conclusion

This paper elaborated that Smart Grid communication infrastructure requires low-weight and highly secure cryptographic techniques and particulary the ones for encryption. Also, it is shown that certain communications channels are highly noisy. Taking into account the addressed scenario, this paper points out to a light-weight and highly secure stream cipher encryption technique which employs the unavoidable channel noise for enhancing the cryptographic security.

This paper shows an alternative approach for encryption which is based on joint employment of pseudorandomness, randomness and dedicated coding. The approach is based on an enhanced underlying LPN problem. The LPN problem based schemes provide a background for simple and efficient designs in terms of code-size as well as time and space requirements. This makes them prime candidates for light-weight devices like RFID tags, which are too weak to implement standard cryptographic primitives like the block cipher AES.

This paper points out to a particular light weight symmetric encryption as a component for reduction of the following overheads: (i) implementation overhead; (ii) processing overhead, and (iii) power consumption overhead. At the same time the considered encryption provides a high level of provable security required because of possible high impacts of the broken encryption algorithms and particularly in the scenarios which assume pre-distribution of the symmetric secret keys.

The algorithm proposed for the encryption, within noisy communication channels of the Smart Grid communication infrastructure, provides the required light-weightiness and high security. Particularly, note that the encryption/decryption operations are based only on simple mod 2 additions and table look-up operations which directly imply low complexity (as well as the related overheads) implementation, processing and power-consumption. In order to illustrate its simplicity, the considered encryption is preliminary compared with AES. Quantitative consideration of the complexities as well as an in-details comparison of the proposed approach with the traditional ones is highly dependable on the particular instantiations of implementation constraints, and so is out of the scope of this paper. Accordingly, the paper yields a construction and application framework which is a proposal for an alternative encryption approach which contributes to reduction of the security overheads in certain application scenarios. This proposal could be considered as a background for the planed experimental implementation and analysis.

ACKNOWLEDGMENT

Parts of this work were supported by NSF grant ECCF-0129081.

The authors thank to the anonymous reviewers for their comments and suggestions resulting in an improved presentation of the addressed issues.

REFERENCES

[1] B. Applebaum, D. Cash, C. Peikert and A. Sahai, ”Fast Cryptographic Primitives and Circular-Secure Encryption Based on Hard Learning Problems”, CRYPTO 2009, *Lecture Notes in Computer Science*, vol. 5677, pp. 595-618, Aug. 2009.

[2] J. Anatory, N. Theethayi, R. Thottappillil, M. M. Kissaka, and N. H. Mvungi, ”Broadband Power-Line Communications: The Channel Capacity Analysis,” *IEEE Transactions on Power Delivery*, vol. 23, no. 1, pp. 164-170, Jan. 2008.

[3] E.R. Berlekamp, R.J. McEliece, and H.C.A. van Tilborg, “On the Inherent Intractability of Certain Coding Problems”, *IEEE Trans. Info. Theory*, vol. 24, pp. 384-386, 1978.

[4] T.W. Cusick and P. Stanica, *Cryptographic Boolean Functions and Applications*. Academic Press (Elsevier), San Diaego, USA, 2009.

[5] M.M. Fouda, Z.M. Fadlullah, N. Kato, Rongxing Lu and Xuemin Shen, ”A Lightweight Message Authentication Scheme for Smart Grid Communications”, *IEEE Transactions on Smart Grid*, vol. 2, pp. 675-685, dec. 2011.

[6] ”Guidelines for Smart Grid Cyber Security: Vol. 1, Smart Grid Cyber Security Strategy, Architecture, and High-Level Requirements”, NIST TR 7628, August 2010.

[7] M. Katayama, T. Yamazato, and H. Okada, ”A Mathematical Model of Noise in Narrowband Power Line Communication Systems,” *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 7, pp. 1267-1276, Jul 2006.

[8] Y.S. Khiabani, S. Wei, J. Yuan, and J. Wang, ”Enhancement of Secrecy of Block Ciphered Systems by Deliberate Noise”, *IEEE Transactions on Information Forensics and Security*, vol. 7, pp. 1604-1613, Oct 2012.

[9] E. Kiltz, K. Pietrzak, D. Cash, A. Jain, and D. Venturi, ”Efficient Authentication from Hard Learning Problems”, EUROCRYPT 2011, *Lecture Notes in Computer Science*, vol. 6632, pp. 7-26, 2011.

[10] J. Massey, “Some Applications of Source Coding in Cryptography”, *European Transactions on Telecommunications*, vol. 5, pp. 421-429, July-August 1994.

[11] M.J. Mihaljevic and H. Imai, ”An approach for stream ciphers design based on joint computing over random and secret data,” *Computing*, vol. 85, pp. 153-168, June 2009.

[12] M.J. Mihaljevi´, H. Imai, M. David, K. Kobara and H. Watanabe, ”On Advanced Cryptographic Techniques for Information Security of Smart Grid AMI”, *CSIIRW 2011*, Oak Ridge National Laboratory, Tennessee, USA, 11-14 Oct. 2011, Proceedings, ACM International Conferences Series, Article no. 64, 4 pages, March 2012.

[13] M.J. Mihaljević, S. Gangopadhyay, G. Paul and H. Imai, “State Recovery of Grain-v1 Employing Normality Order of the Filter Function”, *IET Information Security*, vol. 6, no. 2, pp. 55-64, June 2012

[14] M.J. Mihaljević, S. Gangopadhyay, G. Paul and H. Imai, “Internal State Recovery of Keystream Generator LILI-128 Based on a Novel Weakness of the Employed Boolean Function”, *Information Processing Letters*, vol. 112, no. 21, pp. 805-810, November 2012.

[15] M.J. Mihaljević, ”An Approach for Light-Weight Encryption Employing Dedicated Coding”, *IEEE GLOBECOM 2012, CISS*, Anaheim CA, USA, 03-07 Dec. 2012, Proceedings, pp. 892-898.

[16] M. Qiu, H. Su, Z. Ming and T. Yang, ”Balance of Security Strength and Energy for a PMU Monitoring System in Smart Grid”, *IEEE Communications Magazine*, pp. 142-149, May 2012.

[17] D.A. Spielman, ”Linear-time encodable and decodable error-correcting codes”. *IEEE Trans. Information Theory*, vol. 42, No 6, pp. 1723-1732, 1996.

[18] A.D. Wyner, “The wire-tap channel”, *Bell Systems Technical Journal*, vol. 54, pp. 1355-1387, Oct. 1975.

Yet Another Bounded Broadcasting for Random Key Predistribution Schemes in Wireless Sensor Networks

Aishwarya Mishra
School of Information Technology
Illinois State University
Normal IL 61790 USA
amishra@ilstu.edu

Tibor Gyires
School of Information Technology
Illinois State University
Normal IL 61790 USA
tbgyires@ilstu.edu

Yongning Tang
School of Information Technology
Illinois State University
Normal IL 61790 USA
ytang@ilstu.edu

***Abstract*—Wireless Sensor Networks (WSNs) have many promising applications involving unattended deployment in hostile territories. Random key predistribution schemes (RKPS) have been proposed to secure these networks. RKPS require broadcasting within the secured sensor network for key discovery and key revocation. Unbounded broadcasting in RKPS could incur large transmission and computational overheads and may not be sustainable on sensor node platforms, due to their limited power resources. Since the requests are triggered by unauthenticated nodes, this broadcasting can be exploited by a sabotaging adversary to deliberately exhaust the power on the sensor nodes and prevent them from performing their intended function. Enforcing the maximum value of the TTL (MAXTTL) on all nodes of the sensor networks can be an effective approach to mitigating this potential threat if it does not impede the function of the RKPS key discovery and revocation. In this paper, we model the RKPS sensor network as a Random Geometric Graph (RGG) and investigate the upper bounds on RGG diameter as guidance for MAXTTL on all RKPS key discovery and key revocation broadcasts. The simulation results show that our approach is practical and does not impede its function.**

Keywords- sensor networks, random key predistribution, graph diameter, random graph, theoretical bound.

I. Introduction

Wireless sensor networks (WSNs) comprise of a large population of inexpensive battery-powered sensor nodes that are deployed randomly in a large area. Each node communicates through a wireless radio interface with other neighboring nodes within its wireless transmission range. Consequently, communicating sensors form a wireless ad-hoc network transmitting real-time physical measurements in its deployment area. WSNs have promising applications that require unattended deployment such as environment monitoring and military operations in hostile territory. These applications motivate research in securing WSNs.

Among the proposed WSN security schemes, Random Key Pre-distribution Scheme (RKPS) [1], [2] has shown to be an effective approach that guarantees any pair of neighboring nodes in a WSN would be able to build a secure connection using symmetric cryptography. Modeling a sensor network as a random graph allows RKPS to apply Erdős-Rényi random graph theory to choose an optimal keyring size for a given keypool size. Keyring size is chosen such that each sensor node is able to authenticate at least a fraction of its neighboring nodes, and set up secure connections to these authenticated nodes to form a trust graph. Each sensor node can later authenticate the remaining untrusted neighboring nodes by flooding the trust graph with authentication requests. We term this mechanism as secured flooding since this flooding occurs along the edges of the trust graph only.

Unbounded broadcasting in RKPS may excessively consume computational and transmission power on each sensor node for authentication and retransmissions. An adversary can exploit this weakness to inject bogus authentication requests into a WSN incurring large performance hits on the network over time. These performance hits constitute a Denial-of-Service attack that can be used to sabotage a WSN by draining sensor power and preventing the network from carrying authentic traffic.

Flooding in ad-hoc networks is typically controlled by a TTL value to ensure that a packet will not be forwarded indefinitely within the networks. Prior research [1], [4], [6], [7], [9]–[12] has presented several studies reporting empirical observations and theoretical analysis on the diameter of a WSN. However, they either lack rigorous and repeatable results [1], [6], [7], or have to base upon different assumptions [4], [9]–[12] that may not be applicable for all WSNs.

In this paper, we study a more applicable modeling based on Random Geometric Graph (RGG) to identify the diameter of a WSN for bounding broadcasting in RKPS. We propose a MAXTTL value setting on each sensor node to ensure that packets with TTL values above MAXTTL cannot be injected into the network. The RGG based modeling can more accurately represent a WSN, and thus derive a more applicable MAXTTL value to mitigate excessive power consumption. The simulation results also show our approach is practical and accurate.

The rest of the paper is organized as the following. Section II discusses the related research work. Section III

ISBN: 978-1-61208-245-5

describes how results on the upper bound of the diameter of Random Geometric Graph (RGG) can be used to derive the value of MAXTTL for a sensor network enabled with RKPS. Section IV describes RKPS in detail and reviews the application of both Erdős-Rényi graph theory and RGG theory relevant to RKPS modeling. Section V present our simulation design and results respectively. Finally, Section VI concludes the paper with future directions.

II. Related Work

In this section, we review prior research that has either proposed guidance on the TTL values for authentication requests, or has some bearing on the derivation of MAXTTL. We also review research on modeling RKPS deployment using RGG theory and the work on upper bound of RGG diameter.

RKPS [1] and its variations have been widely used for securing WSNs, which has been reviewed in [5]. Since this paper is to address a fundamental problem for RKPS, we base our work on a generalized model of the basic RKPS detailed in the next section. This model includes all elements of the scheme that have remained invariant in the derived schemes.

Prior research in [1], [6], [7] had presented empirical observations that the keypath lengths do not exceed a constant number for their simulated WSNs with 1000 to 10000 nodes. However, there is no formal mathematical guidance that could characterize the relationship of the TTL values and the size of WSNs. For example, the first reference to use of a TTL for secured flooding [6] only provided the observation on the average lengths of the keypaths based on the simulation results on a limited node population.

Recent research in RKPS has applied RGG theory to model the highly clustered topology of a practical sensor network deployment. While RGG models the connectivity graph with high fidelity, the presence of edges in the trust graph depends upon the probability with which any two neighboring nodes share a common key. Consequently, RGG with unreliable links has been explored for modeling the trust graph. Di Peitro et al., defined the cryptographs in [8] that model the trust graph as an intersection of a Erdős-Rényi random graph with a RGG graph modeling the deployment. References [9]–[12] modeled the trust graph as a RGG graph, with the presence of edges governed by a Bernoulli function.

It is notable that the authors of [10] also proved connectivity of the RGG with edge probability modulated by a uniform random intersection graph, which shows a theoretical model of the random key predistribution under the full visibility assumption [11]. We discuss these results more formally in the next section.

The authors of this paper have introduced the problem of MAXTTL in [4] and applied theoretical upper bound on the diameter of Erdős-Rényi random graphs to solve it for the full visibility case. In this paper we tackle the practical limited visibility case where a sensor node can only communicate with nodes within their transmission range. This limits their visibility to a much smaller subset of nodes within the network. In our modeling, we applied results from several papers [13], [14] on the application of random graph theory to sensor networks, which discusses the application of Erdős-Rényi random graph theory to RKPS in the context of sensor networks and produces validating results for specific ranges of its parameters. We also used guidance from [14] that discusses the construction of a high performance simulation and allowed us to validate our simulation design.

III. MAXTTL For Secure Flooding

In Figure 1, we show a model of a sample sensor network deployment implementing RKPS, where each sensor node plotted as a node vertex in the graph is surrounded by a circle representing its transmission range. The two overlaid graphs on this model represent the transmission connectivity and secure connectivity respectively. The lighter edged graph among the node vertices represents the connectivity graph formed among a node within its transmission range. The darker edges form the trust graph representing secure connectivity among neighboring node vertices in the connectivity graph. Secure connectivity can be achieved if two neighboring nodes share a common key within their keyrings. Note that the trust graph is a subgraph of the connectivity graph.

While there are more economical broadcasting schemes for ad-hoc networks, flooding may be necessary to ensure speedy and fault tolerant communication of security information in RKPS. In particular, two important protocols in RKPS for authentication and key revocation rely upon secured flooding to accomplish their functions. Authentications typically occur immediately after deployment of a WSN and before the sensor nodes can securely communicate to initialize more optimized broadcasting protocols based on the dominant set in the topology. At this stage the secured sensor network may not be fully connected and a gossiping based broadcast may not reach an authentication node and return back within a bounded time. Key revocation in RKPS aims to remove keys of compromised sensors from the network and also requires speedy announcement of compromised nodes that can be executed by secured flooding in a secure and fault tolerant manner.

As mentioned in the introduction, RKPS secured flooding can have large computational and power overheads that can be exploited to launch DoS attacks. This can be mitigated by setting a maximum limit on the TTL (MAXTTL) on each sensor before deployment. A node receiving a flooded packet will ensure that the contained TTL less than MAXTTL before forwarding it. This will ensure that an adversary

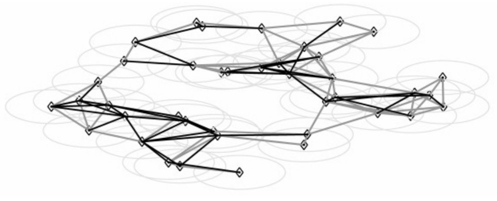


Figure 1: A sensor network secured with RKPS.

injecting packets with long TTL can only inflict limited damage to a RKPS sensor network.

Recent research in RKPS based schemes has utilized RGG for modeling the deployment of a sensor network. MAXTTL can be derived on the basis of the theory related to the upper bound on the diameter of RGG. Diameter of a random graph is the longest of all the shortest paths between every pair of vertices in the graph. Deriving MAXTTL from this value would allow a secured flooding request (SFR) to adequately cover all shortest paths of a connected RGG, without impeding their function.

MAXTTL would also ensure the economy of the RKPS scheme since it would prevent SFRs from traveling on longer redundant paths and cycles within the network. To establish that the paths longer than the diameter are present in a RGG, we observe that longer paths exist between two nodes connected to the nodes between which the diameter exists. By induction, it can be deduced that the diameter can be included in the path between any two pair of nodes accessible by the nodes between which the diameter exists.

Delinquent packets with large TTLs may also get forwarded indefinitely in cycles. The only other solution to prevent forwarding of packets in cycles is to enforce duplicate checking of each packet on every sensor node. Typical secure duplicate checking would require that each sensor spend a prohibitive amount of computational resources for calculating the hashcode for each packet it receives. To prevent cycles of a length l each sensor will need to store a comparable (l) number of hashcodes. There is evidence to suggest that the length of the longest cycles on large random graphs is $O(n)$ [3]. The memory, if it were available could be better used to increase the number of keys in sensor keyrings. We can therefore safely assume its absence.

IV. RKPS Model And Theoretical Analysis

In this section, we first analyze RKPS to show how Erdős-Rényi random graph theory is applicable to choose the size of the keyring for a keypool based on the network size and deployment density. Subsequently, we introduce theory related to RGG connectivity and the analytical results on the upper bound of its diameter that can be directly used to calculate an optimal MAXTTL.

A. Generalized RKPS Model

RKPS predistributes random subsets of keys (keyring) from a large pool of keys (keypool) on each sensor node. Any two keyrings share a common key with a small probability and after deployment each sensor attempts to establish trust with its neighbors by discovering common key(s) through keyrequests. A keyrequest contains a list of key identifiers which uniquely identify each key in a requesting node's keyring. A neighboring node receiving the keyrequest will attempt to find a key in its own keyring. If successful the node will respond back by encrypting a random number with the identified common key (challenge), which must be decrypted by the requesting node and sent back as plain text (response) to complete the authentication. Subsequently, the identified common key can be used to negotiate a shared session encryption key.

Due to the limited memory available on each sensor, the keyring are only large enough to allow a fraction of neighbors to successfully identify common keys in a keyrequest. If a receiving sensor is unable to identify common keys in a keyrequest, it resorts to a path key establishment mechanism (PKEM), where it forwards the keyrequest to the neighbors it is securely connected to. These secure neighboring sensors will in turn either authenticate the keyrequest or forward it to their secure neighbors which will repeat the process until some sensor able to authenticate the keyrequest responds back. RKPS choice for keypool and keyring sizes also ensures that every sensor is securely connected to the rest of the sensor network and the keyrequests sent by it will propagate throughout the network.

A repeatedly forwarded keyrequest constitutes a path through the network, where each node within the path trusts the next node in the path, termed as a keypath [1]. For a single PKEM execution multiple keypaths emanate from the node requesting PKEM authentication of a single keyrequest. Consequently, a large number of the connected sensor nodes within the network will spend power in computation and communication to authenticate a single keyrequest.

RKPS chooses the keyring and keypool sizes such that the secure network formed by direct authentications of the neighboring sensor nodes forms a connected Erdős-Rényi graph, and a keyrequest sent by any node would be forwarded to all nodes within the network. The deployment model of the sensor network is generally assumed to be uniformly random and the neighboring nodes of any particular sensor node after deployment cannot be predicted beforehand. This requires that any sensor node within the network should be able to connect with any other node if they happen to be deployed in each other's neighborhood.

RKPS models the a sensor network in the form of a connected Erdős-Rényi random graph represented by $G(n,p)$, where n is the number of vertices and p represents the probability with which a vertex is connected to any other

ISBN: 978-1-61208-245-5

vertex in the graph. Erdős-Rényi graph theory introduced in [15] proves that $G(n,p)$ where the value of p is derived according to Eq. 1 will be connected with the probability $P[G(n,p)\ \textit{is connected}]$ shown in Eq. 2. Authors in [1], suggested choice of the common parameter C_C such that $P[G(n,p)\ \textit{is connected}]$ is close to 1.0 in Eq. 2. Figure 2 indicates the value of CC, for the desired values of $P[G_{(n,p)}\ \textit{is connected}]$.

$$if \quad p = \frac{\ln(n)}{n} + \frac{C_C}{n} \tag{1}$$

$$then \lim_{n \to \inf} P(G_{(n,p)}\ \textit{is connected}) = e^{e^{-C_C}} \tag{2}$$

where C_C is a constant.

Formally, to design a connected $G_{(n,p)}$, we choose value of CC in Eq. 1, such that $P[G_{(n,p)}\ \textit{is connected}]$ in Eq. 2 is close to 1.0.

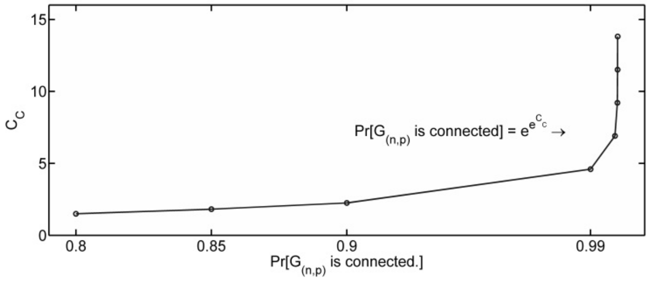


Figure 2: Values of C_C for desired probability of connectivity in Eq. 2.

Prior research in [1] on RKPS identified the desired range for CC is between 8 and 16, as shown in Figure 2. The value obtained for p can be used subsequently to calculate the keyring size (k) for a given keypool size (K) to ensure that the RKPS sensor network is connected with high probability.

B. Full Visibility vs Limited Visibility

It is essential to note, however, that the Erdős-Rényi random graph theory assumes that any node within the graph can be connected to another one, i.e., every node can see any others within the network (full visibility model). However, in practical sensor networks, a sensor node is only connected to a subset of the n vertices, $n_a \ll n$, that represents the expected number of neighboring nodes of a sensor within its communication range (limited visibility model). In order to overcome this practical limitation, the work in [1] proposed scaling p to the effective probability p_a, such that the average degree d_{avg} of the deployed sensors in the network remain equal to the expected degrees of a vertex in the equivalent $G_{(n,p)}$ as indicated in Eq. 3. Note that the p_a represents the probability with which a sensor network will be connected to any node within its neighborhood and this is the probability that will be used to calculate k and K subsequently.

$$d_{avg} = (n_a - 1)p_a = np \tag{3}$$

The value of p_a, calculated from Eq. 3 is used to derive the keyring size k, from Eq. 4 for a given keypool size K.

$$p_a = 1 - \frac{(K-k)!^2}{K!\,(K-2k)!} \tag{4}$$

C. Deployment Modeling of RKPS with RGG

While the original scheme only models the average degree of the connectivity graph, more recent research in key predistribution schemes has formally modeled the connectivity graph as a Random Geometric Graph. Study of Random Geometric Graph (RGG) theory began with [16] and has been adopted generally to study the practical deployment of ad-hoc networks on a planner surface. We borrow the definition from [17] to define the generalized form of RGG, quoted as follows. Let $X_1, \cdots, X_n$ be independent, uniformly distributed random points in the unit cube $[0,1]_d$, where d represents the number of dimensions. The set of vertices of the graph $G_n(r_n)$ is $V = 1, \cdots, n$ while two points i and j are connected by and edge if and only if the Euclidian distance between X_i and X_j does not exceed a positive parameter r_n, i.e., $E = \{(i,j) \parallel X_i - X_j \parallel < r_n\}$ where $\parallel . \parallel$ denotes the Euclidean norm.

Note that by definition the $G_n(r_n)$ is generalized over multiple dimensions d, however the two dimensional case is of specific interest to modeling the spatial deployment of sensor nodes on planner field. The Euclidean norm in this case becomes the distance between any two nodes and r_n corresponds to the transmission range of each node. The points described in RGG correspond to the location of each sensor node uniformly distributed on a unit area $(d = 2)$ that can be modeled as a unit square or unit circle without loss of generality. In the context of modeling the practical sensor network deployment the results obtained from RGG theory can be trivially scaled to the actual deployment area.

D. Connectivity of RGG

While RGG models the graph connectivity with high fidelity, the links of the trust graph are also modulated by the probability with which two neighboring nodes share a common key. As a result, RGG with unreliable links has been explored for modeling the trust graph [10] established the following result for the connectivity of RGG, where p_l and p_n represent the probability of the link and node presence respectively.

$$r = \sqrt{\frac{\ln n + C}{n p_l p_n \pi}} | n, C \to \infty, p_l, p_n \in [0,1] \tag{5}$$

More recently, research in [11] has investigated routing in a practical sensor network deployment using Random Geometric Graph with randomly deleted edges. They formally proved results for the following conditions.

$$if \quad \pi p_l r^2 \geq C\frac{\ln n}{n} | C > 8, p_l \in [0,1], r \in (0, 1/\sqrt{\pi}) \tag{6}$$

$$r \geq c\sqrt{\frac{\ln n}{n}} | c > 1.598, p_l \in [0,1], r \in (0, 1/\sqrt{\pi}) \tag{7}$$

Then RGG $G_n(r, p_l)$ is connected with probability tending to 1 as $n \to \infty$, p_l is again the probability of link presence.

E. Network Connectivity Requirement in RGG

Xue et al., in [18] showed that in a two-dimensional RGG where the nodes are distributed uniformly, the number of neighbors of each node need to grow like $\Theta(\log n)$ if the network is connected. Further they also showed analytically that for a RGG where each node is connected to less than $0.074 \log n$ (lower bound) nodes the network is asymptotically disconnected. However, if the nodes are connected to greater than $5.1774 \log n$ (upper bound) nearest neighbors, then the network will be asymptotically connected. Finally, Balister et al., in [19] improved the lower bound to $0.3043 \log n$ and the upper bound to $0.5139 \log n$.

We note [18] and [19] rely upon a Poisson distribution of nodes on a unit disk. Low or rare events [20] allow the approximation of the binomial distribution with a Poisson distribution.

F. RGG Diameter

Ellis et al., [17] have shown two important asymptotic results on the diameters of RGG. Let $\phi(n) \to \infty$ be non-negative. There exists an absolute constant $K > 0$ such that if

$$r \geq \sqrt{\frac{\ln n + \phi(n)}{n}} | n, \phi(n) \to \infty \tag{8}$$

then the unit disk random graph $G(n, r)$ is connected with diameter denoted by $D(G_{(n,r)}) < K(2/r)$. They also derived the value of K analytically to 129.27. In Theorem 7 [17] they prove a still lower bound for the following conditions.

Let

$$r = c\sqrt{\frac{\ln n}{n}} | c \geq 2.26164 \tag{9}$$

Then the unit disk random graph $G_{(r,n)}$ is connected with diameter

$$D(G_{(n,r)}) \leq (4 + o(1))/r \tag{10}$$

V. Simulation Results

We constructed a simulation model to verify the diameter of the trust graph generated in RKPS scheme using direct authentication. The simulation generates random topologies for sensor networks with limited visibility by varying the number of nodes from 1000 to 12000, and calculated the corresponding keyring sizes from a keypool of 100000. The visibility range of each sensor is calculated on the basis of Eq. 9 since it provides an elegant result which can be used to directly calculate the diameter of the random graph.

Our simulation model closely follows the guidance from [14]. The keying size are derived based on the guidance from [1], and allows for variations in the sensor network deployment densities through node range variation according to Eq. 9. We have also taken into account the boundary effect identified in [14] and eliminated it from our final results.

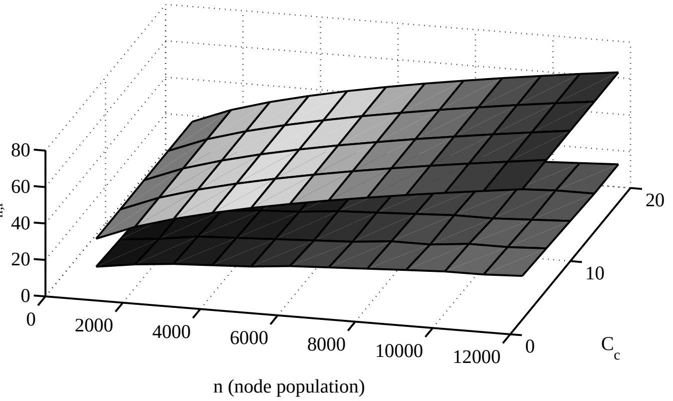


Figure 3: Simulation results and asymptotic predictions for a 12000 node sensor network.

Figure 3 shows a plot of our simulations on MATLAB and Java. The upper surface represents the calculated upper bound. The lower surface represents the actual diameter of the simulated sensor network. The actual diameter of the network is consistently smaller than the prediction for the diameter by a wide margin. However, it is notable that the upper bound on the diameter is much higher than the actual diameter of the network. This indicates that tighter theoretical bounds on the diameter of the Random Geometric Graphs, and consequently the MAXTTL are possible.

However, we recommend keeping a wide margin between the actual MAXTTL and predicted MAXTTL value. The actual value used in practical deployments should be set higher than the predicted value. This is to accommodate the fact that the trust graph is not a perfect Random Geometric Graph. It may have many absent edges between neighboring nodes at the beginning of the deployment, when secure trust relationships have not been established. Theory on faulty Random Geometric Graphs [**?**] is still nascent and upper bound on its diameter may provide a more accurate prediction for the MAXTTL.

Figure 4 shows the prediction of the diameter for sensor networks for large node populations $O(10^6)$. We observe that the diameter is relatively small and grows slowly with the node population $O(log(n)/n)$.

Diameter of the network increases very slowly with network size and remains constant for large ranges of node populations. This shows promise in the extensibility and graceful degradation of a sensor network deployment,

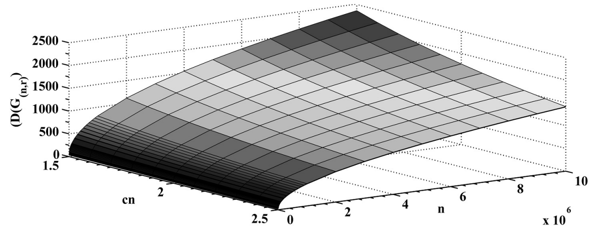


Figure 4: Long range predictions for the sensor network diameters.

even if the MAXTTL value is locked as a constant before deployment. On the other hand, this shows that controlling the TTL would only provide limited control over the number of nodes visited by a keyrequest and therefore MAXTTL alone may not be well suited for precise control of power consumption for SFRs. The consequent power consumption of PKEM is not precise and the number of transmissions increase rapidly with each increment in TTL value.

VI. Discussion and Conclusion

While we have utilized asymptotic results on RGG graph theory, formal proofs for asymptotic diameters on faulty RGGs is still an open problem. Our simulation results indicate that either the upper bound on RGG diameters holds well for faulty RGGs as well and there is further scope for a tighter upper bound on the RGG diameter. Future theoretical results for tighter bounds on the diameter of RGGs and specifically for faulty RGGs would provide precise bounds applicable to the problem of MAXTTL.

Some form of fault tolerant gossiping may eventually be considered to reduce the transmission overhead and the power consumption of PKEM, however the latency would increase in this case. This would also make the network more vulnerable to worm hole attacks where an adversary is able to exploit the latency between two different parts of the network to launch various attacks.

Finally we hope to trigger a discussion of the problem of secure broadcasting as applied to RKPS, and the analytical modelling of its overhead. We believe that this overhead is unique to RKPS based schemes and may prove to be prohibitive in large networks. Competing public key cryptography like Elliptic Curve do not require a broadcast for key discovery with lower overhead than RSA and that may eventually be more feasible with development in technology.

References

[1] L. Eschenauer and V. D. Gligor, ”A key-management scheme for distributed sensor networks”, in the Proceedings of the 9th ACM conference on Computer and communications security, Washington, DC, USA, 2002.

[2] C. Blundo, A. D. Santis, A. Herzberg, S. Kutten, U. Vaccaro, and M. Yung, ”Perfectly-Secure Key Distribution for Dynamic Conferences”, in the Proceedings of the 12th Annual International Cryptology Conference on Advances in Cryptology, 1993.

[3] B. Bollobs, Random graphs: Cambridge University Press, 2001.

[4] A. Mishra, T. Gyires, and Y. Tang, ”Towards A Theoretically Bounded Path Key Establishment Mechanism in Wireless Sensor Networks”, in the Proceedings of the Eleventh International Conference on Networks, Saint Gilles, Reunion, 2012.

[5] Y. Xiao, V. K. Rayi, B. Sun, X. Du, F. Hu, and M. Galloway, ”A survey of key management schemes in wireless sensor networks”, Comput. Commun., vol. 30, pp. 2314-2341, 2007.

[6] J. Hwang and Y. Kim, ”Revisiting random key pre-distribution schemes for wireless sensor networks”, in the Proceedings of the 2nd ACM workshop on Security of ad hoc and sensor networks, Washington DC, USA, 2004.

[7] C. F. C. Aldar, ”A graph theoretic approach for optimizing key pre-distribution in wireless sensor networks”, in the Proceedings of the 7th international conference on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, Seoul, Korea, 2009.

[8] R. D. Pietro, L. V. Mancini, A. Mei, A. Panconesi, and J. Radhakrishnan, ”Redoubtable Sensor Networks”, ACM Trans. Inf. Syst. Secur., vol. 11, pp. 1-22, 2008.

[9] O. Ya?an, ”Performance of the Eschenauer-Gligor key distribution scheme under an ON-OFF channel”, IEEE Transactions on Information Theory, vol. November, 2011.

[10] Y. Chih-Wei, W. Peng-Jun, L. Xiang-Yang, and O. Frieder, ”Asymptotic distribution of the number of isolated nodes in wireless ad hoc networks with Bernoulli nodes”, Communications, IEEE Transactions on, vol. 54, pp. 510-517, 2006.

[11] K. K. a. K. Rybarczyk, ”Geometric Graphs with Randomly Deleted Edges - Connectivity and Routing Protocols”, 2011.

[12] B. Y. Seyit Ahmet Camtepe, Moti Yung, ”Expander graph based key distribution mechanisms in wireless sensor networks”, IEEE International Conference on Communications (2006), vol. June 2006, pp. 2262-2267, 2006.

[13] T. M. Vu, R. Safavi-Naini, and C. Williamson, ”On applicability of random graphs for modeling random key predistribution for wireless sensor networks”, in the Proceedings of the 12th international conference on Stabilization, safety, and security of distributed systems, NewYork, NY, USA, 2010.

[14] T. M. Vu, C. Williamson, and R. Safavi-Naini, ”Simulation modeling of secure wireless sensor networks”, in the Proceedings of the Fourth International ICST Conference on Performance Evaluation Methodologies and Tools, Pisa, Italy, 2009.

[15] P. Erdős and A. Rényi, ”On the evolution of random graph”, Publications of the Mathematical Institute of the Hungarian Academy of Sciences 5: 17-61, 1960.

[16] M. Penrose, Random geometric graphs: Oxford University Press, 2003.

[17] R. B. Ellis, X. Jia, and C. Yan, ”On random points in the unit disk”, Random Struct. Algorithms, vol. 29, pp. 14-25, 2006.

[18] F. Xue and P. R. Kumar, ”The number of neighbors needed for connectivity of wireless networks”, Wirel. Netw., vol. 10, pp. 169-181, 2004.

[19] P. Balister, B. Bollobs, A. Sarkar, and M. Walters, ”Connectivity of random k-nearest-neighbour graphs”, Adv. in Appl. Probab., vol. 37, pp. 1-24, 2005.

[20] A. Papoulis and S. U. Pillai, Probability, Random Variables, and Stochastic Processes: McGraw-Hill, 2002.

[21] J. Diaz, J. Petit, and M. Serna, ”Faulty random geometric networks,” Parallel Processing Letters, vol. 10, pp. 343-357, 2000.

ADOPTATION OF WEIL PAIRING IBE FOR SECURE FILE SHARING

Cheong Hyeon Choi
Security Lab., MIS Dept. School of Business
Kwangwoon University
Seoul, Korea
e-mail: chchoi@kw.ac.kr

***Abstract*—Competitive enterprise has lots of secret files in digital form, which is vulnerable to illegal online activity. Our target network accepts authorized users on registered machines. In general it is known that the traditional public-key scheme is suitable to strong authentication and encryption, but it is inferior to the IBE (Identity Based Encryption) scheme in terms of performance and certificate management. Thus we improve the WP (Weil Pairing) IBE scheme so as to be suitable to our SFS (secure file sharing) system with certificate-less key revocation as normalizing public identity. In addition, the man-in-the-middle attack is negligible because the security is based on hard ECDLP (elliptic curve discrete logarithm problem). However, in the perspective of performance, our WP IBE is bound to the complexity of modular-ADD. In the perspective of security, our SFS network is close with single authorized server and multiple registered clients. In addition, the SFS architecture is fortunately useful in DRM network and P2P file sharing network.**

Keywords-IBE; Weil Pairing; Diffie-Hellman; DLP.

I. INTRODUCTION

In general, corporate secrets are accessible inside a restrictive area proofed against exposure in the building. In real, secrets such as manufacturing blueprints are confined so as to be accessible to only CEO or authorized technicians through registered machines in a restrictive area. Such network is composed of a single server and multiple clients, where confidential files are stored in the server and disseminated to the registered clients. Its cryptographical scheme must satisfy stronger security constraints than traditional schemes. Communication between the server and the clients is asymmetric with respect to the cryptographical functions and the exchanged messages [12]. Our target network called SFS (Secure File Sharing) satisfies these properties.

There are a couple of commercial networks similar to our SFS architecture. The first is P2P (Peer-to-Peer) network, which is formed with asymmetric connections, where its file server knows the client keeping a registered file, and redirects a file request to the corresponding client [19]. The second is DRM (Digital Right Management) network, which consists of single DRM server with copyrighted digital materials and multiple purchasers (clients). The DRM server must verify both a purchaser and its digital player with so-called two-factor signature, then package the purchased materials using encryption algorithm [22]. In this network, the important security issue is authenticity [15].

It is desirable to use the public-key scheme for such security issue. Nevertheless, we note that the traditional public-key scheme is inappropriate in the perspectives of performance and certificate management. In 1984, Shamir suggested Identity Based Encryption (IBE) scheme to lessen a burden of certificate management [1]. The IBE needs no more public-key certificate since any string {0,1}* can be a public key. In the early stage, it was proposed to use an e-mail address as a public key. The early IBE, however, might expose key material to adversary since the public ID and crypto-functions are open without authentication. In order to avoid such exposure risk, we combine a public identity and a period as normalizing to temporal public key [2]. As the result, the normalized public key will be ineffective after its valid period, which means the key revocation [15], called as *certificate-less key revocation*.

We modify the WP (Weil Pairing) IBE scheme for better performance and stronger security so as to be appropriate to our SFS network, which is characterized as follows:

① One-to-many: Data transmission between single file server and multiple file consumers

② CS model: Asymmetric communication between the file server and consumers (clients).

③ Closedness: All links of the network formed by secure channels.

④ Strong authenticity: Adoption of multi-factor authentication (signature).

In Section II, we review the related work on the IBE scheme, the crypto-functions of PKG (Private Key Generator) and the complexity of crypto-functions for performance analysis. Section III addresses the SFS system architecture and its core protocols. Specifically we mention the system setup and the client registration. Section IV is concerned with how to improve our SFS WP IBE. Section V concludes the performance issue, unsolved problems of our works and its future.

II. RELATED WORK

A. IBE (Identity Based Encryption) scheme

In the IBE [7], the public key is generated from the public identity combined with an unique ID and a time period which may have an effect on key revocation [1]. Boneh and Franklin proposed the WP ECC (elliptic curve cryptography) scheme among efficient pairing-based ECC [8]. The WP IBE scheme is much more efficient because its key size is much shorter and is based on the bilinearity over the Field F_p , which is the Group law to replace multiplicative Group with additive Group [8]. In spite of the shorter key size, the ECC 256-bit public-key provides the same level of security as the 3024-bit RSA-based public key [2].

Nevertheless, security flaw of the WP IBE is relevant to the public identity which is literally open to the public including adversaries. In addition, it is difficult to hide the crypto-functions like *encryption*, *decryption and key generation* from adversaries. Thus, the WP IBE must be based on the random oracle model, which prevents the adversary from successful guessing key-related things using queries to the oracle with a public identity. In order to have the scheme secured against chosen ciphertext attacks (CCA), the oracle's responses to the quesries must satisfy onewayness and randomness to hide details of the crypto-functions [17][20].

B. Weil Pairing (WP) ECC

The WP is one of the implemented ECC schemes, being formed with the Group of points over an elliptic curve. Let p be a prime and E be the elliptic curve of Weiser Strass equation $y^2 = x^3 + 1$ over the Filed F_p . The set of rational points $E(F_p) = \{(x, y) \in F_p \times F_p : (x, y) \in E\}$ forms a cyclic Group of order $p-1$. The set of points of order q defined as $p = 12q - 1$ forms the cyclic sub-Group G_1 of which the generator is ρ . Let G_2 be the sub-Group of F_{p^2} of order q^2 [6].

The WP is based on the bilinear map $\hat{e}: G_1 \times G_1 \rightarrow G_2$ between two cyclic Groups G_1 , G_2 of order q , q^2 respectively with following properties [6]:

① *Bilinearity:* $P, Q \in G_1$ and $a, b \in \mathbb{Z}_q$, then $\hat{e}(aP, bQ) = \hat{e}(P, Q)^{ab}$

② *Computability:* $P, Q \in G_1$, then the efficient algorithm for $\hat{e}(P, Q) \in G_2$ exists.

③ *Non-degenerate:* $\hat{e}(\rho, \rho) \in F^*_{p^2}$ is a generator of G_2 .

The WP satisfies Diffie-Hellman problem (DHP) assumption [21]: even if an eavesdropper observed g^x, g^y in the key exchange protocol, the eavesdropper cannot compute the secret key g^{xy} with easy means, while legitimate parties can do. Simultaneously, the decisional DHP(DDHP) to decide whether g^{xy} comes from g^x and g^y , given g^x, g^y and g^{xy} , is hard. But, in the WP, DDHP becomes easy with the bilinear map as follows:

$$\hat{e}(g^x, g^y) = \hat{e}(xg, yg) = \hat{e}(g, xyg) = \hat{e}(g, g^{xy}) .$$

It is proved that DHP is equivalent to the hard discrete logarithm problem (DLP) to compute $a = Log_{\rho}P$ with $P(= a\rho, a \in \mathbb{Z}_p)$ and ρ [21]. Here, we insist that the WP is enough secure to protect secret values because its security is based on ECDLP (elliptic curve discrete logarithm problem) [21].

C. WP IBE crypto-functions

The IBE scheme is characterized by four randomized algorithms: *setup(), extract(), encrypt(), decrypt()* [2][9], defined as follows:

① $setup(k) \rightarrow (s, parm)$

② $key\_extract(parm, s, ID) \rightarrow d$

③ $encrypt(parm, ID, M) \rightarrow C$

④ $decrypt(parm, C, d) \rightarrow M$

Here, $k, parm, s$ are the seed value, the security parameter like a prime order, and the system wide master key $s \in \mathbb{Z}_q$ respectively. In addition, ID, d, M and C are: $ID \in \{0,1\}^*$ acting as a public identity, d as the corresponding private key, $M = \{0,1\}^n$ as a plain message and C as its ciphertext respectively.

Specifically, the WP IBE scheme produces the security parameter $parm$ summarized as follows:

$parm =< q, G_1, G_2, \hat{e}(), n, \rho, S, H_1, H_2 >$, where q is the prime order of Group, n is the message block length, the system wide public key is $S(= s\rho)$, ρ is the generator of the cyclic Group G_1 , where s is the system wide master key kept in secret on PKG (Private Key Generator). In addition, H_1 and H_2 are the hash functions with onewayness and randomness which scramble the hash algorithm so as to behave like a random oracle, defined as follows:

$$H_1 : \{0,1\}^* \rightarrow G_1,\ H_2 : G_2 \rightarrow \{0,1\}^n$$

The private key d_{ID} is computed from the public identity ID as follows: the public key Q_{ID} is normalized with the public identity ID as $Q_{ID} = H_1(ID) \in G_1$, and the private key is computed as $d_{ID} = sQ_{ID}$. The encryption of a message M with Q_{ID} is done as follows:

① Computing $g_{ID} = \hat{e}(Q_{ID}, S) \in G_2$

② Choosing a prime number $r \in Z_q^*$,

③ Generating the ciphertext like follows:

$$C = < r\rho, M \oplus H_2(g_{ID}{}^r) >$$

The decryption of the cyphertext C is done as follows [2][4][11]:

① $M \oplus H_2(g_{ID}{}^r) \oplus H_2(e(d_{ID}, r\rho))$

② $M \oplus H_2(g_{ID}{}^r) \oplus H_2(g_{ID}{}^r) = M$ since

$$\hat{e}(d_{ID}, r\rho) = \hat{e}(sQ_{ID}, r\rho) = \hat{e}(Q_{ID}, s\rho)^r = g_{ID}{}^r$$

D. WP Signature

The WP authentication process is quite different from the traditional public-key scheme. It is much faster since the signature is based on the pairing-based additive Group and its verification belongs to DDHP, as follows [6][10]:

Signer:

① Generating a temporary key pair :

$$(r, R), R = r\rho, r \in \mathbb{Z}_p$$

② Generating a signed digest: $T = rS + H(M, R)d_{ID}$

③ Generating the signature: $Sign = < R, T >$

Verifier:

④ Computing $v = \hat{e}(R + H(M, R)Q_{ID}, S)$,

⑤ Computing $u = \hat{e}(T, \rho)$ as

$$\hat{e}(rS + H(M, R)d_{ID}, \rho) =$$
$$\hat{e}(rs\rho + H(M, R)sQ_{ID}, \rho) =$$
$$\hat{e}(r\rho + H(M, R)Q_{ID}, s\rho) =$$
$$\hat{e}(R + H(M, R)Q_{ID}, S) = v$$

⑥ If $u \equiv v$, then it is verified.

M and $H(.)$ are the message and the hash function respectively.

E. Performance consideration.

Table 1. Basic Operation's complexity [18]

Algorithm	Input/output	Running Time
INT-DIV	*a/N (N >0)*	$O(\|a\| \cdot \|N\|)$
MOD	*a mod N (N >0)*	$O(\|a\| \cdot \|N\|)$
EXT-GCD	*a, b ((a, b)≠(0, 0))*	$O(\|a\| \cdot \|b\|)$
MOD-ADD	$a+b \bmod N\ (a, b \in Z_N)$	$O(\|N\|)$
MOD-MULT	$ab \bmod N\ (a, b \in Z_N)$	$O(\|N\|^2)$
MOD-INV	$ab=1 \pmod N\ (a \in Z_N)$	$O(\|N\|^2)$
MOD-EXP	$a^n \bmod N\ (a \in Z_N)$	$O(\|n\| \cdot \|N\|^2)$
EXP_G	$a^n \in G\ (a \in G)$	*2|n| G-operations*

It was known that the ECC took less time to generate the key pair and digital signature than RSA, however, the only ECC signature verification is much costly (Table 2) [13][14]. In addition, *setup()* need much time to generate E points over F_p . Supposed that participants already had the security parameter produced by *setup(),* generation of the private key d_{ID} is bound to WP-time (Weil pairing time-complexity) $O(a^{(p+1)/4} \pmod p)$ of a small prime p, which is the Euclid-criteria (Section IV.B) algorithm complexity in [6].

III. SFS ARCHITECTURE AND PROTOCOL

A. Architecture

Our SFS system architecture is shown in Figure 1.

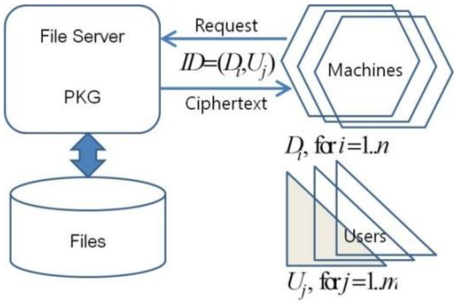


Figure 1. Architecture for Secure File Server with PKG.

As seen in Figure 1, the SFS architecture is composed of the single file server playing the role in PKG and the secret file manager, and the multiple clients of which a client means a pair of a legal user and a registered machine. A secret file can be exchanged only between the file server and a client. After completing manipulation of the secret file, the machine must eliminate the secret file from the machine, and only the file server can store the secret file.

P2P file sharing system using Gnutella protocol [19] is similar to this architecture except that the files stay at the clients registered in the network and the server keeps only file location information [19]. The architecture for DRM system is also similar to Figure 1, except that the clients are purchasers and the server is a distributor of copyrighted material.

In our SFS system, the file server is the supplier of a secret file and the clients are consumers identified by authorized user-ID U_j on registered machine-ID D_i . The public identity ID is formed by the ID pair (D_i, U_j) .

B. Protocol

The SFS system takes three stages for the lifetime; the *setup stage* is first along with *initial registration* of all the machines and users, the next is the *file dissemination stage*, where transfers an encrypted file with digital signature from the server to the client against the man-in-the-middle attack. The last is the *evolution stage* for joining and releasing users and machines in time. For the lifetime, there is the *setup stage* once initially, and the *file dissemination stage* and the *evolution stage* are repetitive.

A secret file is *created*, and then is *registered* at the file server. The file can be *distributed* on the SFS network and *modified* by the owner. This is the lifecycle of a secret file, depicted in Figure 2.

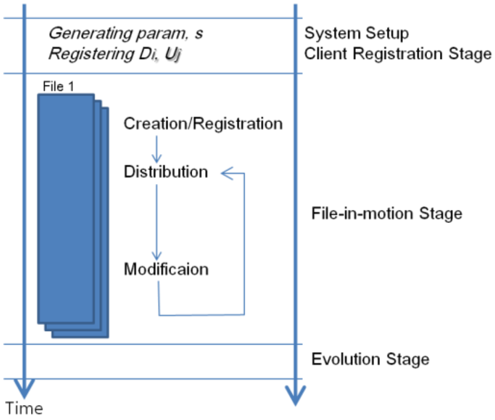


Figure 2. Life Cycle of File sharing system

Setup

The *setup(.)* generates the security parameter and the master key as follows (refer to section II):

① $param =< q, n, G_1, G_2, \hat{e}(), \rho, S, H_1, H_2 >$

② $s \in \mathbb{Z}_q$

param is distributed to authorized machines. The system wide private-key s is kept at a safe place in the server. S is the system wide public-key, ρ is the generator of the cyclic Group G_1 of order q relevant to prime number in $\mathbb{Z}_q^*$. n is the signature size. Two hash functions provide randomized hashing defined as follows:

$H_1 : \{0,1\}^* \rightarrow G_1,\ H_2 : G_2 \rightarrow \{0,1\}^n$.

Registration

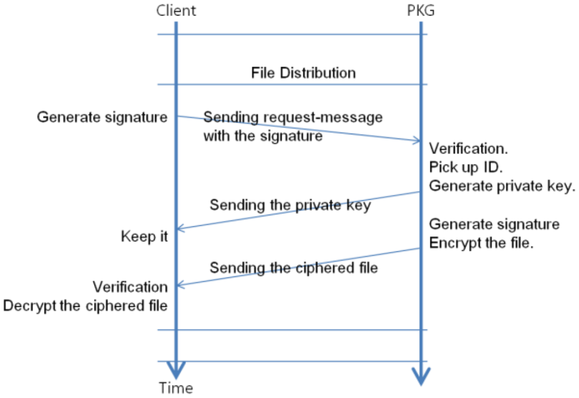
The SFS network limits the members to the machines and the users registered prior to the *file-in-motion* stage of Figure 2. Member is identified with its public identity ID_{entity} . In the *file distribution*, ID_{client} in the *request-message* is formed as combining two members' ID D_i and U_j .

① $ID_{entity} = [D_i \mid U_j]$

② $ID_{client} =< D_i \,\|\, U_j >$

Distribution

In the *file distribution*, two messages are exchanged between the server and the client. One is the *request-message* requesting a secret file from the server, and another is the *file-message* conveying the encrypted file to the client. Always the sender must authenticate the signature on the messages. ID_{client} is signed with a temporal private key. The sender attaches the signature on the *request-message*; refer to Figure 3.



Figurte 3. File Distribution

With ECC key pair (P, x) where $P = xQ,\ 1 \le x \le n-1$, the early ECC signature scheme is [14]:

① Choosing a random number $1 \le k \le n-1$.

② Computing point $kQ = (x_1, y_1) = X$.

③ Computing $r = x_1 \pmod{n}$

④ Computing $k^{-1} (\mathrm{mod\ n})$

⑤ Computing $e \leftarrow SHA(m)$

⑥ Signing as $s = k^{-1}(e + xr) \pmod{n}$.

⑦ Generating the signature tuple $< r, s >$.

Our digital signature scheme is improved as follows:

sign()

① Choosing a prime number $r \in \mathbb{Z}_q$ as a temporal secret key.

② Computing $R = r\rho \ \in G_1$ as the public key.

③ Generating $ID = D_i \,\|\, U_j \ \in G_1$

④ Signing as $\Sigma_{ID} = rID \ \in G_1$.

⑤ Generating the signature tuple $< R, ID, \Sigma_{ID} >$.

⑥ Adding the tuple to the *request-message*

We remind that ECC arithmetic is based on the modular operation, and given $< R, ID, \Sigma_{ID} >$, finding r is ECDLP (Elliptic Curve Discrete Logarithm Problem).

The ECC verification is [14]:

① Computing $e \leftarrow SHA(m)$

② Computing $w = s^{-1}(\bmod\ \mathrm{n})$

③ Computing $u_1 = ew\ (\bmod\ \mathrm{n})$, $u_2 = rw\ (\bmod\ \mathrm{n})$

④ Computing $X = u_1 Q + u_2 P$.

⑤ Computing point $v = x_1\ (\bmod\ n)$.

⑥ Accepting if $r = v$

verify()

① If $\hat{e}(R, ID)\ (= \hat{e}(r\rho, ID) = \hat{e}(\rho, rID)) \equiv \hat{e}(\rho, \Sigma_{ID})$ is true, the verification is successful. Otherwise, it is failed. Here ρ is the generator of the security parameter.

Here, remember that $\hat{e}(R, xID) \neq \hat{e}(\rho, x\Sigma_{ID})$ because $xID\ (\bmod\ q) \neq ID$. Finding such x is ECDLP

IV. IMPROVEMENT OF WP IBE

The reason that our WP scheme is much better than RSA or the early ECC is that the WP is based on the Bilinearity. The IBE security is relevant to the public identity open to the public. We assume the selective-ID security [5] is satisfied in SFS, in which the target ID is specified in advance before the master public-key is published [11].

Improvement of our scheme is threefold: the selective ID normalization, the two-factor signature, and the certificate-less key revocation.

A. Public ID Normalization

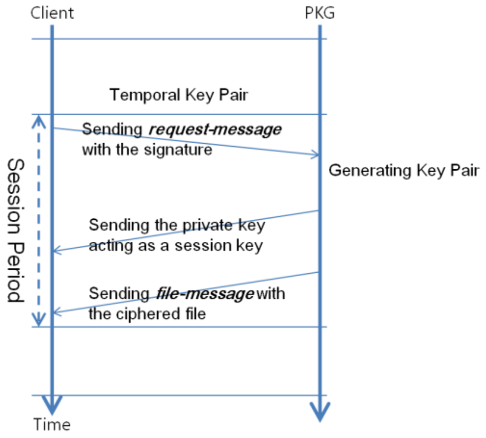

Client
PKG
Temporal Key Pair
Sending *request-message* with the signature
Generating Key Pair
Sending the private key acting as a session key
Sending *file-message* with the ciphered file
Session Period
Time

Figure 4. Session Period

In the IBE, public ID is identical for the lifetime and the adversary can obtain key pairs to guess the target key adaptively. For IBE security, we modify constant IBE public ID into temporal ID for *certificate-less key revocation*. The SFS system generates the public key after normalizing public identity ID as follows:

① Client: generating $ID_{client} = D_i \parallel U_j$

② PKG: normalizing the public identity ID as ID_{client}^{time} $= ID_{client} \parallel time_{dur}$.

③ PKG: hashing ID into $a \in \mathbb{Z}_q$, as $a = H(ID)$

The format of $time_{dur}$ is denoted as *year:mon:day:hour:min.* Wild * in $time_{dur}$ means an entire period of public identity ID. For instance if $time_{dur}$ is *year:mon:day:hour:***, the public identity ID is invalid at the next hour. This ID is called a *temporal client identity*, denoted as ID_{client}^{time}.

B. Key Generation

We mentioned that the private key generation is triggered by the *request-message* of a client (Figure 3). The key pair is similar to the session key valid for the session from the *request-message* through its *file-message* (Figure 4).

The key generation algorithm of WP ECC is generated from the well-known function *MapToPoint*, as follows [3][6][20]:

① $x \leftarrow H(ID_{client}^{time} \parallel j)\ (\bmod\ p)$ at $p = 12q - 1$, $j = 0$

② $a \leftarrow x^3 + 1\ (\bmod\ p)$

③ If $a^{(p-1)/2} = 1 \bmod p$, $y \leftarrow a^{(p+1)/4}\ (\bmod\ p)$; $Q \leftarrow 12(x, y)$ -- Euler Criteria

④ Otherwise, $j = j + 1$, do the first step ①.

We modify the *MapToPoint* algorithm [6] for the key generation of the SFS system. In the SFS, the public key is denoted as Q_{ID} relevant to the client's public identity ID. The private key d_{ID} is generated as $d_{ID} = sQ_{ID}$. It is common that the private key d_{ID} is distributed to the client using secure channels like TLS or VPN. Therefore, note that there is no difficulty in SFS network since all connections in SFS are protected by the IBE based VPN.

In the performance perspective, the mean number of loops is: $E[n] = \sum_{n=1}^{\infty} \frac{n}{2^n}$ since the probability satisfying the Euler Criteria in *MapToPoint* algorithm is 1/2. Thus the loop mean $E[n]$ is approximately less than 2. Therefore, the key generation is bound to $O(E[n] \cdot a^{(p+1)/4}) = O(a^{(p+1)/4})$,

ISBN: 978-1-61208-245-5

called "*WP-operation*" complexity [6]. However, since $p(=12q-1)$ is a small odd prime where Group G_1 is of order q , the key generation is approximate to the complexity *MOD-EXP* $O(p^2)$ (Table 1).

C. Encryption and Decryption including Authentication

If the CCA-proof encryption scheme of Canneti, Halevi and Katz [11] is considered, the SFS suggests the hybrid WP IBE combining the following schemes. The SFS system uses two sets of key pairs: $<R,r>$ for $R = r\rho\ (r \in \mathbb{Z}_q^*)$ and $<Q_{ID}, d_{ID}>$ for $d_{ID} = sQ_{ID}$. The reason is that *Encrypt()* satisfies both integrity and authenticity as well. Let M and C be the plain-message and the cipher-message.

Encryption - Encrypt()

① $g_{ID} = \hat{e}(Q_{ID}, S) \in G_2$

② Modified message $M' =< R, M >$

③ Intermediate cipher message $C' = M' \oplus H_2(g_{ID}{}^r)$

④ Signature $\Sigma_{C'} = rH_1(C')$

⑤ Cipher message $C =< R, C', \Sigma_{C'} >$

Decryption - Decrypt()

① $C' \oplus H_2(\hat{e}(d_{ID}, R)) =$
$M' \oplus H_2(g_{ID}{}^r) \oplus H_2(\hat{e}(d_{ID}, R)) =$
$M' \oplus H_2(g_{ID}{}^r) \oplus H_2(g_{ID}{}^r) = M'$ $< R', M >$ since
$\hat{e}(d_{ID}, R) = \hat{e}(sQ_{ID}, r\rho) = \hat{e}(Q_{ID}, s\rho)^r = \hat{e}(Q_{ID}, S)^r$
$= g_{ID}{}^r$

② If verification $\hat{e}(\Sigma_{C'}, \rho) \equiv \hat{e}(H_1(C'), R)$ is true, C is clean.

③ If $R' = R$ in $M' =< R', M >$, it means that the message integrity is verified and the plain message M is restored.

The *encryption* and *decryption* satisfy confidentiality, integrity, and authenticity as well. However, adversary cannot obtain a hint on any relation between C and M in CCA [16] since randomized hash is applied to C and M . In the performance perspective, this encryption and decryption are bound to $O(g_{ID}{}^r)$, EXP_G , r *WP-operation* (Table 1). $O(g_{ID}{}^r)$ is $O(g_{ID}{}^r \pmod p)) = O(rp^2)$. Since $1 \le r \le p$, therefore, the encryption scheme is bound to $O(p^3)$.

V. PERFORMANCE

The early ECC is known to be better than RSA in terms of key size, signature and encryption. Of course the verification is worse than RSA (Table 2) [13][14].

Table 2. Comparison of the early ECC and RSA [13].

Key Length		Key Generation		Signature		Verification	
ECC	RSA	ECC	RSA	ECC	RSA	ECC	RSA
163	1024	0.08	0.16	0.15	0.01	0.23	0.01
233	2240	0.18	7.47	0.34	0.15	0.51	0.01
283	3072	0.27	9.80	0.59	0.21	0.86	0.01
409	7680	0.64	133.9	1.18	1.53	1.80	0.01
571	15360	1.44	679.06	3.07	9.20	4.53	0.03

Among ECCs, the WP is known to be more efficient with respect to security and performance. Comparing with RSA, the ECC key length 163-bit is the same level of security as RSA's 1024-bit [6][13][14]. In Table 2, we know that the early ECC is better than RSA, except the verification. The early ECC is compared with our SFS WP scheme in Table 3. Our improved WP scheme is much better.

In the security level, therefore, the WP IBE turned out to be the better public-key cryptography than the early ECC [2].

Table 3. Comparison of SFS WP and ECC

Operation	Our WP		ECC	
	sign	verify	sign	verify
MOD-EXP			1	1
MOD-MULT	2	2	4	4
MOD-ADD			2	2
Hash()			1	1
total	2	2	8	8

We note that the signature scheme of our SFS WP is similar to the verification one in term of the number of operations (Table 3). In the ECC, it is known that the signature takes more time than the encryption in term of the running time. Therefore, we expect that our SFS WP is improved in the perspective of performance.

VI. CONCLUSION AND FUTURE WORK

In the future, we will make a real time test-bed to assess their performance. We think that the WP is difficult to substitute for the conventional public-key scheme in the crypto-process structure. The IBE hash function is assumed to satisfy both onewayness and randomness, however, it is not straightforward to implement such hash function. The IBE random oracle model is not practical and the model is speculative for security proof.

As the pairing-based ECC limits keys over the field, it surprisingly reduces key computation time. Nevertheless, adaptive attacks like IND-ID-CCA [16] are feasible, however, the WP neglects such attacks. In future work, we will improve the signature and the encryption collaboratively in a simple fashion.

REFERENCES

[1] A. Shamir, "Identity-based cryptosystems and signature schemes," Advances in Cryptology, Crypto '84, Lecture Notes in Computer Science, vol. 196, pp. 47-53, Springer-Verlag, 1984.

[2] D. Boneh and M. Franklin, “Identity-based encryption from the Weil pairing,” Advances in Cryptology - Crypto'01, LNCS 2139, pp. 213 - 229, Springer-Verlag, 2001

[3] V. S. Miller and A. K. Lenstra, “Weil Pairing and Its Efficient Calculation,” J. Cryptology (2004) 17: pp. 235–261, 2004.

[4] J. Callas, “Identity-based Encryption with Conventional Public-Key Infrastructure,” PGP Corporation Palo Alto, California, USA, jon@pgp.com.

[5] C. Gentry and A. Silverberg, “Hierarchical ID-Based Cryptography,” Proceedings of ASIACRYPT ’02, LNCS 2501, pp. 548-566, as <http://eprint.iacr.org/2002/056/>.

[6] X. Yi, “An Identity-Based Signature Scheme From the Weil Pairing,” IEEE Communication Letters, vol. 7, no. 2, Feb. 2003.

[7] M. Burmester and Y. Desmedt, “Identity-based key infrastructures,” Proceedings of the IFIP TC11 19th International Information Security Conference (SEC 2004), pp. 167–176, Kluwer, August 2004.

[8] A. Menezes, “An Introduction to Pairing-Based Cryptography,” Contemporary Mathematics, vol. 477, 2009.

[9] Y. Zheng, “Improved public key cryptosystems secure against chosen ciphertext attacks,” Technical Note, University of Wollongong, 1994.

[10] E. Fujisaki and T. Okamoto, “Secure integration of asymmetric and symmetric encryption schemes,” in Advances in Cryptology - Crypto'99, LNCS 1666, pp. 537-554, Springer-Verlag, 1999.

[11] R. Canneti, S. Halevi and J. Katz, “Chosen-ciphertext security from identity-based encryption,” Proc. Of Eurocrypt’04, LNCS 3027, pp. 207-222, Springer-Verlag, May 2–6, 2004, Interlaken, Switzerland.

[12] R. Lu and Z. Cao, “ID-based Encryption Scheme Secure against Chosen Ciphertext Attacks,” Cryptology ePrint Archive http://eprint.iacr.org/2005/355.

[13] N. Jansma and B. Arrendondo, “Performance Comparison of Elliptic Curve and RSA Digital Signatures,” nicj.net/files, Apr. 28, 2004.

[14] K. Gupta and S. Silakari, “ECC over RSA for Asymmetric Encryption: A Review,” International Journal of Computer Science Issues, vol. 8, issue 3, no. 2, pp. 370-375, May 2012.

[15] C. Gentry, “Certificate-Based Encryption and the Certificate Revocation Problem,” Proceedings of EUROCRYPT’03, LNCS 2656, pp.272-293, May 2003, Varsaw, Poland

[16] M. Bellare, A. Desai, D. Pointcheval, and P. Rogaway, “Relations among notions of security for public-key encryption schemes,” CRYPTO '98, volume 1462 of LNCS, pp. 26-45, Aug. 23-27, 1998, Santa Barbara, USA.

[17] M. Bellare and P. Rogaway, “Random oracles are practical: a paradigm for designing efficient protocols,” First ACM Conference on Computer and Communications Security, ACM, 1993.

[18] M. Bellare and P. Rogaway, “Introduction to Modern Cryptography,” Sep. 21, 2005, citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.124, [retrieved: Sep. 2012].

[19] D. Bickson and D. Malkhi, “A Study of Privacy in File Sharing Networks,” 2004, http://citeseerx.ist.psu.edu/viewdoc/similar?doi=10.1.1.7.3157, [retrieved: Sep. 2012].

[20] P. Yang, T. Kitagawa, G. Hanaoka, R. Zhang, K. Matsuura, and H. Imai, “Applying Fujisaki-Okamoto to Identity-Based Encryption,” Lecture Notes in Computer Science, 2006, Volume 3857/2006, pp. 183-192, DOI: 10.1007/11617983_18.

[21] A. Joux and K. Nguyen, “Separating Decision Diffie-Hellman from Diffie-Hellman in cryptographic groups,” (2001), http://eprint.iacr.org/2001/003.ps.gz, [retrieved: Sep. 2012].

[22] K. Park, H. Hwang, C. Lee, and S. Min, “DRM Technology Status and Contents Distribution Infrastructure Construction,” Journal of KIISE, vol. 23, no. 8, pp. 8-14, Sep. 2005. (in Korean).

Data Security Model for Cloud Computing

Eman M.Mohamed
Department of Computer Science, Menofia University
Faculty of computers and information
Egypt
emanhabib_1987@yahoo.com

Hatem S.Abdelkader
Department of Information Systems, Menofia University
Faculty of computers and information
Egypt
hatem6803@yahoo.com

Sherif El-Etriby
Department of Computer Science, Menofia University
Faculty of computers and information
Egypt
El_etriby100@yahoo.com

***Abstract*— From the perspective of data security, which has always been an important aspect of quality of service, Cloud computing focuses a new challenging security threats. Therefore, a data security model must solve the most challenges of cloud computing security. The proposed data security model provides a single default gateway as a platform. It used to secure sensitive user data across multiple public and private cloud applications, including salesforce, Chatter, Gmail, and Amazon Web Services, without influencing functionality or performance. Default gateway platform encrypts sensitive data automatically in a real time before sending to the cloud storage without breaking cloud application. It did not effect on user functionality and visibility. If an unauthorized person gets data from cloud storage, he only sees encrypted data. If authorized person accesses successfully in his cloud, the data is decrypted in real time for your use. The default gateway platform must contain strong and fast encryption algorithm, file integrity, malware detection, firewall, tokenization and more. This paper interested about authentication, stronger and faster encryption algorithm, and file integrity.**

Keywords- Cloud computing; Data Security model in cloud computing; Randomness testing; Cryptography for cloud computing; One Time Password (OTP).

I. INTRODUCTION

In the traditional model of computing, both data and software are fully contained on the user's computer; in cloud computing, the user's computer may contain almost no software or data (perhaps a minimal operating system and web browser, display terminal for processes occurring on a network).

Cloud computing is based on five attributes: multi-tenancy (shared resources), massive scalability, elasticity, pay as you go, and self-provisioning of resources, it makes new advances in processors, Virtualization technology, disk storage, broadband Internet connection, and fast, inexpensive servers have combined to make the cloud a more compelling solution.

The main attributes of cloud computing are illustrated as follows [1]:

- Multi-tenancy (shared resources): Cloud computing is based on a business model in which resources are shared (i.e., multiple users use the same resource) at the network level, host level, and application level.
- Massive scalability: Cloud computing provides the ability to scale to tens of thousands of systems, as well as the ability to massively scale bandwidth and storage space
- Elasticity: Users can rapidly increase and decrease their computing resources as needed.
- Pay as you used: Users to pay for only the resources they actually use and for only the time they require them.
- Self-provisioning of resources: Users self-provision resources, such as additional systems (processing capability, software, storage) and network resources.

Cloud computing can be confused with distributed system, grid computing, utility computing, service oriented architecture, web application, web 2.0, broadband network, browser as a platform, Virtualization, and free/open software [2].

Cloud computing is a natural evolution of the widespread adoption of virtualization, service-oriented architecture, autonomic, and utility computing [3]. Details are abstracted from end-users, who no longer have a need for expertise in, or control over, the technology infrastructure "in the cloud" that supports them as shown in figure 1.

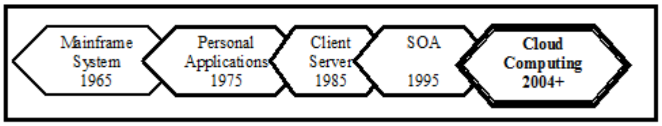


Figure 1. Evolution of cloud computing

Cloud services exhibit five essential characteristics that demonstrate their relation to, and differences from, traditional computing approaches such as On-demand self-service, Broad network access, Resource pooling, Rapid elasticity, and Measured service [4].

Cloud computing often leverages Massive scale, Homogeneity, Virtualization, Resilient computing (no stop computing), Low cost/free software, Geographic distribution, Service orientation Software and Advanced security technologies [4].

The main objective of this paper is to enhance data security model for cloud computing. The proposed data security model solves cloud user security problems, help cloud provider to select the most suitable encryption algorithm to its cloud. We also help user cloud to select the highest security encryption algorithm.

The proposed data security model is composed of three-phase defense system structure, in which each floor performs its own duty to ensure that the data security of cloud. The first phase is responsible for strong authentication. It applies the OTP (one time password) as a two-factor authentication system. OTP provides high security because it used one password in a session and cannot be cracked. The second phase selects the stronger and a faster encryption algorithm by proposing algorithm called "Evaluation algorithm". This algorithm used for selected eight modern encryption techniques namely: RC4, RC6, MARS, AES, DES, 3DES, Two-Fish, and Blowfish. The evaluation has performed for those encryption algorithms according to randomness testing by using NIST statistical testing. This evaluation uses Pseudo Random Number Generator (PRNG) to determine the most suitable. This evaluation algorithm performed at Amazon EC2 Micro Instance cloud computing environment. In addition, this phase checks the integrity of user data. It encourages cloud users to encrypt his data by using "TrueCrypt" software or proposed software called "CloudCrypt V.10". The third phase, ensure fast recovery of user data.

The paper is organized as follows, in section 2 cloud computing architecture is defined. Cloud computing security is discussed in section 3, in section 4 Methodology is described, finally in section 5 interruptions of the results are described.

II. CLOUD COMPUTING ARCHITECTURE

A. *Cloud computing service models*

- Cloud Software as a Service (SaaS): Application and Information clouds, Use provider’s applications over a network, cloud provider examples Zoho, Salesforce.com, Google Apps.
- Cloud Platform as a Service (PaaS): Development clouds, Deploy customer-created applications to a cloud, cloud provider examples Windows Azure, Google App Engine, Aptana Cloud.
- Cloud Infrastructure as a Service (IaaS): Infrastructure clouds, Rent processing, storage, network capacity, and other fundamental computing resources, Dropbox, Amazon Web Services, Mozy, Akamai.

B. *Cloud computing deployment models*

- Private cloud : Enterprise owned or leased
- Community cloud: Shared infrastructure for specific community
- Public cloud: Sold to the public, mega-scale infrastructure
- Hybrid cloud: Composition of two or more clouds

C. *Cloud computing sub-services models* [12]

- IaaS: DataBase-as-a-Service (DBaaS): DBaaS allows the access and use of a database management system as a service.
- PaaS: Storage-as-a-Service (STaaS): STaaS involves the delivery of data storage as a service, including database-like services, often billed on a utility computing basis, e.g., per gigabyte per month.
- SaaS: Communications-as-a-Service (CaaS) : CaaS is the delivery of an enterprise communications solution, such as Voice over IP, instant messaging, and video conferencing applications as a service.
- SaaS: SECurity-as-a-Service (SECaaS): SECaaS is the security of business networks and mobile networks through the Internet for events, database, application, transaction, and system incidents.
- SaaS: Monitoring-as-a-Service (MaaS): MaaS refers to the delivery of second-tier infrastructure components, such as log management and asset tracking, as a service.
- PaaS: Desktop-as-a-Service (DTaaS): DTaaS is the decoupling of a user’s physical machine from the desktop and software he or she uses to work.
- IaaS: Compute Capacity-as-a-Service (CCaaS) : CCaaS is the provision of “raw” computing resource, typically used in the execution of mathematically complex models from either a single “supercomputer” resource or a large number of distributed computing resources where the task performs well.

D. *Cloud computing benefits*

Lower computer costs, improved performance, reduced software costs, instant software updates, improved document format compatibility, unlimited storage capacity, device independence, and increased data reliability

E. *Cloud computing drawbacks*

Requires a constant Internet connection, does not work well with low-speed connections, can be slow, features might be limited, stored data might not be secure, and stored data can be lost.

F. *Cloud computing providers*

Amazon Web Services (AWS) –include Amazon S3, Amazon EC2, Amazon Simple-DB, Amazon SQS, Amazon FPS, and others. Salesforce.com – Delivers businesses over the internet using the software as a service model. Google Apps - Software-as-a-service for business email, information sharing and security. And others providers such as Microsoft Azure Services Platform, Proof-point, Sun Open Cloud Platform, Workday and so on.

ISBN: 978-1-61208-245-5

III. CLOUD COMPUTING SECURITY

With cloud computing, all your data is stored on the cloud. So cloud users ask some questions like: How secure is the cloud? Can unauthorized users gain access to your confidential data?

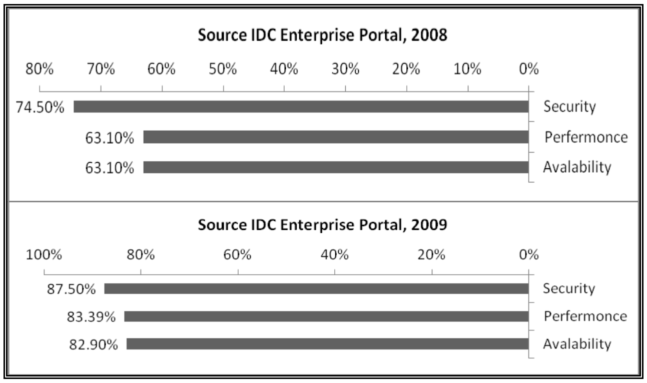


Figure 2. Security is a major concern to cloud computing[28]

Cloud computing companies say that data is secure, but it is too early to be completely sure of that. Only time will tell if your data is secure in the cloud. Cloud security concerns arising which both customer data and program are residing in provider premises. Security is always a major concern in Open System Architectures as shown in figure 2.

While cost and ease of use are two great benefits of cloud computing, there are significant security concerns that need to be addressed when considering moving critical applications and sensitive data to public and shared cloud environments. To address these concerns, the cloud provider must develop sufficient controls to provide the same or a greater level of security than the organization would have if the cloud were not used.

There are three types of data in cloud computing. The first type is a data in transit (transmission data), the second data at rest (storage data), and finally data in processing (processing data).

Clouds are massively complex systems can be reduced to simple primitives that are replicated thousands of times and common functional units, These complexities create many issues related to security as well as all aspects of Cloud computing. So users always worry about its data and ask where the data is? And who has access?. Every cloud provider encrypts the data in three types according to table 1.

TABLE I. DATA SECURITY [ENCRYPTION] IN CLOUD COMPUTING

Storage	**Processing**	**Transmission**
Symmetric encryption	***Homomophric encryption***	***Secret socket layer SSL encryption***
AES-DES-3DES-Blowfish-MARS…	Unpadded RSA- ElGamal …	SSL 1.0- SSL 3.0-SSL 3.1-SSL 3.2...

IV. METHODOLOGY

Security of data and trust problem has always been a primary and challenging issue in cloud computing. This section describes a proposed data security model in cloud computing. In addition, focuses on enhancing security by using an OTP authentication system, check data integrity by using hashing algorithms, encrypt data automatically with the highest strong/ fast encryption algorithm and finally ensure the fast recovery of data.

A. Proposed data at rest security model

The proposed data security model used three-level defense system structure, in which each floor performs its own duty to ensure that the data security of cloud as shown in figure 3.

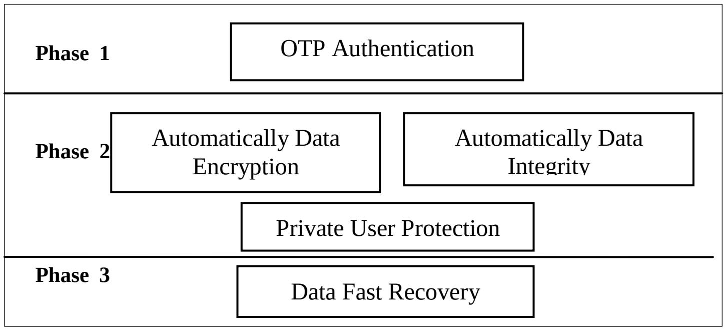


Figure 3. Proposed data security model in cloud computing

The first phase: strong authentication is achieved by using OTP.

The second phase: data are encrypted automatically by using strong/fast encryption algorithm. In addition to encrypt data, users can encrypt his sensitive data by using TrueCrypt software or proposed software CloudCrypt V.10. CloudCrypt software uses eight modern/strong encryption algorithms. Finally, data integrity is achieved by using hashing algorithms.

The third phase: fast recovery of user data is achieved in this phase.

The three phases are implemented in default gateway. As shown in figure 4. The proposed data security model provides a single default gateway as a platform to secure sensitive customer data across multiple public and private cloud applications, including salesforce, Gmail, and Amazon Web Services, without affecting functionality or performance.

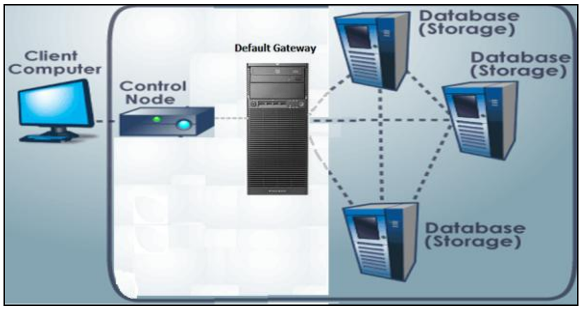


Figure 4. How data stored in the cloud by using the proposed data security model?

Default gateway platform tasks:

ISBN: 978-1-61208-245-5

- Encrypt sensitive data automatically on a real time before sending to the cloud without breaking cloud application.
- The default gateway platform did not effect on user functionality and visibility.
- If an unauthorized person gets data from cloud storage, he can see the encrypted data.
- If authorized person access success in his cloud, the data is decrypted in real time for your use.
- The default gateway platform must contain Strong/Fast Encryption Algorithm.
- The default gateway platform must contain File integrity.
- The default gateway platform must contain Malware detection, Firewall, Tokenization and more.

Proposed data security model implemented and applied to cloudsim 3.0 by using HDFS architecture and Amazon web services (S3 and EC2).

In this paper, automatically encryption, integrity, fast recovery and private user encryption all are achieved in the proposed data security model.

B. *Implementation details*

1) In first phase, Authentication:

a) The cloud user select company, then create an account

b) Cloud provider upload user information in DB in cloud storage

c) Cloud Provider confirms user with his username and password

d) Cloud user request login page

e) The cloud provider displays login screen

f) Cloud user login with username and password

g) A cloud provider check is valid username and password by searching in DB in cloud storage. If user information not valid display error message else display reserve a PC page.

h) Cloud user reserves your PC

2) OTP authentication steps:

a) Cloud user enters passphrase, challenge and sequence number for OTP authentication

b) Cloud user generates an OTP

c) The cloud provider generates the OTP temporary DB based on user information

d) Cloud user login with OTP

e) A cloud provider check is valid OTP by searching in temporary DB for OTP in cloud storage. If OTP not valid display error message else display user PC page.

3) In second phase, Private user protection

a) Before adding data, cloud user can encrypt data by using TrueCrypt or CloudCrypt software's.

b) In second phase, Automatic data encryption

c) Cloud user adds data.

d) Cloud server encrypt data automatically by using fast/strong encryption algorithm that selected based on an evaluation algorithm for the cloud company

4) In second phase, Automatic check data integrity

a) The cloud server generates file hash value

b) Cloud server store data with its hash value

c) When a cloud user requests his data, cloud server decrypt data automatically, check integrity by check the hash value.

5) In third phase, fast recovery of data

a) Finally, cloud server retrieves data with message of file integrity.

C. *Proposed Evaluation Algorithm*

We use NIST statistical tests to get the highest security encryption algorithm from eight algorithms namely RC4, RC6, MARS, AES, DES, 3DES, Two-Fish, and Blowfish as shown in figure 6. NIST Developed to test the randomness of binary sequences produced by either hardware or software based cryptographic random or pseudorandom number generators.

NIST statistical tests has 16 test namely The Frequency (Mon-obit) Test, Frequency Test within a Block, The Runs Test, Tests for the Longest-Run-of-Ones in a Block, The Binary Matrix Rank Test, The Discrete Fourier Transform (Spectral) Test, The Non-overlapping Template Matching Test, The Overlapping Template Matching Test, Maurer's "Universal Statistical" Test, The Linear Complexity Test, The Serial Test, The Approximate Entropy Test, The Cumulative Sums (Cusums) Test, The Random Excursions Test, and The Random Excursions Variant Test.

We also compare between eight encryption algorithms based on speed of encryption to achieve faster recovery.

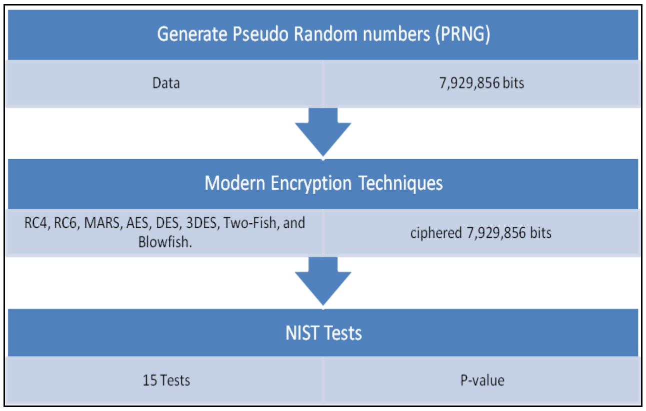

Generate Pseudo Random numbers (PRNG)
Data
7,929,856 bits
Modern Encryption Techniques
RC4, RC6, MARS, AES, DES, 3DES, Two-Fish, and Blowfish.
ciphered 7,929,856 bits
NIST Tests
15 Tests
P-value

Figure 5. Steps to select the highest encryption algorithm

We use Amazon EC2 as a case study of our software. Amazon EC2 Load your image onto S3 and register it. Boot your image from the Web Service. Open up the required ports for your image. Connect to your image through SSH. And finally execute your application.

For our experiment in a cloud computing environment, we use Micro Instances of this Amazon EC2 family, provide a small amount of consistent CPU resources, they are well suited for lower throughput applications, 613 MB memory, up to 2 EC2 Compute Units (for short periodic bursts), EBS (Elastic Block Store) storage only from 1GB to 1TB, 64-bit platform, low I/O Performance, t1.micro API name, We use Ubuntu Linux to run NIST Statistical test package [9-11].

D. *Selection the highest encryption algorithm steps*

Sign up for Amazon web service to create an account. Lunch Micro instance Windows (64 bit) Amazon EC2. Connect to Amazon EC2 windows Micro Instance. Generate 128 plain stream sequences as PRNG, each sequence is 7,929,856 bits in length (991232 bytes in length) and key stream (length of key 128 bits). Apply cryptography algorithms to get ciphers text. Lunch Micro instance Amazon EC2 Ubuntu Linux Connect to Amazon EC2 Ubuntu Linux Micro instance Run NIST statistical tests for each sequence to eight encryption algorithms to get P-value Compare P-value to 0.01, if p-value less than 0.01 then reject the sequence.

We compare between eight encryption methods based on P-value, Rejection rate and finally based on time consuming for each method.

We have 128 sequences (128-cipher text) for each eight-encryption algorithm.

Each sequence has 7,929,856 bits in length (991232 bytes in length). Additionally, the P-values reported in the tables can find in the *results.txt* files for each of the individual test – not in the *finalAnalysisReport.txt* file in NIST package.

The P - value represents the probability of observing the value of the test statistic which is more extreme in the direction of non-randomness. P-value measures the support for the randomness hypothesis on the basis of a particular test Rejection Rate number of rejected sequences (P-value less than significance level α may be equal 0.01 or 0.1 or 0.05). The higher P-Value the better and vice versa with rejection rate, the lower the better [19].

For each statistical test, a set of P-values (corresponding to the set of sequences) is produced. For a fixed significance level α, a certain percentage of P-values are expected to indicate failure. For example, if the significance level is chosen to be 0.01 (i.e., $\alpha \geq 0.01$), then about 1 % of the sequences are expected to fail. A sequence passes a statistical test whenever the P-value $\geq \alpha$ and fails otherwise.

We produce P-value, which small P-value(less than 0.01) support non-randomness. For example, if the sample consists of 128 sequences, the rejection rate should not exceed 4.657, or simply expressed 4 sequences with α = 0.01. The maximum number of rejections was computed using the formula [20]:

$$\text{Rejection rate \#}= s\left(\alpha + 3\sqrt{\frac{\alpha(1-\alpha)}{s}}\right) \qquad (1)$$

Where s is the sample size and is α is the significance level is chosen to be 0.01.

V. SIMULATION RESULTS

In this section, we show and describe the simulation results of the proposed data security model.

A. *OTP Authentication*

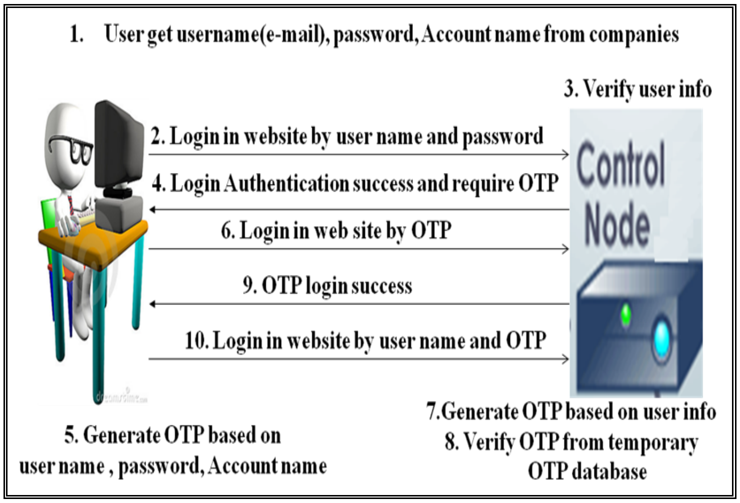

1. User get username(e-mail), password, Account name from companies
3. Verify user info
2. Login in website by user name and password
4. Login Authentication success and require OTP
6. Login in web site by OTP
9. OTP login success
10. Login in website by user name and OTP
Control Node
5. Generate OTP based on user name , password, Account name
7.Generate OTP based on user info
8. Verify OTP from temporary OTP database

Figure 6. OTP authentication in PDSM

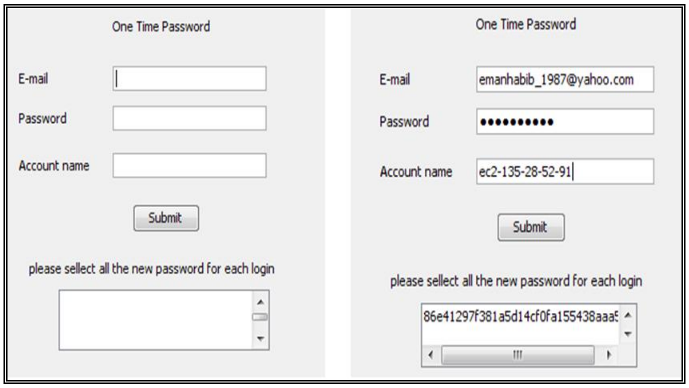
OTP System steps as shown in figure 6.

The users connect to the cloud provider. Then the user gets the username (e-mail), password and finally account password.

Users login to the cloud provider website by getting username (e-mail), password and account password.

Cloud node controller verifies user info. If user info is true, controller-node send that login authentication success and require OTP.

OTP generation software used to generate OTP as shown in figure 7.


One Time Password

E-mail

Password

Account name

Submit

please sellect all the new password for each login

One Time Password

E-mail: emanhabib_1987@yahoo.com

Password: ●●●●●●●●●●

Account name: ec2-135-28-52-91

Submit

please sellect all the new password for each login

86e41297f381a5d14cf0fa155438aaa5

Figure 7. Proposed software for OTP Generation

Users generate OTP by using MD5 hash function and sequence number based on user name, password and account password.

Then users login to cloud website with OTP as shown in figure 8.

ISBN: 978-1-61208-245-5

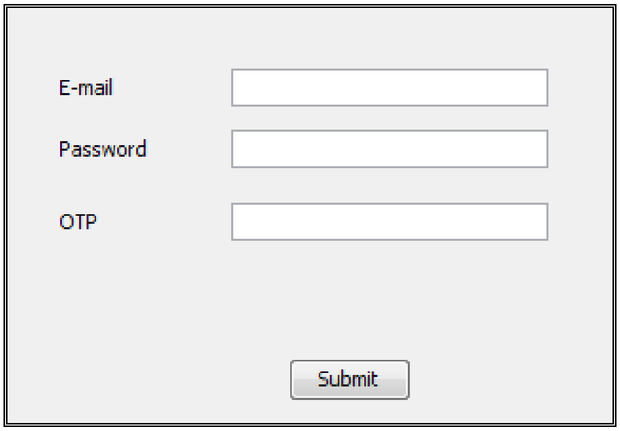


Figure 8. proposed OTP login screen

The cloud controller node generates 1000 OTP based on user info by using the MD5 hash function. Then the cloud controller saves 1000 OTP in the temporary OTP database.

The cloud controller verifies user OTP from the temporary OTP database.

If OTP is true, send OTP login success.

We have compared password space with different password schemas; we can identify the most secure approaches with respect to brute force attack as shown in table 2. This table shows the comparison of the password space and password length for popular user authentication schemas for cloud computing. The next table shows that the approach presented by us is both more secured and the easiest to remember. At the same time, it is relatively fast to produce during an authentication procedure as shown in figure 9 and figure 10.

TABLE II. PASSWORD SPACE COMPARISON

Authentication System	Alpha bet	Password Length	Password space size	Entropy bits
Static password	82	12	92.4 * 1021	22.96
PIN number	10	12	1 * 1012	12
OTP	40	30	1.15 * 1048	48.06

We must remember that, a one-time password (OTP) is a password that is valid for only one login session or transaction. OTPs avoid a number of shortcomings that are associated with traditional (static) passwords. The most important shortcoming that is addressed by OTPs is that, in contrast to static passwords, they are not vulnerable to replay attacks. This means that, if a potential intruder manages to record an OTP that was already used to log into a service or to conduct a transaction, he or she will not be able to abuse it since it will be no longer valid. On the downside, OTPs are difficult for human beings to memorize. Therefore they require additional technology in order to work.

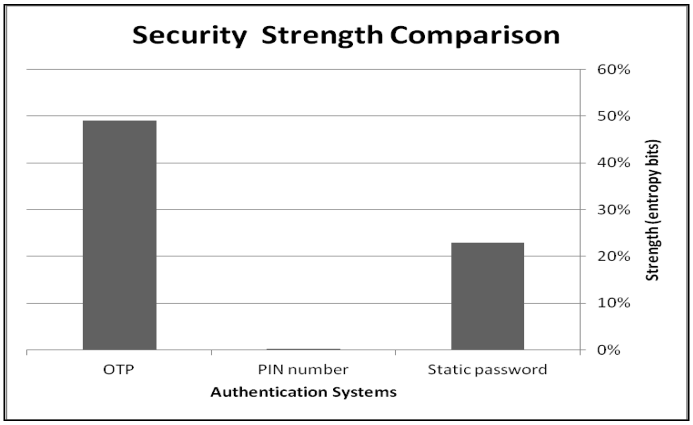


Figure 9. Security strength comparison based on entropy bits

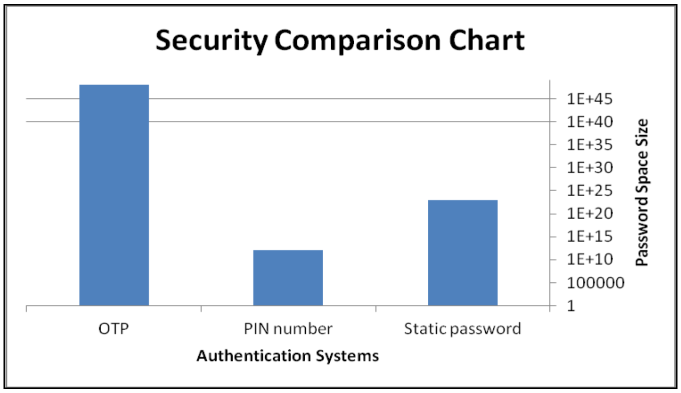


Figure 10. Security strength comparison based on password space size

Benefits of OTP in cloud computing

- OTP offers strong two-factor authentication,
- The OTP is unique to this session and cannot be used again
- OTP offers strong security because they cannot be guessed or hacked
- Provides protection from unauthorized access
- Easier to use for the employee than complex frequently changing passwords
- Easy to deploy for the administrator
- Good first step to strong authentication in an organization
- Low cost way to deploy strong authentication

B. Evaluation Algorithm Results

In this paper, we select the strongest and a the fastest encryption algorithm by proposing algorithm called "Evaluation algorithm". This algorithm used for selecting eight modern encryption techniques namely: RC4, RC6, MARS, AES, DES, 3DES, Two-Fish, and Blowfish. The evaluation has performed for those encryption algorithms according to randomness testing by using NIST statistical testing. This evaluation uses Pseudo Random Number

Generator (PRNG) to determine the most suitable. This evaluation algorithm performed at Amazon EC2 Micro Instance cloud computing environment.

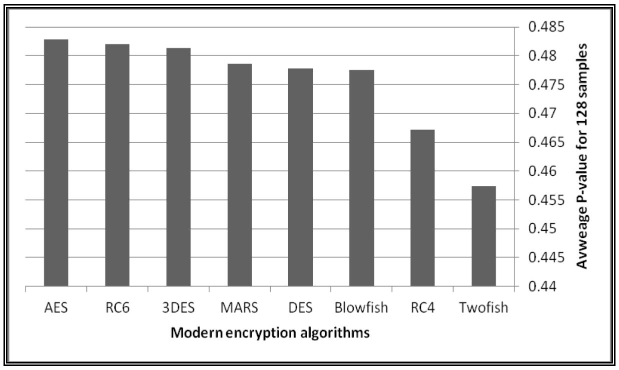


Figure 11. Amazon EC2 Average P-value for eight modern encryption algorithms based on 16 NIST test

Experimental results for this comparison point are shown in figure 11 to indicate the highest security for modern encryption techniques. The results show the superiority of the AES algorithm over other algorithms in terms of the P-value. Another point can be noticed here; that RC6 requires more P-value than all algorithms except AES. A third point can be noticed here; that 3DES has an advantage over other DES, RC4, MARS, 3DES and Twofish in terms of P-value. Finally, it is found that Twofish has low security when compared with other algorithms.

Experimental results for this comparison point are shown Figure 12 to indicate the speed of encryption/decryption. The results show the superiority of the Blowfish algorithm over other algorithms in terms of the processing time. Another point can be noticed here; that AES requires less time than all algorithms except Blowfish. A third point can be noticed here; that RC4 has an advantage over other DES, RC6, MARS, 3DES and Twofish in terms of time consumption . A fourth point can be noticed here; that 3DES has low performance in terms of power consumption when compared with DES. It always requires more time than DES because of its triple phase encryption characteristics. Finally, it is found that Twofish has low performance when compared with other algorithms.

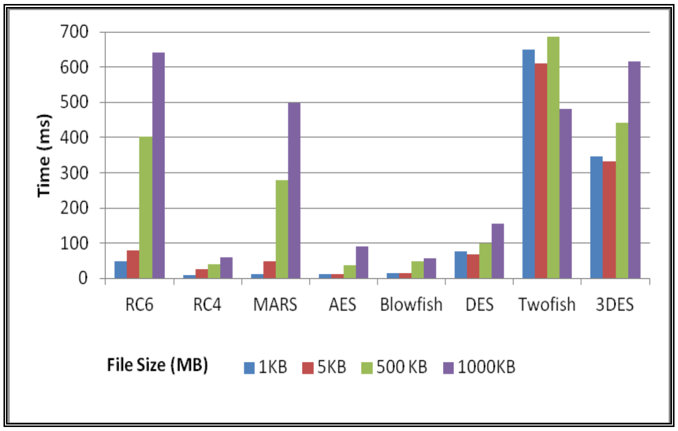


Figure 12. Encryption/decryption comparison with different size in Amazon EC2

C. Private User protection

Amazon web services encourage user's to encrypt sensitive data by using TrueCrypt software. A new computer software program is implemented to encrypt data before storing in cloud storage devices. This software enables users to choose from eight encryption techniques namely: AES, DES, 3DES, RC4, RC6, Twofish, Blowfish, and MARS as shown in figure 13.

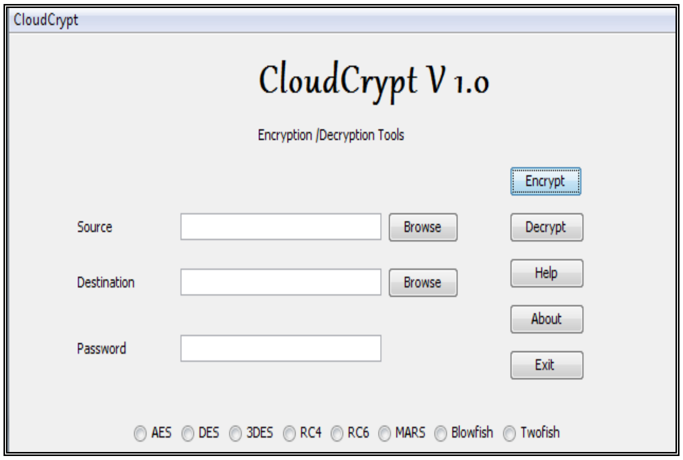


Figure 13. Proposed encryption software CloudCrypt at runtime in Amazon EC2

D. Ensuring Integrity

This is an extra concern for customers that now they have to worry about how to keep data hidden from auditors. The actual problem of "trust" remains the same. In order to avoid third party auditors in this chain, this paper propose that the integrity check of data stored in the cloud can be checked on customer's side. This integrity check can be done by using cryptographic hash functions.

For integrity check, we have to think about a simple solution that is feasible and easy to implement for a common user. The trust problem between Cloud storage and customer can be solved, if users can check the integrity of data

ISBN: 978-1-61208-245-5

themselves instead of renting an auditing service to do the same. This can be achieved by hashing the data on user's side and storing the hash values in the cloud with the original data. As shown in figure 14. This figure presents the overview of the scheme

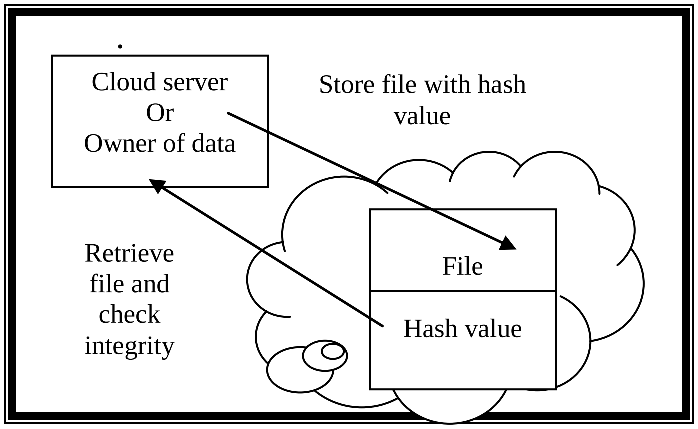
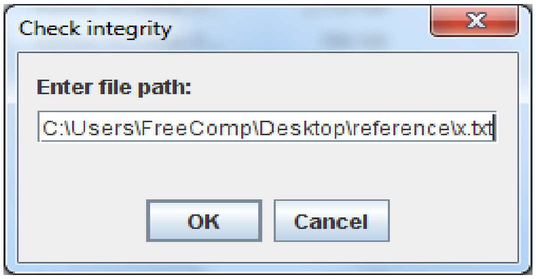


Figure 14. Overview of integrity check with hash functions

Integrity Check using Hash Function steps

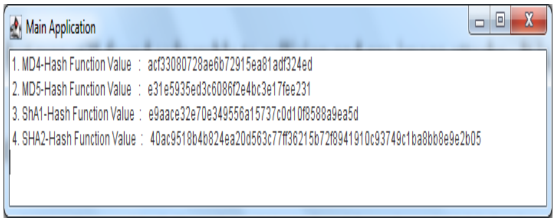
- The program takes file path that as shown in figure 15.

- The program computes a four-hash values in this file based on the four hash functions (MD4, MD5, SHA-1, and SHA-2) as shown in figure 16.



Enter file path:
C:\Users\FreeComp\Desktop\reference\x.txt

Figure 15. Screen shot of Check integrity program



```
1. MD4-Hash Function Value  :  acf33080728ae6b72915ea81adf324ed
2. MD5-Hash Function Value  :  e31e5935ed3c6086f2e4bc3e17fee231
3. ShA1-Hash Function Value  :  e9aace32e70e349556a15737c0d10f8588a9ea5d
4. SHA2-Hash Function Value  :  40ac9518b4b824ea20d563c77ff36215b72f8941910c93749c1ba8bb8e9e2b05
```

Figure 16. Check integrity program calculating hash values

- When users store data in cloud storage devices, server store filled with four hash values.
- When a user retrieve data file, server generate four hash values
- Server check integrity by comparing new four hash values with stored four hash values.

The following are the advantages of using the utility:

- Not much implementation effort required.
- Cost effective and more secured.
- Does not require much time to compute the hash values.
- Flexible enough to change the security level as required.
- Not much space required to store the hash values.

VI. CONCLUSION

According to the simulation results, in the authentication phase in the proposed data security model, OTP is used as two-factor authentication software. OTP archived more password strength security than other authentication systems (BIN and static password). This appears by comparing between OTP, BIN, and static password authentication systems based on the space time size and entropy bits.

From the simulation results of the second phase in the proposed data security model, test the proposed system in Ubuntu Amazon Micro Instance EC2, and from randomness and performance evaluation to eight modern encryption algorithms AES is the best encryption algorithm in Ubuntu Amazon Micro Instance EC2. In addition to the randomness and performance evaluation, data integrity must be ensured. Moreover, the proposed data security model encourages users to use true-crypt to encrypt his/her sensitive data.

From the comparison and performance evaluation, fast recovery of data achieved to the user. These appear in the proposed data security model third phase.

From the comparison and performance evaluation, cloud computing depend on some condition however it has advanced security technologies rather than traditional desktop. The summarized results of proposed data security model are shown in table 3.

TABLE III. SUMMARIZED RESULTS OF THE PROPOSED DATA SECURITY MODEL IN CLOUD COMPUTING

Features	**Description**
Authentication	OTP Authentication System (mathematical generation)
Provider Encryption	Software implemented to select the highest security and faster encryption algorithm based on NIST statistical tests. This software select AES algorithm to Micro Instance ubunto Amazon EC2 with Amazon S3.
Private user Encryption	TrueCrypt system or proposed software CloudCrypt v.10
Data integrity	Hashing- MD5- MD4-SHA-1-SHA-2
Data fast recovery	Based on decryption algorithm speed
Key management	User keys not stored in provider control domain

REFERENCES

[1] Center Of The Protection Of National Infrastructure CPNI by Deloitte "Information Security Briefing 01/2010 Cloud Computing", p.71 , Published March 2010.

[2] lan Foster, Yong Zhao, Ioan Raicu, Shiyong Lu, " Cloud Computing and Grid Computing 360-Degree Compared " Grid Computing Environments Workshop, 2008. GCE '08 p.10, published 16 Nov 2008.

ISBN: 978-1-61208-245-5

[3] Jeremy Geelan, "Twenty-One Experts Define Cloud Computing", cloud computing journal, published january 24, 2009.

[4] National Institute of Science and Technology."The NIST Definition of Cloud Computing".p.7. Retrieved July 24 2011.

[5] Ngongang Guy Mollet, "Cloud Computing Security" Thesis, p. 34 + 2 appendices Published April 11, 2011.

[6] Rajkumar Buyya, Chee Shin Yeo, and Srikumar Venugopal, " Market-Oriented Cloud Computing: Vision, Hype, and Reality for Delivering IT Services as Computing Utilities" Department of Computer Science and Software Engineering, University of Melbourne, Australia. p. 9. Retrieved July 31 2008.

[7] Mladen A. Vouk "Cloud Computing- Issues, Research and mplementations" Journal of Computing and Information Technology -CIT 16, 2008, 4, 235–246

[8] Dai Yuefa, Wu Bo, Gu Yaqiang, Zhang Quan, Tang Chaojing"Data Security Model for Cloud Computing"Proceedings of the 2009 international Workshop on Information Security and Application IWISA 2009) Qingdao, China, November 21-22, 2009.

[9] Amazon EC2 API ," Amazon Elastic Compute Cloud Developer Guide " http://docs.amazonwebservices.com/AWSEC2/2006-10-01/DeveloperGuide/Amazon Elastic Compute Cloud Developer Guide, published 2006-10-01

[10] Amazon Web Services, "Amazon Simple Storage Service Developer Guide , " http://docs.amazonwebservices.com/AmazonS3/2006-03-01/ Amazon Simple Storage Service Developer Guide , published 2006-03-01

[11] Amazon Web Services," Overview of Security Processes" http://aws.typepad.com/aws/2009/08/introducing-amazon-virtual-private-cloud-vpc.html, September 2009.

[12] Cloud Security Alliance Guidance, "Security Guidance For Critical Areas of Focus In Cloud Computing V1.0", www.cloudsecurityalliance.org/guidance/csaguide.v1.0.pdf, published April 2009

[13] Cloud Security Alliance Guidance," Security Guidance For Critical Areas of Focus In Cloud Computing V2.1", www.cloudsecurityalliance.org/guidance/csaguide.v2.1.pdf, published Dec 2009

[14] Cloud Security Alliance Guidance, "Security Guidance For Critical Areas of Focus In Cloud Computing V3.0 , www.cloudsecurityalliance.org/guidance/csaguide.v3.0.pdf published 11/14/2011

[15] Luis M. Vaquero1, Luis Rodero-Merino1 , Juan Caceres1, Maik Lindner2 "A Break in the Clouds: Towards a Cloud Definition ", ACM SIGCOMM Computer Communication Review, Vol 39, Number 1, published January 2009

[16] John W. Rittinghouse James F. Ransome "Cloud Computing Implementation, Management, and Security" book, published 17 Aug 2009.

[17] Mark Baker, "An Introduction and Overview of Cloud Computing",43 slides, published 19th May, 09 . http://acet.rdg.ac.uk/~mab/Talks/Clouds-La-Coruna09/Talk.ppt

[18] Cloud Security Alliance "Top Threats to Cloud Computing V1.0" , March 2010.

[19] Andrew Rukhin, Juan Soto, James Nechvatal, Miles Smid,: "A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications", April 2010 .

[20] Affiliation Juan Soto, National Institute of Standards and Technology 100 Bureau Drive, Stop 8930 Gaithersburg "Randomness Testing of the Advanced Encryption Standard Candidate Algorithms".

[21] Carolynn Burwick c , Don Coppersmith "The MARS Encryption Algorithm " , published August 27, 1999

[22] Dawson, Helen Gustafson, Matt Henricksen, Bill Millan. " Evaluation of RC4 Stream Cipher , Information Security Research Centre Queensland University of Technology", July 31, 2002

[23] W.Stallings, "Cryptography and Network Security 4th Ed," Prentice Hall , 2005,PP. 58-309 .

[24] Coppersmith, D. "The Data Encryption Standard (DES) and Its Strength Against Attacks."I BM Journal of Research and Development, May 1994,pp. 243 -250.

[25] Daemen, J., and Rijmen, V. "Rijndael: The Advanced Encryption Standard."D r. Dobb's Journal, March 2001,PP. 137-139.

[26] Bruce Schneier. "The Blowfish Encryption Algorithm Retrieved ",October 25, 2008,

[27] John Kelseyy Doug Whitingz David Wagnerx Chris Hall, "Two_sh: A 128-Bit Block Cipher" , Niels Ferguson k , 15 June 1998

[28] http://www.csrc.nist.gov/groups/SNS/cloud-computing/cloud-computing-v26.ppt.

Energy Efficiency and Cost Optimization of OTDR Supervision Systems for Monitoring Optical Fiber Infrastructures

J. Montalvo, José A. Torrijos, R. Cantó, I. Berberana
Access Evolution
Telefónica I+D
Madrid, Spain
jmg@tid.es

***Abstract*—A new optical fiber supervision architecture based on Multi-Wavelength Optical Time Domain Reflectometer (OTDR) and hybrid active/passive fiber-optic cross-connect system (FOCS) using Wavelength Division Multiplexing (WDM) multiplexer-demultiplexers is reported. The results of our study show that up to 60% cost reduction and 50% energy savings can be obtained using a 4 wavelengths OTDR-based supervision system. Experimental validation of the architecture is also reported.**

Keywords-Optical Time Domain Reflectometer (OTDR); Wavelength Division Multiplexing (WDM); optical fiber, optical supervision.

I. INTRODUCTION

In today's Passive Optical Network (PON) systems, the physical infrastructure is not entirely visible to the Network Management System (NMS). As a direct consequence, a physical failure is not detected before creating service outage in upper layers, which in turn may lead to tremendous loss in business for the operators. These arguments have been gaining importance as the warranty on the quality of the infrastructure becomes a deciding factor in the strongly competitive market place [1]. The aim of preventive maintenance is to detect any kind of deterioration in the network that can cause suspended services and to localize these faults in order to avoid specially trained people deployed with dedicated and often expensive equipments, which increases operation-and-maintenance expenses (OPEX).

PON infrastructure does not only suffer from accidental damages and environmental effects (e.g. water penetration in splice closures) but are also subject to a lot of changes after the network is installed and activated. As an example, the optical access network may not be initially fully loaded; subscribers would be turned up, possibly over an extended period of time [2]. Hence, network operators should continuously be aware if a change noticed by its monitoring system is service oriented or indeed a fault. So, the existing maintenance methods in PONs [3] need to be updated.

The most common maintenance tool employed for troubleshooting in long-haul, point-to-point fiber optic links is an Optical Time Domain Reflectometer (OTDR).

OTDR measurements can also be applied to PON systems, and there are three main approaches for doing that:

- Using dark fibers accompanying the feeder fibers of PONs, physically bypassing the first level of splitting up to the second level splitting. The efficiency of this approach relies on the gathering of information from active elements [4] and the fact that PON fibers can probabilistically share a high percentage of the same fiber cable infrastructure.
- Performing in-line measurements inside each PON fiber infrastructure by multiplexing an OTDR signal inside each PON feeder fiber at the Central Office, using Wavelength Division Multiplexers. In this approach, the OTDR signal is generated by an external optical source at a different wavelength than data signals.
- Performing in-line measurements inside each PON fiber infrastructure by generating the OTDR signal inside the PON transceiver at the OLT.

The management of the optical layer in PON systems is being standardized in [5].

The in-line measurements using the integrated OTDR signal inside the PON data transceiver is a challenging approach [6] whose implementation depends on the transceiver implementation and the physical media dependant layer of a particular PON technology. Even though GPON and EPON standards are completely closed, and XG-PON1 is on its way, no commercial product has appeared up to now for those systems; the technological uncertainty of XG-PON2 and NG-PON2 makes even more difficult the commercial adoption of this approach for PON supervision in a massive way.

On the other hand, external OTDR approaches, either using dark fibers or by multiplexing the OTDR signal inside the PON feeder fibers by using WDMs is an already commercially available tool.

Central Offices (CO) with PON technologies can typically cover between 10 to 50 thousand homes passed (HP), and even higher in Long-Reach PON scenarios [7]. For a splitting ratio of 1:64, this means that there would be required up to 800 PON interfaces from a single CO. Additional fibers used for metro and core systems must also be considered.

In order to share the cost of OTDR measurement equipment between all these PON interfaces, fiber switches are typically used, thus launching the OTDR pulses on a selected fiber at a certain time either in a periodic way (preventive measurement) or on demand (after a detected

Copyright (c) IARIA, 2013. ISBN: 978-1-61208-245-5

alarm), switching to other fibers when required. Fiber Optic Cross-Connected Systems (FOCS) are used to address this need.

The most suitable optical fiber switching technologies for FOCS implementations are Micro-Electro-Mechanical (MEM) switches [8] and opto-mechanical switches [9], being both types active elements which require power supply for operating.

The OTDR implementation inside PON transceivers is a very interesting proposal, but suffers from a high uncertainty due to both technological challenges and slow standardization advances, thus being difficult to have a commercial solution and massive deployment able to address the current and mid-term supervision requirements of FTTx network operators.

Regarding the existing external OTDR approaches,

- they lack of efficient scalability as more fibers are deployed requiring to be monitored as FTTx services become to be massively deployed. As new FTTx feeder fibers are deployed smoothly, a large number of active switches are required increasing the power consumption of the system.
- they require a high number of fibers or electrical supply points to be installed in order to increase the number of test ports of a OTDR supervision system. If a new active switch is installed close to fibers under test, a new electrical supply is required. If the same active switch is installed close to an already available electrical supply, longer fiber links are required to deliver the test signal to the fibers to be monitored.
- they have the risk of blocking a fiber connection in case of power supply failure at a certain switching stage of the fiber-optic cross-connects.

In this paper, we propose a new approach for optical fiber infrastructure supervision from Central Offices, using Tunable/Multiple-Wavelength OTDR and Wavelength Division Multiplexing (WDM) techniques, see Fig. 1, by employing hybrid switching elements which combine active and passive elements, see Fig. 2.

II. Proposed Multi-Wavelength OTDR Supervision System

Instead of having a single wavelength operating OTDR, we propose to use N different wavelengths and WDM passives in the FOCS. By tuning the OTDR wavelength, the WDM passives combined with optical switches will deliver the test signal to the desired fiber under test, which may be part of the PON, metro and/or long haul fiber transmission operator infrastructures.

The proposed solution relies on two key factors:

- OTDR wavelength tunability. The OTDR pulses can be launched at different wavelengths, all of them within the legacy waveband already established for fiber monitoring. In the case of in-line monitoring, this can be the U-band (1625-1675nm) for access systems [10], and any available channel of the employed wavelength grid in WDM metro/core systems. Dark fiber supervision can be performed using arbitrary wavelengths.
- Wavelength multiplexer-demultiplexer filters at some stage/s of the FOCS. The OTDR pulses are delivered to a selected fiber in a passive and inherent way depending on the wavelength of the OTDR. At certain parts of the optical fiber switching system, wavelength demultiplexing of OTDR signals is used as a fiber selection mechanism, instead of mechanically moving input fiber to a desired fiber output or using MEM switches.

A. Supervision NMS operation

The Supervision NMS communicates with the local management system in the CO for monitoring the M fibers of the system. The local management system is a local subsystem performing as interface for the NMS to obtain measurements on specific fibers, thus configuring the OTDR as well as all the switches in all the stages to prepare an optical path for delivering the test signal to a specific fiber to test. The local management should link the inventory information to the physical interconnections of the fiber ports of the different switching stages between them, and with the M fibers under test.

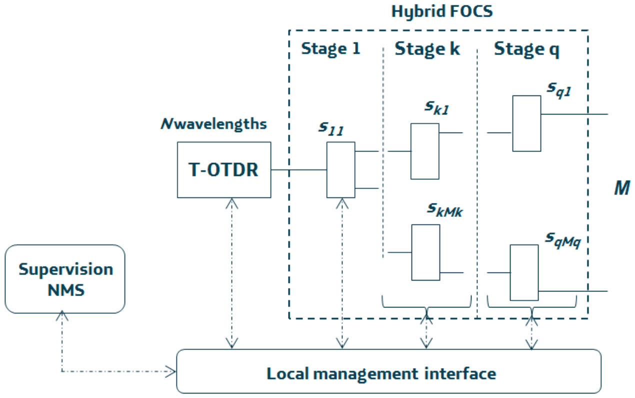


Figure 1. Proposed hybrid Fiber Optic Cross-Connect System (FOCS) for OTDR fiber supervision. T-OTDR: Tunable Optical Time Domain Reflectometer.

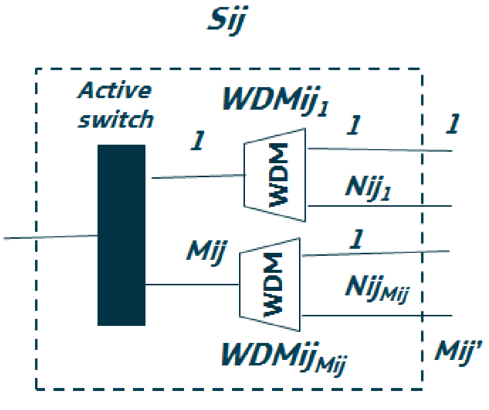


Figure 2. Generic architecture of the proposed hybrid switching element sij.

B. *Switching elements description*

We propose a hybrid switch element sij design, reducing the active switching components and/or using optical wavelength multiplexer-demultiplexers WDMijk, see Fig. 2.

At each of the Mij ports of the active switch, a 1xNijk WDM (k=1...Mij) passive is used. The number of output ports of sij is:

$$Mij' = \sum Nijk \ (k=1...Mij). \qquad (1)$$

This configuration allows two relevant advantages:

- FOCS WDM passive scalability: An increased number of ports (Mij>Mij) can be achieved in a passive and cost-effective way by increasing each sij output port of an active switch with Nijk ports (Nijk ⩾1).
- Reduced active components in switching elements sij. In case that an increase in the number of output ports is not desired for a particular FOCS design, it is possible to reduce the value of Mij and use WDM filters to keep the total number of desired outputs. This allows a reduction in the cost and energy consumption of the switching elements in the FOCS.

The proposed hybrid switching elements can also be implemented in a totally passive way (without active switch). In that particular case, an active switch is completely replaced by a passive wavelength multiplexer-demultiplexer.

Fiber selectors sij can either replace a design with less active elements while keeping the same number of fiber outputs (M) for system cost and power consumption reduction, or they can increase the number of output fibers in a cost effective way using the WDM scalability approach of the invention without increasing the power consumption.

The WDM filters may slightly increase the insertion loss of the OTDR signal through the FOCS with regards to active switches, see Tab. 1. Nevertheless this should not be considered a restriction in most cases, as the loss of dynamic range of OTDR will not be very significant compared with the total range (~40dB) or could be compensated with an increased measurement time (more averaging) or wider test pulses.

TABLE I. TYPICAL INSERTION LOSSES IN COARSE WDM (C-BAND) WITHOUT CONNECTORS

#channels	Active switches	Thin Film Filters (TFF) WDM
2	0.5-1.5	1.4-1.8
4	0.6-1.7	1,6-2.0
8	0.6-1.7	1.8-2.5
16	0.6-1.7	3.8-4.5
32	0.6-1.7	4.8-5.5
40	0.6-2.3	5.2-6.0

III. CAPITAL EXPENDITURE (CAPEX) AND POWER EFFICIENCY ANALYSIS OF A 4 WAVELENGTHS OTDR AND FOCS DESIGN

In this section, a conventional FOCS implementation using a single wavelength OTDR and active switches is compared with the proposed alternative approach using a 4 wavelength OTDR (N=4) and 1x4 ports multiplexers/demultiplexers (Nijk=4) design example, see Fig. 3.

The comparative analysis has been performed for different number of fibers under test (16 to 1024). For each number of test fibers, the proposed FOCS uses an active switch with a 4 times lower number of outputs that in a conventional system, and adds ports in groups of 4 by cascading 1x4 WDMs. In the case of GPON and EPON deployments, the proposed approach can be deployed as any already available PON supervision solution with external OTDR, using commercial triplexers in the Central Office, which combine optical data and OTDR signals.

TABLE II. COST AND POWER MODEL PARAMETERS

Element	Cost (a.u.)	Power Cons. (W)
1-wavelenth OTDR	1.00	40
4-wavelengths OTDR	2.09	40
Active switch 1x4	0.82	0.6
Active switch 1x8	1.12	0.6
Active switch 1x32	3.21	2.8
Active switch 1x64	5.22	5.8
4-channels Mux/Demux	0.04	0.0

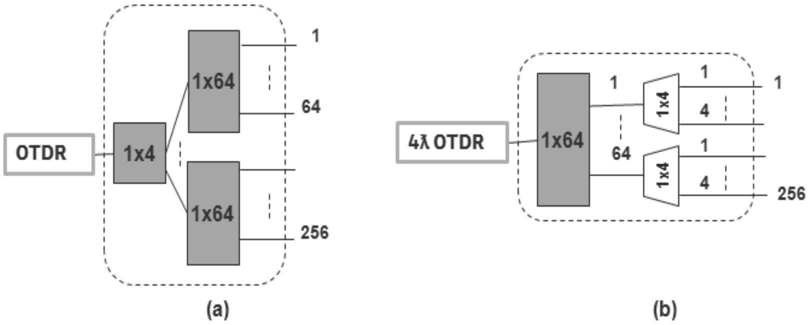


Figure 3. Centralized OTDR supervision system (a) with active FOCS and (b) proposed system with hybrid FOCS (4 wavelengths) for 256 supervision fibers.

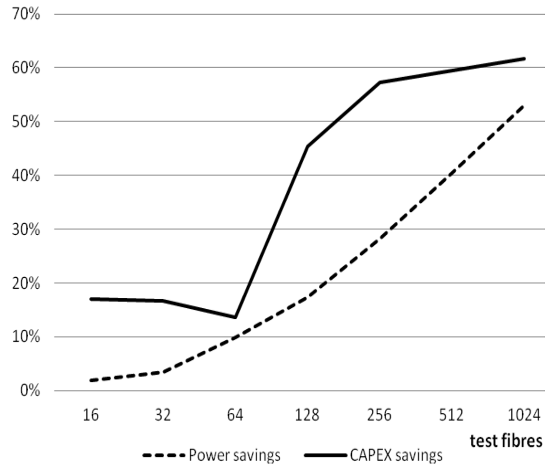


Figure 4. Savings (%) of proposed system versus number of test fibers with regards to a single wavelength supervision system.

Table II shows the values of the parameters used in the cost and power model employed for the cost and energy efficiency analysis. Due to the high cost of active switches with regards to passive Mux/Demuxes, reducing by 4 times the number of ports allows to significantly reduce the total cost, even adding a large number of Mux/Demux components. The use of passive components for fiber switching also achieves a relevant power consumption reduction.

As shown in Fig. 4 (dashed line), the proposed 4 wavelengths system can save up to 50% of power consumption of a conventional FOCS with a single-wavelength OTDR. For a number of test fibers smaller or equal than 64, the CAPEX savings keep below 20% due to the impact of the high cost of the OTDR with regards to the total system CAPEX. For a number of test fibers higher that 64, the proposed WDM FOCS with Multiple-Wavelength OTDR can save more than 20% and up to 62% of the total system CAPEX.

The penalty of the OTDR dynamic range is typically reduced around 1.0 dB, which is a very low value compared with the total dynamic range (~41dB) of the considered OTDR modules.

IV. Experimental Validation

In order to experimentally validate the principle of concept of the proposed Multi-Wavelength OTDR supervision system, a laboratory setup has been prepared emulating a Central Office monitoring four Single Mode Fiber (SFM 10/125μm) coils, whose lengths are 10027m (L1), 19850m (L2), 24330m (L3) and 40216m (L4).

The Multi-Wavelength OTDR system has been implemented using a CWDM OTDR module with λ1=1551nm, λ2=1571nm, λ3=1591nm and λ4=1611nm selectable nominal center wavelengths for measurements.

A Coarse WDM TFF multiplexer-demultiplexer operating in the same four CWDM channels of the OTDR module was employed as a totally passive fiber switch (s11, M11=1, N111=4, see Fig. 2). All fiber connectors were SC/APC type. Maximum insertion loss of the CWDM TFF mux/demux is 2.0 dB according to its specifications datasheet.

In the management system database, the supervision wavelengths λ1, λ2, λ3, λ4 are assigned to fibers L1, L2, L3 and L4, respectively, by connecting the corresponding ports of the CWDM TFF multiplexer-demultiplexer to each fiber.

TABLE III. Measured Transmission Losses of a 4 Channels CWDM Mux/Demux TFF

Wavelength (nm)	**Measured Transmission Loss (dB)**
1551	0.67
1571	0.67
1591	1.45
1611	1.38

Successful measurements using 100ns pulses and 30s of acquisition time were obtained using the novel Multi-Wavelength OTDR supervision architecture. Clear traces with 0.18dB/km propagation loss were obtained, see Fig. 6.

For L1 and L2, the end of fibers was measured at around 10 km and 20 km, as expected.

At L3 and L4, high connection losses of 7.7 dB and 3.8 dB were detected at 20km from the Central Office. These losses generally appear when dirty connections appear at intermediate Central Offices along an optical path. End of fibers L3 and L4 were also clearly detected at around 24.3 and 40.2 km from the Central Office, as expected.

In order to evaluate the loss of dynamic range in the OTDR due to the transmission loss of the 4-channel CWDM multiplexer-demultiplexer, an alternative setup using a 20 km fiber coil connected to the OTDR module, followed by the CWDM multiplexer-demultiplexer and a cascaded 20km fiber coil at each output port was employed.

The obtained traces are shown in Fig. 5, where the connection losses at 20km correspond to the transmission loss of the 4-channel WDM multiplexer-demultiplexer.

The transmission losses are shown in Table III with values close to typical specifications and well below the maximum loss of 2 dB of the CWDM TFF mux/demux. The reduction of less than 1.5 dB of the OTDR dynamic range is negligible with regards to an obtained range of around 20 dB in the measurements, even with only 30 seconds of acquisition time, and being 41 dB the maximum dynamic range of the CWDM OTDR module.

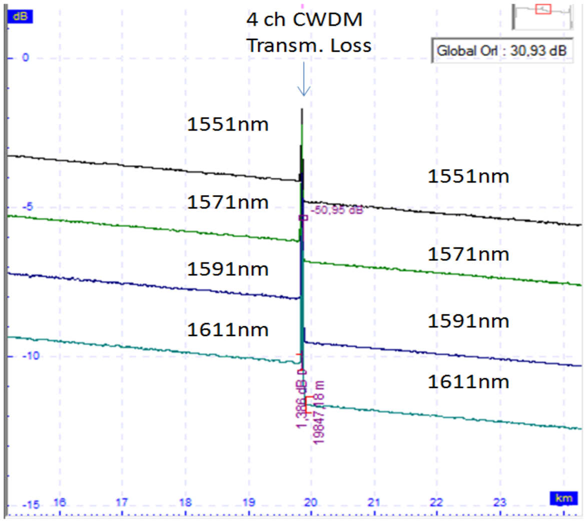


Figure 5. Measurement of transmission loss of the 4channels CWDM.

V. Summary and Conclusions

A new reflectometric system and FOCS approach has been presented for physical layer supervision of fiber optic infrastructures from the Central Offices, with high potential for CAPEX savings and energy efficiency.

It is based on the use of Tunable or Multi-Wavelength OTDR measurement equipment and WDM components used as passive fiber switches. By selecting the operating wavelength of the OTDR, the test signal can be delivered to a desired fiber under test of the network operator fiber infrastructure.

Compared to external OTDR solutions with a single wavelength, already available in the market, the proposed approach keeps the same architecture and maintenance processes than the existing products, being the additional feature a selectable OTDR operating wavelength assigned to different test ports in an automatic way.

The proposed approach enhances energy and cost efficiency of fiber infrastructure supervision systems, what can be especially interesting in the case of massive PON deployments with in-line reflectometry for physical layer supervision, which have a high number of fiber under test (>64). In the case of in-line OTDR PON supervision, the standard waveband is the U-band [10], so U-band passive components and U-band Multi-Wavelength or Tunable OTDRs should be used. The U-band is specified for monitoring purposes when communication wavelength band extends up to the L-band, thus the proposed approach is compatible not only with GPON and EPON, but even with emerging XG-PON deployments and, in the future, with NGPON2 systems and beyond.

In the case of using reserved dark fibers attached to PON fiber infrastructure for supervision, there is no restriction in the waveband.

For a number of channels of the WDM mux-demuxes higher than 16, the insertion losses of the OTDR test signal may be increased more than 3 dB with regards to a totally active fiber switching approach, so it is recommended a design of the hybrid FOCS employing filters with a lower number of output ports, unless the reduction of the dynamic range can be afforded by the monitoring system.

The overall power consumption of the system is reduced because the FOCS is partially implemented in a passive way. Switching cost is also reduced because a lower number of active elements is required.

The system advantages increase with the number of fibers under test. The most suitable use case is a massive PON deployment with in-line external OTDR supervision, but the proposed system is also applicable to any supervision system using dark fiber or metro-core systems with vacant channels.

In a PON scenario with 10 million Homes Passed and 256 ports FOCS (610 Central Offices), it is estimated that a 4 wavelength OTDR plus a hybrid FOCS system can save several tenths M€ of CAPEX.

From the energy efficiency perspective, energy savings are in the range of 100 MWh/year in the same deployment scenario.

Acknowledgment

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n. 257740 (Network of Excellence “TREND”).

References

[1] A. Teixeira, Giorgio M. T. Beletti, Optical Transmission: The FP7 BONE Project Experience, e-ISBN 978-94-007-1767-1, Springer, 2011.

[2] N. J. Frigo, et al., Centralized in-service OTDR testing in a CWDM business access net-work, IEEE J. Lightw. Technol., vol.22, no.11, pp. 2641-2652, 2004.

[3] K. Yüksel, et al., Optical Layer Monitoring in Passive Optical Networks (PONs): a review. International Conference on Transparent Optical Networks, 2008.

[4] G.984.2 Ammendmend 2. ITU-T. G-PON Physical Media Dependent (PMD) layer specification, 2008.

[5] WT-287 (draft): PON Optical-Layer Management, BroadBand Forum, 2011.

[6] J. Hehmann and T. Pfeiffer., New monitoring concepts for optical access networks, Bell Labs Tech. Journal 13(1), pp. 183-198, 2008.

[7] F. Saliou et al, Energy efficiency scenarios for long reach PON Central Offices, paper OThB2, OFC/NFOEC, 2011.

[8] N. Madamopoulos, et al., Applications of large-scale optical 3D-MEMs switches in fiber-based broadband-access networks, Photonic Network Communications, Vol. 19, 2010.

[9] J.E. Ford, et al., 1xN fiber bundle scanning switch, Optical Fiber Communication Conference, pp. 143-144, 1998.

[10] ITU-T L.66: Optical fiber cable maintenance criteria for in-service fiber testing in access networks, 2007.

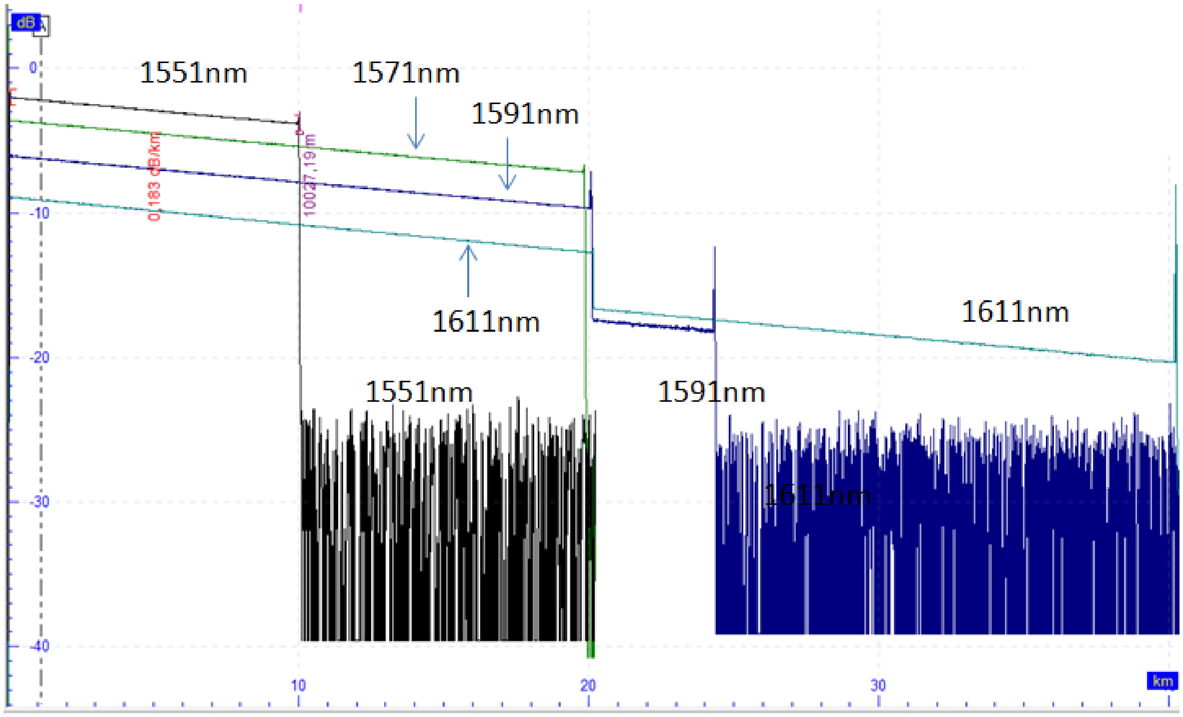


Figure 6. Traces corresponding to the four supervision fibers obtained from the CO with the Multi-Wavelength OTDR and the CWDM passive mux/demux as FOCS.

Provisioning and Resource Allocation for Green Clouds

Guilherme Arthur Geronimo, Jorge Werner, Carlos Becker Westphall, Carla Merkle Westphall, Leonardo Defenti
Networks and Management Laboratory, LRG
Federal University of Santa Catarina, UFSC
Florianópolis, Brazil
E-Mail:{*arthur,jorge,westphal,carla,ldefenti*}*@lrg.ufsc.br*

***Abstract*—The aim of Green Cloud Computing is to achieve a balance between the resource consumption and quality of service. In order to achieve this objective and to maintain the flexibility of the cloud, dynamic provisioning and allocation strategies are needed to regulate the internal settings of the cloud to address oscillatory peaks of workload. In this context, we propose strategies to optimize the use of the cloud resources without decreasing the availability. This work introduces two hybrid strategies based on a distributed system management model, describes the base strategies, operation principles, tests, and presents the results. We combine existing strategies to search their benefits. To test them, we extended CloudSim to simulate the organization model upon which we were based and to implement the strategies, using this improved version to validate our solution. Achieving a consumption reduction up to 87% comparing Standard Clouds with Green Clouds, and up to 52% comparing the proposed strategy with other Green Cloud Strategy.**

***Keywords*-Green Clouds; Provisioning; Resource Allocation.**

I. Introduction

This paper proposes to improve the sustainability of Private Clouds, suggesting new strategies for provisioning and allocation for physical machines **(PMs)** and virtual machines **(VMs)**, transforming the cloud into Green Cloud and Hybrid Cloud, when needed [1]. Green Clouds crave the resource economy for the components that belongs to it. To do so, we adopt the positive characteristics of multiple existing strategies [2], developing a hybrid strategy that, in our scope, aims to address:

- A sustainable solution to mitigate peaks in environments with rapid changes and unpredictable workload.
- Optimizing the estimated data center infrastructure without compromising the availability of services, during the workload peaks.
- Improving the balance between the sustainability of infrastructure and availability of Services Layer Agreements (SLAs).

This work was based in the university data center reality, which disposes of multiple services suffering often with unexpected workload peaks, whether from attacks on servers or overuse of services in a short time.

A. Motivation

The motivation for this work can be summarized in the following points:

- **Energy saving**: Murugesan [3] says ”Energy saving is just one of the motivational topics within IT environments greens.”. We highlight the following points: (1) the reduction of monthly data center operating expenses (OPEX), (2) the reduction of carbon emissions into the atmosphere (depending on the country), and (3) extending the lifespan of Uninterruptible Power Supply (UPS) [4].
- **Availability of Services**: Given the recent wave of offering products, components, and elements in the form of services (*aaS), a series of pre-defined agreements between stakeholders aimed at governing the behavior of the service that will be supplied / provided is needed [5]. According to administrators in the area of information technology, the alarming factor is agreements that provide for the availability of such rates, usually 99.9% of the time or more. Thus, the question is how to provide this availability rate while consuming little power.
- **Variation Workload**: In environments with multiple services, the prediction of workload is a very complex factor. Historical data is used most to predict future needs and behaviors. However, abrupt changes are unpredictable and end up causing unavailability of the provided services. The need to find new ways to deal with these sudden changes in the workload is evident.
- **Delayed Activation**: Activation and deactivation of resources are a common technique for reduce power consumption, but the time required to complete this process is a problem that can cause some unavailability of the services provided, generating contractual fines.
- **Public Clouds**: Given the growing amount of public clouds and the development of communication methods among clouds, like Open Cloud Consortium [6], and Open Cloud Computing Interface [7], it became possible to use multiple public clouds as extensions of a single private cloud. We considered this as an alternative resource to implement new Green Cloud strategies.

B. Objective

Thus, we aim to propose an allocation strategy to private clouds and a provisioning strategy for Green Clouds, which suits the oscillatory workload and unexpected peaks. We will focus on finding a solution that consumes low power and generates acceptable request losses.

This paper is organized as follows: Section 2 brings the state of the art sorted by gaps found. Section 3 explains under which model the strategies were based. Section 4 presents the proposal, the idea behind the strategies, their pros and cons and where each one should be applied and not applied, tests, and the results. In Section 5, we conclude this paper and address some future works.

II. State Of The Art

About energy consumption, the paper [8] uses a Dynamic Voltage Frequency Scaling (DVFS) strategy to decrease the energy consumption in PMs used as virtualization hosts. It adapts the clock frequency of the CPUs with the real usage of the PMs. It decreases the frequency in idle nodes and increases when is needed. But the problem is that, the major energy consumption is not in the CPU, but in the other parts of the PM, so to really decrease the energy consumption you need to turn off the PMs.

The workload balance strategy for clusters in [9], tries to achieve a lower energy consumption unbalancing the cluster workload, generating idle nodes and turning off them. Extending this idea for Cloud Computing this don't work very well in cases that the cloud is fully loaded (like in Deny-of-Service attack) and the ”unbalance” can not be done. This way, we saw the necessity of VMs migrations between clouds as mandatory function, to avoid this kind case.

The paper [10] tries to decrease the hosting costs in public and/or federated clouds using the costs and fines in contracts as metrics to better allocate the resources. But it limits itself in migrating VMs, between clouds, in a pool of pre-hired Clouds. This way, we foreseen that we also could considerer the resource consumption as a metric to allocate the VMs.

III. Model

The concept of combining Organization Theory and complex distributed computing environments is not new. Foster [11] already proposed the idea of virtual organizations (VOs) as a set of individuals and / or institutions defined by such sharing rules in grid computing environments. This work concludes that VOs have the potential to radically change the way we use computers to solve problems, as well as the web has changed the way of information exchange.

Following this analogy, we have a similar view: Management Systems based on the Organization Theory, providing the means to describe why / how elements of the cloud environment should behave to achieve global system objectives, which are (among others): optimum performance, reduced operating costs, appointment of dependence, service level agreements, and energy efficiency.

This organizational structures, proposed in [12], allows the network managers to understand the interaction between the Cloud elements, how their behavior is influenced in the organization, the impact of actions on macro and micro structures and vice versa, as the macro-level processes allow and restrict activities at the micro level. It aims to provide computational models to classify, predict, and understand these interactions and their influence on the environment.

Managing Cloud through the principles of the Organization Theory provides the possibility for an automatic configuration management system, since adding a new element (e.g., Virtual Machines, Physical Machines, Uninterrupted Power Supply, Air Conditioning) is just a matter of adding a new service on the Management Group.

The proposed strategies are based on a pro-active managemnt of Clouds, which is based on the distribution of responsibilities in holes, as seen in Figure 1. The responsibility of management of the cloud elements is distributed among several agents, separated in holes, and each agent controls individually, a Cloud element that suits him.

IV. Proposal

For the conscious resource provisioning of the data center, we propose a hybrid strategy that uses public cloud as an external resource used to mitigate probable Service-level Agreements (SLA) breaches due to unexpected workload peaks. In parallel, to the optimal use of local resources, we propose a strategy of dynamic reconfiguration of the VMs attributes, allocated in the data center. Given the distributed model presented in the previous section, we use the Cloud simulation tool CloudSim [14] to simulate the university data center environment. In order to simulate a distribution faithful to reality and also stressful to the infrastructure, we decided to take two distribution workloads: a real distribution, derived from monitoring requests of the institution websites, as shown in Figure 2, and another distribution derived from a mathematical oscillatory model, as shown in Figure 3. We defined these strategies with the goal of obtaining different approaches. One approach is using real environments, which would have results that intend to reflect the reality. Another approach is biased, striving for correlating the trends of the load with the results inferred.

A. Allocation

The resource allocation strategy is a proposal that introduces a composition of two other approaches: (1) the migration of VMs, which aims to focus on the processing of cloud, and (2) the Dynamic Reconfiguration of VMs, which aims to relocate dynamically the resources used by the VMs.

1) VMs Migration Strategy: This strategy aims to reduce power consumption by disabling the idle PMs of the Cloud. To induce idleness in the PMs, the VMs are migrated and

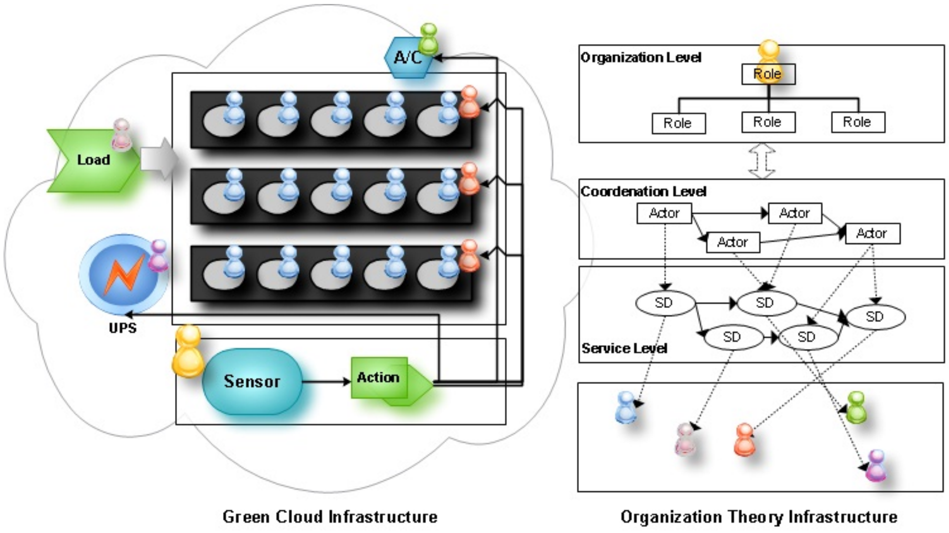


Figure 1. Model Based in Organization Theory [13]

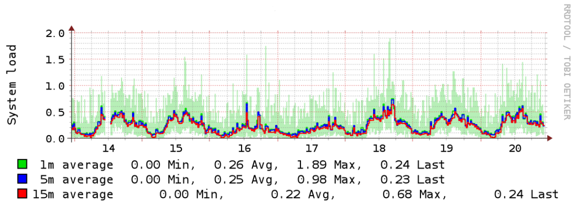


Figure 2. Real Workload Distribution

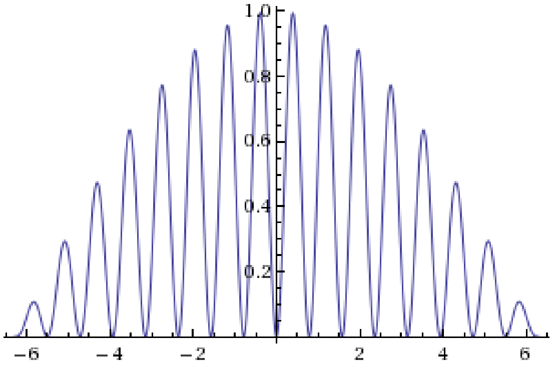


Figure 3. Oscillatory Workload Distribution

concentrated in a few PMs. This way, the cloud manager can disable the empty PMs, reducing the consumption of the data center. However, it is understood that, for the optimal results of the strategy, it must be used in conjunction with another strategy, that is a strategy that permits hosting more VMs in less PMs, generating more idle PMs.

2) VMs Dynamic Reconfiguration Strategy: Seeking the improvement of the previous strategy, this strategy is an alternative optimization to dynamically shrink the VM. It adjusts the parameters of the VM [15], without migrating it or turning it off. For example, we can increase or decrease the parameters of CPU and memory allocated. Thus, the VMs would adapt according to the demand at that moment.

3) Tests Results: To simulate the strategies we used a Cloud simulator tool developed in Melbourne, CloudSim [14]. But, in order to achieve the simulations that we need, we made some modifications in the code [13], allowing to simulate the distributions patterns and the scenario definition. Four scenarios were simulated in order to seek the comparative analysis between ordinary cloud (Scenario 1), the existing methods (Scenarios: 2 and 3), and the proposed approach (Scenario 4). Those were:

- No strategies;
- Migrating VMs Strategy;
- Reconfiguring the VMs Strategy;
- Reconfiguring and migrating VMs Strategy.

At the simulations, we gathered behavior, sustainability, and availability metrics, such as the number of idle PMs, total energy consumption, and number of SLA breaches. The graph in Figure 4 represents the energy consumption in a scenario with 100 PMs without strategies implemented. In this, the power consumption is regularly during the whole period, since all the VMs and PMs were activated during the period.

Table I shows the results of the simulations. It tells what strategies were used in each scenario and what percentage (approximate) reduction was obtained, compared to the scenario without strategies.

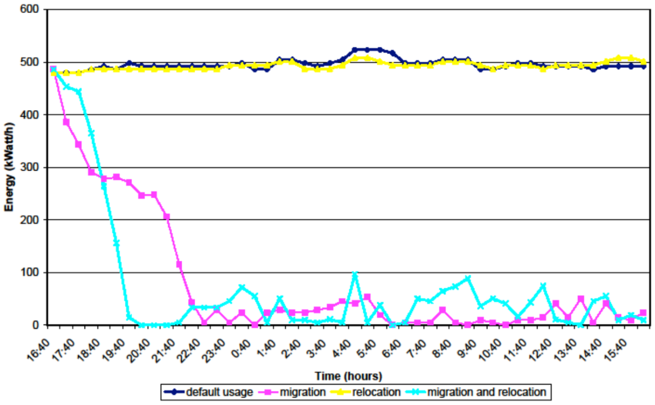


Figure 4. Energy Consumption of Scenarios 1-4 [2]

Table I
RESULTS OF ALLOCATION'S SCENARIOS

Scenario	Reconf. Strategy	Mig. Strategy	Consumption	Timeout
1	No	No	-	-
2	No	Yes	84.3%	8.0%
3	Yes	No	0.4%	-
4	Yes	Yes	87.2%	7.3%

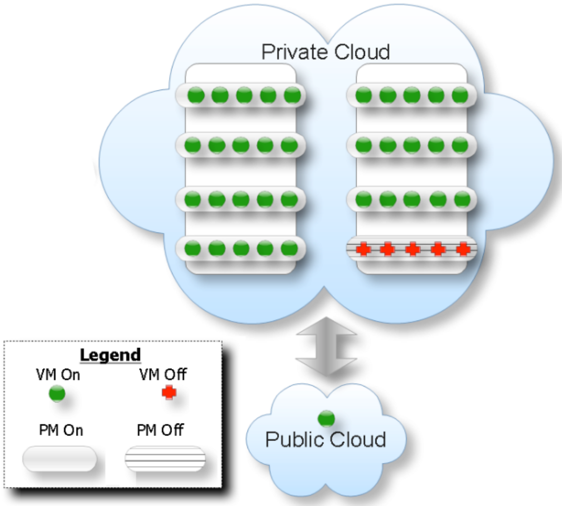


Figure 5. Hybrid Strategy [2]

B. Provisioning

The hybrid strategy is based on the merge of two other strategies, the OnDemand strategy (OD) and the Spare Resources strategy (SR). It tries to be the middle ground between the two, enjoying the strengths of both sides. It aims to present a power consumption lower than the SR strategy and a wider availability than the OD strategy.

1) On Demand Strategy: The principle of OD strategy is to activate the resources when they are needed. In our case, when a service reaches a saturation threshold, new VMs would be instantiated. When there is no more space to instantiate new VMs, new PMs would be activated to host the new VMs. The opposite also applies; when a threshold of idleness is reached, the idle VMs and PMs are disabled.

This strategy proved to be very efficient energetically, since it maintains a minimum amount of active resources. But, it has been shown ineffective in scenarios that had sudden spikes in demand, because the process to activate the resource took too much time, and the requests ended up generating losses.

2) Spare Resource Strategy: To mitigate the problem of requests timeouts, originated by a long activation time of resources, we adopt the strategy SR, whose principle is reserve idle resources ready to be used. In our case, there was always one idle VM ready to process the incoming requests and one idle PM ready to instantiate new VMs. If these resources were used, they were no longer considered idle, and new idle resources were activated. As long as the resources were no longer being used they were disabled. The strategy has been shown effective in remedying unexpected peak demands, but it showed the same behavior OD strategy in cases where demand rose very rapidly; in other words, the idle feature was not enough to process the demand. Another negative point was the energy consumption; since they always had an active and idle resource, the consumption has been greater than the OD strategy.

3) Hybrid Strategy: Seeking the merger of the strengths of the previous strategies and mitigating its shortcomings, we propose a hybrid strategy. This strategy aims to reduce the energy consumption on private cloud and reduce the breakage of SLA's service in general.

As shown in Figure 5, the cloud enables the VMs when the service in question reaches its saturation threshold, just as the OD strategy. When more PMs space is unable to allocate more VMs, it uses the public cloud to host the new VMs while the PM is passing through the activation process. This is to fulfill requests that would be lost during the activation process.

The deactivation process occurs just as the other strategies; however, it is considered that the public cloud is paid by time (usually by hour of processing); so, it disables the VM hosted in the public cloud only when (1) it's idle and (2) it is almost time to complete a full hour of hosting.

4) Tests Results: As previously mentioned, we performed some modifications to the CloudSim code, in order to enable the simulation of scenarios. Before we started the simulation, we defined some variables for the scenario, such as the saturation threshold and idleness, for example. Some of these variables are shown in Table II.

To get an overview of how each strategy would behave in different scenarios, we ran a series of tests which varied (1) the amount of requests and (2) the size of the requests.

To maintain the defined request distribution (explained in the beginning of Section 3), we used multipliers to increase

Table II
SIMULATION'S VARIABLES

Variable	Value
Saturation Threshold (Load 1 minute)	1.0
Idleness Threshold (Load 1 minute)	0.1
Activation VM time (seconds)	10
Activation PM time (seconds)	120
Size of Request (MI)	1000 to 2000
Number of PMs	8
Maximum number of VMs per PMs	5
SLA timeout threshold (seconds)	10

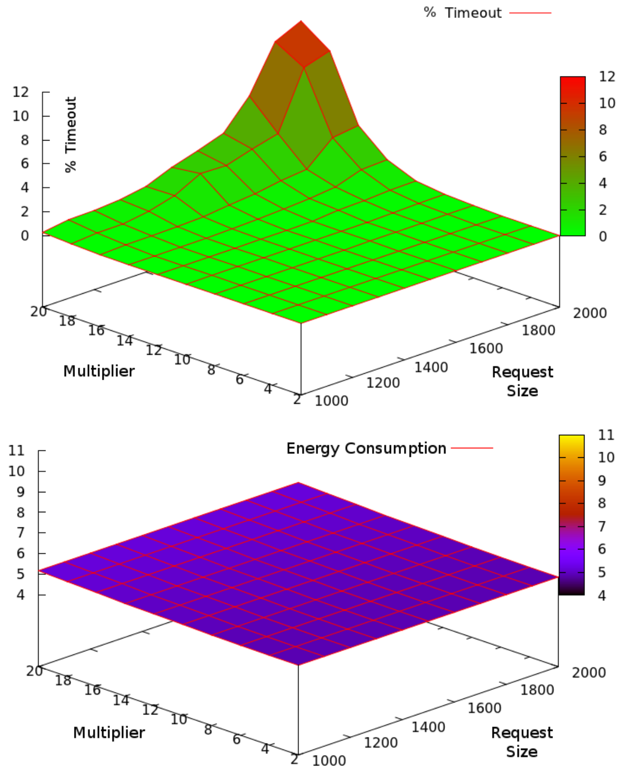


Figure 6. Number of Timeouts (top) and Energy Consumption (bottom) with Hybrid Strategy

the requests. Those multipliers started from 2 to 20 in steps of 2 (2, 4, 6, etc.).

The size of the requests ranged from 1000 to 2000 MI (Millions Instructions), in steps of 100 (1000, 1100, 1200, etc.).

This way, it performed a series of 100 simulations. This test evaluated the power consumption of the private cloud and the total number of timeouts. Figure 6 demonstrates 100 simulations in two images, the percentage of timeouts (top) and the energy consumption of the private cloud (bottom) are plotted.

Table III shows the results obtained in the ”worst case scenario”, by definition, with the multiplier equal to 20 and the request size equal to 2000 MI. Regarding the results in Table III, it took the Hybrid Strategy as a basis of comparison. In this case, the values listed are for hybrid strategy. For example, the hybrid strategy presented 3% less requisition timeouts than the OD strategy.

Table III
HYBRID STRATEGY COMPARED TO THE OTHER STRATEGIES

	OnDemand	Spare
Timeouts	-3 %	+15 %
Consumption	-18 %	-52 %

V. CONCLUSIONS AND FUTURE WORKS

Based on what was presented in the previous sections, and considering the objectives set at the beginning of this paper, we consider the intended goal was achieved. Two strategies for allocation and provisioning, were proposed; both aimed at optimizing the energy resource without sacrificing service availability.

The allocation strategy in private clouds, compared to a normal cloud, demonstrated a 87% reduction in energy consumption. It was observed that this strategy is not effective in scenarios where the workload is oscillating. That's because it ends up generating too much unnecessary reconfigurations and migrations. Despite this, it still shows a significant gain in energy savings when compared to a cloud without any strategy deployed.

The hybrid strategy for provisioning in green clouds, demonstrated a 52% consumption reduction over the SR strategy, and a timeout rate 3% lower than the OD strategy. Thus, we conclude that the use of this strategy is recommended in situations where the activation time of the resource is expensive for the health of SLA. We also identified that using this is not recommended when the public cloud should be used sparingly due to their course or other factors.

As future work, we aim at adding the strategy of Dynamic Reconfiguration of VMs in public clouds. This procedure was not adopted because, during the development of this work, this feature was not a market reality. We also intend to invest in new simulations of the cloud extending the variables (such as DVFS and UPS) and, if possible, explore some artificial intelligence techniques [16] such as Bayesian networks, the recalculation of beliefs. Our PCMONS (Private Cloud Monitoring System), open-source solutions for cloud monitoring and management, also will help to manage green clouds, by automating the instantiation of new resource usage [17].

We foresee, in opposition to unexpected peaks scenarios, work with cloud management based on prior knowledge of the behavior of hosted services. It is believed to be necessary to develop a description language to represent the structure and behavior of a service, enabling the exchange of information between applications for planning, provisioning, and managing the cloud.

References

[1] J. Werner, G. A. Geronimo, C. B. Westphall, F. L. Koch, C. M. Westphall, R. R. Freitas, and A. Fabrin, "Aperfeiçoando a gerência de recursos para nuvens verdes," *INFONOR*, vol. 1, pp. 1–8, 2012.

[2] J. Werner, G. A. Geronimo, C. B. Westphall, F. L. Koch, R. R. Freitas, and C. M. Westphall, "Environment, services and network management for green clouds," *CLEI Electronic Journal*, vol. 15, no. 2, p. 2, 2012.

[3] S. Murugesan, "Harnessing green it: Principles and practices," *IT professional*, vol. 10, no. 1, pp. 24–33, 2008.

[4] R. Buyya, A. Beloglazov, and J. Abawajy, "Energy-Efficient management of data center resources for cloud computing: A vision, architectural elements, and open challenges," in *Proceedings of the 2010 International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA 2010), Las Vegas, USA, July 12*, vol. 15, 2010.

[5] M. A. P. Leandro, T. J. Nascimento, D. R. dos Santos, C. M. Westphall, and C. B. Westphall, "Multi-tenancy authorization system with federated identity for cloud-based environments using shibboleth," in *ICN 2012, The Eleventh International Conference on Networks*, 2012, pp. 88–93.

[6] OpenCC, "Open cloud consortium," 2012. [Online]. Available: http://opencloudconsortium.org/

[7] OCCI, "Open cloud computing interface," 2012. [Online]. Available: http://www.occi-wg.org

[8] G. von Laszewski, L. Wang, A. Younge, and X. He, "Power-aware scheduling of virtual machines in dvfs-enabled clusters," in *Cluster Computing and Workshops, 2009. CLUSTER '09. IEEE International Conference on*, 31 2009-sept. 4 2009, pp. 1 –10.

[9] E. Pinheiro, R. Bianchini, E. Carrera, and T. Heath, "Load balancing and unbalancing for power and performance in cluster-based systems," in *Workshop on Compilers and Operating Systems for Low Power*, vol. 180. Citeseer, 2001, pp. 182–195.

[10] H. A. Franke, "Uma abordagem de acordo de nível de serviço para computação em nuvens," PPGCC/UFSC, 2010.

[11] I. Foster, Y. Zhao, I. Raicu, and S. Lu, "Cloud computing and grid computing 360-degree compared," in *Grid Computing Environments Workshop, 2008. GCE 08*, nov. 2008, pp. 1–10.

[12] J. Werner, G. A. Geronimo, C. B. Westphall, F. L. Koch, and R. R. Freitas, "Um modelo integrado de gestão de recursos para as nuvens verdes," in *CLEI 2011*, vol. 1, 2011, pp. 1–15.

[13] Werner, J. and Geronimo, G. A. and Westphall, C. B. and Koch, F. L. and Freitas, R. R., "Simulator improvements to validate the green cloud computing approach," *LANOMS Latin American Network Operations and Management Symposium*, vol. 1, pp. 1–8, 2011.

[14] R. Buyya, "Modeling and simulation of scalable cloud computing environments and the cloudsim toolkit: Challenges and opportunities," in *HPCS 2009. International Conference on*. IEEE, 2009, pp. 1–11.

[15] T. Wood, P. Shenoy, A. Venkataramani, and M. Yousif, "Sandpiper: Black-box and gray-box resource management for virtual machines," *Comput. Netw.*, vol. 53, no. 17, pp. 2923–2938, Dec. 2009. [Online]. Available: http://dx.doi.org/10.1016/j.comnet.2009.04.014

[16] F. L. Koch and C. B. Westphall, "Decentralized network management using distributed artificial intelligence," *Journal of Network and Systems Management*, vol. 9, pp. 375–388, 2001, 10.1023/A:1012976206591. [Online]. Available: http://dx.doi.org/10.1023/A:1012976206591

[17] S. A. de Chaves, R. B. Uriarte, and C. B. Westphall, "Toward an architecture for monitoring private clouds," *Communications Magazine, IEEE*, vol. 49, no. 12, pp. 130 –137, December 2011.

A Validation Model of Data Input for Web Services

Rafael Bosse Brinhosa, Carla Merkle Westphall, Carlos Becker Westphall, Daniel Ricardo dos Santos, Fabio Grezele
Post Graduate Program in Computer Science
Federal University of Santa Catarina
Florianópolis, Brazil
{brinhosa,carlamw,westphal,danielrs,fgrezele}@inf.ufsc.br

Abstract— **Web services inherited many well-known security problems of Web applications and brought new ones. Major data breaches today are consequences of bad input validation at the application level. This paper presents a way to implement an input validation model for Web services which can be used to prevent cross-site scripting and SQL injection through the use of predefined models which specify valid inputs. The proposed WSIVM (Web Services Input Validation Model) consists of an XML schema, an XML specification, and a module for performing input validation according to the schema. A case study showing the effectiveness and performance of this mechanism is also presented.**

Keywords—security; Web service; input validation; SOA

I. INTRODUCTION

Different technologies for collaboration and information sharing are emerging and therefore new forms of interaction are evolving and creating new requirements for the development of distributed applications. Enterprises are experiencing increased collaboration and information sharing and a greater need for the use of distributed and computational resources [1].

The paradigm of *Services Oriented Architecture* (SOA) has transformed the Internet from a data repository to a services repository [2]. In SOA style, an application is composed of reusable services that are integrated through standardized interfaces.

Web services technology based on the use of open standards facilitates information exchange, interoperability, and software reuse, and is therefore considered a major component of SOA. Web services are software components that can be discovered and used to implement applications. Web services are suitable to integrate heterogeneous systems because they make extensive use of XML (*Extensible Markup Language*) [3]. The Web service interface, for example, is described using a language based on XML, called WSDL (*Web Services Description Language*). Furthermore, communication among parts of a distributed application is carried out using SOAP (*Simple Object Access Protocol*) messages which are XML-based.

The Internet makes many Web services available for use: it is possible to obtain information about the weather, stock exchange, and postal codes [4] or to provide information to the federal government [5].

While the SOA paradigm provides cost savings by eliminating redundant efforts through software reuse, security is a major concern according to the Gartner Research Institute [6].

To implement security in Web services, various standards and specifications have been created. However, the correct use of standards alone does not guarantee that the right level of security will be achieved [7]–[9].

A report by the SANS Institute (*SysAdmin, Audit, Network, Security*) [10] lists the major risks to cybersecurity, and the OWASP community (*Open Web Application Security Project*) [11] states that validation of data input can be one of the most effective controls for Web applications security.

The validation of data input [12] [13] is a set of controls that an application should carry out on the lexical and syntactic aspects, type checking, integrity, and origin of data. The lack of these controls has become a major problem for software because interfaces exposed to the Internet could be easily exploited by malicious users.

Thus, in the SOA and Web services environment, improving security mechanisms by the use of more robust data validation has become essential [11] [13].

This paper proposes a model for validation of data input in Web services, providing protection against attacks based on malicious input. The proposed model is called WSIVM (*Web Services Input Validation Model*), and is an input validation mechanism composed of an XML schema, an XML specification, and a validation module.

This paper is organized in the following sections: Section 2 presents the related works, Section 3 describes the security problems of Web services, Section 4 presents the proposed model, WSIVM, Section 5 describes the implementation, a case study is shown in Section 6, and Section 7 presents the conclusions of the paper.

II. RELATED WORK

The lack of input validation is a major cause of Web application attacks [11] [13] [14], whether these applications are developed with Web services or with other technologies. This happens because the lack of input validation of data allows multiple attacks listed in [10] [15] to occur. Among the attacks that can be cited are the injection of malicious code by use of SQL (*Structured Query Language*) and *cross-site scripting* (XSS), which allows code execution (scripts) in the client-side browser to perform malicious actions [15].

Many studies have been undertaken to ensure input validation in Web applications, such as Microsoft Anti-Cross Site Scripting Library [16] and the use of Open Source solutions in PHP. However, there are few specific mechanisms for Web services.

Regarding the implementation of security mechanisms for Web services, the MIT (Massachusetts Institute of Technology) has an implementation called WS-Security Wrapper [17], which is an intermediate between the Web service and the client that carries out validation of certain aspects. However, this work was developed to be compatible only with Web services developed on the platform Microsoft.Net v1.1 and does not include features such as validation of predefined data entries.

Wu and Hisada [18] have proposed a token based metadata to validate semantic notation built on top of ESB (Enterprise Service Bus). This approach uses a different method for input string validation using the ESB for implementing SOA security.

A reusable and independent mechanism for data input is very important in the process of creating a secure Web service. The mechanism proposed in this paper, WSIVM, assists in this task differently from other studies examined. First, because it focuses on the aspect of handling of data input, it differs from IAPF (Integrated Application and Protocol Framework) [19], which seeks to address all the security aspects related to Web services. Moreover, other works [3] [20] have focused on the use of existing technologies such as XML encryption to ensure the security of Web services but do not mention input validation.

With respect to input validation aspects in Web applications, there are some works such as [21] that have developed tools that automatically insert the input validation on the server side by eliminating malicious insertions vulnerabilities. However, this approach has a disadvantage in that it produces a great many false positives; that is, the validator may fail by considering a message invalid when in fact the message is valid.

Besides the works already listed there is another category of work focused on developing firewalls, like Web Service Firewall Nedgty [22], which deals with protection against denial of service and stack overflow attacks. XML firewalls, presented in [23], are concerned with validation of the structure of XML content but not the content itself. Reference [24] mentions protection against SQL injection through an XML schema and a precompiled blacklist of SQL commands, an approach which tends to produce many false positives; however, details about the effectiveness of this work with more extensive tests are not presented.

Among the related works it can be seen that there is a lack of studies specifically addressing input validation for Web services.

III. Security Issues in Web Services

Web services create new security risks for organizations because old methods of protection such as firewalls and antivirus applications are not able to protect them. Common firewalls that act in the networking layer allow the normal flow of HTTP (Hypertext Transfer Protocol) requests without blocking these flows, because they are designed to make use of HTTP using port 80.

In addition, the Web services functionality is exposed through WSDL files since from the descriptions of the methods and variables of the WSDL file, important information can be obtained in order to accomplish an attack known as WSDL scanning [1].

There are attacks which are directly related to data manipulation: XSS and SQL injection. Reference [25] classifies two types of XSS attacks: first order and second order.

In a first-order XSS attack, the vulnerability results from the application inserting part of the user input on the page itself. The malicious user uses social engineering to convince the victim to click on a URL that contains malicious HTML/JavaScript code. The user's browser displays the HTML page and runs the JavaScript that was part of a malicious URL received, resulting in the theft of session cookies or other sensitive data from the user. This type of attack can hardly be done against Web services.

In the second-order XSS attacks, vulnerability results from the storage of malicious entries by the user in the application database, and then when the HTML page is accessed, the code runs and is shown to the victims (for example, on social network pages). Second-order attacks are more difficult to avoid because the application needs to validate or sanitize inputs, which may contain executable script code. In the context of Web services, by presenting unvalidated data directly to the user, Web services can be attacked. For example, by making use of AJAX (Asynchronous JavaScript and XML), data Web services provided by third parties that may be contaminated can be obtained. Using, for example, the command `document.write(xmlhttp.responseText)`, if the answer to this AJAX call made to a Web service contains HTML and JavaScript data, these data will be interpreted and executed, posing a risk to the user.

Code injection attacks (SQL injection) work through malicious inputs aimed at the execution of SQL commands in the database [15], [25], [26].

In Web services that do not have proper exception handling, the error message may contain valuable data for the attacker to use. Thus, through trial and error, the attacker can find which database technology is being used, tables that can be explored, and all the necessary information to make an attack. SQL injection attacks can increase the privileges, and thus it is possible to run in administrator mode on the compromised server. It is possible to test whether a Web service is vulnerable by sending SOAP requests with properly handled parameters. For example, by sending “`''"1=1  –`” as a parameter for a particular service, it is possible to obtain in return the outputs of Figs. 1 and 2. These figures show two examples of responses of error outputs from the database that were returned by the server. These responses can be used to discover details of the database and to send new requests to the database, allowing more details to be acquired, and to carry out more complex commands in the database, which can result in elevation of privilege, injection of files, and theft or destruction of the database.

Through the output response in Fig. 1 it is possible to identify that it is a MySQL database and that SQL command injection was performed, because of the type of error returned. It can be concluded based on the output response in Fig. 2 that the name of one of the columns of the database is `ItemId`, because of the syntax of the SQL `select` command, and it can be deduced that the database is Microsoft SQL Server by observing the syntax of `stored procedure` “`dbo.`”.

```
ERROR: The query was not accomplished.
Description: 1064 - You have an error in
your SQL syntax; check the manual that
corresponds to your MySQL server version
for the right syntax to use near '1=1'' at
line 1
```

Figure 1. Example of output response 1.

```
Line 11: Incorrect syntax near '')) or
ItemId in (select ItemId from
dbo.GetItemParents('4''. Unclosed
quotation mark before the character
string ')) ) ) > 0 '.
```

Figure 2. Example of output response 2.

The *Blind SQL injection* attack is a type of SQL injection in which the results are not displayed to the attacker. It is very commonly used against Web services, because many servers prevent the error messages generated by the service from reaching the user. In Web services an HTTP 500 error is usually returned when an attempt at *blind SQL injection* is performed; however, there are techniques such as measurement of response times of the server that can be used to determine the parameters necessary to perform the attack successfully.

Although it is difficult to obtain reliable information on security incidents and data breaches as reported in the book of Adam Shostack and Andrew Stewart [27], in *Databreaches* table [28] it is possible to observe that the major attacks in cases of data theft that occurred were related with the injection of malicious data. In the table given in [28], for example, when accessing the URL that exists on the date of the incident of the entity Heartland Payment Systems, in which 130 million data were lost, it is possible to read the reason for the incident: SQL injection.

IV. WSIVM – Web Services Input Validation Model

In this section, we describe the WSIVM (Web Services Input Validation Model) which is proposed to validate input data to provide security for Web services. Initially the operation of Web services is described without the use of the model. After that the operation of Web services is explained using the proposed model.

A. *Operation of Web Services without WSIVM*

In the traditional way, a customer finds the Web service he or she needs through research into repositories of Web services. The repositories store the references in the form of Web services in UDDI (Universal Description, Discovery, and Integration) format. UDDI is a standard protocol that specifies a method of publishing and discovering directories of services in a service-oriented architecture.

After the discovery the client sends a request to the Web service. The Web service returns the response to the client containing the result of his or her request.

In this traditional process, the client request is made directly to the Web service, which processes the inputs and returns the result. The standard used for this exchange of messages is the XML-based SOAP format.

We assume that the Web service executes the user's entry without any kind of validation and that the user input is part of the SQL query described in (1).

SELECT name, age FROM clients WHERE name=**input**; (1)

If the user provides as input the name "Paul", which is a valid entry for the name, the Web service returns as answer the name and age of the customer "Paul". However, if the user provides the following malicious input string: “`Paul' UNION SELECT name, password FROM clients; –`”, the SQL query would be represented in (2).

SELECT name FROM clients WHERE name=‘Paul’
UNION SELECT name, password FROM clients; -- (2)

As the Web service does not perform validation of input represented in (2), it returns the secret value of "account password" to the user that sent the malicious request to the Web service. For the WSDL, the request is valid, because it is a string as specified in the service description. However, confidence in user input and lack of validation of this input resulted in a vulnerability that provided sensitive data.

B. Operation of Web Services with WSIVM

The model WSIVM (Web Services Input Validation Model) proposes to validate input data to provide security for Web services.

The proposed model has several advantages compared with the input validation normally done in an application, since: (a) it prevents the waste of server processing with invalid messages, (b) it reduces the possibility of denial of service using content of messages, and (c) it is independent of the technology used for the internal development of services.

With WSIVM, the user makes the request in the usual way, however, that request is validated by WSIVM (Fig. 3).

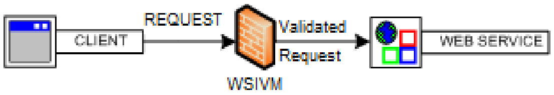


Figure 3. SOAP request with WSIVM.

If a malicious request is sent, as shown in the example represented in the entry (2), instead of sending the password, the WSIVM validation mechanism validates the request and returns a generic error to the user. Thus the Web service does not receive the malicious request (Fig. 4).

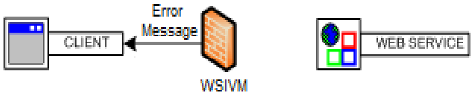


Figure 4. WSIVM blocks malicious request.

In more detail, what happens when a message arrives for validation in WSIVM is that the entry submitted by the user is validated through a module that was developed with the XML specification that is on the server.

This avoids unnecessary consumption of server resources, so, considering the example, the Web service does not execute the SQL query if a malicious request is sent.

The interaction of the components of WSIVM is represented in Fig. 5.

The *WSIVMModule* is a module responsible for calling the other components.

The *WSIVMValidator* maps the SOAP message, obtaining the fields of the body of the message, and sends it to the *WSIVMXMLLoader*.

The *WSIVMXMLLoader* loads the elements and the rules specified in the XML specification and checks the validity or invalidity of the response with the *WSIVMVerifier*.

The *WSIVMVerifier* contains all the pre-defined rules for validation of entries and is responsible for validating these entries.

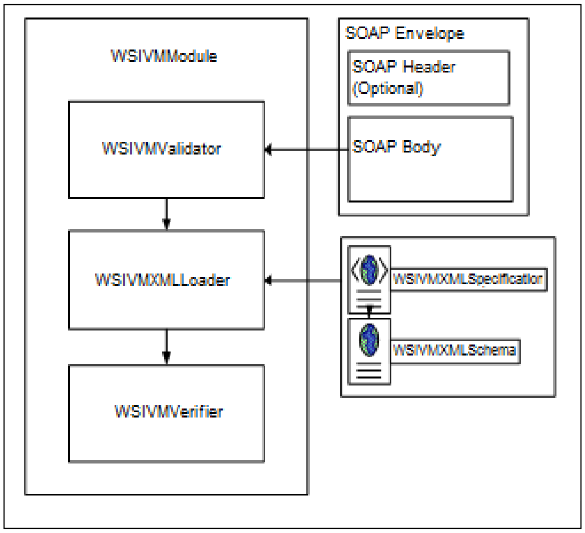


Figure 5. Operation of WSIVM [29].

V. DEVELOPMENT OF THE IMPLEMENTATION

The implementation of the model was developed using the Apache Tomcat Web server and Apache Axis2 framework for SOAP messages [30] (Fig. 6). Apache Axis2 was chosen for the implementation of this work due to its extensibility through modules and the ease of intercepting SOAP messages through the modules.

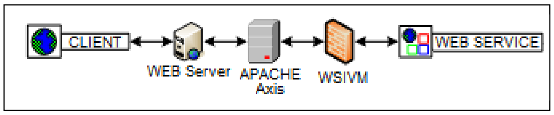


Figure 6. WSIVM.

To implement the validation module for Apache Axis2 the Rampart module was used, which is the mode of extension of Apache Axis2.

The phase of interception can be specified in the file `Module.xml`. It was chosen to intercept the message in the phase `PreDispatch`, which is the phase immediately preceding the sending of the message and its processing by the Web service.

The implementation operation is as follows: the customer, which can be an application, a Web page, or any mechanism capable of communicating with a Web service, sends a message to the Web service. This message passes through the Web server, that is, the Apache Tomcat. The Web server sends this SOAP message to the Apache Axis. The Apache Axis sends the message to be analyzed by WSIVM. After the message is parsed, if it is held to be invalid, an error message is sent to the customer by WSIVM. If it is considered valid, it is usually transmitted for processing by the Web service, and the result is returned to the client (Fig. 6).

The following is a detailed description of the WSIVM model components: WSIVMXMLSchema, WSIVMXMLSpecification, and WSIVM Rampart module.

WSIVMXMLSchema is the specification of the validation schema of entries. It defines the format of the XML specification and the valid attributes.

WSIVMXMLSpecification is the specification of validation of entries. It specifies valid parameters according to a set of predefined attributes and is used for validating user input. Among the possible parameters are the entries:

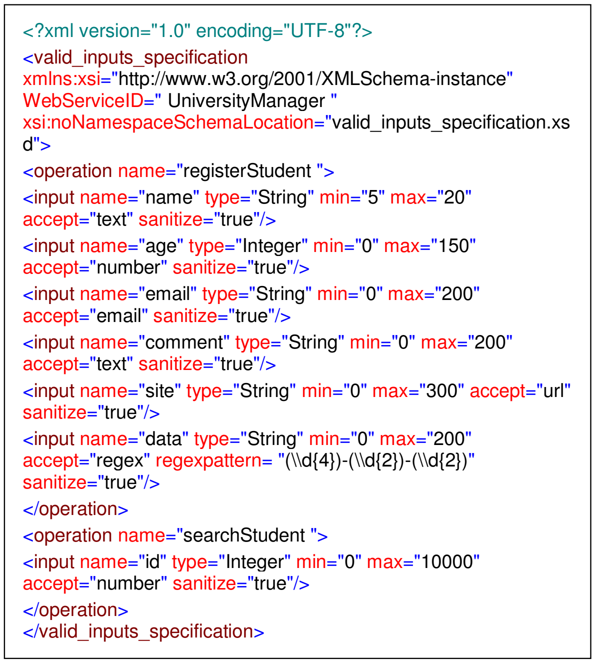
- **OperationName:** the name of the operation or function displayed in the Web service referred to in the validation;
- **SanitizeOperation:** defines whether the parameters of this operation or function can be reformulated if necessary for the removal of characters that are not accepted;
- **ParamName:** the name of the parameter or field referred to in the validation;
- **Allowed:** an allowed field type, which is valid (text, html, html+java-script, email, number, and all);
- **Length:** specifies the exact size of the field;
- **Maxsize:** specifies the maximum field size;
- **Minsize:** specifies the minimum field size;
- **Nillable:** determines whether or not it is possible that the field is null (true or false);
- **regEx:** allows a regular expression to be specified for validation.

The *WSIVM Rampart module* is the main component of the mechanism implemented. It is a module for Apache Axis 2, which receives data from the client and validates these data according to XML specification, calling Java classes to perform validations. This module consists of a `wsivm.mar` file that has the following compressed components:

- module.xml: contains a description of the module, the class that will carry out the validation, and the phase in which that validation will occur;
- MANIFEST.MF: a Java manifest file;
- Java classes related to input validation: WSIVMModule, WSIVMValidator, SIVMXMLLoader, and WSIVMVerifier.

VI. Case Study

As a case study, a hypothetical system of registration of students for a university named `UniversityManager` was developed. The system comprises a client application called `ClientManager` and a server (a Web service) named `UniversityManager` (Fig. 7).



```
<?xml version="1.0" encoding="UTF-8"?>
<valid_inputs_specification
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
WebServiceID=" UniversityManager "
xsi:noNamespaceSchemaLocation="valid_inputs_specification.xsd">
<operation name="registerStudent ">
<input name="name" type="String" min="5" max="20"
accept="text" sanitize="true"/>
<input name="age" type="Integer" min="0" max="150"
accept="number" sanitize="true"/>
<input name="email" type="String" min="0" max="200"
accept="email" sanitize="true"/>
<input name="comment" type="String" min="0" max="200"
accept="text" sanitize="true"/>
<input name="site" type="String" min="0" max="300" accept="url"
sanitize="true"/>
<input name="data" type="String" min="0" max="200"
accept="regex" regexpattern= "(\\d{4})-(\\d{2})-(\\d{2})"
sanitize="true"/>
</operation>
<operation name="searchStudent ">
<input name="id" type="Integer" min="0" max="10000"
accept="number" sanitize="true"/>
</operation>
</valid_inputs_specification>
```

Figure 7. WSIVMXMLSpecification – UniversityManager [29].

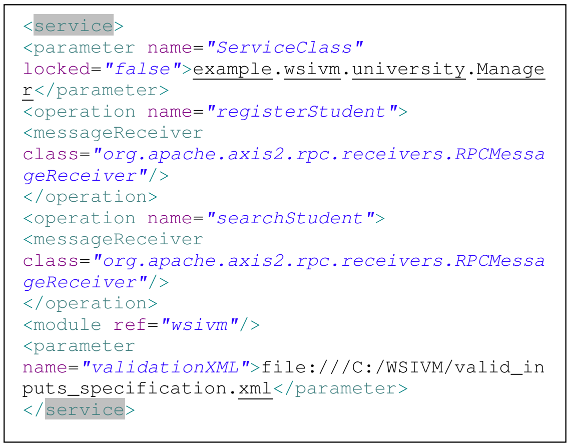
For the development of the Web service, Java language was used and performance tests were conducted using the program soapUI [30]. For the development of the Web service `UniversityManager`, a class with the operations `searchStudent` and `registerStudent` and a class to handle the operations of the database were created. This Web service was developed without any input validation in the operations of Java classes, purposely leaving the validation to WSIVM.

The `searchStudent` operation receives a registration number (ID) that must be an integer that is greater than zero and no more than 10000 and returns the student record containing a *String* with his or her information. The `registerStudent` operation receives the information on the student, which must not contain HTML or Javascript code, and registers it on the MySQL database. In the database a `student's` table is created with the following fields: ID (auto-incrementing identifier), name, age, email, comment, site, and birthday.

After creating the Web service and the database, a Java class called `managerTest` was created to test them locally.

To operate the WSIVM, WSIVMXMLSpecification – UniversityManager was specified according to the standard model WSIVMXMLSchema and describes the parameters for validation of entries.

This way, the `Services.xml` file required for Apache Axis 2 was created (Fig. 8).

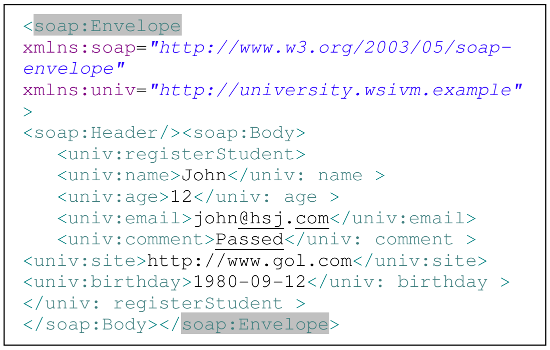


```
<service>
<parameter name="ServiceClass"
locked="false">example.wsivm.university.Manage
r</parameter>
<operation name="registerStudent">
<messageReceiver
class="org.apache.axis2.rpc.receivers.RPCMessa
geReceiver"/>
</operation>
<operation name="searchStudent">
<messageReceiver
class="org.apache.axis2.rpc.receivers.RPCMessa
geReceiver"/>
</operation>
<module ref="wsivm"/>
<parameter
name="validationXML">file:///C:/WSIVM/valid_in
puts_specification.xml</parameter>
</service>
```

Figure 8. Services.xml – UniversityManager.

A package named `Gerenciador.aar`, containing the class Manager, MySQL, the MANIFEST.MF descriptor, and the services.xml file in the META-INF folder, was created.

In this experiment, two tests were performed: one using the WSIVM input validation model and the other without using it. The following scenario was configured to perform the tests: 150 users are started gradually with a user booting every 2 seconds. The test runs for 300 seconds (5 minutes). The database is clean in order to analyze the number of operations for registration of students that are carried out successfully.



```
<soap:Envelope
xmlns:soap="http://www.w3.org/2003/05/soap-
envelope"
xmlns:univ="http://university.wsivm.example"
>
<soap:Header/><soap:Body>
   <univ:registerStudent>
   <univ:name>John</univ: name >
   <univ:age>12</univ: age >
   <univ:email>john@hsj.com</univ:email>
   <univ:comment>Passed</univ: comment >
<univ:site>http://www.gol.com</univ:site>
<univ:birthday>1980-09-12</univ: birthday >
</univ: registerStudent >
</soap:Body></soap:Envelope>
```

Figure 9. Example of SOAP message sent by soapUI.

SoapUI offers a friendly interface for testing. The tests are performed by making direct calls to the Web service. The SOAP message is sent as shown in the example in Fig. 9.

Fig. 10 shows the graph of results of response times of the tests with and without input validation. The X-axis shows the elapsed time of the test and the Y-axis shows the value of the response time in milliseconds.

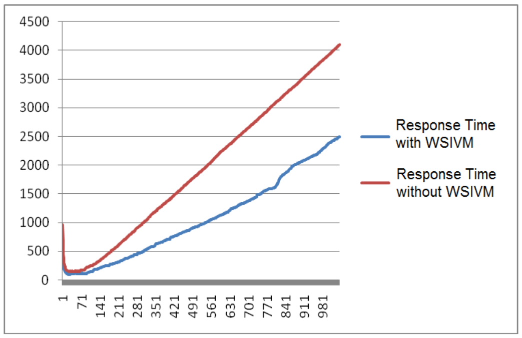


Figure 10. Response times with and without the use of WSIVM.

Fig. 11 shows the graph of the results of throughput tests with and without the WSIVM input validation. The X-axis shows the elapsed time of the test and the Y-axis shows the number of bytes per second (B/s).

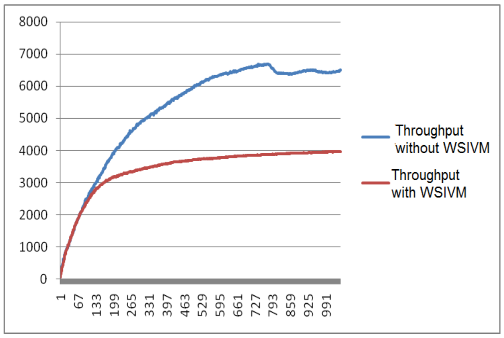


Figure 11. Throughput with and without the WSIVM validation.

It can be observed that the rate of transfer of bytes per second (B/s) or throughput falls considerably with the use of WSIVM.

In Table 1, the calculations that appear in the “Total” row in each of the columns were carried out as follows: the entry in the "Without WSIVM" column is subtracted from that in the "With WSIVM" column and this value is divided by the value of the column "-Without WSIVM".

TABLE I. CONSOLIDATED RESULTS FOR THE CASE STUDY [29].

Comparison	Min. Time	Max. Time	Average Time	Transferred Bytes	Bytes per second (*throughput*)	Insertions in the Database
Without WSIVM	35 ms	27848 ms	2494,85 ms	1974195 B	6506 B/s	10078
With WSIVM	64 ms	13346 ms	4541,24 ms	1236330 B	4012 B/s	5134
Total	83%	-52%	82%	-37%	-38%	-49%

The test results show that when WSIVM is used, a significant increase (82%) in the average response times can be observed, the total throughput decreases by 38%, and the number of students registered in the database decreases by 49% from 10,078 to 5134.

The time spent in the processing of XML messages had an impact on the performance of transactions, which are validated one by one and compared with the rules specified for valid entries. The interpretation of the messages is a costly task in terms of processing and memory requirements, so the validation is done before the processing by the Web service and the return of the response.

There was also a decrease in the total number of bytes transferred because messages took longer to process and therefore a smaller number of messages were processed and the number of responses was lower.

In this case, due to the time required for validation of each message, the application was able to process fewer messages in the same period of time, resulting in fewer insertions of students in the database.

A decrease in performance was expected due to the time required to go through the validation of XML trees in order to validate fields, which is often costly in terms of processing. However, preventing the insertion of invalid data by validating fields can compensate for the loss of performance. This performance loss can be addressed in future work: tests using other mechanisms for interpreting XML files may be carried out as well as tests of the use of a Web service that requires more processing, demonstrating the gain with less waste due to processing of invalid messages. Even so, the protection of services obtained through the use of the model is an advantage that should be considered.

In the tests that were performed no improper entry has been processed since the environment was properly configured to filter invalid entries.

VII. CONCLUSION AND FUTURE WORK

Because unevaluated data entry is the biggest challenge for any application development team in the Web environment and is the source of security problems in many applications [11] [15], a reusable and independent mechanism for data entry validation such as the WSIMV proposed in this paper is an important contribution to the security of Web services.

The WSIVM focuses on validation of data entry, allowing only valid entries to be accepted, since it is based on the white list approach, in which only predefined values are accepted and others are considered invalid.

This model is particularly interesting for the case of Web services that require processing of large amounts of data entries, because by ensuring that only valid entries are accepted it avoids the waste of processing by the application.

Carrying out input validation using the presented model is a solution for legacy applications that were not designed with validation of input data, since carrying out validation at entry points to the Web service decreases the need for a greater number of changes in the existing application, reducing development costs.

Moreover, according to Tsipenyuk et al. [31], the white list approach is more reliable than the blacklist. In the blacklist approach all values are considered valid unless explicitly specified. This approach has some problems; for example, if the validation of a field that does not contain HTML code is desired and a blacklist is created based on the current version of HTML, in the case of new versions, this list may no longer be considered valid.

The white list approach used in WSIVM results in a reduction in false positives and is a more reliable means of validating data entries. In contrast, the work reported in [21] has the disadvantage of obtaining large false positives; that is, the validator may fail by considering a message invalid when in fact the message is valid.

The number of false positives and true positives or false negatives will depend on the WSIVM XML Specification defined. More restrict regular expression specifications could have a negative impact on false positive numbers. The framework provides the specification to be customized according to the Web Service requirements and needs.

This study found a solution for the prevention of data injection attacks in Web services, providing a reusable protection mechanism which prevents the processing of malicious calls and is able to provide validation of input data regardless of the implementation of the Web service that uses this solution.

In the case study, it was observed that improved security had a negative impact on the performance of the developed Web service, which is quite common in security research. However, the validation of inputs reduces the possibility of inserting invalid data and thus prevents attacks that would stop the correct execution of the Web service, offsetting the decrease in performance.

Using SQLmap, SQLninja or Acunetix or majority of available dynamic black-box security tools to test was not considered because most of these tools do not support web services testing.

Tests would be limited to the kind of web service or to the specification, the contribution of the framework with its inherited flexibility is supposed to be more valuable than tests on specific situations, however as the model advances new tests and comparisons will be proposed.

Our previous work published in Brinhosa et al. [29] is a reduced version of these research results. Here, in this paper, we presented in a detailed way: security issues in web services, the WSIVM model as well as the case study development and results obtained with tests.

There are different aspects that can be addressed in future work: (a) optimization of the implementation to improve the performance of the proposed model; (b) development of a semi-automatic generator of security specifications from WSDL; (c) verification of SOAP messages and paths in XPath format; (d) use of artificial intelligence or an anomaly detection system; and (e) making a feedback loop filter validation of invalid entries.

REFERENCES

[1] A. Belapurkar, A. Chakrabarti, H. Ponnapalli, N. Varadarajan, S. Padmanabhuni, and S. Sundarrajan, *Distributed Systems Security Issues, Processes and Solutions*. Hoboken, NJ: John Wiley and Sons, 2009.

[2] M. Q. Saleem, J. Jaafar, and M. F. Hassan, "Model driven security frameworks for addressing security problems of service oriented architecture," in *Proc. 2010 Int. Symp. Information Technology (ITSim)*, June 15–17, vol. 3, pp. 1341–1346.

[3] N. A. Nordbotten, "XML and Web services security standards," *Communications Surveys & Tutorials, IEEE*, vol. 11, no. 3, pp. 4–21, 2009.

[4] CEP. (2011). *CEPWebService*. Available: http://www.i-stream.com.br/webservices/cep.asmx. [retrieved: November, 2012]

[5] SIORG. (2009, Oct.). "Sistema de informações organizacionais do governo federal – SIORG – descrição do Web service SIORG versão 2.0". Available: http://catalogo.governoeletronico.gov.br/arquivos/Documentos/SIORG-DocumentacaoWebServicev.2-091006.pdf. [retrieved: November, 2012]

[6] J. Feiman, "Security in the SOA world: methodologies and practices," *Enterprise Integration Summit*, Sao Paulo, Brazil, Apr. 13–14, 2010. Available: http://www.gartner.com.br/tecnologias_empresariais/pdfs/brl37l_a4.pdf. [retrieved: November, 2012]

[7] J. Viega and J. Epstein, "Why applying standards to Web services is not enough," *IEEE Security & Privacy*, New York, NY, vol. 4, no. 4, pp. 25–31, 2006.

[8] M. Jensen, N. Gruschka, and R. Herkenhöner, "A survey of attacks on Web services," *Computer Science – Research and Development*, vol. 24, no. 4, pp. 185–197, 2009.

[9] S. Lakshminarayanan, "Interoperable security standards for Web services," *IT Professional*, vol. 12, no. 5, pp. 42–47, Sept./Oct. 2010.

[10] SANS. (2011). *The Top Cyber Security Risks*. Available: http://www.sans.org/top-cyber-security-risks/. [retrieved: November, 2012]

[11] OWASP. (2011). "OWASP code review guide. Codereview-Input validation." Available: http://www.owasp.org/index.php/Codereview-Input_Validation. [retrieved: November, 2012]

[12] E. Bertino, L. Martino, F. Paci, and A. Squicciarini, *Security for Web Services and Service-Oriented Architectures*. New York: Springer-Verlag, 2009.

[13] T. Scholte, D. Balzarotti, and E. Kirda, "Quo vadis? A study of the evolution of input validation vulnerabilities in Web applications," in *Proc. Int. Conference on Financial Cryptography and Data Security '11*, St. Lucia, 2011.

[14] CENZIC. (2009). *Web Application Security Trends Report*. Available: http://www.cenzic.com/downloads/Cenzic_AppSecTrends_Q1-Q2-2009.pdf. [retrieved: November, 2012]

[15] OWASP. (2010). "OWASP top 10 Web application security risks". Available: http://www.owasp.org/index.php/Category:OWASP_Top_Ten_Project. [retrieved: November, 2012]

[16] Microsoft. (2012). "Microsoft Anti-Cross Site Scripting Library V4.2" Available: http://www.microsoft.com/en-us/download/details.aspx?id=28589. [retrieved: November, 2012]

[17] SSA (Sosnoski Software Associates Ltd). (2007). *WS-Security Wrapper*. Available: http://wsswrapper.sourceforge.net/. [retrieved: November, 2012]

[18] R. Wu and M. Hisada, "SOA Web Security and Applications", *Technology*, vol. 9, n^{o}. 2, p. 163-177, 2010.

[19] N. Sidharth and J. Liu, "A framework for enhancing Web services security," in *Proc. 31st Ann. Int. Computer Software and Applications Conf., 2007, COMPSAC 2007*, Jul. 24–27, vol. 1, pp. 23–30.

[20] L. Sun and Y. Li, "XML and Web services security," in *Proc. 12th Int. Conf. Computer Supported Cooperative Work in Design*, *CSCWD 2008*, April 16–18, pp. 765–770.

[21] J. Lin and J. Chen, "An automated mechanism for secure input handling," *Journal of Computers*, vol. 4, no. 9, pp. 837–844, 2009.

[22] R. Bebawy, H. Sabry, S. El-Kassas, Y. Hanna, and Y. Youssef, "Nedgty: Web services firewall," in *Proc. IEEE Int. Conf. Web Services – ICWS*, Orlando, pp. 597–601, 2005.

[23] A. Blyth, "An architecture for an XML enabled firewall," *International Journal of Network Security*, vol. 8, no. 1, pp. 31–36, 2009, ISSN 1816–3548.

[24] Y. Loh, W. Yau, C. Wong, and W. Ho, "Design and implementation of an XML firewall," in *Proc. 2006 Int. Conf. Computational Intelligence and Security*, Nov. 3–6, vol. 2, pp. 1147–1150.

[25] A. Kieyzun, P. J. Guo, K. Jayaraman, and M. D. Ernst, "Automatic creation of SQL injection and cross-site scripting attacks," in *Proc. 31st Int. Conf. Software Engineering (ICSE '09)*, IEEE Computer Society, Washington, DC, USA, pp. 199–209.

[26] J. Clarke, *SQL Injection Attacks and Defense*. Syngress Media Inc., 2009.

[27] A. Shostack and A. Stewart, *The New School of Information Security,* Boston: Addison-Wesley, 2008.

[28] Databreaches. (2009). "Top 10 worst data losses or breaches, updated." Available: http://www.databreaches.net/?p=7691. [retrieved: November, 2012]

[29] R. B. Brinhosa, C. M. Westphall, and C. B. Westphall, "Proposal and Development of the Web Services Input Validation Model, " *in Proc. IEEE Network Operations and Management Symposium (NOMS 2012), Maui, Hi, USA, pp. 643-646.*

[30] Apache. (2012). "Welcome to Apache Axis2/Java," Available: http://axis.apache.org/axis2/java/core/. [retrieved: November, 2012]

[31] K. Tsipenyuk, B. Chess, and G. McGraw, "Seven pernicious kingdoms: a taxonomy of software security errors," *Security & Privacy, IEEE*, vol. 3, no. 6, pp. 81–84, 2005.

A Method for Overlay Network Latency Estimation from Previous Observation

Weihua Sun
Nara Institute of
Science and Technology
Ikoma, Japan 630–0192
sunweihua@is.naist.jp

Naoki Shibata
Nara Institute of
Science and Technology
Ikoma, Japan 630–0192
n-sibata@is.naist.jp

Keiichi Yasumoto
Nara Institute of
Science and Technology
Ikoma, Japan 630–0192
yasumoto@is.naist.jp

Masaaki Mori
Shiga University
1-1-1 Banba, Hikone,
Shiga, Japan 522–8522
mori@biwako.shiga-u.ac.jp

Abstract*—Estimation of the qualities of overlay links is useful for optimizing overlay networks on the Internet. Existing estimation methods requires sending large quantities of probe packets between two nodes, and the software for measurements have to be executed at both of the end nodes. Accurate measurements require many probe packets to be sent, and other communication can be disrupted by significantly increased network traffic. In this paper, we propose a link quality estimation method based on supervised learning from the previous observation of other similar links. Our method does not need to exchange probe packets, estimation can be quickly made to know qualities of many overlay links without wasting bandwidth and processing time on many nodes. We conducted evaluation of our method on PlanetLab, and our method showed better performance on path latency estimation than estimating results from geographical distance between the two end nodes. *(Abstract)

Keywords-link quality; PlanetLab; Estimation; Learning Algorithm.

I. Introduction

In order to construct an efficient Peer-to-peer (P2P) overlay network, we need to know the link quality of overlay links, and several methods for estimating link qualities such as available bandwidth, packet-loss rate and latency between peers on the Internet have been proposed. This kind of estimation methods are also useful in client-server applications.

Existing estimation methods requires sending large quantities of probe packets between two nodes. Pathload[1] assumes that a periodic packet stream shows an increasing trend when the stream's transmission rate is higher than the available bandwidth, and it measures the available bandwidth between two nodes. Abing [2] estimates the capacity of a path (bottleneck bandwidth) based on the observed the dispersion experienced by two back-to-back packets. These methods require measurement software to be executed at both of the end nodes. Since accurate measurements require many probe packets to be sent, other communication can be disrupted by significantly increased network traffic. Moreover, in order to make more accurate measurement of link qualities, more probe packets need to be sent into the network. If we could estimate link qualities between each pair of nodes on the Internet. Estimation of link qualities is useful for optimizing overlay networks on the Internet. However, the number of overlay links is the square of the number of peers, it is difficult to estimate all the link qualities using the tools discussed above, since the packets for estimation between a pair of nodes can disrupt measurements between other nodes. In general, the network delay is considered to increase as the geographical distance or the number of routers in the route increases. However, due to the disproportionate data flow, large delay occurs at some specified routers. Also, there are detour of physical communication links by geographical or political reasons. Because of these reasons, link qualities are considered to be attributed to the geographical positions of the two end nodes, rather than just the geographical distance between the nodes. We also need to consider the varying conditions of congestion, and that the situation can suddenly change. However our observation tells that most of the links usually have relatively stable available bandwidths and delays. Since most people use the Internet in the daytime, there should be constant periodical changes of link qualities. Thus, we assume that we can estimate the degree of congestion of an overlay link from periodical observation of the link qualities in the past.

We first discuss these assumptions by conducting experiments on observing link qualities of PlanetLab nodes, and show that the assumptions stated above are probable. Then, we explain our proposed method based on supervised learning for estimation of overlay link qualities from qualities observed in the past. Our method takes account of the geographical locations of end nodes to estimate the link qualities. Our experiments on PlanetLab showed that our method has good performance on path latency estimation. The estimation based on just geographical distance showed large error, especially when the distance is shorter than 2000 km. The proposed method achieved high estimation accuracy in that range. We have shown a part of our results in a work-in-progress paper [3], and we show detailed experimental results and discussion in this paper.

In Section II, we provide an overview of related works, while in Section. III we present a preliminary discussion on how accurately we are able to estimate link qualities from those previously observed. We propose a method for estimating link qualities based on a supervised learning algorithm in Section IV and present the results of experiments on PlanetLab to demonstrate the accuracy of the proposed method in Section

ISBN: 978-1-61208-245-5

V. Finally, our conclusions are given in Section VI.

II. Related Research and Contribution

Previously, in the field of wide area networks, many approaches were proposed to measure and estimate the delay and bandwidth between end nodes. Accurate estimation of the available bandwidth is important for throughput optimization between end nodes, overlay network routing, peer-to-peer file distribution, traffic engineering, and capacity planning. In this section, we discuss the measurement and prediction methods with respect to the available path bandwidth between end nodes.

A. Bandwidth Measurement Method

There are three different metrics for path bandwidth between end nodes: (1) capacity (maximum bandwidth), (2) available bandwidth (maximum unused bandwidth), (3) TCP throughput/bulk transfer capacity (maximum achievable bandwidth). The existing four measurement methodologies are:

- **VPS (Variable Packet Size probing)** is a method to estimate link capacity by measuring the round-trip time; that is, calculating the serialization delay of various sized packets sent from a sender node to a receiver node.
- **PPTD (Packet Pair/Train Dispersion)** is a method for measuring the capacity of the path between end nodes. Letting a sender node continuously send uniform sized packet pairs or trains to a receiver node, this approach calculates the maximum link serialization delay in the path to estimate the minimum link capacity (bottleneck) by measuring the dispersion of the received packet times.
- **SLoPS (Self-Loading Periodic Streams)** is a method for measuring available bandwidth. While a sender node continuously sends uniform sized packets to a receiver node with transmission rate R, SLoPS observes the variation in delay for each packet at the destination node, and measures whether R is greater than A. By adjusting the transmission rate R, SLoPS estimates the available bandwidth A.
- **TOPP (Trains of Packet Pairs)** measures capacity and available bandwidth by transmitting data at a particular transmission rate for a specified number of packet pairs. Unlike SLoPS, TOPP estimates the available bandwidth by increasing the transmission rate linearly and observing the arrival delay.

Other tools that have been proposed and implemented are **Pathchar, Clink**, and **Pchar** for measuring link capacity, **Brpobe, Nettimer, Pathrate** and **Sprobe** for measuring path capacity, **Cprobe, Pathload, IGI**, and **pathChirp** for measuring available bandwidth, and **Treno, Cap, TTCP, NetPerf, Iperf** for measuring TCP throughput. As reported in [4], **Pathload** and **pathChirp** showed better performance than **Abing, Spruce**, and **Iperf** on a high-speed network testbed.

Most of the above tools focus on measuring the average available bandwidth, but do not consider bandwidth variation. Therefore, the authors in [6] proposed a method to measure bandwidth variation. Moreover, with the goal of estimating the bandwidth without causing excessive traffic, a method was proposed in [5] to estimate capacity and available bandwidth without congesting the minimum capacity link in the path.

Most of the existing bandwidth measurement methods and tools work by exchanging probe packets between sender and receiver nodes. Although these methods are useful for accurate bandwidth measurement, they generate traffic while measuring bandwidth. SLoPS and TOPP, in particular, cause temporary congestion of the minimum capacity link. Consequently, in a large scale P2P network with millions of nodes, these methods may cause serious deterioration in the network performance.

B. Bandwidth/Latency Prediction Method

Various network traffic prediction models have been proposed. In networks, similar traffic patterns with long time intervals are said to be self-similar, while those patterns with short time intervals are called multi-fractal. In [8], a method was proposed to predict network traffic at several time steps in advance, based on past measured traffic information. Moreover, the authors in [7] improved the method in [8], by proposing a new ARIMA/GARCH model that predicts network traffic with higher accuracy. In this model, self-similarity and multi-fractals can be predicted by utilizing short-range and long-range dependencies. Through comparison experiments with real network traffic, the authors showed that network traffic can be predicted with reasonable accuracy.

These models aim to predict future traffic from previous detailed measurements. Moreover, the models can be used to predict the available bandwidth and latency by separately measuring the capacity of the path between the end nodes. Similar to the above methods, the method in [9] accurately estimates the latency of the path between end nodes based on traffic measurement. However, because detailed measurements are needed in advance, the models are not suitable for estimating bandwidth/latency at low cost owing to the additional traffic generated.

C. Contribution

The traffic prediction model makes use of the self-similarity and multi-fractal properties of traffic. By applying these characteristics to the different nature of similar paths, link qualities (including end to end delay, available bandwidth, and so on) can be predicted using fewer a priori measurement results.

In this paper, by considering the similarity of paths, we propose an overlay link quality prediction method, which assumes that similar paths have similar characteristics. To the best of the authors' knowledge, there is no other prediction method that, like ours, does not require much bandwidth. Moreover, we have implemented the proposed method in PlanetLab and evaluated the performance thereof.

III. Preliminary Experiments And Observation

In this section, we first describe the results of two preliminary experiments. In the first experiment, we observed the fluctuations in link quality over time, while in the second, we investigated the relation between route (overlay link) similarity

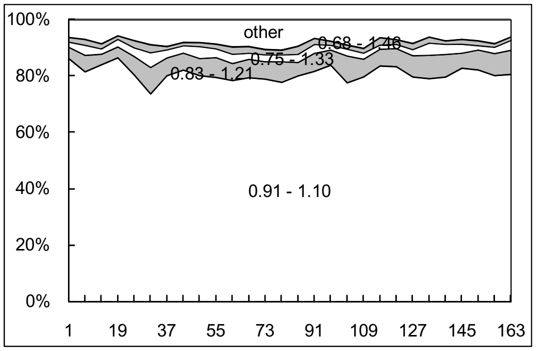


Fig. 1. Observed fluctuation of latency (X axis = time)

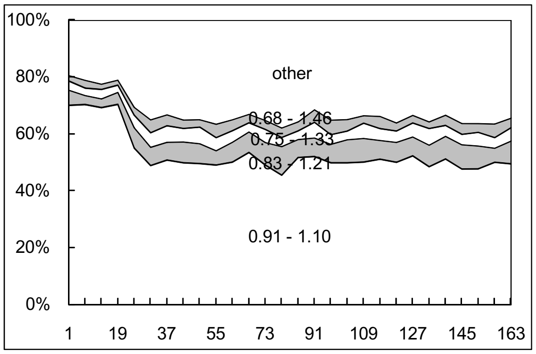


Fig. 2. Observed fluctuation of available BW (X axis = time)

and the difference in link qualities. The amount of traffic on the Internet changes continuously, influenced both by the day of the week and the season. We observed the actual fluctuations in link quality on PlanetLab. In the subsequent subsections, we describe the configuration of the experiments, the definition of route similarity, and the results of these experiments.

Observation of PlanetLab: We observed the fluctuations in available bandwidth and latency between nodes in PlanetLab over 7 days starting on 20th January 2011. We created 500 random pairs among the nodes in PlanetLab and measured the available bandwidth and latency using Pathload and ping every hour. About 63000 valid data records were obtained.

Fig. 1 shows a stacked bar graph of the observed latency at each time divided by the latency observed at the beginning. The bottom series indicates the ratio of routes where the observed latency divided by the latency observed at the beginning was between 0.91 and 1.1. The second series indicates the ratio of overlay links with latencies between 0.83 and 1.21 times. Fig. 2 shows the results for bandwidth. From Fig. 1 it is clear that for 80% of the routes, the fluctuation in latency was less than 10%, and this ratio did not change for the whole week. Fig. 2 shows that for 70% of the routes, the fluctuation in bandwidth was less than 10% for 20 hours from the beginning of the experiment. It also shows that for half the routes, bandwidth fluctuation was less than 10% for the week. We did not observe daily periodic fluctuations in bandwidth or latency.

A. Relation between Route Similarity and the Difference in Link Quality

It would be convenient if we could estimate the link quality of an unknown overlay link on which no link quality observations have been made. To realize such a method, we first define similarity between two overlay links based on geographical distance. There are free databases from which we can find the geographical location of nodes from their IP addresses, and thus it is easy to locate the geographical position of nodes on the Internet. We also show the measurement results for link similarity and the difference in link quality.

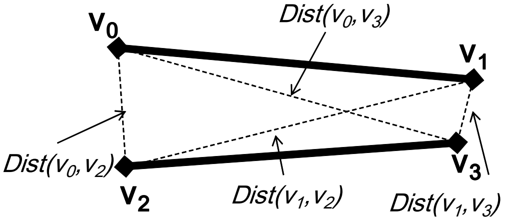
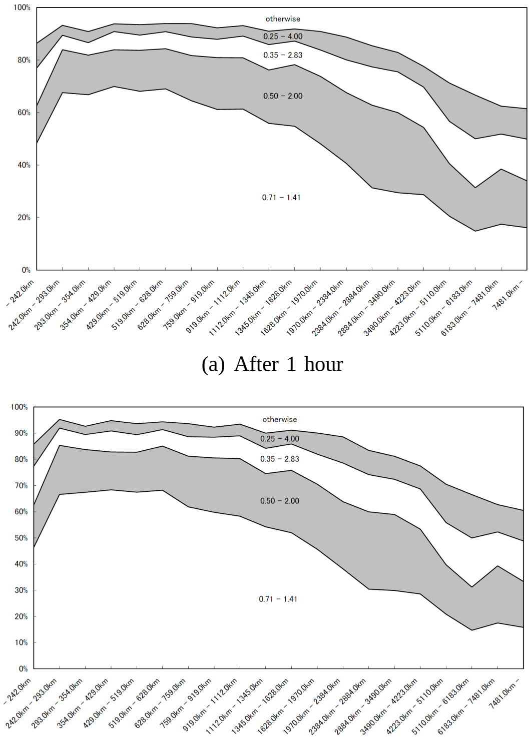


Fig. 3. Similarity Definition



(a) After 1 hour

(b) After 6 days

Fig. 4. Estimated latency by Proposed method, X axis = distance(Km)

Route Similarity: As shown in Fig. 3, the route similarity $geo(v_0, v_1, v_2, v_3)$ between two routes is defined as the minimum value between $dist(v_0, v_2) + dist(v_1, v_3)$ and $dist(v_0, v_3) + dist(v_1, v_2)$, where $dist(v_0, v_1)$ denotes the geographical distance between v_0 and v_1.

Measurements on PlanetLab: We created 500 random pairs of nodes on PlanetLab, and investigated the relation between similarity as defined above and the observed latency. Fig. 4(a) shows the relation between link similarity and latency fluctuation one hour after the first measurement was made, while Fig. 4(b) shows the results obtained six days after the first measurement. We can see that these two graphs are almost identical, and that there is almost no change in the fluctuation over time. We can also see that the amount of fluctuation decreases with more similar routes. With the sum of the

ISBN: 978-1-61208-245-5

distance less than 600 km, the fluctuation is within 50% to 200% for 80% of the routes.

We also performed similar experiments on bandwidth, but did not observe any relation between route similarity and fluctuation. This seems to be due to the fact that the available bandwidth is usually limited by the bandwidth for the last hop rather than that for the entire backbone. However, we are still investigating finding an appropriate similarity definition for estimating the correct bandwidth.

IV. Overlay Link Quality Estimation Method

In this section, we propose an overlay link quality estimation method based on the results of the preliminary experiments in Section III. In the proposed method, (1) a centralized server periodically collects, from various peers in the P2P network, quality information of overlay links they have observed, and (2) the quality of a given overlay link is estimated from the information of previously observed overlay links based on the weighted k-nearest neighbor ($WKNN$) algorithm, which is one of the supervised learning techniques.

A. Preliminaries

1) Weighted k-nearest neighbor algorithm: The WKNN method uses training samples expressed as pairs of an object and a real number, and learns a function that maps an arbitrary object to a real number. In our proposed method, the object and real number correspond to an overlay link and latency, respectively.

To use the WKNN algorithm, the following two functions must be given: (1) a function to calculate the distance between two objects; and (2) a function that assigns a weight to each object.

In the WKNN algorithm, learning is carried out using all training samples (the training set) stored in memory. When estimating a real number for an input object, WKNN selects the k samples in the training set geographically closest to the input object, and estimates a real number by calculating the weighted average of the k samples with their weights.

2) Assumptions, estimation target, and algorithm outline: We aim to apply the proposed method to estimate overlay link quality in a P2P application such as video streaming. We assume that the P2P application consists of a central server and many peers (users). In the application, each peer observes the quality of the overlay links directly connected to other peers and periodically sends the observed information to the server. In this study, we have designed the learning algorithm as a centralized one, but it could easily be implemented as a distributed algorithm using, e.g., a distributed hash table.

The proposed algorithm is executed on the server and estimates the quality of a given overlay link by applying the WKNN method to the previously observed quality information collected by the server. As described in Section III-A, we could not find any correlation between link similarity and the observed available bandwidth. Thus, we focus mainly on overlay link latency as the quality estimation target in this study.

Each peer sends the server a query to estimate the quality of the specified overlay links. When the server receives a query, it estimates the quality of the given links based on the proposed algorithm and sends the estimated result back to the peer.

The server carries out learning and estimation. In the WKNN algorithm, the server performs learning using all training samples stored in its memory. As time progresses, the number of training samples increases and more memory space is required. To limit the required memory size, when the number of training samples exceeds a predefined threshold, the oldest samples are deleted from memory.

The size of a message that a peer exchanges with the server (to upload the observed link quality, send a query for link quality estimation, or receive the estimation result) is at most 200 bytes since it contains only an overlay link together with the associated quality.

The server has a table that maps IP addresses to geographic coordinates as explained in Section III-A.

B. Learning algorithm

The proposed algorithm consists of two phases: (i) a learning phase, and (ii) an estimation phase. We describe these phases in detail below.

1) Learning phase: We assume that each peer participating in a target application communicates frequently with other peers participating in the same application, e.g., to realize video P2P streaming.

In the proposed algorithm, each peer performs the following steps:

- When the peer communicates with other peers, it measures the quality of the overlay links to those peers.
- The peer periodically sends the quality of overlay links observed during the current period to the server. The message contains the IP addresses of both ends of each overlay link and the measured latency.

When the server receives the observed quality of an overlay link from a peer, it stores the data –that is, the IP addresses of the end nodes of the overlay link and the latency, in its memory.

When the amount of data exceeds a predefined threshold, the server removes the oldest data from its memory.

2) Estimation phase: When a peer wishes to know the quality of an overlay link, it sends the server a query specifying the IP addresses of the end points of the link. When the server receives the query, it estimates the quality of the specified link as follows:

- The server selects the k closest training samples from the training set.
- It calculates the weight of each selected sample as explained in Section IV-B4.
- It calculates the weighted average of the latency of the selected k samples.
- It sends the calculated result to the peer that originally sent the query.

3) Estimation example: Let us suppose that peer n_0 has sent the server a query regarding the overlay link between itself and peer n_1. When the server receives the query, it selects the k training samples closest to the overlay link between n_0 and n_1 based on the distance function defined in Section III-A. Let us suppose that $k = 2$ and overlay links r_1 and r_2 have been selected. Then, the server calculates the weights of r_1 and r_2 according to the method in Section IV-B4. Let the weights for r_1 and r_2 be 1 and 2, respectively. Let us also suppose that the previously observed latencies of r_1 and r_2 are 3 and 4, respectively. Finally, the server obtains the value $(1 \times 3 + 2 \times 4)/(3 + 4) = 1.57$ as the weighted average and sends this value as a reply to peer n_0.

4) Weight function: In Section III-A, we defined the similarity between two overlay links observed at the same time. In general, this similarity should be defined between two links observed at different times. However, as explained in Section III-A, the variation in latency with time is rather small. Thus, we use the similarity function defined for two links observed at the same time in the proposed algorithm.

According to the measurement results presented in Section III, more than 80% of overlay links experience a latency variation between 0.71 and 1.41 times the initial measured latency. Thus, we define the weight function as follows:

$$Weight(u_s, u_d, v_s, v_d)0.7 - \frac{0.3}{5000} \cdot geo(u_s, u_d, v_s, v_d) \quad (1)$$

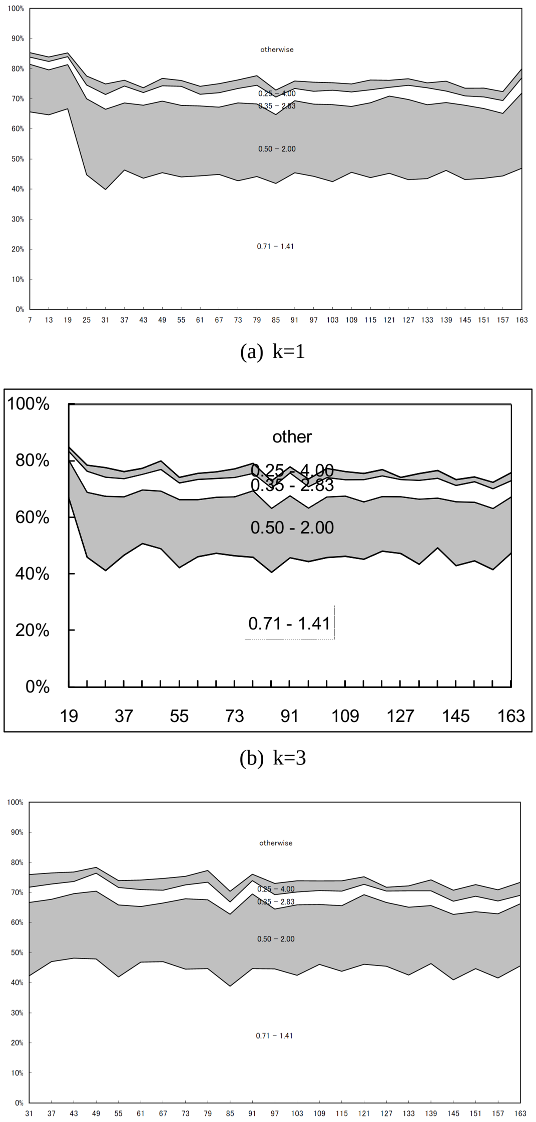
where (u_s, u_d) and (v_s, v_d) are the geographic coordinates of the target overlay link and the training sample, respectively.

V. Evaluation

In this section, we evaluate the estimation accuracy of the proposed method. According to the underlying principle of the proposed method, the estimation accuracy depends on the distance and time from the measured path. The greater the difference in time or distance is, the worse is the estimation accuracy. With respect to available bandwidth, we investigated the relationship over time and estimation accuracy. With respect to latency, we investigated the relationship over the distance between paths and estimation accuracy.

A. Evaluation of Available Bandwidth

As described above, despite the paths being similar, no correlation with available bandwidth was observed. In this experiment, using the k measured results of both cases of one measurement per day and one measurement per hour on a certain path, we investigated the estimation accuracy when varying k and the elapsed time from the last measurement. The results are shown in Figs. 5(a)-5(c). According to these figures, the observed estimation accuracy corresponds to the results of the preliminary experiments. However, the estimation accuracy did not improve even when increasing k.



(a) k=1

(b) k=3

(c) k=5

Fig. 5. Estimated bandwidth, X axis = time

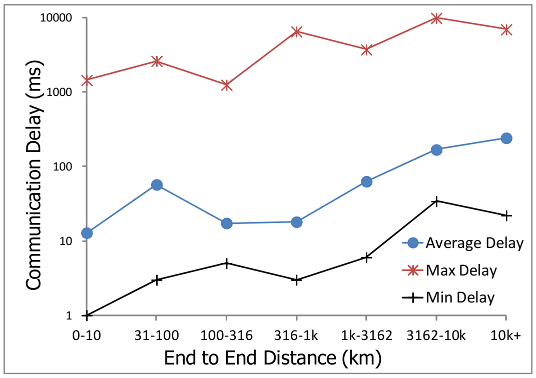


Fig. 6. The average, maximum and minimum path delay

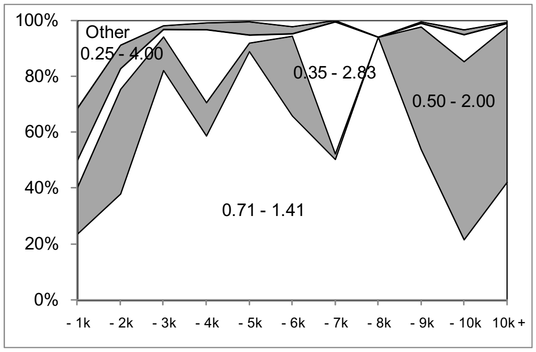


Fig. 7. Accuracy of Estimated delay based on path length (X axis = distance (KM))

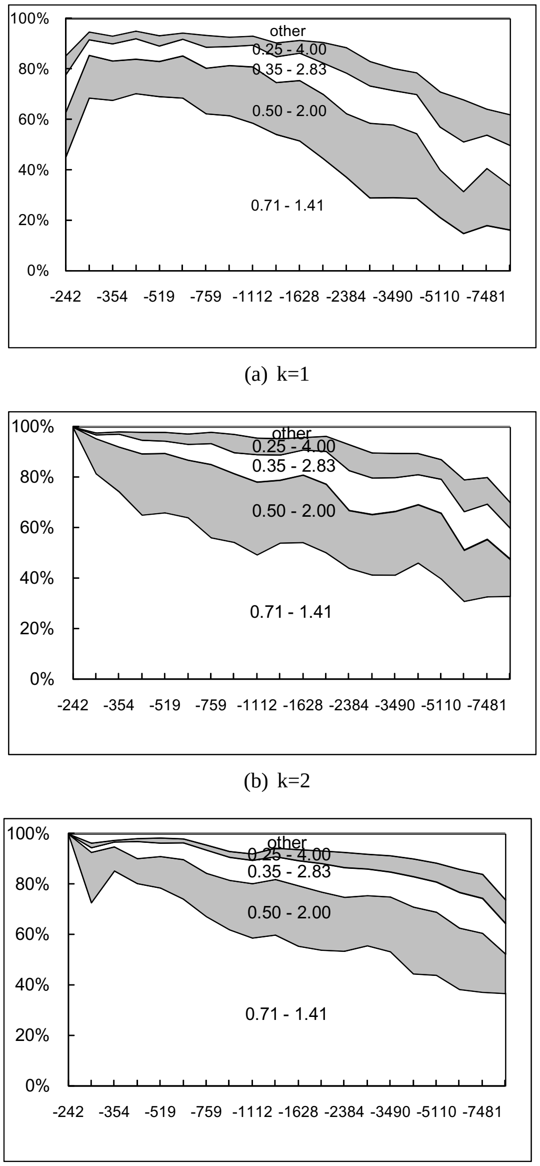
1) Estimation based on link distance: For comparison with the proposed method, we used a delay estimation method based on link distance. Fig. 6 shows the relationship between link length and delay. According to this result, the average delay of a path increases roughly in proportion to the distance. However, the maximum and minimum delays do not follow this trend. The delay calculated from the average delay is 0.019 ms/km. The results of applying this value to the delay estimation method are shown in Fig. 7. Obviously, the estimation accuracy is low when the link distance is less than 2000 km.

2) The proposed method: In this experiment, we investigated the accuracy of measuring path latency based on the measured latency results of k different paths six days previously. We investigated the estimation accuracy for a number of distance functions by varying k. The results are shown in Figs. 8(a) – 8(c).

According to these figures, accurate estimation was observed. The estimation accuracy improved as k increased. In particular, we confirmed that the estimation accuracy (0.71–1.41 and 0.5–2.0) is very high for medium and short distances, respectively.

VI. Conclusion

We proposed a learning-based overlay link quality estimation method that uses the quality observed for other links in the past. With respect to latency, by defining geographical similarity between overlay links, the proposed method achieves good estimation accuracy. With respect to bandwidth, we found that there is no correlation between overlay links with close geographical similarity. In the future, we intend devising a new similarity metric to accurately estimate overlay link bandwidth taking into account domain type, connecting ISP, and so on.



(a) k=1

(b) k=2

(c) k=3

Fig. 8. Estimated latency by Proposed method, X axis = distance(Km)

References

[1] M. Jain and C. Dovrolis, “End-to-end available bandwidth: measurement methodology, dynamics, and relation with TCP throughput,” in IEEE/ACM Transactions on Networking, vol. 11, Issue 4, pp. 537–549, 2003.

[2] J. Navratil and R. L. Cottrell, “ABwE:A practical approach to available bandwidth estimation,” in Proc. of Passive and Active Measurement Workshop (PAM’03), in-cdrom, 2003.

[3] W. Sun, N. Shibata, K. Yasumoto, and M. Mori, “Estimation of overlay link quality from previously observed link qualities,” in Proc. of The 10th Annual IEEE Consumer Communications Networking Conference (CCNC’13), pp.582-585, Jan. 2013.

[4] A. Shriram, M. Murray, Y. Hyun, N. Brownlee, A. Broido, M. Fomenkov, and K. Claffy, “Comparison of public end-to-end bandwidth estimation tools on high-speed links,” in Proc. of of Passive and Active Measurement Workshop (PAM’05), pp.306–320, 2005.

[5] S. Kang, X. Liu, M. Dai, D. L, and D. Loguinov, “Packet-pair bandwidth estimation: stochastic analysis of a single congested node,” in Proc. of 20th IEEE International Conference on Network Protocols (ICNP’04), pp.316–325, 2004.

[6] M. Jain and C. Dovrolis, “End-to-end estimation of the available bandwidth variation range,” in Proc. of the 2005 ACM SIGMETRICS international conference on Measurement and modeling of computer systems, pp.265-276, 2005.

[7] B. Zhou, D. He, Z. Sun, and W. H. Ng, “Network traffic modeling and prediction with ARIMA/GARCH,” In Proc. of 3rd international working conference: performance modelling and evaluation of heterogeneous networks (HET-NETs’05), in-cdrom, 2005.

[8] A. Sang and S. Li, “A predictability analysis of network traffic,” in Proc. of IEEE INFOCOM 2000, pp.342–351, 2000.

[9] N. Hariri, B. Hariri, and S. Shirmohammadi, “A distributed measurement scheme for Internet latency estimation,” in Proc. of IEEE Trans. on Instrumentation and Measurement, 60 (5): pp.1594–1603, 2011

An Intelligent Approach to an Efficient Internet Network Management

Antonio Martín
Department of Electronic Technology
Seville University
Seville, Spain
toni@us.es

Carlos Leon, Alejandro López
Escuela Superior de Ingeniería Informática
Seville University
Seville, Spain
cleon@us.es, alexlopez@us.es

***Abstract*— Telecommunications networks are continuously growing in scale and complexity, and the amount of information and services provided more day to day. The management of the resulting networks gets additional important and time-critical. More advanced tools are needed to support this activity. In this paper we describe the design and implementation of a management platform using Artificial Intelligent reasoning technique. This paper explores intelligent agent architecture to make the argument for an intelligent solution as opposed to traditional methods. We propose a new paradigm where the intelligent network management is integrated into the conceptual repository of management information. This study focuses on an intelligent framework and a language for formalizing knowledge management descriptions and combining them with existing Internet management model. Based on the present proposal and Internet management model, we describe the design and implementation of an integrated intelligent management platform named ALATUS. We have tested our system on real data to the fault diagnostic in a university wireless network. The results of a validation show a significant improvement with respect to the number of rules and the error rate in other traditional systems.**

Keywords-SNMP; MIB; IAs; Management Network.

I. Introduction

Due to the increasing complexity and heterogeneity of networks and services, many efforts have been made to develop intelligent techniques for management. Current communications networks support a large demand of services, which the traditional model of network management is inadequate. Classic management architectures are not well suited for low-bandwidth or disconnected operation. Traditional management frameworks such us OSI (Open System Interconnection) and SNMP (Simple Network Management Protocol) are capital approaches where a manager uses distributed agents to collect management information. But these strategies have drawbacks, in particular due to the lack of extensibility and scalability on very large networks. This restriction is coming from the inability of a centralized manager to handle huge amounts of management information across geographically distributed sites, which is also expensive in resources. A key technology for operating large heterogeneous data transmission is network intelligent management [1].

Several authors have addressed these problems along the past years [2][3][4][5][6] resulting in ad-hoc and partial solutions typically based on management distribution and intelligent delegation. There is no full scale of integration of IA (Intelligent Agent) applied to SNMP network management architectures. The main contribution of this paper lies in the proposal for a framework that integrates management object specifications and the knowledge of expert systems. This study draws on the theory, experimentation and findings of the SNMP management model and the integration management knowledge to obtain an efficient network control. The goals are to improve insight and understanding of network management, and present an alternative distributed management network model. This architecture is based on the idea that a main manager can delegate the control to several IAs agents thus improving scalability and efficiency of a management network through intelligence distribution and intelligent management actions. In order to develop it a language to formalize the knowledge base description in Internet management model is discussed.

The scope of this work covers why IAs are very well suited to meet the network management requirements and how an intelligent-agent-based system can be applied to achieve intelligent network management. The study addresses problems that traditional network management poses and makes the argument for the solution based on IAs agents to these problem areas. Section II examines Internet management network model, including concepts and major approaches. Section III presents capabilities required for an efficient network management and current shortcomings of SNMP model. Section IV gives the formulation of proposal and a schema of the various stages in the system development. Section V presents a conceptual, high-level intelligent agent named ALATUS and summarizes the performance of the research. Finally we outline the conclusion and future works.

II. Internet Management Network Overview

According to ISO (International Organization for Standardization), the network management model defines a conceptual architecture for managing all communication entities within a network. There are several organizations which have developed services, protocols and architectures for network management. The most important organizations are: ISO, which was the first one and started as part of its OSI program [7], Telecommunication Management Network (TMN), developed by International Telecommunication Union (ITU), and Internet Model by the Internet Engineering Task Force (IEFT), figure 1.

Copyright (c) IARIA, 2013.   ISBN: 978-1-61208-245-5

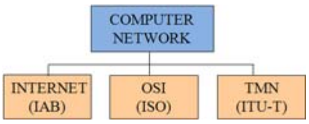


Figure 1. Management Models.

A main concept in management networks is the managed object, which is an abstract view of a logical or physical resource which must be controlled. The managed objects provide necessary operations for the running, monitoring and control of the telecommunications network. These operations are realized through the use the Common Management Information Protocol (CMIP) to OSI model and SNMP in the Internet model. SNMP is one of the widely accepted protocols to manage and monitor network elements and operates in the application layer in Internet protocol suite, commonly known as TCP/IP (Transmission Control Protocol/Internet Protocol) suite [8].

The SNMP framework is based on the principle of minimally simple agents and complex managers. The managed object provides an abstract view of a real resource, and the agent provides a management view of their underlying logical and physical resources, such as transport connections to the managing applications. For a right running, the management processes involved will take on one of two possible roles, figure 2.

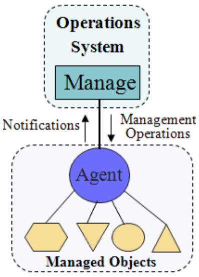


Figure 2. Manafe/Agent roles.

In SNMP uses, one manager has the task of monitoring or managing a group of hosts on a computer network, so it is responsible for network management processes. An agent performs the management operations on the managed objects according to the request of the manager, and reports notifications that managed objects produce.

A. Information Management Repository

SNMP objects represent single, atomic data elements that may be read or written in order to effect the operation of the associated resource. This set of managed object classes and instances under the control of an agent is known as Management Information Base (MIB), an abstraction of network resources, properties and states for the purpose of management. SNMP information modeling principles are collectively referred to as the Structure of Management Information (SMI) and are specified in (RFC1155) for SNMPv1 and in (RFC1902) for SNMPv2. The MIB is written in Abstract Syntax Notation 1 (ASN.1), a standard syntax for data types and values, which is maintained by ISO [9]. ASN.1 is a language for describing structured information, which is widely used in the specification of communication protocols. ASN.1 describes the relevant information and its structure at a high level and need not be unduly concerned with how it is represented while in transit.

The manager and agent use the MIB with a relatively small set of commands to exchange information. When an SNMP device sends a trap, each data object in the message is identified with a number string called an Object Identifier (OID). A SNMP manager knows the value of an object/characteristic, such as the state of an alarm point, the system name, or the element uptime. All these information will assemble in a GET packet that includes the OID for each object/characteristic of interest.

The MIB is organized in a tree structure with individual variables, such as point status or description, being represented like leaves on the branches. The MIB is an ASCII text file that describes SNMP network elements as a list of data objects. It is like a dictionary of the SNMP language where every object, related to an SNMP message, must be listed in the MIB. The first step towards development of an agent is to define its MIB. The steps involved in developing the MIB file are [10]:

- Data identification: To identify data or objects which need to be managed using IA methods, laying them out in the form of scalar or tabular objects. In that way, all knowledge associated to a specific managed resource is categorized.

- Data definition: Construct ASN.1 MIB definitions for the IA. In this study, we define different ASN.1 types of knowledge related to the network resources. For this purpose, we have used an editor tool. Editors can help in MIB design hiding the unimportant details of the MIB syntax rules, clauses etc.

In our work a framework for the inclusion of formal Knowledge Management descriptions into MIB specifications has been proposed. An object-oriented logic programming language is presented, which can be used in conjunction with the framework to specify knowledge management of a managed object.

III. Intelligent Management Agents

IA technology has the capability to distribute intelligence throughout network and perform intelligent management functions dynamically. This provides efficiency and flexibility, and cuts down on bandwidth constriction and overloading on a single central control. An important goal is convergence on solutions despite of incomplete or inconsistent knowledge or data. IAs can seek to cooperate to solve problems using task and domain-level protocols actively and dynamically. IAs learn the normal behaviour of each measurement variable and add intelligent knowledge to the management network resources. IA is based on three essential properties: autonomy or self-government independence; communication, which is the ability to speak with a peer; and cooperation in order to create a collaborative environment to work together.

So as to improve the quality of the IA description and the resulting implementations, a formal method for specifying knowledge is desirable. Due to IA is based on management knowledge, formal knowledge descriptions helps to make easier for an engineer to understand the complete information model and to derive a valid, consistent and compatible implementation. In the next section knowledge management using ASN.1 notation is modelled. It defines the formal specification of managed object types and the associated access mechanisms [11].

An IA is essentially a self-contained software program module that is programmed to carry out specific actions on behalf of a human user or another software entity in a certain software environment. Every SNMP IA maintains an information database which describes the managed device parameters and the knowledge base where all relevant information used with management purposes (data, rules, cases, and relationships) is stored. IA can perform actions such as to collect management information about its local environment, store and retrieve management information as it is defined in the MIB, execute specified tasks or collaborate with other agents. These actions are conducted in an autonomous way that requires little or no human intervention, figure 3.

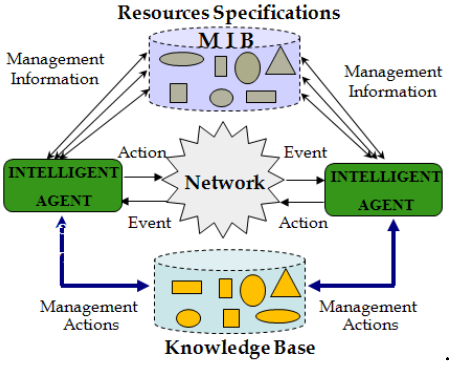


Figure 3. SNMP Intelligent Agents.

In our proposal IAs are the administrative systems and have the task of monitoring or managing a group of devices on a network. Each IA executes, at all times, a software component called an agent which reports management actions via SNMP to the managed objects. This component software is supplied with the right MIB file. The SNMP IA can correctly interpret alarm data from any device that supports SNMP and executes the corresponding management actions.

IV. Internet Management Model Extension

Although SNMP SMI identifies how resources are represented and named within the MIB, there is no mechanism provided by SNMP to enable agents to operate with management knowledge. In this section, we face the problem of specifying knowledge transfer and how to describe this knowledge in the abstract syntax notation in order to show rules to create the MIB, to improve insight, and understanding of network management. The structure of SMI, an SNMP standard, defines the structure of the MIB information and the allowable data types. The philosophy behind SMI is to encourage simplicity and extensibility within the MIB. When we are planning monitoring SNMP resources, it is necessary to be able to read MIBs so that we can get a realistic idea of what management capabilities we have available. Just looking at the physical components of a device will not tell us what kind of events and knowledge we can get from it. So the MIB is the guide to the real capabilities of an SNMP device. It is usual that a manufacturer adds a new component or functionality to a device without describe it in the MIB. In fact, nowadays, a lot of devices have sketchy MIBs that do not fully support all their functions. However, the object description in the MIB is a capital point in order to combine all their properties and management actions.

Whit the purpose of achieving an SNMP IA the knowledge base and the MIB have been joined, figure 4.

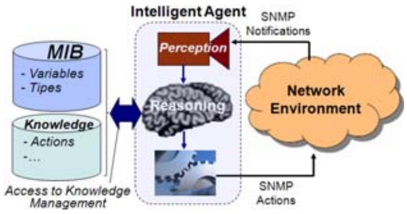


Figure 4. Intelligent Agent Architecture.

SNMP SMI defines a specification for describe the properties of new object types, while ASN.1 is used to specify the object syntaxes and the tabular structure. ASN.1 is specifically designed for communication between dissimilar agents, thus it is the same for every system and it can be used for describing almost anything. In this work we have used ASN.1 notation to extend the SNMP model by integrating the knowledge base in the MIB. Once a term is defined in ASN.1, it can be used as a building block to make other terms. ASN.1 defines each term as a sequence of components, some of which may be sequences themselves. ASN.1 provides basic types like boolean, null, integer, real, etc., and type constructors that can be used to define new types: CHOICE, SEQUENCE, SET, and SEQUENCE OF and SET OF. We take advantage of ASN.1 flexibility and extensibility to apply a top-down approach to our problem.

First we consider the problem as a whole to describe the most general data types and second we concentrate on more specific types which are referenced in general data types, in order to show a set of rules, which allows intelligent actions. There are different knowledge representation techniques to structure knowledge. In our study, we are going to represent the knowledge management in production rules or simply rules. Rules specify actions for the inference engine with the purpose of taking them when the premise or conditions in the rule are true. Rules are expressed as “IF-THEN” statements, which are relatively simple, very powerful and a natural way to represent expert knowledge [12]. A major feature of a rule-based system is its modularity and modifiability that allow an incremental improvement and fine tuning of the system with virtually no degradation of performance. The next definition shows the new elements of the knowledge type using ASN.1 notation.

```
Know_Rule ::= SEQUENCE {
    priority     INTEGER,
    condition ConditionType,
    action       ActionType    }
```

The priority of the element is defined using the primitive element “integer”. However condition and action elements are defined as new types named “ConditionType” and “ActionType” respectively. We can do this because these types are defined in another sequence like so:

```
ConditionType ::= SEQUENCE {
    variable  OCTET STRING,
    operator OperatorType,
    value       INTEGER             }

 OpertatorType ::=SET { equal                      [1] INTEGER,
                         not equal              [2] INTEGER,
                          less than               [3] INTEGER,
                        …           }
```

When a type is defined, a name to reference it in another type assignment should be given to it.

```
ActionsType ::= SEQUENCE      {
    executemode         Modetype,
    command  OCTET STRING,
    arguments OCTET STRING    }

Modetype ::=SET { user         [1] INTEGER,
                  privileged  [2] INTEGER,
                  global        [3] INTEGER,
                  interface    [4] INTEGER     }
```

In ASN.1, the concept of information object class is used to represent formally properties uncovered by the notions of type and value in particular [13]. An information object class is a group of things sharing common characteristics. We have used object class to interpret the semantic links between types and values in a management action definition.

```
ACTION ::= CLASS {
                  &code INTEGER UNIQUE,
                 &Argument-type,
                 &Return-result-type }
```

The class has three parts: an identification code to specify the function which must be executed by the remote application: an argument with a value that conforms to the ASN.1 type of the argument; and the IA, which receives a value that conforms to the result type if the function execution was successful, otherwise one specified error message. The block WITH SYNTAX defines a more user-friendly syntax to denote the objects of this class.

```
WITH SYNTAX {ACTION CODE &code
    TAKES AN ARGUMENT OF TYPE &Argument-type
     AND RETURNS A VALUE OF TYPE &Return-result-type}
```

The following definition is an example of expert rules integration in the SNMP IA proposed standard. It defines an IA named *accessPoint* corresponding to a real device in the network. *AccessPoint* IA contains all the specifications and knowledge corresponding to the device. These units offer the convenience of multiple functions such as establishing radio channels, controlling signals, monitoring stations, monitoring alarm conditions, controlling logic to activate operations in response to commands received over said communications network, and so on.

```
accessPoint IA-OBTECT-TYPE                        -- Object
                SYNTAX  SEQUENCE OF accessPointEntry
                ACCESS  not-accessible
                STATUS  currrent
                DESCRIPTION  "Access point Slot Table"
::= { accessPointsSatus 2}

accessPointEntry IA-OBJECT-TYPE              -- Instance
                SYNTAX  AccessPointEntry,
                ACCESS  not-accessible
                STATUS  currrent
                DESCRIPTION "An entry in accessPoint"
          INDEX {accessPointNbr, … }
::= { accessPoint 1 }

AccessPointEntry :=SEQUENCE{  -- SEQUENCE statement
       accessPointNbr                Unsigned32, -- index 1
       accessPointStatus              INTEGER,
        …
      accessPointLinkDown   LinkDown_Rule   -- index 4}
```

The expert rule used in the SNMP IA specification is *accessPoinLinkDown*. This expert rule is used to capture and detect anomalies or defects of operations produced in the access point device and suggest the necessary measures for solving the problem. When a mistake occurs, the rule goes to the agenda system. This rule is fired when the conditions are right: “The physical link on one of the switch (controller) ports is down”. The rule provides recommendations on how to solve the failures.

```
Link_Down_Action ACTION ::= CLASS {CODE 12,
    ARGUMENT  Port "{0}" is down on Switch "{1}."
     RETURN “Troubleshoot physical network connectivity to the
affected port”}
```

Using this methodology, we can define all the knowledge management in a specific domain network and add these new types in a MB module. The module will constitute the complete knowledge specification of the management network.

V. Development Environment

Whit the purpose of improving insight and understanding of network management, we have developed a system named ALATUS based on SNMP IAs. The following system exemplifies how network topology information, resources properties and management information may be used to develop dynamically an intelligent diagnostic when errors occur.

We have studied an example of alarm detection and incident resolution concerning a private wireless network of the University of Seville. It provides wireless access using this technology in all of its departments where ReInUS (Radio network of University of Seville) facilitates access to connection. ReInUS allows the University community to connect to its network using WiFi technology: libraries, classrooms, departments, leisure rooms, open spaces, etc. Today, the University of Seville has 1200 wireless access points distributed in five campus. Every access point provides different capabilities to maximize wireless LAN performance, security, reliability and scalability.

ALATUS system works together with the access points providing real-time monitoring, management functions and supporting simultaneous data forwarding. The intelligent alarm management system will not just report that there is a problem, but the location of the problem, provide instructions for corrective action and correct the situation automatically. Fault identification involves testing the hypothetical faulty components and repair by taking intelligent actions. Advanced features just described can make the difference between a minor incident and a higher downtime. In ALATUS system, an IA agent works collecting information from the resources, in order to detect the network anomalies that typically precede a fault. It creates knowledge, about the network node, stored in MIB, which saves the management knowledge and a set of variables related to that node in particular, figure 5.

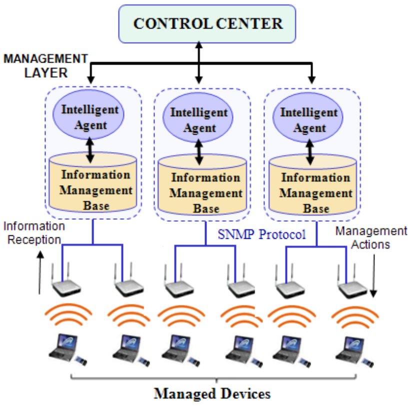


Figure 5. Wireless Network Architecture

The runtime system is defined as an SNMP application, according to the SNMPv3 architecture. The system has two major components [14]:

- Graphic Unser Interface (GUI): In our system, we have implemented a GUI written in Java running on a server who controls the whole embedded system. It is a set of I/O handling routines for managing the system and allows the management of the system by the user. To access the interface only requires a Web browser such as Explorer o Mozilla. The GUI controls the inference engine and manages system input and output.
- Inference engine. The distributed IA may be decomposed in two blocks: the SNMP entity, which implements the MIB, and the runtime system, capable of executing the scripts. Inference engine sees the runtime system as a set of distributed IAs. IAs emphasize autonomy and learning with the purpose of performing management tasks in the managed resources. In order to distribute intelligence in the network, ALATUS has been developed according to rule-based expert systems technique. This knowledge representation technique has played an important role in modern intelligent systems and their applications in strategic goal like setting, planning, design, scheduling, fault monitoring, diagnosis and so on. Conventional rule-based expert systems use human expert knowledge to solve real-world problems that normally would require human intelligence [15]. The management operations are modeled as scripts written in an intelligent language like CLIPS (C Language Integrated Production System) and associated with the MIB variables of a specific resource. CLIPS is a public domain software tool which provides a complete environment to build rules and/or objects based on expert systems.

A. *System Validation*

Validation constitutes an inherent part of the knowledge based expert system development and is intrinsically linked to the development cycle. ALATUS has been validated with respect to the following aspects: system validation using test cases, validation on site and validation against human experts. To verify the system, we have feed it with arbitrary amount of real alarms at random for more than one year. We have analyzed the evolution of the incidents from April 2011 to July 2012. We can appreciate a negative trend in the number of incidents resolved and how ALATUS system improves this negative trend with its implementation since October 2011. The result of this analysis is included on Table 1.

TABLE I. Prototype Testing Anakysis

	Apr.11	Jul.11	Oct.11	Jan.12	Apr.12	Jul.12
Initial Events	458	543	745	879	907	947
Resolv. Incidents	358	426	558	681	700	786
Warning to Oper.	311	346	498	305	201	101

Overall, figure 6 reflects an increase of total events, which causes a higher number of incidents in the university wireless network. But as time progresses, we also observe an increase of solve alarms and less alarms to operator.

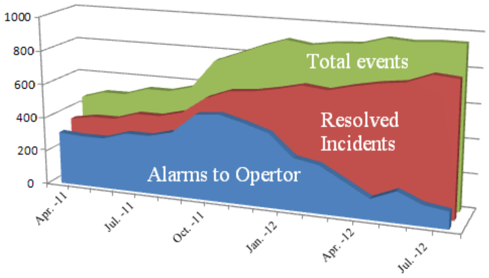


Figure 6. Prototype results

From these result, it can be said that ALATUS system reduces the need for additional dedicated monitoring software, thereby the cost and complexity of WLAN I reduced. This solution has produced excellent results which, after extensive field-testing, has proved to be able to filter 95% of produced alarms with a precision of 93% in locating, with about a 96% rate of success.

VI. CONCLUSION AND FUTURE WORKS

In this study, we have proposed that IA technologies are a future leading solution to face the current management network problems. We have seen that actual management systems are not able to solve questions explained in the initial parts of this paper. Until now, the managed objects are not able to use the information that the base of knowledge collects from management operations. The necessary requirements of area analysed must undertake those aspects.

This paper introduces an original contribution to include management knowledge coming from the network devices themselves into the specifications of the MIB. To formalize the main proposal of this work, a language to describe the knowledge base descriptions in Internet management network has been introduced. A number of questions raised from the design of a language have been discussed, and a general framework for the inclusion of formal knowledge management in SMI has also been introduced.

This research has showed an useful and interesting modular approach in the development of a knowledge based integrated expert system which can be quite powerful in tackling the huge and wide subject on diagnosis of common problems in management network. An integrated knowledge process is developed to guarantee the whole engineering procedure which uses expert rules as knowledge representation technique. This demonstrated that ALATUS is capable to specify the knowledge of a reasonably sized information model.

In our work, where knowledge is brought onto networks based on Internet model, can be reviewed like a first step toward automated management by using intelligent agents on SNMP. One guideline of our future work is to improve the agent's performance. We are also studying deeply how to incorporate the previous knowledge available at a network node. In that sense, we plan to get further investigating the feasibility and limitations of other knowledge representation techniques such as semantic networks, Bayesian networks and ontology engineering. In addition to the fault detection functional area, currently we are also studying possibilities of expanding the scope of our successful tool to other functional areas such as accounting, configuration, performance and security management.

REFERENCES

[1] N. Agoulmine, “Autonomic network management principles: from concepts to applications”, Burlington, Academic Press/Elsevier, 2011

[2] Z. Wang, Y. Wang & G. Shao, “Research and Design of Network Servers Monitoring System Based on SNMP”, First International Workshop on Education Technology and Computer Science, 2009

[3] C. Zhu, L. Shu, T. Hara, L. Wang, S. Nishio & L.T. Yang, “A survey on communication and data management issues in mobile sensor networks”, Wireless Communications and Mobile Computing, Wirel. Commun. Mob. Comput., John Wiley & Sons, Ltd, 2011

[4] A. Martín, C. León, J. Luque & I. Monedero, “A framework for development of integrated intelligent knowledge for management of telecommunication networks”, Expert Systems with Applications, Volume 39, Issue 10, Pages 9264-9274, 2012.

[5] S. Yang, Y. Chang, “An active and intelligent network management system with ontology-based and multi-agent techniques”, Expert Systems with Applications, Volume 38, Issue 8, 2011.

[6] H. Xu & Debao Xiao, “A Common Ontology-Based Intelligent Configuration Management Model for IP Network Devices”, First International Conference on Innovative Computing, Information and Control - Volume I (ICICIC'06), 2006

[7] J. Ding, “Advances in network management”, Boca Raton: CRC Press, 2010

[8] D. Nikolaou & T. Wolf, “Architecture of network systems”, Burlington, Morgan Kaufmann, 2011

[9] J. Larmouth, “ASN.1 complete,” San Francisco, Cal. Morgan Kaufmann, 2000

[10] L. Walsh, “SNMP MIB Handbook: essential guide to MIB development, use and diagnosis” Stanwood, WA : Wyndham, 2008

[11] A. Pedai, I Astrov, "Multi-Rate Expert Systems in Supply Chain Simulation for Telecommunication Industry," Wireless Communications, Networking and Mobile Computing, 2008. WiCOM '08. 4th International Conference on , 2008

[12] J. Liebowitz, “Knowledge management handbook, collaboration and social networking” Boca Raton, Fla. : CRC Press, 2012

[13] S. D. J. McArthur, S. M. Strachan, and G. Jahn, “The design of a multiagent transformer condition monitoring system,” IEEE Trans. Power Syst., vol. 19, no. 4, pp. 1845–1852, Nov. 2004.

[14] S. Sahin, M.R. Tolun, Y. Baykal, "Expert System for Access Telecommunication Networks," Computer Science and Information Engineering, 2009 WRI World Congress on , vol.4, no., 2009

[15] A. Lezhenko, S. Kuznetsov, I Kuznetsov, "Techniques and Methods for "Smart" Processing of Information Flows in Applications of Information and Telecommunication Technologies: The Use of Expert Systems for Data Mining," Cyberworlds (CW), 2011 International Conference on , vol., no., pp.168-172, 4-6 Oct. 2011

Mass Configuration of Network Devices in Industrial Environments

György Kálmán
ABB Corporate Research
Norway
gyorgy.kalman@no.abb.com

***Abstract*—Industrial Ethernet offers greater flexibility and potentially lower deployment costs than traditional fieldbuses. Although similar, the configuration and engineering of these networks need considerable effort. This paper presents the lessons learned from an approach to bulk configuration of an Ethernet infrastructure for industrial applications. The paper presents different approaches and decisions taken for the proof of concept implementation. The paper gives an overview about the issues related to representation and generation of configuration data, support of multiple vendors in the engineering phase and also during operation. An outlook for possible improvements and promising features is given.**

***Index Terms*—industrial Ethernet; infrastructure; switch; configuration; life cycle; multi vendor**

I. Introduction

A modern industrial communication system contains a considerable amount of nodes interconnected with Ethernet and current trends point towards moving the Ethernet connectivity down to instrument level. Having an all-Ethernet infrastructure offers several advantages over traditional fieldbus-based or Ethernet-fieldbus mixed networks. These include simpler deployment by using the same connectors and wires over the whole network, ample bandwidth, wide range of communication hardware and easy connectivity towards office networks or the internet.

One of the main drawbacks is a result of the inherently different network topology compared to office environments. In industry, the bus-like structure has proven to reduce costs with cutting cabling need. In these scenarios, the backbone is usually composed as a ring and the devices, subnetworks or other devices are connected to this with small switches (up to approx. 10 ports).

As the networks are built with mainly these small switches, the whole installation typically contains a magnitude more devices than a comparable office network (e.g.¨a bigger refinery can have several hundreds). During engineering and Factory Acceptance Test (FAT), the effort of configuring these devices is high and severely influences the competetiveness. In the majority of cases, the actual configuration of the devices can be described with setting port-VLAN allocations, RSTP priorities, SNMP parameters and performance monitoring. These steps currently require manual work, which is incresing cost during engineering and also leads to increased resource usage during FAT as configuration errors may happen.

The use of small switches in addition results in issues associated with e.g. Quality of Service (QoS) and management.

A. Topology

A key area, where industrial networks do differ considerably from their office counterparts is the topology used. In an office environment the network is structured to resemble an equalized tree as much as possible. Also, high port density switches are used to lower the hierarchy levels in the network.

The industrial environment, as stated earlier, resembles more a bus-like topology. Ring-based redundancy solutions [4], traditional planning and cabling cost both force network engineering towards the use of rings as backbone and small switches to connect the few nodes which are located close.

Ring structures are beneficial for redundancy, but are problematic for traffic engineering. These rings aggregate traffic and force longer paths in the network than in a comparable office counterpart.

B. Network segmentation

The traffic aggregation of rings do cause other issues too, especially if multi- or broadcast traffic is involved. In a typical installation, several industrial protocols are in use. In order to reach a more stable network and avoid that nodes are receiving unnecessary traffic, these networks are often segmented into several Virtual LANs (VLANs).

C. Configuration and Maintenance

Current industry practice builds on a detailed network drawing and unit-to-unit configuration of the network devices as part of the deployment. Here in most cases the built-in web configuration solutions of the different vendors are used, although some provide their tools for own product lines e.g. Hirschmann HiVision or Cisco Network Assistant, which support configuration of multiple units.

From the engineering viewpoint, setting up these devices one-by-one is a great risk, as the chance of human error is high. This risk is mitigated with additional resources, meaning more work hours to check the actual setup [5].

From maintenance viewpoint, this situation is even worse. Most installations have a long life expectancy and therefore future maintenance engineers will either face 10-15 year old web interfaces if they have to modify something or the the problems associated with replacing the old device and migrating the configuration to a new one.

ISBN: 978-1-61208-245-5

II. Motivation

The motivation behind this work was to reduce engineering costs and to explore possible solutions for provider-independent configuration representation and setup of multiple devices at the same time [6], [7]. The potential cost reduction in the engineering phase is expected to reach 20-25% of the total cost, not counting the life cycle support.

The review of a project portfolio revealed that in most installations 2-3 vendors are involved in supplying network infrastructure based on various preferences. Although the planning of the network is done independently from the actual manufacturers, the configuration and acceptance checks do depend on per vendor knowledge and tools.

The expected result of the research task was in addition to explore possible solutions, to create a proof-of-concept tool, which can compose, deploy and modify configuration of one or multiple Ethernet switches in the same work session.

In long-term, the vision of a common configuration and management tool was defined, where planning, configuration, as-planned checking, monitoring and life cycle management was provided. Such a tool could offer a common interface to plan a network with defining the segmentation and port distribution (this covers the current network drawing step), generate configuration for the devices (which is done typically by engineers), deploy and then through discovery, check that the network has the same structure as planned (for example the VLANs are set up correctly). During operation, the tool could read out the current configuration from a device and upload it to a replacement unit, even if these are from different manufacturers.

III. Background

To explore configuration possibilities, remote configuration features of selected product lines were reviewed:

- *RuggedCom RS9xx* [2]: This switch line supports configuration update using a builtin Trivial FTP (TFTP) client or server, depending on requirements. In addition, Secure Copy (SCP), terminal with Command Line Interface (CLI) and Simple Network Management Protocol (SNMP) is supported for file and configuration manipulation. As all of the reviewed managed switches, this unit offers a web interface. A vendor-specific tool for monitoring is available[1].
- *Hirschmann RSRxx* [1]: This switch line offers a TFTP Client, CLI access through telnet or the web interface, a java-based web interface, SCP file transfers and a proprietary Automatic Configuration Adapter. This adapter, if physically connected to the device, uploads or downloads configuration enabling easy replacement from the same vendor.
- *Moxa EDS-508* [3]: Has TFTP server and client, CLI, SCP transfers and offers a web interface. A proprietary Auto-Backup Configurator is offered for backup and restore, allowing easy replacement from the same vendor.

The research also showed that SNMP is supported on all units, although the features were focused on monitoring and not on configuration.

The review showed considerable differences between web interface structures and the available options. The differences were big enough to limit reuse of configuration knowledge and proved to support the initial assumption about cost reduction potential.

Configuration data was accessible on all devices as structured text files, which were human readable and could be a base for the configuration tool design. In figure 1, the expected coverage of a configuration tool is shown. The objective was to allow up- and downloading, manipulation and storage of configuration information.

A. Multiple unit configuration

One of the most important features was to check the feasibility of configuring multiple units in the same time and to explore the possible issues.

As part of the planning, a feature set was identified, which were set the same on all devices or could be calculated automatically. An example is the selection of the Rapid Spanning Tree (RSTP) root bridge.

Other questions were risen in connection with the long paths and rings used in these networks. It was assumed, that depending on the behavior of the devices, the configuration might need to be topology aware.

The user interface was also a crucial point, as the objective was to reduce engineering cost, which pointed towards a simpler interface then most of the switches offered. This request was supported by, that only a handful of features needed to be set and most of the parameters were left at factory defaults.

IV. Proof-of-Concept

The implementation was focused on a subset of the possible features. Based on feedback from engineering, configuration of multiple devices and support of multiple vendors were selected as key features, which should be supported by a simple user interface. In figures 2 and 3, the test user interface is shown for single- and multi-unit mode.

The planned system was designed to cover tasks associated with configuration and deployment stages of the engineering process. To ensure that options, which are not being used by the system are preserved, the tool only replaces relevant parts of the original configuration files with new data (figure 4).

A. Requirements

- vendor independent, simple user interface
- remote configuration of one or multiple devices
- life cycle support with configuration versioning and cloning
- configure selected features

[1] RuggedNMS

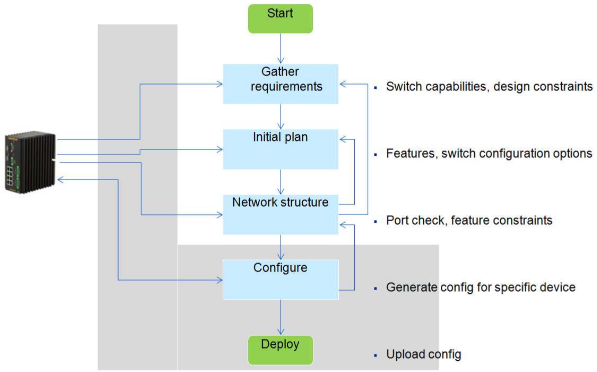
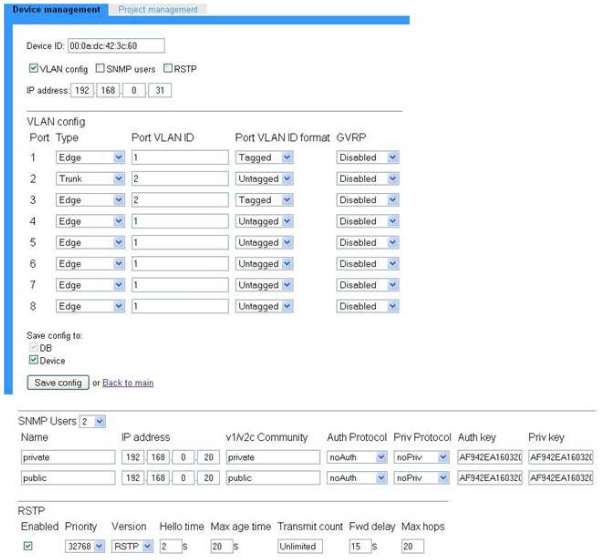

Start
Gather requirements
Initial plan
Network structure
Configure
Deploy
Switch capabilities, design constraints
Features, switch configuration options
Port check, feature constraints
Generate config for specific device
Upload config

Fig. 1. Proof of concept coverage



Device ID: 00:0a:dc:42:3c:60

☑ VLAN config ☐ SNMP users ☐ RSTP

IP address: 192 . 168 . 0 . 31

VLAN config

Port	Type	Port VLAN ID	Port VLAN ID format	GVRP
1	Edge	1	Tagged	Disabled
2	Trunk	2	Untagged	Disabled
3	Edge	2	Tagged	Disabled
4	Edge	1	Untagged	Disabled
5	Edge	1	Untagged	Disabled
6	Edge	1	Untagged	Disabled
7	Edge	1	Untagged	Disabled
8	Edge	1	Untagged	Disabled

Save config to:
☑ DB
☑ Device

Save config or Back to main

SNMP Users 2

Name	IP address	v1/v2c Community	Auth Protocol	Priv Protocol	Auth key	Priv key
private	192 . 168 . 0 . 20	private	noAuth	noPriv	AF942EA16032(	AF942EA16032(
public	192 . 168 . 0 . 20	public	noAuth	noPriv	AF942EA16032(	AF942EA16032(

RSTP

Enabled	Priority	Version	Hello time	Max age time	Transmit count	Fwd delay	Max hops
☑	32768	RSTP	2 s	20 s	Unlimited	15 s	20

Fig. 2. Single unit configuration

B. Features

A subset of available features on the switches was selected based on experiences from engineering. This set was planned to cover most of the engineering needs without resulting in a complex interface.

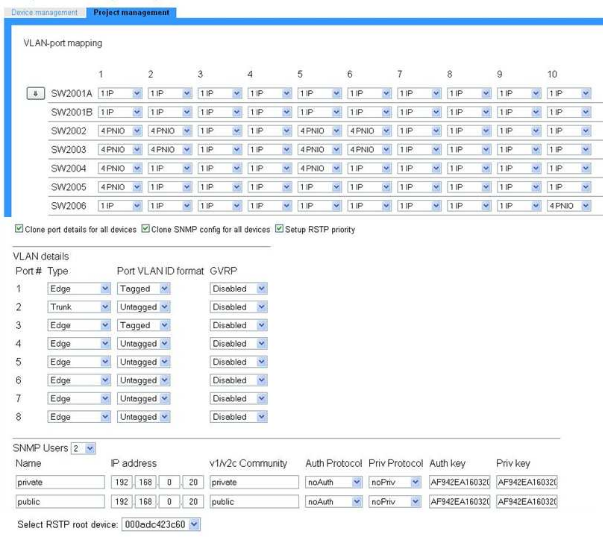
The feature set was defined for both single and multi mode as:

- *Host IP*: to be able to set the device's IP
- *Port-based VLAN*: allow the setup of per-port VLANs
- *SNMP setup*: configure SNMP access rights and community memberships
- *Spanning Tree*: select STP protocol and allow changes in bridge priority

To support documentation, an automatic network documentation generator function was also included.

A single unit configuration section was included for practical purposes and served also as a testbed for checking the configuration generation capabilities.

The system was designed so that it would preserve changes made outside the configuration tool (thus allowing device specific configuration for features not covered by the tool), so the composition of the configuration data was implemented in a way, that it is only changing the relevant part and keeps the rest of the data untouched.



VLAN-port mapping

	1	2	3	4	5	6	7	8	9	10
SW2001A	1 IP	1 IP	1 IP	1 IP	1 IP	1 IP	1 IP	1 IP	1 IP	1 IP
SW2001B	1 IP	1 IP	1 IP	1 IP	1 IP	1 IP	1 IP	1 IP	1 IP	1 IP
SW2002	4 PNIO	4 PNIO	1 IP	1 IP	4 PNIO	4 PNIO	1 IP	1 IP	1 IP	1 IP
SW2003	4 PNIO	4 PNIO	1 IP	1 IP	4 PNIO	4 PNIO	1 IP	1 IP	1 IP	1 IP
SW2004	4 PNIO	1 IP	1 IP	1 IP	4 PNIO	1 IP	1 IP	1 IP	1 IP	1 IP
SW2005	4 PNIO	1 IP	1 IP	1 IP	1 IP	1 IP	1 IP	1 IP	1 IP	1 IP
SW2006	1 IP	1 IP	1 IP	1 IP	1 IP	1 IP	1 IP	1 IP	1 IP	4 PNIO

☑ Clone port details for all devices ☑ Clone SNMP config for all devices ☑ Setup RSTP priority

VLAN details

Port #	Type	Port VLAN ID format	GVRP
1	Edge	Tagged	Disabled
2	Trunk	Untagged	Disabled
3	Edge	Tagged	Disabled
4	Edge	Untagged	Disabled
5	Edge	Untagged	Disabled
6	Edge	Untagged	Disabled
7	Edge	Untagged	Disabled
8	Edge	Untagged	Disabled

SNMP Users 2

Name	IP address	v1/v2c Community	Auth Protocol	Priv Protocol	Auth key	Priv key
private	192 . 168 . 0 . 20	private	noAuth	noPriv	AF942EA16032(	AF942EA16032(
public	192 . 168 . 0 . 20	public	noAuth	noPriv	AF942EA16032(	AF942EA16032(

Select RSTP root device: 000adc423c60

Fig. 3. Multiple unit configuration

C. Multiple vendor support

Enabling support for multiple vendors has risen several issues, which were not foreseen. Even if all the switches covered were complying the same IEEE standards, the actual implementation and availability of features depends on the vendor.

As a result, a vendor independent representation of the configuration data was needed and the configuration generation process had to be split into storage, representation and actual configuration data.

D. Multiple device support

Configuring multiple devices in one session was considered as the most important feature, as this would result in the highest cost cut. Covering multiple devices also meant, that the difference between the per unit web interfaces and the configuration tool might be the most emphasized.

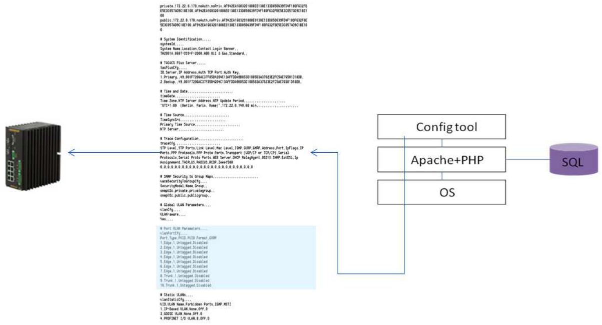


Fig. 4. Composition of the configuration

For the IP configuration and VLAN settings, a matrix of switches and VLANs was generated. This offers a single-screen overview of a typical network in the evaluated projects. The drawback of this representation is, that if a large number of ports and switches are used, the size of the matrix is getting large. This limitation was found acceptable in this case, as in a typical industrial environment low port count switches are used, so adding more switches will result in a longer matrix, but the width will stay limited.

The tool offers cloning of the port and SNMP settings to all devices and setting the root RSTP bridge.

E. Connectivity

The review of connectivity methods has shown, that it is problematic to choose one specific solution. Even in case of just the three product lines reviewed, different protocols turned out to be easier to use.

For the proof of concept, for one vendor (RuggedCom) TFTP was chosen for up- and downloading configuration data. For the other vendor (Hirschmann), CLI-based configuration and telnet. While being aware, that none of these protocols provide secure transfers, this requirement was relaxed for the current version. This decision was supported by that the tool is intended to be used during engineering, where these networks operate as isolated islands.

V. Lessons learned

There are several important issues which were identified during the evaluation and development of the configuration tool.

A. Vendor independent configuration data

In order to support multiple vendors, the configuration data needs to be stored in an independent format. Generation of the appropriate configuration file or script depends on the vendor's implementation and there might be considerable differences.

Changes between vendors in most of the cases results in information loss about the configuration of the device. An example is the support of vendor specific spanning tree protocols. The use of these proprietary protocols is beneficial if the network is homogenous, but might cause problems if multiple vendors are present. If the original configuration was set up e.g. to use RuggedCom's eRSTP and the device is replaced with an other manufacturer's switch, the configuration tool has to fall back on e.g. RSTP, as that is the nearest standard protocol which is supported by the new device.

If later the device is changed back (e.g. a device needed to be taken out from the network and was temporarly replaced by an other), if the configuration storage depends on the vendor, then only RSTP will be used even if eRSTP is available, as the migration process will only create a representation of the current configuration in the new device.

B. Topology-awareness

An interesting issue with configuration was raised while the tests of the multiple unit configuration were executed. In single unit mode, were no problems, the configuration was updated, the device was reset and after some seconds, network operation was restored. The same happened if multiple units were configured in an office-like topology (equalized tree), where only a few levels of switches were involved and the longest path was 3-4 hops. In case of industry-typical rings, anomalies and connectivity errors happened.

The investigation showed, that while the update operation itself is done in a fraction of a second and it takes approximately 2-3 seconds for a device to reset, this was too short to update all members of the ring. In the tree topology, the devices were updated before the first unit decided to reset. In the ring, however, these resets happened before all members were updated. The result was that the network was falling into fractions and in some cases one had to approach each *lost* device separately.

As a result, it was identified, that it would be beneficial if in case of multiple unit configuration, the update would be done with respect to the topology, starting from the leaves and progressing upwards in the tree. The same approach can be used in rings, as these will be represented as a long unbalanced tree (in normal operation RSTP is disabling the redundant link to avoid a loop).

C. Identical configurations

Although the switches used in this work were not the most complex units available, it turned out to be a complicated task to reach exactly the same configuration on devices from different vendors.

A typical example is the configuration of a trunk port. In one case, this option was available directly on the webinterface and in the configuration file, but on a different switch at least 6 commands in the CLI were required.

An other example is the above mentioned case of RSTP. In practice, all major vendors have their own enhancements to RSTP to achieve better convergence times. This also means, that these proprietary solutions can only used on homogenous fractions of the network. If a device is replaced by a device from a different vendor can result in weaker performance, as all the units have to fall back on the first standard solution, which in most cases will be RSTP.

VI. Future work

Future work will focus on topology awareness to mitigate the restart anomalies. By having a representation of the network, also as planned checks could be executed and discovery of previously unknown networks can be enabled. This functionality extends the usage area of the tool towards the network management systems.

Security is an important topic and for the current stage of the work, no emphasis was put on this field. While some possible threats were identified, currently leaving a device open for remote configuration does in most cases result in vulnerabilities. A possible area of research here is how to provide an easy to use interface while having a detailed logging system and secure connections to the devices with respect to the limited possibilities.

VII. Conclusion

The background research of this paper has shown, that there is a considerable potential to cut costs in network engineering if appropriate tools are available. Although network management software are available and widely used in office environments, their resource need and cost render them unrealistic for industrial deployment.

The paper has shown a proof of concept implementation of a configuration tool, which can partially automate the setup of Ethernet switches. The main difference compared to proprietary solutions is, that this tool supports multiple vendors and with a vendor independent representation of configuration data, also allows future extensions.

Testing of the tool revealed several issues associated with device configuration, especially related to problems caused by the topology and the complexity of generating identical configurations for switches from different vendors.

References

[1] Hirschmann, *Reference Manual, Command Line Interface Industrial Ethernet Gigabit Switch* Release 7.0, Hirschmann, 2011.

[2] RuggedCom, *Rugged Operating System v3.10 User Guide*, Ruggedcom, 2012. January 19.

[3] Moxa, *Datasheet, EDS-508*, Moxa, 2010. May 5.

[4] Kleineberg et.al., *Automatic device configuration for Ethernet ring redundancy protocols*, ETFA 2009.

[5] Rojas, C.; Morell, P., *Guidelines for Industrial Ethernet infrastructure implementation: A control engineer's guide*, IEEE IAS/PCA 2010.

[6] Imtiaz et.al., *A novel method for auto configuration of Realtime Ethernet Networks*, ETFA 2008.

[7] Reinhart et.al., *Automatic Configuration (Plug & Produce) of Industrial Ethernet Networks*, INDUSCON 2010.

Adaptive Traffic Dependent Fuzzy-based Vertical Handover for Wireless Mobile Networks

Thanachai Thumthawatworn and Anjum Pervez
Faculty of Engineering, Science and the
Built Environment
London South Bank University
London, United Kingdom
{thumthat, perveza}@lsbu.ac.uk

Pratit Santiprabhob
Intelligent Systems Research Laboratory
Faculty of Science and Technology
Assumption University
Bangkok, Thailand
pratit@scitech.au.edu

***Abstract*—An intelligent handover decision system is necessary for heterogeneous wireless mobile networks to fulfill user's expectations in terms of the quality of services. With emerging real-time services, including multiple QoS parameters in handover decision process seems essential. In this paper, fuzzy logic is applied to enhance the intelligence of the handover decision engine. An adaptive traffic dependent fuzzy-based handover decision system (ATD-HDS), which employs multiple decision engines each optimized to a specific traffic type, is presented. The results show that, compared to a monolithic fuzzy-based handover decision system, the proposed ATD-HDS significantly improves the decision quality and algorithm execution time.**

***Keywords*-fuzzy logic; handover; traffic dependent; adaptive; wireless; heterogeneous**

I. Introduction

Heterogeneous wireless mobile networks require interconnections of diverse wireless technologies such as WLAN, WiMAX and Cellular mobile networks as illustrated in Fig. 1. Mobile users expect seamless services over a wide area of mobility, with adequate quality and favourable price. In order to satisfy the above requirements, multiple handovers often become necessary. A handover may take place in a homogeneous network environment (horizontal handover) or in a heterogeneous network environment (vertical handover). In either case some form of decision mechanism needs to exit within the mobile device.

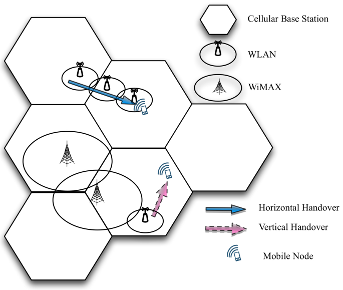


Fig. 1. Architecture of Heterogeneous Wireless Networks

A horizontal handover decision [1] is normally a straightforward process as the decision can simply be based on the received signal strength (RSS). However, due to varied characteristics of different wireless networks, a simple RSS based decision cannot achieve the required results in a vertical handover decision process. Clearly there is a need for a much more intelligent handover decision system (HDS) in heterogeneous network environment [2].

Numerous fuzzy logic based solutions, which enhance intelligence for vertical handovers, have been presented in the literature [3], [4]. However, in most of the existing work only a limited number of decision parameters are considered. This restriction seems to be due to the fact that as the number of decision parameters increases, the number of fuzzy rules increases significantly, which leads to computational complexity and very long execution time.

Nevertheless, for a more realistic evaluation of a vertical HDS, an increased number of decision parameters must be considered. Furthermore, due to the growing demand for real-time services (VoIP, video streaming, etc.), the decision parameters concerned with the QoS requirements (latency, jitter and packet loss) are an essential part of this work.

In this paper, we are presenting an adaptive traffic dependent fuzzy-based HDS. The HDS consists of three dedicated decision engines; each optimized to a given traffic stream. The traffic streams assumed are: Constant Bit Rate (CBR), Variable Bit Rate (VBR) and Available Bit Rate (ABR). The fact that each traffic type has different QoS requirements has been taken into account in the design of decision engines. In doing so, the total number of fuzzy rules have been reduced.

The performance of the proposed approach is compared, in terms of the decision quality and execution time, with a conventional monolithic fuzzy-based HDS design and Simple Additive Weighting (SAW). Simulation results show an improvement of over 39% in the handover performance and a reduction of up to 90% in algorithm execution time in certain scenarios.

The paper is organized as follows. The related work is presented in section 2. Section 3 presents a monolithic fuzzy-based HDS. In section 4, an adaptive traffic dependent fuzzy-based HDS is presented. Handover decision system designs are given in section 5. Section 6 gives simulation results and comparison between different HDS designs. Conclusions and future work is given in section 7.

II. Related Work

As has been stated previously, numerous fuzzy-based solutions for vertical handover decision systems have been proposed in the literature. A fuzzy-based vertical handover decision algorithm, which assumes interconnection between WLAN and WMAN, is proposed in [5]. The decision parameters considered are: RSS, data rate, usage cost and user preference. The main aim of this work is to minimize the number of handovers and the results presented are encouraging.

In a more recent work [6], minimization of the number of handovers is considered whilst assuming RSS, data rate and usage cost as the primary decision parameters. The results show that the proposed algorithm can dramatically reduce the total number of handovers.

A fuzzy-based handover decision for interconnection between WLAN and WiMAX is proposed in [7]. The decision parameters considered are: RSS, data rate, and distance. The main aim of this work is to minimize the percentage packet loss, which is achieved successfully.

In all the above solutions, only the data rate is assumed to be the QoS related decision parameter. However, recognizing the importance of including other QoS parameters such as latency, jitter and packet loss in the decision process, a great deal of effort has been directed to evaluate the performance of a HDS in the presence of multiple QoS parameters.

In [8], [9], bit error rate (BER) and RSS are considered in their fuzzy-related decision algorithm. The results show improvement in terms of the number of handover reduction. In [10], a fuzzy-based vertical handover algorithm taking data rate, delay and BER (along with other parameters such as cost and security) into consideration is proposed. The algorithm improves the process of wireless network selection, thus avoiding unnecessary handovers.

Authors in [11] have proposed a QoS aware fuzzy rule based vertical handover mechanism that considers data rate, latency, jitter and BER. The proposed work is found to be effective for selecting a wireless network that meets the requirements of different applications. The results show a reduction in average end-to-end delay and yield a moderate average bandwidth.

It seems that although it is important to extend the number of decision parameters (which must include the QoS parameters), it is often not done due to computational complexity, which results in unacceptably long execution time. Thus, a new approach is needed that allows an extended number of decision parameters to be included, considers QoS and minimizes the execution time.

III. Monolithic Fuzzy-based HDS

A. Architecture of Fuzzy System

The architecture of a fuzzy system is shown in Fig. 2. It comprises four components. Fuzzifier converts crisp inputs into fuzzified data. Rule base contains IF-THEN rules, which are required by the Fuzzy Inference System (FIS). FIS generates aggregated fuzzified data, based on fuzzy inference method used. Defuzzifier converts the aggregated fuzzified data into a scalar value (score). The score is then used by the application.

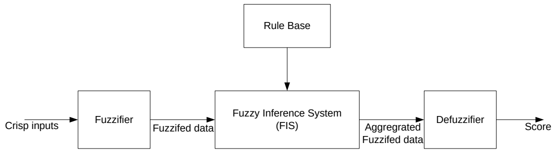


Fig. 2. Architecture of Fuzzy System

B. Development of Monolithic Fuzzy-based HDS

In this study we have taken six decision parameters: data rate (DR), usage price (PR), battery life (BA), latency (LA), jitter (JI) and packet loss (PL). The corresponding input fuzzy sets are denoted by $\widetilde{DR}$, $\widetilde{PR}$, $\widetilde{BA}$, $\widetilde{LA}$, $\widetilde{JI}$ and $\widetilde{PL}$.

Each fuzzy set has three fuzzy memberships (low, medium and high). With this combination the total number of rules = 3^6 = 729. Each rule is then assigned a decision output, which is based on expert knowledge. This process formulates an output fuzzy set, $\tilde{Z}$, which contains seven fuzzy memberships defined as: very low (VL), low (L), medium-low (ML), medium (M), medium-high (MH), high (H) and very high (VH)). Triangular functions are used to express the fuzzy memberships in both input and output fuzzy sets.

The crisp inputs (the values for each of the six parameters offered by the mobile node and individual wireless networks within heterogeneous network environment) are fuzzified and provided to FIS. There are two well-known fuzzy inference systems, namely, Mamdani [12] and Sugeno [13]. However, Mamdani FIS is used in this work as it is known to be well suited to capture expert knowledge [14].

The aggregated fuzzified data, $\mu\tilde{Z}_{mono}$, is given by:

$$\begin{aligned}\mu\tilde{Z}_{mono}(y) =& max_k[min[\mu\widetilde{DR}^k(datarate),\\ & \mu\widetilde{LA}^k(latency), \mu\widetilde{JI}^k(jitter),\\ & \mu\widetilde{PL}^k(packetloss), \mu\widetilde{PR}^k(price),\\ & \mu\widetilde{BA}^k(battery)]],\\ & fork = 1, 2, 3, \ldots, 729 \end{aligned} \tag{1}$$

where k is the total number of rules.

Defuzzifier then converts the aggregated fuzzified data into crisp value (score). The final score, $Score_{mono}$, is calculated using a centroid method given by:

$$Score_{mono} = \frac{\int \mu\tilde{Z}_{mono}(y).ydy}{\int \mu\tilde{Z}_{mono}(y)dy} \tag{2}$$

This score is then used to make the handover decision.

IV. Adaptive Traffic Dependent Fuzzy-based HDS

From the above work, we note that extending the number of decision parameters to six (with three memberships), a monolithic decision engine generates 729 rules. This raises the question of execution time. Furthermore, the membership functions used are fixed for all types of traffic streams.

In order to deal with the above two issues, we are proposing a new adaptive traffic dependent fuzzy-based HDS (ATD-HDS), in which multiple decision engines are employed, the number of rules is reduced by considering the QoS requirements [15] for each traffic type and the FMFs are tailored to match the characteristics of the incoming traffic.

The system consists of three dedicated fuzzy-based decision engines, each matched to one of the three traffic types, namely, CBR, VBR and ABR. The Engine Selector (ES) first identifies the traffic type and then selects the corresponding decision engine to carry out the network selection process. The general architecture is shown in Fig. 3

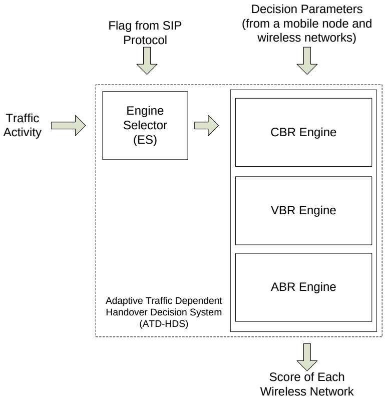

Flag from SIP Protocol
Decision Parameters (from a mobile node and wireless networks)
Traffic Activity
Engine Selector (ES)
CBR Engine
VBR Engine
ABR Engine
Adaptive Traffic Dependent Handover Decision System (ATD-HDS)
Score of Each Wireless Network

Fig. 3. Architecture of ATD-HDS

The ES periodically sniffs incoming packets with sufficient frequency to detect traffic activity. The traffic type is identified by receiving a flag from the application layer. This can be obtained from a commonly used session initiation protocol (SIP), which runs at the application layer. SIP has the ability to differentiate between CBR and VBR traffics. Thus, the ES selects one of the three engines using the following logic (as shown in Fig. 4):

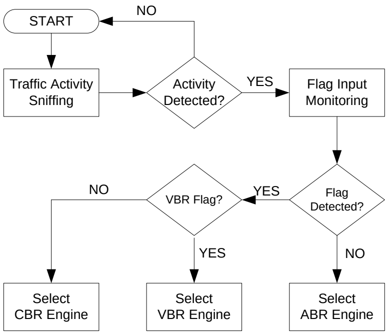


Fig. 4. Logic of Engine Selection

- Traffic activity is present and CBR flag is received - select CBR engine.
- Traffic activity is present and VBR flag is received - select VBR engine.
- Traffic activity is present but NO flag is received - select ABR engine

The number of rules is reduced by including latency, jitter and packet loss for CBR engine, latency and packet loss for the VBR engine and only packet loss for ABR decision engine. By this matching process the total number of rules becomes 729, 243 and 81 for CBR, VBR and ABR decision engines respectively.

The quality of decision is enhanced by using a combination of triangular and trapezoidal functions to express the fuzzy memberships. Furthermore, the fuzzy membership functions (FMFs) are tailored according to the QoS requirements of each traffic type.

V. Handover Decision System Designs

Three HDS designs are produced: Monolithic design 1 (MD1) , Monolithic design 2 (MD2) and ATD design.

MD1 is a conventional design with 6 decision parameters and all FMFs are triangular, with no regard to the incoming traffic type. MD2 has 6 decision parameters but the FMFs are a combination of triangular and trapezoidal functions, which are tailored to the most QoS-sensitive traffic (in Fig. 5).

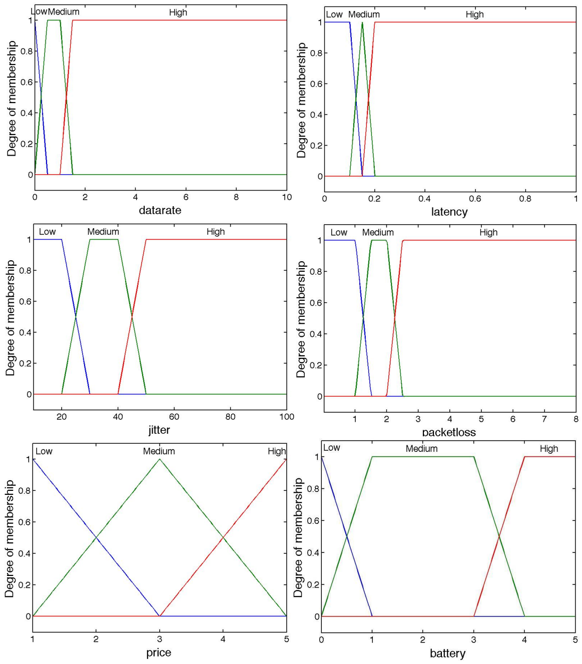


Fig. 5. FMFs Specific to CBR Traffic

ATD design has a variable number of decision parameters and employs a combination of triangular and trapezoidal functions, which are tailored to the incoming traffic, as shown in Fig. 5, 6 and 7 (noting that Fig. 5 is common to MD2 and ATD designs).

A small portion of the fuzzy rules for CBR traffic is given (in table I) to illustrate the general idea. As the number of decision parameters is fixed in MD1 and MD2, the same set of rules is used for all traffic types (CBR, VBR and ABR), and in all the three HDS designs. In contrast, the number of

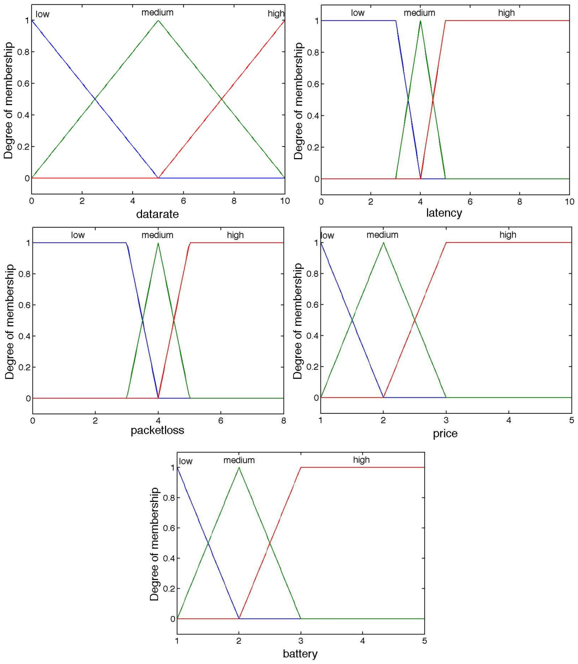


Fig. 6. FMFs Specific to VBR Traffic

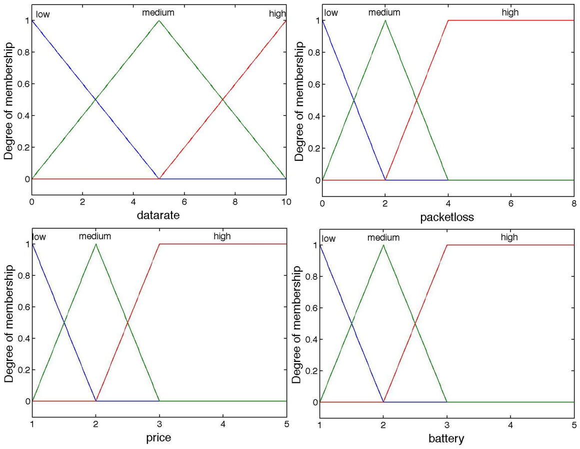


Fig. 7. FMFs Specific to ABR Traffic

decision parameters is variable in ATD design, so different sets of rules are used for different traffic types, as can be seen from tables I

VI. Results and Discussion

To evaluate handover decision performance of the three fuzzy-based HDS designs, we have assumed three wireless network technologies (WLAN, WiMAX and Cellular) and three traffic models (VoIP, video streaming and file transfer). Their QoS requirements are given in [15]

The performance criteria are the percentage success, which is measured in terms of the number of times (expressed as a percentage) the HDS selected the wireless network that had the highest score among the three and fully satisfied the QoS requirements, and the execution time.

TABLE I
Fuzzy Rules for Each Individual Traffic Type

Fuzzy Rules for CBR Traffic							
No.	DR	LA	JI	PL	PR	BA	Output
1	Low	Low	Low	Low	Low	Low	M
2	Low	Low	Low	Low	Low	Medium	MH
:	:	:	:	:	:	:	:
729	High	High	High	High	High	High	ML
Fuzzy Rules for VBR Traffic							
No.	DR	LA	JI	PL	PR	BA	Output
1	Low	Low	▨	Low	Low	Low	ML
2	Low	Low	▨	Low	Low	Medium	M
:	:	:	▨	:	:	:	:
243	High	High	▨	High	High	High	VL
Fuzzy Rules for ABR Traffic							
No.	DR	LA	JI	PL	PR	BA	Output
1	Low	▨	▨	Low	Low	Low	ML
2	Low	▨	▨	Low	Low	Medium	M
:	:	▨	▨	:	:	:	:
81	High	▨	▨	High	High	High	ML

A. Performance Measurement

Crisp input value for each of the decision parameters (with the exception of usage price) is randomly selected from the range given in table II, and used in all the three HDS designs in the case of VoIP traffic.

Similarly, in the case of video streaming, crisp values are randomly selected from table III and used in ATD, whereas MD1 and MD2 also need a value for jitter (JI), which is taken from table II.

In the case of file transfer traffic, crisp values are randomly selected from table IV and used in ATD. JI is taken from table II and used in MD1 and MD2, and latency (LA) is taken from table III and used in MD1 and MD2.

The usage price for individual technologies is set at a fixed value and assumed to be incremental (i.e. WLAN to be least expensive and Cellular to be most expensive [16]).

The range of values for decision parameters in tables II, III and IV are taken either from real-life tests or commonly used standards [17]–[21].

TABLE II
Decision Parameters for CBR Traffic

Network	DR (Mbps)	LA (ms)	JI (ms)	PL (%)	BA (hrs)	PR (p/min)
WLAN	1 - 8	0-300	0-50	0-1.5	2.5 - 5	1
WiMAX	3 - 6				0.55x(2.5-5)	2
Cellular	1 - 5				0.74x(2.5-5)	3

TABLE III
Decision Parameters for VBR Traffic

Network	DR (Mbps)	LA (s)	JI (ms)	PL (%)	BA (hrs)	PR (p/min)
WLAN	1 - 8	0-7	▨	0-7	2.5 - 5	1
WiMAX	3 - 6				0.55x(2.5-5)	2
Cellular	1 - 5				0.74x(2.5-5)	3

The three HDS designs are simulated using Fuzzy Logic tool on MATLAB platform. Each of the three HDS designs is evaluated by running the algorithm for 200 times for each

TABLE IV
DECISION PARAMETERS FOR ABR TRAFFIC

Network	DR (Mbps)	LA (s)	JI (ms)	PL (%)	BA (hrs)	PR (p/min)
WLAN	1 - 8				2.5 - 5	1
WiMAX	3 - 6			0-7	0.55x(2.5-5)	2
Cellular	1 - 5				0.74x(2.5-5)	3

traffic type. The above procedure was repeated to evaluate the performance of Simple Additive Weighting (SAW) for a comparison purpose. Simulation results are shown in Fig. 8, 9 and 10.

In the case of VoIP traffic (Fig. 8), the performance of MD2 and ATD design have exactly the same performance. This result is expected as identical rules and FMFs are employed. However, there is an improvement of 37.4% compared with MD1, which employs the same rules but fixed FMFs, and an improvement of 38.78% when compared with SAW.

It is interesting to note in Fig. 9 that the performance of MD2 is slightly worse than MD1, in the case of video streaming traffic. As FMFs in MD2 are tailored for the most QoS-sensitive traffic, relatively less QoS-sensitive traffic is penalized. The performance of ATD design, on the other hand, is 24.17%, 22.03% and 30.58% better than SAW, MD1 and MD2, respectively.

In the case of file transfer traffic, Fig. 10, a similar picture emerges when comparing the performance of MD1 and MD2. However, the performance of ATD is comparable with MD1 and is slightly better than SAW. This result suggests that the tailoring of FMFs is more beneficial when the number of QoS parameters is increased.

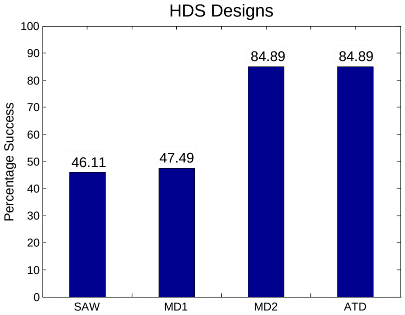


Fig. 8. Network Selection Performance - VoIP

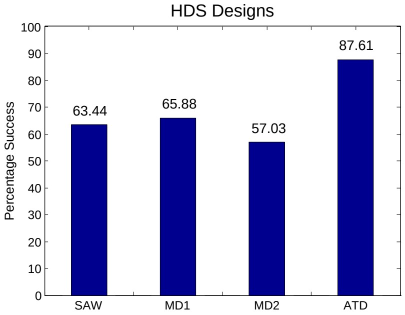


Fig. 9. Network Selection Performance - Video Streaming

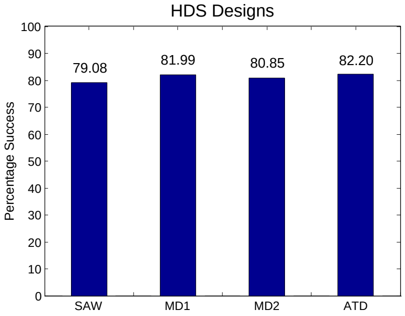


Fig. 10. Network Selection Performance - FTP

B. Algorithm Execution Time

As has been mentioned in section 1, minimization of the execution time (τ) is an essential requirement for real-time applications. We have evaluated τ for the three HDS designs and SAW. The evaluation was carried out on a 2.13GHz Intel Core 2 Duo with 4GB memory.

The results in Fig. 11 clearly show that our proposed ATD design significantly reduces τ for relatively less QoS-sensitive traffics. The reduction achieved in τ is 70.05% and 90.37% for video streaming and file transfer traffic, respectively. However, in the case of VoIP, there is no significant reduction in the value of τ. It is to be expected as all the three HDS designs employ the same number of decision parameters. The execution time of SAW is lower than that of ATD.

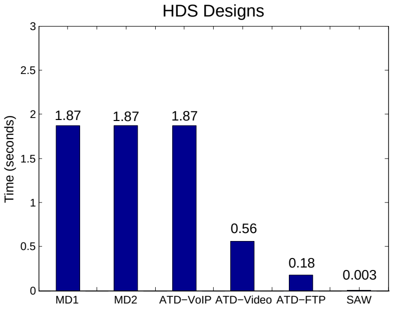


Fig. 11. Algorithm Execution Time

C. Battery Consumption Analysis

Our comparison of fuzzy-based algorithms with SAW algorithm reveals that the superiority of fuzzy-based designs comes at a price, i.e. the algorithm execution time of even the best (ATD) fuzzy algorithm is much higher than that required by the SAW. This raises the issue of power consumption and the recharging frequency for the battery. In order to address these issues we have made some projections based on the data available to us.

Our simulations were carried out on MATLAB platform using Intel processor of 65watts rating. The longest execution time (worst case) for our fuzzy algorithms is 1.87 seconds (Fig. 11). Therefore, the power consumption for the worst case = 65x1.87 = 121.55 watt-seconds or 0.033 watt-hours. Now the battery capacity of a modern smart phone is around 5.5 watt-hour. Thus, a smart phone can execute the above algorithm around 166 times (this does not include the power consumption of other components) before the battery needs recharging.

If we now consider a processor that is actually used in mobile devices (e.g. ARM Cortex A series of 1.3 watts rating), the estimated power consumption reduces to 0.000675 watt-hour. Assuming the same battery as above, a smart phone can execute the algorithm for over 8,000 times before the need for recharging. Further improvements will come from the fact that an actual mobile device is likely to use dedicated and embedded software instead of MATLAB platform to run fuzzy algorithm. This will further reduce execution time and hence the power consumption.

VII. Conclusions and Future Work

We have proposed an adaptive traffic dependent handover decision system to deal with the restriction imposed on the number of decision parameters mainly due to the fact that as the number of decision parameters increases, the number of fuzzy rules increases to a very large value, resulting in computational complexity and an unacceptably long execution time. In our approach multiple decision engines, each optimized to a specific traffic type, have been suggested. The optimization has been achieved by tailoring FMFs to match the QoS requirements of each individual traffic type. The number of fuzzy rules has been reduced, compared with a conventional monolithic decision engine, by including only those QoS parameters that are essential for a given traffic type.

For evaluation and comparison purposes, three handover decision system designs (Monolithic design 1, Monolithic design 2 and ATD) have been developed. Assuming a heterogeneous networking environment and three traffic types (VoIP, video streaming and file transfer), simulation results have been produced to compare the performance of the three HDS designs and SAW in term of the decision quality and execution time.

In terms of the decision quality, the simulation results show that ATD design gives an improvement of 37.4% and 22.03% for VoIP and video streaming traffics respectively, when compared with Monolithic design 1. The performance of ATD is comparable with others in the case of file transfer traffic.

In terms of the execution time, the results show that ATD design gives an improvement of over 90% and 70% in case of file transfer and video streaming traffics respectively, when compared with Monolithic design. The performance of Monolithic and ATD designs is comparable in the case of VoIP traffic. Battery consumption analysis suggests that the power consumption of the proposed algorithm is unlikely to have a major impact on the battery life in real-life implementation.

Future work will investigate possibilities of further reduction in computational complexity and hence execution time for the handover process.

Acknowledgement

The authors would like to express their thanks to the Faculty of Engineering, Science and the Built Environment, London South Bank University, UK and Assumption University, Thailand for the support provided to carry out this research.

References

[1] G. Pollini, “Trends handover design,” *IEEE Communications Magazine*, (34)3, 1996.

[2] T. Thumthawatworn and A. Pervez, “Multi-level rule-based handover framework for heterogeneous wireless networks,” *Wireless Advanced (WiAD), 2010 6th Conference on*, pp. 1–6, 2010.

[3] S. J. Wu, ”Fuzzy-based handover decision scheme for next-generation heterogeneous wireless networks,” *Journal of Convergence Technology*, vol. 6, no. 4, pp. 285–297, 2011.

[4] C. G. Patil and M. T. Kolte, ”An approach for optimization of handoff algorithm using fuzzy logic system,” *Int. Journal of Computer Science and Communication*, vol. 2, no. 1, pp. 113–118, 2011.

[5] Q. He, ”A fuzzy logic based vertical handoff decision algorithm between WWAN and WLAN,” *Networking and Digital Society, Int. Conference on*, pp. 561–564, 2010.

[6] A. alhan and C. eken, ”An optimum vertical handoff decision algorithm based on adaptive fuzzy logic and genetic algorithm,” *Wireless Personal Communication*, vol. 64, no. 4, pp. 647–664, 2012

[7] T. Jun, Z. Ying Jiang, Z. Zhi, Y. Zhi Wei, and C. Zhi Lan, “Performance analysis of vertical handoff in wifi and wimax heterogeneous networks,” *Computer Network and Multimedia Technology, 2009. CNMT 2009. International Symposium on*, pp. 1 –15, jan. 2009.

[8] K. C. Foong, C. T. Chee and L. S. Wei, ”Adaptive network fuzzy inference system (ANFIS) handoff algorithm,” *Future Computer and Communication, Conference on*, pp. 195–198, 2009.

[9] A. Calhan and C. Ceken, ”An adaptive neuro-fuzzy based vertical handoff decision algorithm for wireless heterogeneous networks,” *Personal Indoor and Mobile Radio Network (PIMRC), 2010 21st Int. Sym. on*, pp. 2271–2276, 2010.

[10] Y. Chen, J. Ai and Z. Tan, ”An access network selection algorithm based on hierarchy analysis and fuzzy evaluation,” *Wireless Communications and Signal Processing (WCSP), 2009 International Conference on*, pp. 1–5, 2009.

[11] K. Vasu, S. Maheshwari, S. Mahapatra and C. S. Kumar, ”QoS aware fuzzy rule based vertical handoff decision algorithm for wireless heterogeneous networks,” *17th National Conference on Communication (NCC)*, 2011.

[12] E. Mamdani and S. Assilian, “An experiment in linguistic synthesis with a fuzzy logic controller,” *International Journal of Man-Machine Studies*, vol. 7, no. 1, pp. 1 – 13, 1975.

[13] T. Takagi and M. Sugeno, “Fuzzy identification of systems and its applications to modeling and control,” *Systems Man And Cybernetics, IEEE Transactions On*, vol. 15, no. 1, pp. 116–132, 1985.

[14] A. Kaur and A. Kaur,”Comparison of mamdani-type and sugeno-type fuzzy inference systems for air conditioning system,” *Soft Computing and Engineering, Int. Journal of*, vol. 2, no. 2, 2012.

[15] T. Szigeti and C. Hattingh, ”Quality of Service Design Overview,” *CISCO Press*, [Online] Available: http://www.ciscopress.com/articles/article.asp?p=357102&seqNum=3

[16] B. J. Chang and J. F. Chen,Cross-layer-based adaptive vertical handoff with predictive RSS in heterogeneous wireless networks, *Vehicular Technology, IEEE Transactions on*, vol. 57, no. 6, pp. 3679–3692, 2008.

[17] Real-life Speed Test for WLAN [Online] Available: http://www.pantip.com/cafe/mbk/topic/T11689772.html

[18] Real-life Speed Test for WiMAX [Online] Available: http://www.clear.com/coverage

[19] Real-life Speed Test for Cellular [Online] Available: http://www.pantip.com/cafe/mbk/topic/T11594482.html

[20] Battery Life Testing [Online] Available: http://www.anandtech.com/Show/Index/4643?c Page=3&all=False&sort=0&page=7&slug=htc-evo-3d-vs-motorola-photon-4g- best-sprint-phone

[21] QoS Concepts, [Online] Available: http://www.cisco.com/en/US/docs/net_mgmt/ ip_solution_center/3.0/qos/user/guide/concepts.html

Consensus Problem in Stochastic Network Systems with Switched Topology, Noise and Delay

Natalia Amelina
Faculty of Mathematics and Mechanics
St.Petersburg State University
St.Petersburg, Russia
Department of Telematics
Norwegian University of Science and Technology
Trondheim, Norway
Email: natalia_amelina@mail.ru

Alexander Fradkov
Faculty of Mathematics and Mechanics
St.Petersburg State University
St.Petersburg, Russia
Institute of Problems in Mechanical Engineering
St.Petersburg, Russia
Email: fradkov@mail.ru

***Abstract*—This paper deals with the problem of achieving consensus in decentralized stochastic network with switched topology and noise and delays in measurements. To solve the consensus problem of the group of interacting agents it was supposed to use the stochastic approximation type algorithm with the step-size non-decreasing to zero. Simulation results show the quality of the algorithm.**

***Index Terms*—consensus problem; stochastic networks; discrete systems; network systems.**

I. INTRODUCTION

The problems of control and distributed interaction in dynamical networks attracted more and more attention during last decade. A number of survey papers [1], [2], monographs [3], [4], [5], special issues of journals [6], [7], [8] and edited volumes [9], [10] are published. An interest is driven by applications to multiprocessor networks, transportation networks, production networks, coordinated control of motion of flying vehicles, submarines and mobile robots, distributed systems of control of power networks, complex crystal lattices, and nanostructured objects. In the presence of stochastic disturbances and noise, the stochastic gradient-like (stochastic approximation) methods have been used [11], [12], [13], [14], [15], [16].

Despite of large number of publications, satisfactory solutions are obtained only for a restricted class of problems by now. Such factors as nonlinearity of agent dynamics switching topology, noisy and delayed measurements may significantly complicate the solution. Additional important factors are limited transmission rate in the channels and quantizing (discretization) phenomenon. In presence of various disturbing factors, asymptotically exact consensus may be hard to achieve, especially in time-varying environment [17]. It those cases, approximate consensus problems should be examined. In [18], the approximate consensus problem in multi-agent stochastic systems with noisy information about the current state of the nodes and randomly switched topology for agents with nonlinear dynamics is considered.

This work is an extension of [18] to the case of multi-agent systems with delays in measurements. Following [18], we adopt an approach to analysis of stochastic multi-agent systems based on using the averaged models of system dynamics: the so-called method of averaged (discrete or continuous) models [19], [20], [21], [22], [23], [24].

The paper is organized as follows. In Section II, the basic concepts are introduced and an approximate mean-square consensus problem is posed. In Section III, the basic assumptions are described. The main results are presented in Section IV. In Section V, an example of computer network is given and the simulation results are provided. Section VI presents the conclusion.

II. PRELIMINARIES: CONSENSUS PROBLEM ON GRAPHS

Consider a dynamic network of a set of agents (nodes) $N = \{1,2,\ldots,n\}$.

Graph (N,E) is defined by N and set edges E. Define the *set of neighbors* of node i as $N^i = \{j : (j,i) \in E\}$, i.e. the set of nodes with edges incoming to i. Associate with each edge $(j,i) \in E$ a weight $a^{i,j} > 0$ and denote *adjacency matrix (or connectivity matrix)* $A = [a^{i,j}]$ of graph, denoted hereinafter as $\mathcal{G}_A$ (hereinafter the index of variables shows the corresponding number of nodes). Define the *weighted in-degree* of node i as the i-th row sum of A: $d^i(A) = \sum_{j=1}^{n} a^{i,j}$.

Endow each node $i \in N$ at time $t = 0,1,2\ldots,T$ with a time-varying state $x_t^i \in \mathbb{R}$ with dynamics

$$x_{t+1}^i = x_t^i + f^i(x_t^i, u_t^i), \tag{1}$$

where $f^i(\cdot,\cdot) : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ are some functions that depend on the states in the previous time x_t^i and on control actions $u_t^i \in \mathbb{R}$.

We consider the network (multi-agent) system consisting of dynamic agents with inputs u_t^i, outputs $y_t^{i,i}$ and states x_t^i.

Nodes i and j *agree* in a network at time t if and only if $x_t^i = x_t^j$.

The consensus problem is the agreement of all nodes in network, i.e., we have to find a control protocol that drives all states to the same constant steady-state values: $x_t^i = x_t^j \;\; \forall i,j \in N, i \neq j$.

We assume that the structure of links of the dynamic network is modeled by a sequence of directed graphs $\{(N,E_t)\}_{t\geq 0}$, where $E_t \subset E$ change in time. If $(j,i) \in E_t$, then we say that node i at time t obtains information from the node j for the purposes of feedback control. Denote A_t as adjacency matrix corresponding to E_t; $E_{\max} = \{(j,i) : \sup_{t\geq 0} a_t^{i,j} > 0\}$ is the maximum set communication links.

To form its control strategy each node uses its own state (possibly noisy)

$$y_t^{i,i} = x_t^i + w_t^{i,i}, \tag{2}$$

and if $N_t^i \neq \emptyset$, noisy measurements of its neighbors states

$$y_t^{i,j} = x^j_{t-d_t^{i,j}} + w_t^{i,j},\ j \in N_t^i, \tag{3}$$

where $w_t^{i,i}, w_t^{i,j}$ is the noise, $0 \leq d_t^{i,j} \leq \bar{d}$ is integer-valued delay, $\bar{d}$ is a maximal delay.

Since the system starts working at $t = 0$ so implicit requirement to set of neighbors would be: $j \in N_t^i \Rightarrow t - d_t^{i,j} \geq 0$. We put $w_t^{i,j} = 0$ for all other pairs of i, j and denote $\bar{w}_t \in \mathbb{R}^{n^2}$ as a vector (matrix $n \times n$ which is written in rows as a vector) consisting of elements $w_t^{i,j}$, $i, j \in N$.

The control algorithm (protocol), called the *local voting protocol*, is given by

$$u_t^i = \alpha_t \sum_{j\in \bar{N}_t^i} b_t^{i,j}(y_t^{i,j} - y_t^{i,i}), \tag{4}$$

where $\alpha_t > 0$ are step-sizes of control protocol, $b_t^{i,j} > 0\ \forall j \in \bar{N}_t^i$. We set $b_t^{i,j} = 0$ for other pairs i, j and denote $B_t = [b_t^{i,j}]$ as the matrix of control protocol.

For the vector or matrix M denote the Frobenius norm: $||M|| = [Tr(M^{\mathrm{T}}M)]^{1/2}$, where $Tr(\cdot)$ is a trace (sum of the diagonal elements) of matrix. In some cases for matrix A the vector norm (square root of the sum of the squares of all its elements) will be used, which we denote as $||A||_2$.

The n nodes to achieve *asymptotic mean square consensus* if $\mathrm{E}||x_t^i||^2 < \infty, t = 0, 1, \ldots,\ i \in N$ and there exists a random variable $x^\star$ such that $\overline{\lim}_{t\to\infty} \mathrm{E}||x_t^i - x^\star||^2 = 0$ for $i \in N$.

The n nodes to achieve *ε-consensus* if $\mathrm{E}||x_t^i||^2 < \infty,\ i \in N$, and there exists a random variable $x^\star$ such that $\mathrm{E}||x_t^i - x^\star||^2 \leq \varepsilon$ for all $i \in N$.

III. MAIN ASSUMPTIONS

Let $(\Omega, \mathscr{F}, P)$ be the underlying probability space. Let E be symbol of mathematical expectation and E_x be conditional expectation under the condition x.

In the formulation of further results, we assume that the following conditions are satisfied.

A1. $\forall i \in N$ functions $f^i(x,u)$ are Lipschitz in x and u: $|f^i(x,u) - f^i(x',u')| \leq L_1(L_x|x-x'| + |u-u'|)$, the growth rate is bounded: $|f^i(x,u)|^2 \leq L_2(L_c + L_x|x|^2 + |u|^2)$, and for any fixed x the function $f^i(x,\cdot)$ is such that $\mathrm{E}_x f^i(x,u) = f^i(x, \mathrm{E}_x u)$;

Remark. A typical case when this condition holds is the case when $f^i(x,u)$ is linear in control.

A2. **a)** $\forall i \in N, j \in N^i$ the noises $w_t^{i,j}$ are centered, independent and have bounded variance: $E(w_t^{i,j})^2 \leq \sigma_w^2$.

b) $\forall i \in N, j \in N^i$ the appearances of variable edges (j,i) in the graph $\mathscr{G}_{A_t}$ are independent random events with probability $p_a^{i,j}$ (i.e., matrices A_t are independent, identically distributed random matrices).

c) $\forall i \in N, j \in N^i$ weights $b_t^{i,j}$ in the control protocol are bounded random variables: $\underline{b} \leq b_t^{i,j} \leq \bar{b}$ with probability 1, and there exist limits $b^{i,j} = \lim_{t\to\infty} \mathrm{E}b_t^{i,j}$.

d) $\forall i \in N, j \in N^i$ there exists a finite quantity $\bar{d} \in \mathbb{N}$: $d_t^{i,j} \leq \bar{d}$ with probability 1 and integer-valued delay $d_t^{i,j}$ — independent, identically distributed random variables taking values $k = 0, \ldots, \bar{d}$ with probability $p_k^{i,j}$.

Moreover, all of these random variables and matrices are independent of each other and their components have a limited variance.

If $\bar{d} > 0$ we add new nodes to the current network topology $n\bar{d}$. We add new "fictitious" agents with states at time t equal to the corresponding states of the "real" agents at the previous $\bar{d}$ time: $t-1, t-2, \ldots, t-\bar{d}$.

Denote $\bar{n} = n(\bar{d}+1)$. Matrix $A_{\max}$ of size $\bar{n} \times \bar{n}$ is denoted as:

$$a_{\max}^{i,j} = p_{j\div\bar{d}}^{i,j \operatorname{mod} \bar{d}}\, p_a^{i,j \operatorname{mod} \bar{d}}\, b^{i,j \operatorname{mod} \bar{d}},\ i \in N,\ j = 1, 2, \ldots, \bar{n},$$

$$a_{\max}^{i,j} = 0,\ i = n+1, n+2, \ldots, \bar{n},\ j = 1, 2, \ldots, \bar{n}.$$

Here, the operation mod is a remainder of the division, and $\div$ is division without a remainder.

Note that if $\bar{d} = 0$ so this definition of network topology (of matrix $A_{\max}$ of size $n \times n$) is as follows

$$a_{\max}^{i,j} = p_a^{i,j} b^{i,j},\ i \in N,\ j \in N.$$

If we consider the sequence of random matrices $\bar{A}_t$ with elements that define the connections at time t, then all of them are identically distributed and the matrix $A_{\max}$ is in fact its expectation (averaging).

We assume that the following condition is satisfied for the network topology matrix:

A3. Graph $(N, E_{\max})$ has a spanning tree, and for any edge $(j,i) \in E_{\max}$ among the elements $a_{\max}^{i,j}, a_{\max}^{i,j+n}, \ldots, a_{\max}^{i,j+\bar{d}n}$ of the matrix $A_{\max}$ there exists at least one non-zero.

For $t = 1, 2, \ldots$ we define an increasing sequence of σ-algebras of probability of events $\tilde{\mathscr{F}}_t$, generated by random elements $A_1, \ldots, A_{t-1}$; $d_1^{i,j}, \ldots, d_{t-1}^{i,j}$, $b_1^{i,j}, \ldots, b_{t-1}^{i,j}, w_1^{i,j}, \ldots, w_t^{i,j}$, $i, j \in N$, and $\mathscr{F}_t = \sigma\{\mathscr{F}_t^A, A_t; b_t^{i,j}, d_t^{i,j}, i, j \in N\}$.

For a random variable Q and σ-algebra of probability event $\mathscr{F}$ we use the notation $\mathrm{E}_{\mathscr{F}} Q$ for the conditional expectation Q with respect to σ-algebra $\mathscr{F}$.

Note that the random variables $\bar{x}_t$ are measurable with respect σ-algebra $\mathscr{F}_{t-1}$, i.e. $\mathrm{E}_{\mathscr{F}_{t-1}} \bar{x}_t = \bar{x}_t$.

IV. ANALYSIS OF THE CLOSED LOOP SYSTEM DYNAMICS

Denote $\bar{x}_t = [x_t^1; \ldots; x_t^n]$. Let $\bar{x}_t \equiv 0$ for $-\bar{d} \leq t < 0$, and denote $\bar{X}_t \in \mathbb{R}^{n\bar{d}}$ as extended state vector $\bar{X}_t = [\bar{x}_t, \tilde{x}_{t-1}, \ldots, \tilde{x}_{t-\bar{d}}]$, where $\tilde{x}_{t-k}$ is vector consisting of such x_{t-k}^i that $\exists j \in N^i\ \exists k' \geq k : p_{k'}^{i,j} > 0$, i.e. this value with positive probability involved in the formation of at least one of the controls. To simplicity, we assume that so introduced an extended state vector is $\bar{X}_t = [\bar{x}_t, \bar{x}_{t-1}, \dots, \bar{x}_{t-\bar{d}}]$, i.e. it includes all the components with all kinds of delays not exceeding $\bar{d}$.

Rewrite the dynamics of the nodes in vector-matrix form:

$$\bar{X}_{t+1} = U\bar{X}_t + F(\alpha_t, \bar{X}_t, \bar{w}_t), \tag{5}$$

where U is the following matrix of size $\bar{n} \times \bar{n}$:

$$U = \begin{pmatrix} I & 0 & 0 & \dots & 0 \\ I & 0 & 0 & \dots & 0 \\ 0 & I & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I & 0 \end{pmatrix}, \tag{6}$$

where I is the identity matrix of size $n \times n$, and $F(\alpha_t, \bar{X}_t, \bar{w}_t) : \mathbb{R} \times \mathbb{R}^{\bar{n}} \times \mathbb{R}^{n^2} \to \mathbb{R}^{\bar{n}}$ — vector function of the arguments:

$$F(\alpha_t, \bar{X}_t, \bar{w}_t) =$$

$$= \begin{pmatrix} \cdots \\ f^i(x^i_t, \alpha_t \sum_{j \in \bar{N}^i_t} b^{i,j}_t ((x^j_{t-d^{i,j}_t} - x^i_t) + (w^{i,j}_t - w^{i,i}_t))) \\ \cdots \\ 0_{n\bar{d}} \end{pmatrix}, \tag{7}$$

containing a non-zero components only on the first n places.

Consider the corresponding (5) averaged discrete model

$$\bar{Z}_{t+1} = U\bar{Z}_t + G(\alpha_t, \bar{Z}_t),\ \bar{Z}_0 = \bar{X}_0, \tag{8}$$

where

$$G(\alpha, \bar{Z}) = G\begin{pmatrix} & z^1 \\ \alpha, & \vdots \\ & z^{n(\bar{d}+1)} \end{pmatrix} = \begin{pmatrix} \cdots \\ f^i(z^i, \alpha s^i(\bar{Z})) \\ \cdots \\ 0_{n\bar{d}} \end{pmatrix}, \tag{9}$$

$$s^i(\bar{Z}) = \sum_{j \in N^i} p^{i,j}_a b^{i,j} ((\sum_{k=0}^{\bar{d}} p^{i,j}_k z^{j+kn}) - z^i) =$$

$$= -d^i(A_{\max}) z^i + \sum_{j=1}^{\bar{n}} a^{i,j}_{\max} z^j,\ i \in N.$$

It turns out that the trajectory of solutions of the initial system $\{\bar{X}_t\}$ from (5) at time t are close in mean square sense to the average trajectory of the discrete system (8).

Theorem *1:* If conditions **A1**, **A2** are satisfied, **then** there exists $\tilde{\alpha}$ such that for $0 < \alpha_t \leq \bar{\alpha} < \tilde{\alpha}$ the following inequality holds:

$$\mathrm{E} \max_{0 \leq t \leq T} ||\bar{X}_t - \bar{Z}_t||^2 \leq c_1 \tau_T e^{c_2 \tau_T^2} \bar{\alpha}, \tag{10}$$

where $\tau_T = 2^{\bar{d}}(\alpha_0 + \alpha_1 + \ldots + \alpha_{T-1})$, $c_1, c_2 > 0$ are some constants:

$$c_1 = 8n\left(\tilde{c} + \hat{c}(\frac{nL_2L_c + \bar{\alpha}^2\tilde{c}}{c_3} + ||\bar{X}_0||^2)e^{T\ln(c_3+1)}\right),$$

$$\tilde{c} = n^2 L_1^2 \bar{b}^2 \sigma_w^2,\ c_2 = 2^{1-\bar{d}} L_1^2 (\frac{L_x}{\underline{\alpha}} + 2\bar{\alpha}^2 ||\mathscr{L}(A_{\max})||_2^2),$$

$$c_3 = \tilde{d} + L_x(2^{1+\tilde{d}/2}L_1 + L_2) + \bar{\alpha}c',\ \hat{c} = 2L_1^2 n(n-1)\bar{b}^2,$$

$$c' = 2^{1+\tilde{d}/2} L_1 ||\mathscr{L}(A_{\max})||_2 + \bar{\alpha}(L_2 ||\mathscr{L}(A_{\max})||_2^2 + \hat{c}),$$

$\underline{\alpha} = \min_{1 \leq t \leq T} \alpha_t$, $\tilde{d} = 0$ if $\bar{d} = 0$, or $\tilde{d} = 1$ if $\bar{d} > 0$.

Note that in case without delays in the measurement ($\tilde{d} = 0$) and if $L_x = 0$ then constant c_3 which is defined in Theorem 1 is estimated by the value proportional to $\bar{\alpha}$ and therefore constant c_1 is estimated by the value proportional to τ_T, which corresponds to the previously obtained results for this case from [21], [19].

Proof:

Denote

$$v_t = F(\alpha_t, \bar{X}_t, \bar{w}_t) - G(\alpha_t, \bar{X}_t). \tag{11}$$

By condition **A2** averaging with respect to σ-algebras $\mathscr{F}^d_t$ and $\mathscr{F}_t$ yields $\mathrm{E}_{\mathscr{F}_t} v_t = 0$.

To proof Theorem 1, the following facts will be useful.

Proposition *1:*

$$||U\bar{X}||^2 \leq 2^{\bar{d}}||\bar{X}||^2,\ \ldots,\ ||U^{\bar{d}}\bar{X}||^2 \leq 2^{\bar{d}}||\bar{X}||^2,\ \ldots,\ ||U^k\bar{X}||^2 \leq$$

$$\leq 2^{\bar{d}}||\bar{X}||^2,$$

Proof: By the definition of matrix U it is easy to obtain the first inequality, and the rest we get by induction on k and by the following equality

$$\forall k > \bar{d}\ U^k = U^{\bar{d}} = \begin{pmatrix} I & 0 & 0 & \dots & 0 \\ I & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ I & 0 & 0 & \dots & 0 \end{pmatrix}. \tag{12}$$

■

Proposition *2:* By assumptions **A2** the following inequality holds

$$\mathrm{E} \max_{1 \leq t \leq T} ||\sum_{i=1}^{t} v_t||^2 \leq 4n \sum_{t=1}^{T} \mathrm{E}||v_t||^2.$$

Proof:

Under the conditions **A2** random elements v_t are martingale differences, i.e., they are centered with respect to the conditional averaging of the background: $\mathrm{E}_{\mathscr{F}_{t-1}} v_t = 0$. So, Lemma 1 from section 3 of [25] is applicable. The dimension of vectors v_t is $n\bar{d}$, but since only the first n components of vectors v_t are nonzero, then it is possible to use in the estimation the value of n instead of $n\bar{d}$. ■

Proposition *3:* Let the sequence of numbers $\mu_t \geq 0$, $t = 0, 1, \ldots, T$ satisfies the inequalities

$$\mu_{t+1} \leq \bar{\alpha} c_1 \tau_t + c_2 2^{\bar{d}} \tau_t \sum_{k=1}^{t} \gamma_k \mu_k,\ c_1, c_2 \geq 0,$$

then

$$\mu_t \leq c_1 \tau_t e^{c_2 \tau_t^2} \bar{\alpha}.$$

Proof: Statement of Proposition follows directly from the corresponding result in [26] ■

Proposition *4:* [18] For $\bar{z} \in \mathbb{R}^n$ and matrix $A_{\max}$ the following inequality holds $\sum_{i=1}^{n}(\sum_{j \in N^i} a^{i,j}_{\max} z^j)^2 \leq ||A_{\max}||_2^2 ||\bar{z}||^2$.

ISBN: 978-1-61208-245-5

Proposition *5:* [18] $||\bar{s}(\bar{z})||^2 \le 2||\mathscr{L}(A_{\max})||_2^2||\bar{z}||^2$.

Proposition *6:* [18] If **A2** is satisfied then $s^i(\bar{x}) = \frac{1}{\alpha_t}\mathrm{E}_{\mathscr{F}_{t-1}}u_t^i$ and the following inequality holds $\frac{1}{\alpha_t^2}\mathrm{E}_{\mathscr{F}_{t-1}}{u_t^i}^2 \le (n-1)\bar{b}^2||\bar{x}_t - x_t^i\underline{1}||^2 + n\bar{b}^2\sigma_w^2$, $i \in N$.

Proposition *7:* By assumptions **A1**, **A2** yields

$$\mathrm{E}||\bar{X}_t||^2 \le (\frac{2nL_2 + \bar{\alpha}^2\tilde{c}}{c_3} + ||\bar{X}_0||^2)e^{t\ln(c_3+1)}.$$

Proof: We write equation (5) as

$$\bar{X}_{t+1} = U\bar{X}_t + G(\alpha_t, \bar{X}_t) + v_t. \qquad (13)$$

For the squared norm of $\bar{X}_{t+1}$ we have

$$||\bar{X}_{t+1}||^2 = ||U\bar{X}_t + G(\alpha_t,\bar{X}_t)||^2 + 2(U\bar{X}_t + G(\alpha_t,\bar{X}_t))^{\mathrm{T}}v_t + ||v_t||^2. \qquad (14)$$

Taking the conditional expectation of both parts of (14) on σ-algebra $\mathscr{F}_{t-1}$ (i.e. for fixed $\bar{X}_t$) by the centrality of v_t we obtain

$$\mathrm{E}_{\mathscr{F}_{t-1}}||\bar{X}_{t+1}||^2 = ||U\bar{X}_t + G(\alpha_t,\bar{X}_t)||^2 + \mathrm{E}_{\mathscr{F}_{t-1}}||v_t||^2 \le$$
$$\le 2||U\bar{X}_t||^2 + 2||G(\alpha_t,\bar{X}_t)||^2 + \mathrm{E}_{\mathscr{F}_{t-1}}||v_t||^2. \qquad (15)$$

By the form of v_t and Lipschitz in u of functions $f^i(u)$ (by **A1**) for $||v_t||^2$ we have

$$||v_t||^2 = \sum_{i\in N}|f^i(x_t^i, \alpha_t\sum_{j\in\bar{N}_t^i} b_t^{i,j}(x^j_{t-d_t^{i,j}} - x_t^i + w_t^{i,j} - w_t^{i,i}))-$$
$$-f^i(x_t^i, \alpha_t s_t^i(\bar{X}_t))|^2 \le L_1^2||\bar{u}_t - \alpha_t^2\bar{s}_t||^2.$$

Under the conditions **A2**, random variables $\mathrm{E}_{\mathscr{F}_{t-1}}u_t^i$, $i \in N$ satisfy the conditions of Proposition 6

$$\mathrm{E}_{\mathscr{F}_{t-1}}||v_t||^2 = \alpha_t^2L_1^2(2n(n-1)\bar{b}^2||\bar{X}_t||^2 + n^2\bar{b}^2\sigma_w^2). \qquad (16)$$

Consistently evaluating all three summands on the right hand side of (15) and taking into account the results of Propositions 1, 5 and 6, we deduce

$$\mathrm{E}_{\mathscr{F}_t}||\bar{X}_{t+1}||^2 \le 2^{\tilde{d}}||\bar{X}_t||^2 + 2^{1+\tilde{d}/2}||\bar{X}_t||L_1(L_x||\bar{X}_t|| + \alpha_t||\bar{s}||)+$$
$$+L_2(nL_c + L_x||\bar{X}_t||^2 + \alpha_t^2||\bar{s}||^2) + \alpha_t^2L_1^2(2n(n-1)\bar{b}^2||\bar{X}_t||^2+$$
$$+n^2\bar{b}^2\sigma_w^2) \le (2^{\tilde{d}} + 2^{1+\tilde{d}/2}L_1L_x + L_2L_x + \alpha_t2^{1+\tilde{d}/2}L_1||\mathscr{L}(A_{\max})||_2+$$
$$+\alpha_t^2(L_2||\mathscr{L}(A_{\max})||_2^2 + 2n(n-1)L_1^2\bar{b}^2))||\bar{X}_t||^2 + nL_2L_c+$$
$$+\alpha_t^2n^2L_1^2\bar{b}^2\sigma_w^2 \le \bar{c} + \bar{c}_3||\bar{X}_t||^2,$$

where $\bar{c} = nL_2L_c + \alpha_t^2\tilde{c}$, $\bar{c}_3 = c_3 + 1$.

By taking unconditional expectation of both parts of this inequality and consistently iterating on t, we obtain Proposition 7

$$\mathrm{E}||\bar{X}_t||^2 \le \bar{c} + \bar{c}_3\mathrm{E}||\bar{X}_{t-1}||^2 \le \bar{c} + \bar{c}\bar{c}_3 + \bar{c}_3^2\mathrm{E}||\bar{X}_{t-2}||^2 \le$$
$$\le \bar{c}(1 + \bar{c}_3 + \bar{c}_3^2 + \ldots + \bar{c}_3^{t-1}) + \bar{c}_3^t||\bar{X}_0||^2 \le \bar{c}\frac{\bar{c}_3^t - 1}{c_3} + \bar{c}_3^t||\bar{X}_0||^2 \le$$
$$\le \left(\frac{\bar{c}}{c_3} + ||\bar{X}_0||^2\right)\bar{c}_3^t \le (\bar{c}_4 + ||\bar{X}_0||^2)e^{t\ln\bar{c}_3},$$

$\bar{c}_4 = \bar{c}/c_3$. ■

Let us turn to the proof of Theorem 1. By iterating equation (5) for $t, t-1, \ldots t-d+1$ we obtain

$$\bar{X}_{t+1} = U\bar{X}_t + G(\alpha_t,\bar{X}_t) + v_t =$$
$$= U^2\bar{X}_{t-1} + UG(\alpha_{t-1},\bar{X}_{t-1}) + G(\alpha_t,\bar{X}_t) + Uv_{t-1} + v_t = \qquad (17)$$
$$= \cdots = U^{t+1}\bar{X}_0 + \sum_{k=0}^{t}U^{t-k}G(\alpha_k,\bar{X}_k) + \sum_{k=0}^{t}U^{t-k}v_k.$$

Similarly we obtain

$$\bar{Z}_{t+1} = U^{t+1}\bar{X}_0 + \sum_{k=0}^{t}U^{t-k}G(\alpha_k,\bar{Z}_k). \qquad (18)$$

Let us estimate $||\bar{X}_t - \bar{Z}_t||^2$, $t = 1,\ldots,T$. By subtracting (18) from (17) and squaring the result we obtain

$$||\bar{X}_t - \bar{Z}_t||^2 = ||\sum_{k=1}^{t}U^{t-k}v_k + \sum_{k=1}^{t}U^{t-k}(G(\alpha_k,\bar{X}_k) - G(\alpha_k,\bar{Z}_k))||^2 \le$$
$$\le 2||\sum_{k=1}^{t}U^{t-k}v_k||^2 + 2||\sum_{k=1}^{t}U^{t-k}(G(\alpha_k,\bar{X}_k) - G(\alpha_k,\bar{Z}_k))||^2 \le$$
$$\le 2||\sum_{k=1}^{t}U^{t-k}v_k||^2 + 2\frac{\tau_t}{2^{\bar{d}}}\sum_{k=1}^{t}\frac{1}{\alpha_t}||U^{t-k}(G(\alpha_k,\bar{X}_k) - G(\alpha_k,\bar{Z}_k))||^2. \qquad (19)$$

For the summands in the second sum of (19) using Propositions 5, 1 and Lipschitz condition $f^i(\cdot,\cdot)$ (assumption **A1**) we obtain

$$||U^{t-k}(G(\alpha_k,\bar{X}_k) - G(\alpha_k,\bar{Z}_k))||^2 \le 2^{\bar{d}}L_1^2\sum_{i=1}^{n}(L_x|x_k^i - z_k^i|+$$
$$+\alpha_k|s(x_k^i) - s(z_k^i)|)^2 \le 2^{1+\bar{d}}L_1^2\sum_{i=1}^{n}L_x|x_k^i - z_k^i|^2 + \alpha_k^2s(x_k^i - z_k^i)^2 \le$$
$$\le 2^{1+\bar{d}}L_1^2(L_x + 2\alpha_k^2||\mathscr{L}(A_{\max})||_2^2)||\bar{X}_k - \bar{Z}_k||^2$$

We take expectation of both parts of (19) and denote $\mu_T = \max_{0\le t\le T}\mathrm{E}||\bar{X}_t - \bar{Z}_t||^2$. By applying Proposition 2 to the first summand and obtained above estimate of the second summand we obtain

$$\mu_T \le 2^{3+\bar{d}}n\sum_{k=1}^{T}\mathrm{E}||v_k||^2 + 2\tau_TL_1^2\sum_{k=1}^{t}(\frac{L_x}{\underline{\alpha}} + 2\alpha_k||\mathscr{L}(A_{\max})||_2^2)\mu_k. \qquad (20)$$

To estimate $\mathrm{E}||v_k||^2$ by using previously obtained relation (16) and the result of Proposition 7 we deduce

$$\mathrm{E}||v_k||^2 \le \alpha_k^2(\tilde{c} + \hat{c}(\bar{c}_4 + ||\bar{X}_0||^2)e^{k\ln(c_3+1)})$$

and hence

$$2^{3+\bar{d}}n\sum_{k=1}^{T}\mathrm{E}||v_k||^2 \le \bar{\alpha}8n\tau_T(\tilde{c} + \hat{c}(\bar{c}_4 + ||\bar{X}_0||^2)e^{T\ln(c_3+1)}). \qquad (21)$$

By the following relation $2^{\bar{d}}\sum_{k=1}^{t}\alpha_k^2 \le \bar{\alpha}2^{\bar{d}}\sum_{k=1}^{t}\alpha_k = \bar{\alpha}\tau_t$, considering estimates (21) from (20), we have

$$\mathrm{E}\mu_T \le \bar{\alpha}c_1\tau_T + c_2\tau_T2^{\bar{d}}\sum_{k=1}^{T}\alpha_k\mathrm{E}\mu_k. \qquad (22)$$

ISBN: 978-1-61208-245-5

From last inequality (22) by applying Proposition 3 we get the conclusion of Theorem 1. ■

Theorem *2:* Let the conditions **A1**, **A2** be satisfied; $0 < \alpha_t \leq \bar{\alpha}$; in averaged discrete system (8) $\frac{\varepsilon}{4}$-consensus is achieved for time T and for constants c_1, c_2 from Theorem 1 the following estimate holds

$$c_1 \tau_T e^{c_2 \tau_T^2} \bar{\alpha} \leq \frac{\varepsilon}{4},$$

then ε-consensus is achieved in stochastic discrete system (5) at time t.

Proof: Denote $x^\star$ as consensus value of discrete system (8). From the first group of conditions of Theorem 2 the conditions of Theorem 1 hold. From other conditions of Theorem 2 and the result of Theorem 1 we obtain

$$\mathrm{E}||\bar{X}_t - x^\star \underline{1}||^2 \leq 2\mathrm{E}||\bar{X}_t - \bar{Z}_t||^2 + 2||\bar{Z}_t - x^\star \underline{1}||^2 \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \leq \varepsilon.$$

■

Consider an important particular case $\forall i \in N\, f^i(x,u) = u$ and $\alpha_t = \alpha = const$, in which the discrete averaged system (8) has the form:

$$\bar{Z}_{t+1} = (I - ((I-U) - \mathscr{L}(\alpha A_{\max})))Z_t. \qquad (23)$$

Theorem *3:* If conditions **A2**, **A3** are satisfied; $\alpha_t = \alpha > 0$; $f^i(x,u) = u$ for any $i \in N$ and condition $\alpha < \frac{1}{d_{\max}}$ for matrix $A_{\max}$ is satisfied, **then** asymptotic mean square consensus for n nodes in averaged discrete system (23).

Moreover if $\frac{\varepsilon}{4}$-consensus is achieved for the time $T(\frac{\varepsilon}{4})$ in averaged discrete system (23) and there exist $T > T(\frac{\varepsilon}{4})$ for which the parameter α provides the condition

$$\bar{C}_1 e^{\bar{C}_2} \alpha \leq \frac{\varepsilon}{4},$$

$$\bar{C}_1 = 8n\left(\tilde{c} + \hat{c}(\frac{\alpha^2\tilde{c}}{c_3} + ||\bar{X}_0||^2)e^{T\ln(c_3+1)}\right)\tau_t,$$

$$\bar{C}_2 = 2^{2-\bar{d}}\alpha^2||\mathscr{L}(A_{\max})||_2^2,\ \tilde{c} = n^2\bar{b}^2\sigma_w^2,\ \hat{c} = 2n(n-1)\bar{b}^2\tau_t^2,$$

$$c_3 = 2^{1+\tilde{d}} + 2\alpha^2(||\mathscr{L}(A_{\max})||_2^2 + \hat{c}),$$

where $\tilde{d} = 0$ if $\bar{d} = 0$, or $\tilde{d} = 1$ if $\bar{d} > 0$.

then ε-consensus at time $t:\ T(\frac{\varepsilon}{4}) \leq t \leq T$ is achieved in stochastic discrete system (5).

Proof:

The result of Theorem 3 is derived from Theorem 2.

All amounts in rows of elements of the matrix $\bar{\mathscr{L}} = (I-U) - \mathscr{L}(\alpha A_{\max})$ are equal to zero and, moreover, all the diagonal elements are positive and equal to the absolute value of the sum of all the other elements in the row, which are negative. Hence the matrix $\bar{\mathscr{L}}$ is the Laplacian of a graph and a vector of 1's $\underline{1}$ is the right eigenvector corresponding to zero eigenvalue.

By condition **A3**, the graph corresponding to the Laplacian $\bar{\mathscr{L}}$ has a spanning tree. By condition **A3** graph of the first n nodes has a spanning tree. And units on $(n+1)$-th diagonal consistently connect $\bar{n}$-th node with $(\bar{n}-\bar{d})$-th node, $(\bar{n}-1)$-th node with $(\bar{n}-\bar{d}-1)$-th and so on. Hence asymptotic consensus is achieved in such a discrete system since the condition $\alpha < \frac{1}{d_{\max}}$ holds by the assumptions of Theorem 3.

To satisfy the conditions of Theorem 2 it remains to show that the constants $\bar{C}_1$ and $\bar{C}_2$ are the same as the corresponding constants from Theorem 1. It follows from the fact that in this case $L_1 = L_2 = 1$, $L_x = L_c = 0$.

■

Note that in [11], under certain assumptions similar to the conditions of Theorem 3, the necessary and sufficient condition for achieving mean square consensus in case when step-sizes α_t tending to zero was proved. More general case of the form of functions $f^i(x_t^i, u_t^i)$ and step-sizes α_t not tending to zero were considered above.

V. EXAMPLE

To illustrate the theoretical results we give an example the computer network.

We consider the system of separation the same type of jobs between different agents for parallel computing with feedback. Denote $N = \{1,\ldots,n\}$ as a set of intelligent agents, each of which serves the incoming requests a first-in-first-out queue. Jobs are received at different times and on different nodes.

At any time t state of agent i, $i = 1,\ldots,n$ is described by two characteristics:

- q_t^i is queue length of the atomic elementary jobs of the node i at time t;
- p_t^i is a productivity of the node i at time t.

The dynamics of each agent are described by

$$q_{t+1}^i = q_t^i - p_t^i + z_t^i + u_t^i;\ i \in N,\ t = 0,1,\ldots,T, \qquad (24)$$

where z_t^i is the new job received by node i at time t, u_t^i is the result of information redistribution between agents, which is obtained by using the selected protocol of information redistribution. In the dynamics we assume that $\sum_i u_t^i = 0$, $t = 0,1,2,\ldots$.

We assume that to form the control strategy each agent $i \in N$ at time t can receive from its neighbors $j \in N_t^i$ the following information:

- the noisy observations about its queue length
$$y_t^{i,i} = q_t^i + w_t^{i,i}, \qquad (25)$$
- the noisy and delayed observations about its neighbors queue length, if $N_t^i \neq \emptyset$
$$y_t^{i,j} = q_{t-d_t^{i,j}}^j + w_t^{i,j},\ j \in N_t^i, \qquad (26)$$
where $w_t^{i,j}$ are noises, $0 \leq d_t^{i,j} \leq \bar{d}$ is integer-valued delay, $\bar{d}$ is a maximal delay,
- the information about its productivity p_t^i and about its neighbors productivity p_t^j, $j \in N_t^i$.

In the stationary case from all possible options for all job redistribution, which are not distributed by the time t, then minimum operation time of the system corresponds to

$$q_t^i/p_t^i = q_t^j/p_t^j,\ \forall i,j\ \in N \qquad (27)$$

So if we take $x_t^i = q_t^i / p_t^i$ as a state of agent i in dynamic network, then the control gain — to achieve consensus in network — will correspond to the optimal job redistribution between agents in the stationary case [27]. Let the fraction $\frac{q_t^i}{p_t^i}$ denote *the load* of agent i at time t. Thus, it is enough to consider the problem of how to keep the equal load of all agents in the network.

Assume that $p_t^i \neq 0 \forall\, i$. Consider the control protocol (4), where $\forall\, i \in N$, $\forall\, t$ denote $\bar{N}_t^i = N_t^i$ and $b_t^{i,j} = p_t^j / p_t^i, \, , j \in N_t^i$.

As an example of such system consider the simulation for the computer network consisting of six computing agents.

We set the initial queue lengths, the productivities of agents and some initial network topology. Let p_j^t be constant $\forall t$.

For the considered case the dynamics of closed loop system (24) with local voting protocol (4) is as follows:

$$x_{t+1}^i = x_t^i - 1 + z_t^i / p_t^i + \alpha_t \sum_{j \in N_t^i} b_t^{i,j} (y_t^{i,j} / p_t^j - y_t^{i,i} / p_t^i). \qquad (28)$$

where α_t are step-sizes of control protocol, $y_t^{i,j}$ noisy and delayed observation about j-th agents queue length, z_t^i is the new job received by agent i at time t.

In Fig. 1, we can see the system operation in nonstationary case with local voting protocol (4). It means that new jobs can come to different nodes during the system work. We can see that the income of new jobs do not affect to the quality of the system work. It is a big advantage of the algorithm.

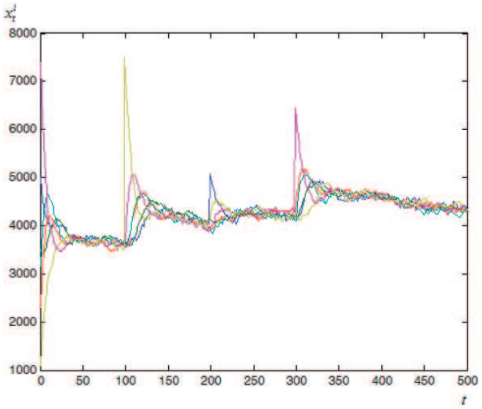


Fig. 1. The dynamics of the agents x_t^i for nonstationary case.

VI. CONCLUSION

In this paper, an approximate consensus problem for networks of nonlinear agents with switching topology, noisy and delayed measurements was studied. In contrast to the existing stochastic approximation-based control algorithms (protocols) local voting protocols with nonvanishing step size are proposed. Nonvanishing (e.g., constant) step size ensures better transients in the time-invariant case and provides bounded error in the case of time-varying loads and agent states. The price to pay is replacement of the almost sure or mean square convergence with an approximate one. To analyze dynamics of the closed loop system the so-called method of the averaged models is used. It allows to reduce complexity of the closed loop system analysis. In the paper new upper bounds for mean square distance between initial system and its approximate average model are proposed. The proposed upper bounds are used to obtain conditions for approximate consensus achievement.

This work was carried out during the tenure of an ERCIM "Alain Bensoussan" Fellowship Programme. The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement No. 246016. Also the work was supported by Russian Federal Program "Cadres" (agreements 8846, 8855), by RFBR (project 11-08-01218) and also by the SPRINT laboratory of SPbSU and Intel Corp.

REFERENCES

[1] R. Olfati-Saber, J. Fax, and R. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–233, 2007.

[2] W. Ren, R. Beard, and E. Atkins, "Information consensus in multivehicle cooperative control," *Control Systems, IEEE*, vol. 27, no. 2, pp. 71–82, 2007.

[3] C. Wu, *Synchronization in complex networks of nonlinear dynamical systems*. World Scientific Publishing Company Incorporated, 2007.

[4] W. Ren and R. Beard, *Distributed consensus in multi-vehicle cooperative control: theory and applications*. Springer, 2007.

[5] F. Bullo, J. Cortés, and S. Martinez, *Distributed control of robotic networks: a mathematical approach to motion coordination algorithms*. Princeton University Press, 2009.

[6] P. Antsaklis and J. Baillieul, "Guest editorial special issue on networked control systems," *Automatic Control, IEEE Transactions on*, vol. 49, no. 9, pp. 1421–1423, 2004.

[7] C. Abdallah and H. Tanner, "Complex networked control systems: introduction to the special section," *Control Systems, IEEE*, vol. 27, no. 4, pp. 30–32, 2007.

[8] P. Antsaklis and J. Baillieul, "Special issue on technology of networked control systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 5–8, 2007.

[9] D. Armbruster, *Networks of Interacting Machines: Production Organization in Complex Industrial Systems and Biological Cells*. World Scientific Publishing Company Incorporated, 2005, vol. 3.

[10] I. Kalaev and E. Melnik, *Decentralized computer control systems*. Rostov on Don: UNC RAN, 2011.

[11] M. Huang, "Stochastic approximation for consensus: a new approach via ergodic backward products," *IEEE Transactions on Automatic Control*, vol. 57, no. 12, pp. 2994–3008, 2012.

[12] R. Olfati-Saber and R. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *Automatic Control, IEEE Transactions on*, vol. 49, no. 9, pp. 1520–1533, 2004.

[13] W. Ren and R. Beard, "Consensus seeking in multiagent systems under dynamically changing interaction topologies," *Automatic Control, IEEE Transactions on*, vol. 50, no. 5, pp. 655–661, 2005.

[14] J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *Automatic Control, IEEE Transactions on*, vol. 31, no. 9, pp. 803–812, 1986.

[15] M. Huang and J. Manton, "Coordination and consensus of networked agents with noisy measurements: stochastic algorithms and asymptotic behavior," *SIAM Journal on Control and Optimization*, vol. 48, no. 1, pp. 134–161, 2009.

[16] T. Li and J. Zhang, "Mean square average-consensus under measurement noises and fixed topologies: Necessary and sufficient conditions," *Automatica*, vol. 45, no. 8, pp. 1929–1936, 2009.

[17] N. Amelina, A. Lada, I. Mayiorov, P. Skobelev, and A. Tsarev, "Cargo transportation models analysis using multi-agent adaptive real-time truck scheduling system," *Problemy Upravleniya*, vol. 6, pp. 31–37, 2011.

[18] N. Amelina, A. Fradkov, and K. Amelin, "Approximate consensus in multi-agent stochastic systems with switched topology and noise," in *Proc. IEEE 2012 Multiconference on Systems and Control (MSC2012)*, Dubrovnik, Croatia, 2012, pp. 445–450.

ISBN: 978-1-61208-245-5

[19] D. Derevitskii and A. Fradkov, *Applied theory of discrete adaptive control systems*. Moscow: Nauka, 1981.

[20] H. Kushner, "Convergence of recursive adaptive and identification procedures via weak convergence theory," *Automatic Control, IEEE Transactions on*, vol. 22, no. 6, pp. 921–930, 1977.

[21] D. Derevitskii and A. Fradkov, "Two models for analysis the dynamics of adaptation algorithms," *Automation and Remote Control*, no. 1, pp. 59–67, 1974.

[22] L. Ljung, "Analysis of recursive stochastic algorithms," *Automatic Control, IEEE Transactions on*, vol. 22, no. 4, pp. 551–575, 1977.

[23] S. Meerkov, "On simplification of slow markovian walks description," *Automation and Remote Control*, no. 3, pp. 6–75, 1972.

[24] A. Fradkov, "Continuous-time averaged models of discrete-time stochastic systems: Survey and open problems," in *Proc. 2011 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC)*. Orlando, Florida, USA: IEEE, 2011, pp. 2076–2081.

[25] I. Gihman and A. Skorohkod, *Stochastic Differential Equations*. Kiev: Nauk. dumka, 1968.

[26] S. Bernstein, "Stochastic difference equations and stochastic differential equations," *Sobr. soch. in 4-th Vol.*, vol. 4, pp. 484–542, 1964.

[27] N. Amelina and A. Fradkov, "Approximate consensus in the dynamic stochastic network with incomplete information and measurement delays," *Automation and Remote Control*, vol. 73, no. 11, pp. 1765–1783, 2012.

Optimization of Achievable Information Rates and Number of Levels in Multilevel Flash Memories

Xiujie Huang*, Aleksandar Kavcic*, Xiao Ma†, Guiqiang Dong‡ and Tong Zhang‡
*Department of Electrical Engineering, University of Hawaii, Honolulu, HI 96822 USA
†Department of Electronics and Communication Engineering, Sun Yat-sen University, Guangzhou, GD 510006 China
‡Department of Electrical, Computer and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180-3590 USA
Email: xiujie@hawaii.edu, maxiao@mail.sysu.edu.cn, dongguiqiang@gmail.com, tong.zhang@ieee.org

***Abstract*—This paper is concerned with channel modeling and capacity evaluation of the multilevel flash memory with m levels. The m-level flash memory channel (m-LFMC) is modeled as an m-amplitude-modulation channel with input-dependent additive Gaussian noise whose standard deviation depends on the channel input. Then the capacity of an m-LFMC is given. The determination of the capacity can be transformed into a two-step optimization problem, which can be numerically solved by an alternating iterative algorithm. This algorithm delivers not only the optimal input/level distribution but also the optimal values of levels. This algorithm also delivers the optimized number of levels at any given voltage-to-deviation ratio. Numerical results are presented to show the consistency with well-known Smith's results for the amplitude-limited AWGN channel and the applicability of the modeling method.**

Keywords-Channel capacity; input-dependent additive Gaussian noise (ID-AGN) channel; multilevel flash memory.

I. Introduction

As the demand for non-volatile data storage increases, flash memories are gaining attention. The original flash memory used only two levels to store one bit in one memory cell. However, a modern mainstream flash memory is a multilevel flash memory (MLFM), which stores more than one bit in one memory cell to improve the storage density and reduce the bit cost of flash memories. The first MLFM product was presented by Bauer *et al.* in [1], which had four levels and stored two bits in one memory cell. Later, MLFMs were investigated and/or designed by many researchers, such as the 4-level MLFM in [2] and the Intel StrataFlash$^{\text{TM}}$ 4-level memory in [3], the 8-level MLFMs in [4, 5] and the 16-level MLFMs in [6, 7].

As the number of levels increases, the capability of the MLFM could be enhanced and the reliability could be decreased. On one hand, due to the complexity of the configuration (including the programming and reading techniques and inter-cell interferences), it is complicated to model precisely the MLFM channel. Hence research on the information-theoretic channel capacity is sporadic, such as [8, 9]. In [8], the MLFM was quantized to different discrete memoryless channels (DMCs) by introducing different reading numbers of reference voltages. By optimizing the reference voltages, the mutual information of DMC could be maximized, and then the achievable rate of the MLFM could be obtained. On the other hand, to guarantee the reliability, two approaches, i.e., signal processing methods and error correcting codes (ECCs), are investigated and applied in MLFMs. Two examples of signal processing methods are data postcompensation and data predistortion [10], which could tolerate cell-to-cell interference in MLFMs. Examples of ECCs include BCH codes [11], Reed-Solomon codes [12, 13], LDPC codes [8, 14], trellis coded modulation [11, 15, 16] and rank modulation [17, 18].

In this paper, we focus on the MLFM with m levels. To answer the question on the information-theoretic capability of the MLFM with m levels, we need to solve a key problem, i.e., channel modeling. The simplest model is the input-independent additive Gaussian noise channel, which is also called amplitude-limited AWGN channel [19, 20]. In [19, 20], Smith proved that the capacity of the amplitude-limited AWGN channel is achieved by a unique discrete random variable taking values on a finite alphabet. Based on the current techniques and configuration, there exist two universal phenomena for the MLFM. One is that the device degrades with age and the degradation varies from cell to cell as mentioned in [3, 21]. The other is the cell-to-cell interference as mentioned in [10, 22]. In this paper, we consider only the former. In this case, we model the m-level flash memory channel (m-LFMC) as an m-amplitude-modulation (m-AM) channel with input-dependent additive Gaussian noise (ID-AGN) whose standard deviation depends on the input (The m-AM with ID-AGN channel can also be regarded as a constrained communication system [23, 24].). Then we give the channel capacity and present a numerical method to evaluate it.

Structure: The remainder of this paper is organized as follows. The channel model of the MLFM with m levels is introduced in Section II, and the channel capacity is given in Section III. Section IV presents an alternating iterative algorithm to evaluate a lower bound on the capacity. Numerical results and discussions are shown in Section V, followed by the conclusion in Section VI.

II. Channel Model

For an MLFM with m levels, each level has an intended *threshold voltage* [1]. Affected by the configuration (including the programming and reading techniques and inter-cell interferences) of the flash memory and device aging, the threshold voltage shift may vary from cell to cell. Hence, each level corresponds to a threshold voltage range [1]. In this

ISBN: 978-1-61208-245-5

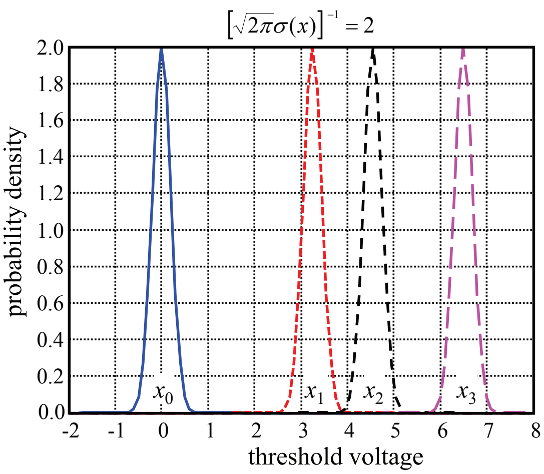


Fig. 1. A threshold voltage distribution model for a 4-LFMC, in which the noise at each level has the same variance $\sigma(x) = \frac{1}{2\sqrt{2\pi}}$.

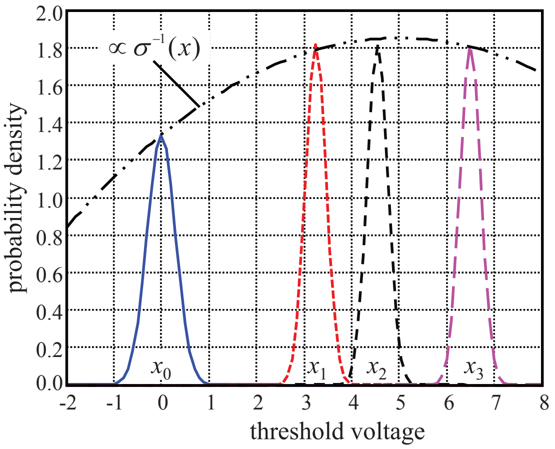


Fig. 2. A threshold voltage distribution model for a 4-LFMC, in which the first level x_0 is the most noisy level while the other three levels have roughly the same noise.

paper, we focus on only the variation caused by device aging. For mathematical modeling, the variation of the threshold voltage is usually approximated by a Gaussian distribution and characterized by its probability density function (pdf). The following example illustrates the models of threshold voltage distributions for a 4-LFMC.

Example 1 (Threshold Voltage Distributions of a 4-LFMC): Consider a 4-LFMC. Let the intended voltages of the four levels be $x_0 = 0$, $x_1 = 3.25$, $x_2 = 4.55$ and $x_3 = 6.5$. By default, the threshold voltage distribution model of the manufactured 4-LFMC is shown in Fig. 1, where the noise at each level has the same variance and the pdf of the output for each level is depicted. As documented in [3,21], the number of electrons of a cell decreases with time and some cells become defective as time elapses, which means that the cell has a long but finite lifetime and the degradation varies from cell to cell. Consequentially, the performance of the 4-LFMC gets gradually worse as the device ages. Suppose that, after three years, the threshold voltage distribution model of the 4-LFMC is shown in Fig. 2, where every level experiences more noise than in Fig. 1 and the first level x_0 is the most noisy level while the other three levels have almost the same noise. Again, suppose that, after five years, the threshold voltage distribution model of the 4-LFMC is shown in Fig. 3, where every level has even more noise than in Fig. 2, while the first level x_0 and the last level x_3 are respectively the most noisy levels. This behavior can be easily modeled by a function $\sigma(x)$, which depends on the age of the device. As shown in Figs. 2 and 3, the dash-dot-dot curve $\left[\sqrt{2\pi}\sigma(x)\right]^{-1}$ is (approximately) the envelope of the peaks of the level-output-pdfs. In Fig. 1, the curve $\sigma(x)$ is assumed to be a constant, i.e., $\sigma(x) = \frac{1}{2\sqrt{2\pi}}$. ❑

Models similar to Figs. 2 and 3 for the 4-LFMC were introduced in [3,11,14]. In [3,11], the model of the 2 bits/cell (i.e., 4-level) NOR flash memory was presented, in which the first level x_0 had the highest noise variance and the last level x_3 had the second highest noise variance while the two middle levels had almost the same noise variances. In [14], the model of a 4-level NAND flash memory was derived by accounting for the cell-to-cell interference, in which the first level x_0 had the highest Gaussian noise and the other three levels had almost the same noises characterized by bounded Gaussian variables.

In this paper, an m-LFMC is modeled as an m-AM channel with ID-AGN. Specifically, it is characterized as follows.

1) Let X, Y and W denote the channel input, the channel output and the channel noise random variables, respectively. They have the relation:
$$Y = X + W. \tag{1}$$

2) The channel input X takes values from a finite alphabet $\mathcal{X}^{(m)} \triangleq \{x_0, x_1, \cdots, x_{m-1}\}$ under the constraint
$$a \le x_0 < x_1 < x_2 < \cdots < x_{m-2} < x_{m-1} \le b \tag{2}$$
where a and b are the respective lowest and highest possible threshold voltages, and their difference is denoted by $V_m \triangleq b - a$. The finite alphabet $\mathcal{X}^{(m)}$ is called

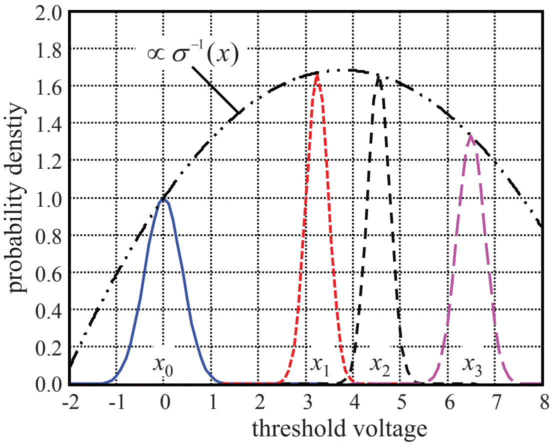


Fig. 3. A threshold voltage distribution model for a 4-LFMC, in which the first level x_0 and the last level x_3 are respectively the most noisy levels while the two middle levels x_1 and x_2 have roughly the same noise.

an m-amplitude-modulation (m-AM) signal set. Denote the collection of all such m-AM signal sets as $\mathscr{X}^{(m)}$, i.e., $\mathcal{X}^{(m)} \in \mathscr{X}^{(m)}$. In the following context, we also use the vector notation $\underline{x}$ to denote the m levels, i.e., $\underline{x} = (x_1, x_2, \cdots, x_{m-1})$.

3) The probability mass function (pmf) of X over $\mathcal{X}^{(m)}$ is denoted by $\underline{p} = (p_0, p_1, \cdots, p_{m-1})$ with $p_i = \Pr(X = x_i)$.
4) The noise W is an input-dependent additive Gaussian noise (ID-AGN). That is, the noise W has mean zero and variance depending on the channel input $x \in \mathcal{X}^{(m)}$, i.e., $W \sim \mathcal{N}\big(0, \sigma^2(x)\big)$. In this paper, the function $\sigma(x)$ is assumed to be continuous and differentiable.

Therefore, the channel transition pdf, i.e., the channel law, is

$$f_{Y|X,\sigma(\cdot)}(y|x) = \frac{1}{\sqrt{2\pi}\sigma(x)} \exp\left\{-\frac{(y-x)^2}{2\sigma^2(x)}\right\}. \quad (3)$$

And the pdf of the channel output Y can be obtained as

$$f_{Y,\sigma(\cdot)}(y) = \sum_{i=0}^{m-1} p_i \, f_{Y|X,\sigma(\cdot)}(y|x_i). \quad (4)$$

Recall Example 1 of 4-AM channels with ID-AGN. At the time of manufacturing, the noise standard deviations for all levels are considered to be constant; see Fig. 1. As the device ages, the noise standard deviations for different levels increase in different extents; see Figs. 2 and 3. That is, the noise standard deviations for an aged device are level-dependent.

III. Channel Capacity

From the last section, we know that the m-LFMC is modeled as an m-AM channel with ID-AGN, parameterized by the m-AM signal set $\mathcal{X}^{(m)}$, the pmf $\underline{p} = (p_0, p_1, \cdots, p_{m-1})$ and the standard deviation function $\sigma(\underline{x})$. Therefore, to express the information-theoretic capacity of the m-LFMC, we introduce a new notation different slightly from the conventional one by inserting the subscript $\big(\mathcal{X}^{(m)}, \sigma(\cdot)\big)$ into the mutual information expression, i.e.,

$$\begin{aligned} &I_{\mathcal{X}^{(m)},\sigma(\cdot)}(X;Y) \\ &\triangleq \sum_{i=0}^{m-1}\int_{-\infty}^{\infty} p_i \, f_{Y|X,\sigma(\cdot)}(y|x_i) \log\left(\frac{f_{Y|X,\sigma(\cdot)}(y|x_i)}{f_{Y,\sigma(\cdot)}(y)}\right) dy. \quad (5)\end{aligned}$$

Definition 1: The capacity of the m-LFMC with standard deviation function $\sigma(\cdot)$ is defined as

$$C_{m,\sigma(\cdot)} \triangleq \sup_{\mathscr{X}^{(m)},\{\underline{p}\}} I_{\mathcal{X}^{(m)},\sigma(\cdot)}(X;Y) \quad (6)$$

where the maximum is taken over all possible m-AM signal sets $\mathcal{X}^{(m)} = \{x_0, x_1, \cdots, x_{m-1}\} \in \mathscr{X}^{(m)}$ satisfying

$$a \le x_0 < x_1 < \cdots < x_{m-2} < x_{m-1} \le b \quad (7)$$

and all possible pmfs $\underline{p} = (p_0, p_1, \cdots, p_{m-1})$ satisfying

$$p_i \ge 0, \text{ and } \sum_{i=0}^{m-1} p_i = 1. \quad (8)$$

❑

Remark 1. Recall Smith's result that the capacity of the amplitude-limited AWGN channel is achieved by a unique discrete random variable taking values on a finite alphabet [19, 20]. The main two differences between the m-AM channel with ID-AGN and the amplitude-limited AWGN channel are: the noise in the former is input-dependent, while in the latter it is independent of inputs; and the number of inputs is fixed to be m in the former, while in the latter the optimal (capacity-achieving) number of inputs is obtained by optimization.

Remark 2. Comparing with Ungerboeck's results of average energy limited AWGN channel with amplitude modulation [25], there are three main differences. First, the m-AM channel with ID-AGN for an m-LFMC is not average energy limited but amplitude limited (in the interval $[a, b]$). Second, the m-AM signal set is not fixed but can be optimized in the evaluation of its capacity. Third, the input distribution is not uniform but can be optimized too.

One of the main objectives in capacity research is numerical evaluation. To this end, a comprehensive understanding is necessary and can provide a methodology of evaluation. The following proposition gives an insight into the capacity $C_{m,\sigma(\cdot)}$ of the m-LFMC.

Proposition 1: When $\underline{x}$ is given, the mutual information $I_{\mathcal{X}^{(m)},\sigma(\cdot)}(X;Y)$ is concave with respect to (w.r.t.) $\underline{p}$; when $\underline{p}$ is given, the mutual information $I_{\mathcal{X}^{(m)},\sigma(\cdot)}(X;Y)$ is continuous and differentiable w.r.t. $\underline{x}$. ❑

Proof sketch: The mutual information is expressed as

$$\begin{aligned} &I_{\mathcal{X}^{(m)},\sigma(\cdot)}(X;Y) \\ &= h_{\mathcal{X}^{(m)},\sigma(\cdot)}(Y) - \sum_{i=0}^{m-1} p_i \log\sigma(x_i) - \frac{1}{2}\log(2\pi e) \quad (9)\end{aligned}$$

since the noise is input-dependent. When $\underline{x}$ is given, due to the linearity of $\sum p_i \log\sigma(x_i)$, we can prove that the mutual information is concave w.r.t. $\underline{p}$ by using the same method as in [26]. When $\underline{p}$ is given, the composition of elementary functions in (5) is continuous and differentiable w.r.t. $\underline{x}$ because $\sigma(x)$ is assumed to be continuous and differentiable. ■

IV. Evaluation of a Lower Bound on Capacity

To evaluate the capacity (6) of the m-LFMC, we turn to a two-step optimization problem

$$\begin{aligned} C_{m,\sigma(\cdot)} = &\sup_{\underline{x}\in[a,b]^m} \sup_{\underline{p}\in[0,1]^m} I_{\mathcal{X}^{(m)},\sigma(\cdot)}(X;Y) \\ &\text{subject to} \begin{cases} a \le x_0 < x_1 < \cdots < x_{m-1} \le b \\ p_i \ge 0, \ i \in \{0, 1, \cdots, m-1\} \\ \sum\limits_{i=0}^{m-1} p_i = 1 \end{cases}. \end{aligned} \quad (10)$$

To solve the two-step optimization problem (10), we turn to two sub-problems.

Sub-problem I.

$$\begin{aligned} C(\underline{x}) = &\max_{\underline{p}\in[0,1]^m} I_{\mathcal{X}^{(m)},\sigma(\cdot)}(X;Y) \\ &\text{subject to} \begin{cases} p_i \ge 0, \ i \in \{0, 1, \cdots, m-1\} \\ \sum\limits_{i=0}^{m-1} p_i = 1 \end{cases}. \end{aligned} \quad (11)$$

When $\underline{x}$ is given, Sub-problem I is a conventional capacity problem for memoryless channel with finite inputs. Due to the concavity of the mutual information w.r.t. $\underline{p}$ shown in

Proposition 1, the well-known algorithm, Blahut-Arimoto algorithm (BAA) [27–29] can be used to solve Sub-problem I.

Sub-problem II.

$$\begin{array}{l} C(\underline{p}) = \max\limits_{\underline{x}\in[a,b]^m} I_{\mathcal{X}^{(m)},\sigma(\cdot)}(X;Y) \\ \text{subject to } a \le x_0 < x_1 < \cdots < x_{m-1} \le b \end{array} . \qquad (12)$$

The Karush-Kuhn-Tucker (KKT) conditions [30] of the sub-problem are that there exists $\mathbf{v}^* = (\underline{x}^*, \lambda^*, \mu^*)$ such that

$$\left\{ \begin{array}{rcll} \left.\frac{\partial I}{\partial x_0}\right|_{\mathbf{v}^*} & = & -\lambda^*, & \\ \left.\frac{\partial I}{\partial x_{m-1}}\right|_{\mathbf{v}^*} & = & \mu^*, & \\ \left.\frac{\partial I}{\partial x_i}\right|_{\mathbf{v}^*} & = & 0, & i \in \{1,2,\cdots,m-2\} \\ x_0^* & \ge & a, & \\ x_{m-1}^* & \le & b, & \\ x_{i-1}^* & < & x_i^*, & i \in \{1,2,\cdots,m-1\} \\ \lambda^* & \ge & 0, & \\ \mu^* & \ge & 0, & \\ \lambda^*(x_0^* - a) & = & 0, & \\ \mu^*(x_{m-1}^* - b) & = & 0. & \end{array} \right. \qquad (13)$$

Note that the solution of (13) may be sub-optimal (a local solution) since the concavity of the mutual information w.r.t. $\underline{x}$ is unknown; see the Appendix for a method to solve (13).

Based on the two sub-problems, an alternating iterative scheme is presented to solve problem (10). At each iteration, the two-stage alternating strategy shown below is employed.

Stage 1. Fix $\underline{x}$. Use BAA to obtain the optimal $\underline{p}^*$

$$\underline{p}^* = \arg\max_{\underline{p}} I_{\mathcal{X}^{(m)},\sigma(\cdot)}(X;Y). \qquad (14)$$

Stage 2. Fix $\underline{p}$. Solve (13) to obtain a better $\underline{x}^*$ such that

$$I_{\mathcal{X}^{(m)},\sigma(\cdot)}(X;Y)\Big|_{\underline{x}^*} \ge I_{\mathcal{X}^{(m)},\sigma(\cdot)}(X;Y)\Big|_{\underline{x}}. \qquad (15)$$

From the discussion of Sub-problem II, $\underline{x}^*$ may be a local solution. This sub-optimality also implies that a lower bound on the capacity $C_{m,\sigma(\cdot)}$ of the m-LFMC is evaluated.

V. Numerical Results and Discussions

In this section, we numerically compute lower bounds on capacities of different m-LFMCs using the alternating iterative scheme of Section IV. We also interpret the results and put them in context with respect to prior work [19, 20].

Let the lowest and highest threshold voltages be $a = 0$ and $b = 6.5$, respectively. Then the difference is $V_m = b - a = 6.5$. We introduce a new parameter $\sigma > 0$ that severs as the varying noise parameter in our computations. Let $q_i(x)$ where $i \in \{1, 2, 3\}$ be continuous and differentiable functions as shown in Fig. 4. We consider three different standard deviation functions $\sigma(x)$, denoted as

$$\sigma_i(x) = q_i(x) \cdot \sigma, \text{ where } i \in \{1, 2, 3\}. \qquad (16)$$

We allow the parameter σ to vary such that the *voltage-to-deviation ratio* (VDR) V_m/σ acts as an effective signal-to-noise ratio. We assume that the intended threshold voltage level x_0 (usually corresponding to the erased state) is 0.

We present results for $m \le 5$, i.e., we consider multilevel flash memory channels with at most 5 levels. We consider three different m-LFMCs whose standard deviation functions are $\sigma_1(x)$, $\sigma_2(x)$ and $\sigma_3(x)$. The lower bounds on capacities of m-LFMCs with deviation functions $\sigma_1(x)$, $\sigma_2(x)$ and $\sigma_3(x)$ are shown in Figs. 5, 6 and 7, respectively.

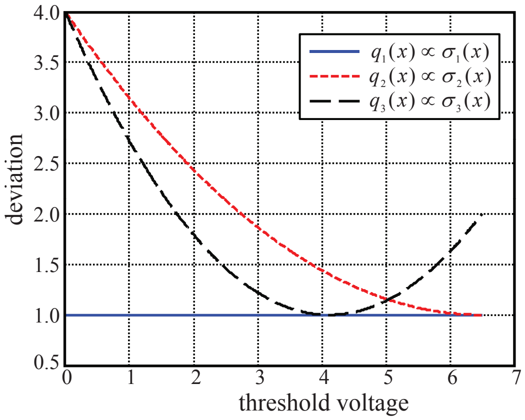


Fig. 4. The standard deviation functions $\sigma_i(x)$ of the input-dependent Gaussian noise W: $\sigma_i(x) \propto q_i(x)$, where $i \in \{1, 2, 3\}$.

From Fig. 5, we make the following observations.

1) When the VDR is less than 10.5 dB, i.e., $20\log_{10}(V_m/\sigma) \le 10.5$ dB, 2-LFMC, 3-LFMC, 4-LFMC and 5-LFMC have the same rates.
2) When the VDR is less than 15.5 dB, 3-LFMC, 4-LFMC and 5-LFMC have the same rates.
3) When the VDR is less than 18.5 dB, 4-LFMC and 5-LFMC have the same rates.

Furthermore, we observe (not explicitly shown in the figure) that in the VDR regime between 10.5 dB and 15.5 dB, the optimized lower bound is achieved with $m^* = 3$ levels, even if, say, the constraint allows up to $m = 5$ levels. This is consistent with prior work [19, 20] which showed that for the amplitude-limited AWGN channel, the capacity is achieved by a discrete channel input distribution over a *finite* alphabet. In other words, for a fixed VDR, there is an optimal number of levels m^* for a given multilevel flash memory channel. Increasing the number of levels m beyond m^* does not further increase the capacity (nor the computed lower bound.)

These observations imply that 2-LFMC, 3-LFMC and 4-LFMC can achieve the capacity

$$C_{\sigma_1(\cdot)} = \max_m C_{m,\sigma_1(\cdot)}$$

in the cases of VDR $\le$ 10.5 dB, 10.5 dB $<$ VDR $\le$ 15.5 dB and 15.5 dB $<$ VDR $\le$ 18.5 dB, respectively. In other words, in the view of capacity, at a given VDR less than 10.5 dB, a 2-LFMC is "optimal"; at a given VDR less than 15.5 dB, a 3-LFMC is "optimal"; at a given VDR less than 18.5 dB, a 4-LFMC is "optimal". Naturally, as the VDR increases, the optimal number of levels doesn't decrease.

Similar conclusions hold for the other two channels with noise standard deviation functions $\sigma_2(x)$ and $\sigma_3(x)$. Namely, even if the constraint is set to be, say, $m = 5$, at low VDRs the optimal number of threshold levels m^* is less than 5. For example, as shown in Fig. 7, the optimal number of levels is $m^* = 4$ in the VDR regime between 13 dB and

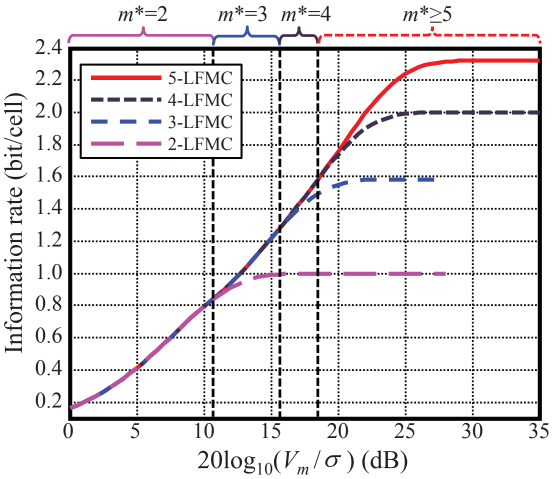


Fig. 5. The information rates of m-LFMCs with $m \in \{2, 3, 4, 5\}$ when the standard deviation function is $\sigma_1(x)$. The numbers m^* on the top of the figure indicate that 2-LFMC, 3-LFMC and 4-LFMC can achieve the (computed) maximum rates in the cases of VDR $\leq$ 10.5 dB, 10.5 dB $<$ VDR $\leq$ 15.5 dB and 15.5 dB $<$ VDR $\leq$ 18.5 dB, respectively.

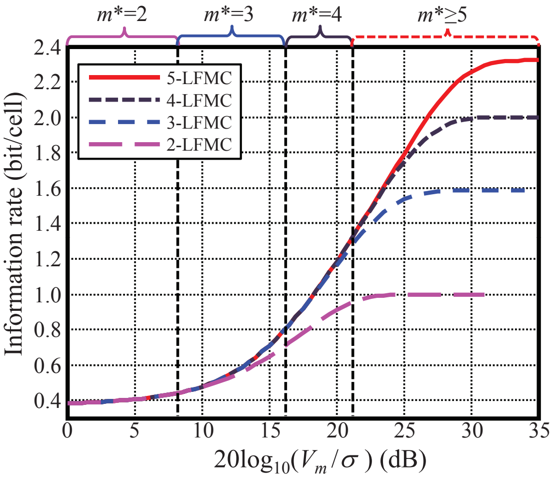


Fig. 6. The information rates of m-LFMCs with $m \in \{2, 3, 4, 5\}$ when the standard deviation function is $\sigma_2(x)$. The numbers m^* on the top of the figure indicate that 2-LFMC, 3-LFMC and 4-LFMC can achieve the (computed) maximum rates in the cases of VDR $\leq$ 8.5 dB, 8.5 dB $<$ VDR $\leq$ 16 dB and 16 dB $<$ VDR $\leq$ 21 dB, respectively.

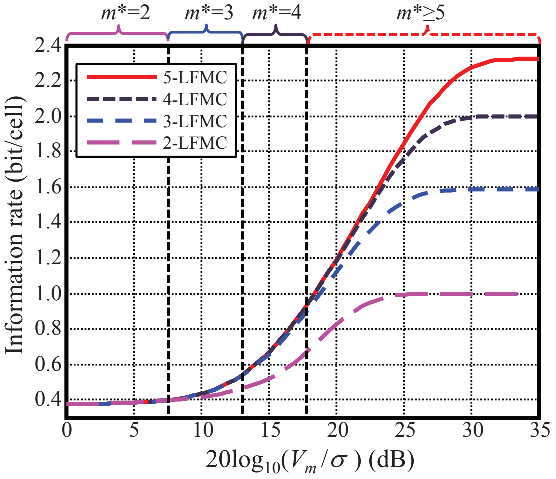


Fig. 7. The achievable rates of m-LFMCs with $m \in \{2, 3, 4, 5\}$ when the standard deviation function is $\sigma_3(x)$. The numbers m^* on the top of the figure indicate that 2-LFMC, 3-LFMC and 4-LFMC can achieve the (computed) maximum rates in the cases of VDR $\leq$ 7.5 dB, 7.5 dB $<$ VDR $\leq$ 13.5 dB and 13.5 dB $<$ VDR $\leq$ 18 dB, respectively.

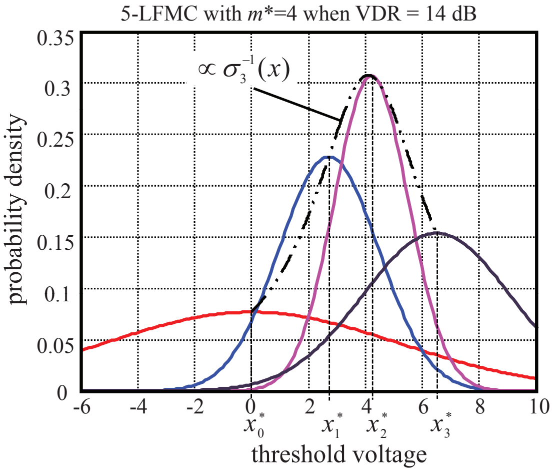


Fig. 8. The pdfs of channel output distributions around the optimal threshold voltage levels when the m-LFMC with the standard deviation function $\sigma_3(x)$ and $m = 5$ is used at VDR $= 14$ dB. The optimal number of levels m^* is 4 with assignment $x_0^* = 0$, $x_1^* \approx 2.718$, $x_2^* \approx 4.212$ and $x_3^* = 6.5$ and pdf $p_0^* \approx 0.274$, $p_1^* \approx 0.171$, $p_2^* \approx 0.271$ and $p_3^* \approx 0.284$.

17.5 dB even when a 5-LFMC with noise standard deviation function $\sigma_3(x)$ is considered. In the case that VDR is equal to 14 dB, using the lower bound optimizing algorithm presented in Section IV, we obtain that the optimal number of levels is $m^* = 4$ with assignment $x_0^* = 0$, $x_1^* \approx 2.718$, $x_2^* \approx 4.212$ and $x_3^* = 6.5$ and pdf $p_0^* \approx 0.274$, $p_1^* \approx 0.171$, $p_2^* \approx 0.271$ and $p_3^* \approx 0.284$, shown in Fig. 8. Again, this is consistent with the literature [19, 20] for the amplitude-limited AWGN channel, even though in m-LFMC the noise standard deviation $\sigma(x)$ is input-dependent.

VI. Conclusions

In this paper, the m-LFMC was modeled as an m-AM channel with ID-AGN, in which the standard deviation of noise depends on the channel input. The capacity of the m-LFMC was given. The determination of the capacity is an optimization problem, which can be transformed into two optimization sub-problems. One can be solved by Blahut-Arimoto algorithm. The other can be solved by finding the solution to KKT conditions. Based on these, an alternating iterative algorithm was presented to evaluate a lower bound on the capacity of the m-LFMC. This algorithm delivered not only the optimal distribution of channel inputs but also the optimal values of channel inputs. Numerical results showed that at any given VDR there exists an optimal value m^* such that the capacity (or its lower bound) is achieved by an m^*-LFMC, and that increasing the number of levels m above m^* does not further increase the information rate for a fixed VDR.

Appendix: Solving KKT Conditions (13)

For convenience, we denote the pdfs $f_{Y|X,\sigma(\cdot)}(y|x_i)$ and $f_{Y,\sigma(\cdot)}(y)$ and the mutual information $I_{\mathcal{X}^{(m)},\sigma(\cdot)}(X;Y)$ as $f(y|x_i)$, $f(y)$ and $I(X;Y)$, respectively.

We compute partial derivatives of the mutual information $I(X;Y)$. To this end, we compute partial derivatives of the transition pdf $f(y|x_i)$ in (3) and the output pdf $f(y)$ in (4) as, for all $i \in \{0, 1, \cdots, m-1\}$,

$$\frac{\partial f(y|x_i)}{\partial x_i} = \begin{cases} f(y|x_i)\left[-\frac{\sigma'(x_i)}{\sigma(x_i)} + \frac{y-x_i}{\sigma^2(x_i)} + \frac{(y-x_i)^2\sigma'(x_i)}{\sigma^3(x_i)}\right], & \text{if } i=j \\ 0, & \text{if } i \neq j \end{cases} \tag{17}$$

where $\sigma'(x_i) \triangleq \frac{d\sigma(x)}{d\,x_i}$ denotes the derivative of $\sigma(x_i)$ w.r.t. x_i. Then, according to (9), partial derivatives of the mutual information are obtained as

$$\begin{aligned}
\frac{\partial}{\partial x_i} I(X;Y) &= -\int_{-\infty}^{\infty} \frac{\partial}{\partial x_i}\left(f(y)\ln f(y)\right)dy - \frac{p_i\sigma'(x_i)}{\sigma(x_i)} \\
&= \left[-\frac{p_i\sigma'(x_i)}{\sigma^3(x_i)}\int_{-\infty}^{\infty} f(y|x_i)\ln f(y)dy\right]\cdot x_i^2 \\
&\quad + \left[\frac{2p_i\sigma'(x_i)}{\sigma^3(x_i)}\int_{-\infty}^{\infty} yf(y|x_i)\ln f(y)dy\right. \\
&\qquad \left. + \frac{p_i}{\sigma^2(x_i)}\int_{-\infty}^{\infty} f(y|x_i)\ln f(y)dy\right]\cdot x_i \\
&\quad + \left[-\frac{p_i\sigma'(x_i)}{\sigma^3(x_i)}\int_{-\infty}^{\infty} y^2 f(y|x_i)\ln f(y)dy\right. \\
&\qquad - \frac{p_i}{\sigma^2(x_i)}\int_{-\infty}^{\infty} yf(y|x_i)\ln f(y)dy \\
&\qquad \left. + \frac{p_i\sigma'(x_i)}{\sigma(x_i)}\int_{-\infty}^{\infty} f(y|x_i)\ln f(y)dy - \frac{p_i\sigma'(x_i)}{\sigma(x_i)}\right] \\
&\triangleq A_i x_i^2 + B_i x_i + C_i.
\end{aligned} \tag{18}$$

Solving KKT conditions (13) is equivalent to finding quantities $(\underline{x}, \lambda, \mu)$ that satisfy the equalities

$$\begin{cases} A_0 x_0^2 + B_0 x_0 + (C_0 + \lambda) = 0 \\ \lambda(x_0 - a) = 0 \end{cases}, \tag{19a}$$

$$\begin{cases} A_{m-1} x_{m-1}^2 + B_{m-1} x_{m-1} + (C_{m-1} - \mu) = 0 \\ \mu(x_{m-1} - b) = 0 \end{cases}, \tag{19b}$$

$$A_i x_i^2 + B_i x_i + C_i = 0, \ i \in \{1, 2, \cdots, m-2\}, \tag{19c}$$

and the inequalities

$$\begin{cases} \lambda \geq 0 \\ \mu \geq 0 \\ a \leq x_0 < x_1 < \cdots < x_{m-2} < x_{m-1} \leq b \end{cases}. \tag{20}$$

Note that all quantities A_i, B_i and C_i depend on the input vector $\underline{x}$ and the standard deviation function $\sigma(\cdot)$ when the pmf $\underline{p}$ is given. To find the solution to the KKT conditions (19) by an iterative method, we assume that quantities A_i, B_i and C_i are independent of x_i. Then Eqns. (19) have at most $9 \times 2^{m-2}$ solutions. Moreover, under the full constraints in (20), the number of solutions may be much less than $9 \times 2^{m-2}$ (This happens in our numerical computations). Based on (19) and (20), we employ an iterative method to find a solution. Suppose that the input vector $\underline{x}^{(k)}$ is known at the beginning of the k-th iteration. Then solve Eqns. (19). Pick those solutions that satisfy all constraints in (20), and from them choose the one with the highest information rate as the improved input vector $\underline{x}^{(k+1)}$.

Acknowledgment

This work was supported by the collaborative NSF grants ECCS-1128705 and ECCS-1128148, and by the NSFC (No. 61172082).

References

[1] M. Bauer, R. Alexis, and et al., “A multilevel-cell 32Mb flash memory,” in *IEEE ISSCC Dig. Tech. Papers*, San Francisco, CA, Feb. 1995, pp. 132–133, 351.

[2] T.-S. Jung, Y.-J. Choi, and et al., “A 117-mm2 3.3-V only 128-Mb multilevel NAND flash memory for mass storage applications,” *IEEE Journal of Solid-State Circuits*, vol. 31, no. 11, pp. 1575–1583, Nov. 1996.

[3] G. Atwood, A. Fazio, D. Mills, and B. Reaves, “Intel StrataFlash$^{\text{TM}}$ memory technology overview,” *Intel Technology Journal*, pp. 1–8, 4th Quarter 1997.

[4] Y. Li, S. Lee, and et al., “A 16Gb 3b/cell NAND flash memory in 56nm with 8MB/s write rate,” in *IEEE ISSCC Dig. Tech. Papers*, San Francisco, CA, Feb. 2008, pp. 506–507.

[5] T. Futatsuyama, N. Fujita, and et al., “A 113mm2 32Gb 3b/cell NAND flash memory,” in *IEEE ISSCC Dig. Tech. Papers*, San Francisco, CA, Feb. 2009, pp. 242–243.

[6] N. Shibata, H. Maejima, and et al., “A 70nm 16Gb 16-level-cell NAND flash memory,” in *IEEE VLSI Circuits*, 2007, pp. 190–191.

[7] C. Trinh, N. Shibata, and et al., “A 5.6MB/s 64Gb 4b/cell NAND flash memory in 43nm CMOS,” in *IEEE ISSCC Dig. Tech. Papers*, San Francisco, CA, Feb. 2009, pp. 246–247, 247a.

[8] J. Wang, T. Courtade, H. Shankar, and R. D. Wesel, “Soft information for LDPC decoding in flash: mutual-information optimized quantization,” in *Proc. IEEE GLOBECOM 2011*, Houston, Texas, USA, Dec. 2011.

[9] A. Jiang, H. Li, and J. Bruck, “On the capacity and programming of flash memories,” *IEEE Trans. Inform. Theory*, vol. 58, no. 3, pp. 1549–1564, Mar. 2012.

[10] G. Dong, S. Li, and T. Zhang, “Using data postcompensation and predistortion to tolerate cell-to-cell interference in MLC NAND flash memory,” *IEEE Trans. Circuits Syst.–I: Reg. Papers*, vol. 57, no. 10, pp. 2718–2728, Oct. 2010.

[11] F. Sun, S. Devarajan, K. Rose, and T. Zhang, “Design of on-chip error correction systems for multilevel NOR and NAND flash memories,” *IET Circuits Devices Syst.*, vol. 1, no. 3, pp. 241–249, 2007.

[12] J. Chen and P. H. Siegel, “Markov processes asymptotically achieve the capacity of finite-state intersymbol interference channels,” *IEEE Trans. Inform. Theory*, vol. 54, no. 3, pp. 1295–1303, Mar. 2008.

[13] B. M. Kurkoshi, “The E8 lattice and error correction in multi-level flash memory,” in *Proc. IEEE International Conference on Communications*, Kyoto, Japan, June 5-9 2011, pp. 1–5.

[14] G. Dong, N. Xie, and T. Zhang, “On the use of soft-decision error-correction codes in NAND flash memory,” *IEEE Trans. Circuits Syst.–I: Reg. Papers*, vol. 58, no. 2, pp. 429–439, Feb. 2011.

[15] H. Lou and C. Sundberg, “Increasing storage capacity in multilevel memory cells by means of communications and signal processing techniques,” *IEE Proc.-Circuits Devices Syst.*, vol. 147, no. 4, pp. 229–236, Aug. 2000.

[16] S. Soldà, D. Vogrig, A. Bevilacqua, A. Gerosa, and A. Neviani, “Analog decoding of trellis coded modulation for multi-level flash memories,” in *Proc. the 2008 IEEE International Symposium on Circuits and Systems (ISCAS 2008)*, Seattle, U.S.A., May 18-21 2008, pp. 744–747.

[17] A. Jiang, R. Mateescu, M. Schwartz, and J. Bruck, “Rank modulation for flash memories,” *IEEE Trans. Inform. Theory*, vol. 55, no. 6, pp. 2659–2673, Jun. 2009.

[18] Z. Wang and J. Bruck, “Partial rank modulation for flash memories,” in *Proc. IEEE Intern. Symp. on Inform. Theory*, Austin, Texas, U.S.A., June 13-18 2010, pp. 864–868.

[19] J. G. Smith, “On the information capacity of peak and average power constrained gaussian channels,” Ph.D. dissertation, University of California, Berkeley, California, Dec. 1969.

[20] ——, “The information capacity of amplitude-and variance-constrained scalar Gaussian channels,” *Information and Control*, vol. 18, pp. 203–219, 1971.

[21] Kingston, “Flash memory guide,” Kingston, Tech. Rep., 2011.

[22] J.-D. Lee, S.-H. Hur, and J.-D. Choi, “Effects of floating-gate interference on NAND flash memory cell operation,” *IEEE Electron Device Letters*, vol. 23, no. 5, pp. 264–266, May 2002.

[23] S. Shamai, “Information theoretic aspects of constrained systems,” in *MSRI Workshop on Information Theory*, Berkeley, California, U.S.A., Feb. 25 - Mar. 1 2002.

[24] ——, “Information theoretic aspects of constrained cell-sites cooperataion,” in *IEEE 26-th Convention of Electrical and Electronics Engineers in Israel*, Eilat, Israel, Nov. 17-20 2010, p. 000086.

[25] G. Ungerboeck, “Channel coding with multilevel/phase signals,” *IEEE Trans. Inform. Theory*, vol. 28, no. 1, pp. 55–67, Jan. 1982.

[26] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley & Sons, Inc, 1991.

[27] S. Arimoto, “An algorithm for computing the capacity of arbitrary discrete memoryless channels,” *IEEE Trans. Inform. Theory*, vol. IT-18, no. 1, pp. 14–20, Jan. 1972.

[28] R. E. Blahut, “Computation of channel capacity and rate distortion functions,” *IEEE Trans. Inform. Theory*, vol. IT-18, no. 4, pp. 460–473, Jul. 1972.

[29] A. Kavčić, “On the capacity of Markov sources over noisy channels,” in *Proc. IEEE GLOBECOM 2001*, vol. 5, San Antonio, TX, USA, Nov. 25-29 2001, pp. 2997–3001.

[30] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge: Cambridge University Press, 2004.

Implementation Challenges in Rich Communication Suite-enhanced (RCS-e)

Nishank Trivedi
Lead Engineer
HCL Technologies Ltd.
Noida, India
Nishank.trivedi@hcl.com

Anurag Jain
Deputy General Manager
HCL Technologies Ltd.
Noida, India
anuragjn@hcl.com

***Abstract* - Everything is going mobile. This evolution is driven by video, cloud-based services, the Internet and machine-to-machine connectivity. It changes how people behave and how they leverage mobility to communicate and to improve their daily lives, through new and existing services. Users now demand connectivity anywhere and anytime with innovative services at minimal cost. Mobile service providers are exploiting the speed of 3G radio networks to offer new SIP-based interactive communications apart from basic Voice over IP (VoIP) services. One of these new emerging initiatives is Rich Communication Suite-enhanced (stated as RCS-e henceforth) which aims to seamlessly unify the communications experience by integrating traditional mobile telephony with new interactive services like presence, instant messaging and content sharing enabled by the enhanced address book of the mobile phone. This paper elaborates the challenges in implementation and roll-out of RCS-e technology, which includes combination of technical and business constraints. Finally, this paper not only compiles the unique challenges but it also touches upon the ways to deal with the implementation hurdles and using various studies and surveys it infers that RCS-e might well be the ideal combat technology against Over The Top applications (OTT apps) which are current leaders in the mobile telephony arena.**

Keywords-RCS-e; Mobile Phone Apps; Rich Communication Suite; Implementation challenges in RCS-e; OTT apps; IMS

I. INTRODUCTION

In today’s world of converged enterprise and consumer-oriented network services for triple play [1], there is a need as well as an opportunity before the operators to provide higher value add in terms of advanced collaboration services and monetize investments in high data intensive networks like IMS (IP Multimedia Subsystem) [2] and LTE (Long Term Evolution) [3]. RCS (Rich Communication Suite) is one of such initiatives led by network operators; network and device vendors that are expected to leverage SIP [4]-based IMS core infrastructure to provide advanced communication and collaboration. RCS shall provide users with an experience beyond voice and Short Message Service (SMS) by providing them with instant messaging, live video sharing and file transfer across any device on any network and with anyone in their mobile address book which is capable of handling RCS enriched data [5]. Figure 1 gives an idea of the future broadband subscription trend. Fixed narrowband voice subscriptions are expected to drop down from near 1100 million subscription lines in year 2008 to 800 million subscriptions in the year 2017 while in the same time period mobile broadband subscriptions are expected to rise steeply from mere 200 million subscriptions to 5 Billion subscription levels.

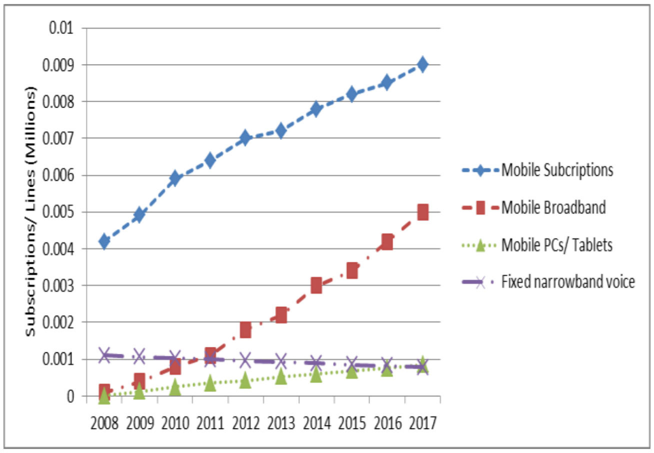


Figure 1-Mobile broadband subscription by 2017 [6]

Looking at the potential opportunity of 5 Billion mobile broadband subscriptions by 2017 [6] and growth in mobile data traffic by 20 times, the network infrastructure vendors are gearing up to this challenge and RCS is one of the technologies which provides a framework to monetize such investments. The RCS Initiative, catering to current growing interests in mobile applications and services, aims at providing an interoperable, convergent and rich communication environment. It uses an incremental and iterative methodology to continually add features, define interoperability conditions and reference guidelines. RCS reuses the capabilities of 3GPP specified IMS core system as the underlying service platform taking care of issues such as authentication, authorization, registration, charging and routing. The interests of the mobile network operator (MNO) or service provider and enterprise are aligning to tap this opportunity for VoLTE (Voice over LTE) as well as roll-out interoperable innovative communication services like rich call, video and social presence.

GSM Association (GSMA) [7] and Open Mobile Alliance (OMA) [8] are the key standard bodies creating specifications for RCS. The RCS documents created by GSMA provide a common, unambiguous reference point for operators and vendors alike to base their RCS implementations on. Up till now, five releases of RCS have been released which have incrementally added features in the RCS portfolio.

For example, RCS Release #1 laid the groundwork for further versions/releases by introducing the concept of voice and chat enrichment and a common evolved address book which facilitates chat and content share. RCS Release #2 aimed at extending the same features as RCS Release #1 to broadband users. RCS Release #3 allowed broadband access devices to be used as the primary devices in absence of mobile devices. RCS Release #4 extended the feature set to Long Term Evolution (LTE) and RCS Release #5 aimed at global interoperability. The RCS releases keep on enriching the feature set and also and continue to refine on existing implementations.

Probably based on the initial feedbacks, GSMA realized that the bulky feature sets offered by RCS were not alone doing enough to lure more users as well as operators/vendors into the RCS arena. Keeping an eye on the market, RCS-e was introduced.

RCS-e is a pruned version of RCS Release #2 which not only reuses capabilities offered in RCS Release #2 but also optimizes & refines its specifications. Several technical prerequisites which were “mandatory” to meet so as to be conformant with RCS Release #2 were changed to “optional” status, some conditions kept intact and while others were dropped. This was done in order to simplify the entry level technical conditions & prerequisites as much as possible so as to encourage the operators to implement RCS-e. Also, this was done so as to boost the market penetration curve twofold because simpler specs meant more operators were willing to invest in it as it offered more clarity and it reduced the go to market time as lightweight specs meant shorter implementation cycle time.

II. TECHNOLOGY OVERVIEW

RCS-e is based on a simple underlying basis of tying down the various available features and making them available from one single access point. It simplifies user experience and offers an easily available enriched communication experience. RCS-e feature availability and usage is driven from a central location known as Enhanced Address Book which as the name suggests, is an evolved version of the traditional contact book present in today’s mobile devices. Enhanced Address Book is a repository of contacts, which can be either RCS-e or the traditional contacts. (Traditional contact is referred to as the contact having just name and number details associated with it, similar to the contacts stored in today’s mobile address books/contact lists.) It enables storage of RCS-e contact and Legacy contacts, all at one place so as to avoid an additional burden on user for maintaining a separate contact list for both types of contacts.

The Enhanced address book allows legacy traditional operations such as ability to dial a contact (whether RCS-e or legacy) or sending a SMS. It also allows using advanced features such as starting a chat session with another RCS-e contact and ability to use multimedia content. Also, a service capability indication of each contact stored is provided which indicates the type of communication possible with each particular contact. In the Enhanced address book, as soon as user selects a contact it would be indicated whether that contact is capable of handling chat session, file transfer and video/image share capability or just simple call functions like SMS/call. For example, an RCS-e contact in the Enhanced Address Book might be shown as capable of handling of image sharing or/and chat capable, in the address book. Since, no such capability is present with non RCS-e contact, it would be depicted appropriately say by graying out the image/video share option, so that user clearly knows which contacts on its contact list are RCS-e or non RCS-e contacts and also what all services the user can avail to reach to a particular contact.

IMS core system forms the base for RCS-e services and enables peer-to-peer communication between RCS-e clients. Further, intercommunication between two RCS-e service providers is made possible using Network-to-Network Interface (NNI) mechanisms as explained in PRD-IR.90 (RCS Interworking Guidelines) [9][10].

Apart from the IMS core system many network entities form the RCS-e ecosystem. For instance, Presence server provides RCS-e clients with the current state information of the buddies in its buddy list. The IM server coupled with message store service provides the RCS-e client with a mix of traditional service such as chatting along with latest services like deferred messaging which ensures that no message sent by the RCS-e client is lost by delivering the message at a later time when the recipient party becomes available [11, 12]. In the same way, the Secure User Plane Location element could be used to exchange geo-location information, as part of Social Presence Information, which is an optional service for RCS-e [13].

A typical protocol message sequence flow for image share as an example is depicted below in Figure 2. User A, wants to share an image file, with Contact B present in its contact list and hence selects it. As soon as the user performs this action, the capability checks followed by session establishment processes are initiated with the help of Session Initiation Protocol. Once the session setup is complete, Message Session Relay protocol is used to send the intended image file from user A to User B. Once, the entire content is transferred from User A to user B, the session is gracefully ended using BYE, a Session Initiation Protocol method.

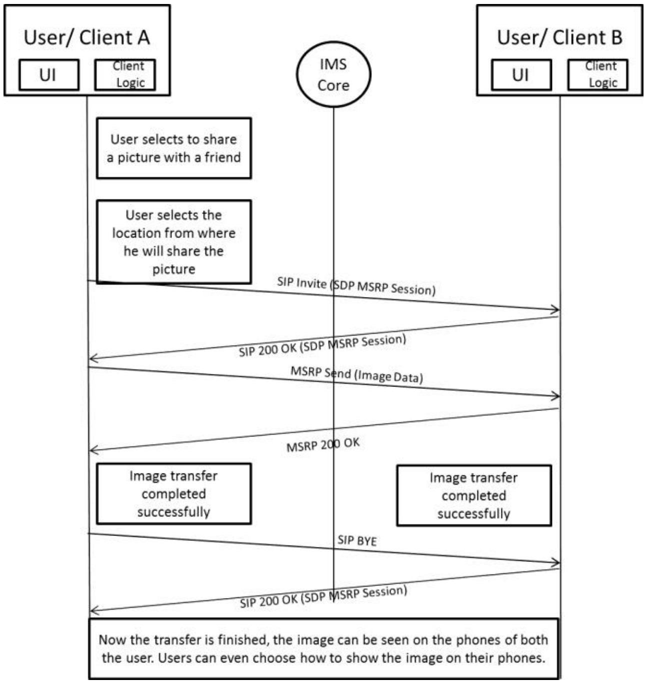

User/ Client A
UI
Client Logic
IMS Core
User/ Client B
UI
Client Logic
User selects to share a picture with a friend
User selects the location from where he will share the picture
SIP Invite (SDP MSRP Session)
SIP 200 OK (SDP MSRP Session)
MSRP Send (Image Data)
MSRP 200 OK
Image transfer completed successfully
Image transfer completed successfully
SIP BYE
SIP 200 OK (SDP MSRP Session)
Now the transfer is finished, the image can be seen on the phones of both the user. Users can even choose how to show the image on their phones.

Figure 2-Message sequence flow for image share scenario

RCS-e as a service, offers a wide mix of traditional and advanced features, which could not only cater to existing needs of mobile users but also offer additional edge in terms of feature availability and ease of use. Going by its portfolio, RCS-e offers a lot of things which are not correctly or fully implemented at present and there are several challenges to be dealt early for successful roll-out. Section III below elaborates some of these challenges.

III. IMPLEMENTATION CHALLENGES

A. Binding to SIM Card

Since, the RCS services are bound to the SIM card in use, the use cases such as roaming appear as serious challenges. Say if a user decides to continue using RCS-e services while in roaming area, he or she might be subjected to heavy roaming data charges on accounts of the RCS-e services he or she is using on the go. Why would then a normal user, who can use his Whatsapp account to connect to a local wifi and use it for messaging at no charge at all, use RCS-e to communicate? In contrast, the Over The Top (OTT) apps are on cloud so as to say and hence their access is not limited or bound by the SIM card which is a clear advantage and a vital one too. It remains interesting to see that how, the current roaming pricing model is tweaked to counter such obvious shortcomings when RCS-e is pitted against OTT apps.

Similarly, in the case where the user makes use of multiple SIM cards based on network availability and tariff needs, multiple RCS-e account information will need to be managed; for example, to have the same buddy list appear on different SIMs and many more associated complexities like these.

B. Inter-operability between operator networks

Wider and large scale IMS deployment, interoperability between different terminal vendor RCS clients and RCS service interworking between operators are the key aims of the RCS Initiative. Looking from a competition perspective, interoperability hasn’t been much of an issue for OTT application market leaders since long. Whether it is Facebook, Whatsapp or Twitter, these apps are tested well and thoroughly by the app developers themselves and hence can boast of high level of interoperability. In fact, some of these apps come pre-embedded into the devices much like what RCS-e proposes to do in future. So, what additional RCS-e can offer to the users in terms of interoperability would be a thing to watch.

Though a common set of GSMA specifications shall be followed by RCS-e enabling operators and application developers, a longer launch cycle time shall be needed due to multiple checks to be performed by the operator to validate inter-operability among different vendor solutions. Longer launch cycle time for RCS-e provides greater opportunity for OTT players to innovate and further impact RCS-e market penetration.

C. Packaging of RCS-e services v/s existing Unified Communication applications

In an enterprise world, equipment vendors and service providers are already investing in unified communication applications which provide mobility. These applications provide inherent capability to make use of advanced collaboration features when the user is in “office mode” or “work mode”. The Unified Communications framework provides a seamless access to IM, Presence, Voice, Video, Social networking, and others on a desktop or mobile device for a registered enterprise user in secured fashion, thereby enhancing productivity and mobility.

RCS-e services for enterprises shall require different packaging and billing model to compete with the Unified communications reach and growth. In some cases there could be overlapping solutions to cater to both personal and official needs.

D. Delay in availability of open APIs

Availability of standard and well known development environment or the Application Programming Interface (APIs) is an essential element in success of any platform or application. A very good example is Android operating system which provides seamless open source access to global developer community to contribute, share and simplify application development and deployment process. This encouraged developers, vendors and operators to flaunt new apps and generate huge interest in Android.

In the same way, RCS-e needs to provide Open APIs to capture the market quickly and offer similar developer experience/reach. Though the RCS-e APIs were made available for developer community just recently, it has to catch –up quickly with tested and scaleable APIs which can provide the required performance.

E. *SIP Inter-Op Issues*

Figure 3 depicts a summary of SIP Inter-op issues in development and launch of new SIP devices across 5 leading original equipment vendors. Though SIP specifications are stable for more than 5 years now, this is actual implementation data for development between 2009 and 2012.

This data gives a historical view of all kind of issues that may originate while introducing extensions to SIP which shall also impact RCS-e devices and associated services.

F. *Lack of standard test specifications or test tools for RCS-e clients*

Original equipment vendors or chipset vendors themselves need to validate the protocol stack compliance with latest set of standards which are still evolving. Availability of standard test specifications and automated test tools remain a challenge for the vendors. One of the silver linings here is that SIP or IMS test frameworks can be extended for the RCS-e standard for control plane validation.

IV. SUMMARY

The key competition to RCS-e is the OTT service. RCS-e forum members, handset equipment vendors, chipset vendors and MNOs have to play a key role in success of RCS journey. In today’s business context and increasing pressure to quickly monetize investments and enhance average revenue per user (ARPU) and user experience it is imperative for stakeholders to collaborate and come up with quick solutions to challenges described above.

Figure 4 gives an idea of how big a threat do operator perceive OTT to be. Nearly three-fourth of the operators surveyed say that because of the usage of OTT IM clients, their revenue takes a hit whereas a mere 12% feel that presence of OTT in the market hardly changes the revenue front for them. This for sure indicates that growing OTT presence is now clearly acknowledged by the operator community.

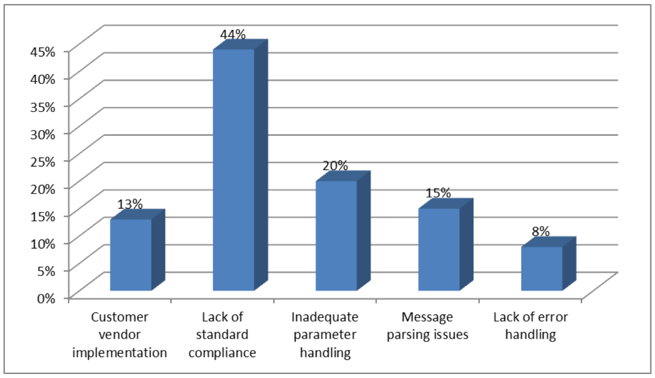


Figure 3-SIP Inter-Op Issues Classification

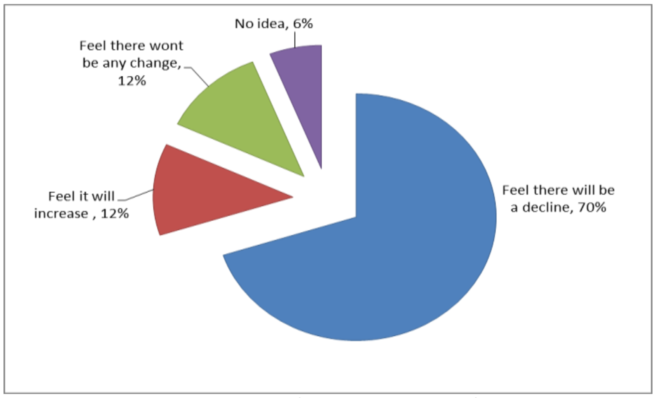


Figure 4-Operators expectation of revenue on usage of OTT IM clients on smart-phones [14]

Keeping mind that RCS-e enabled handsets would be available somewhere late in the year 2012, at present, various operators seem to be having different approaches to handle the OTT threat. Fig. 5 below shows using data collected in year 2011, how various operators across the world are planning to deploy the RCS-e services to combat the OTT threat. Here again, a clear pattern emerges with regards to the urgency with which RCS-e services are being promised. 39 percent of the operators is readying for 2012 year release and another 16 percent of operators are in the process of deploying them. This further substantiates, that operators seems to be betting big on RCS-e to take on OTT applications.

According to another survey [14], 22.6% of respondents also said that they are either offering their own IM client, or partnering with OTT providers. A minority, 6.5%, are trying to either block access to OTT clients or imposing surcharges for using OTT clients via deep packet inspection technology.

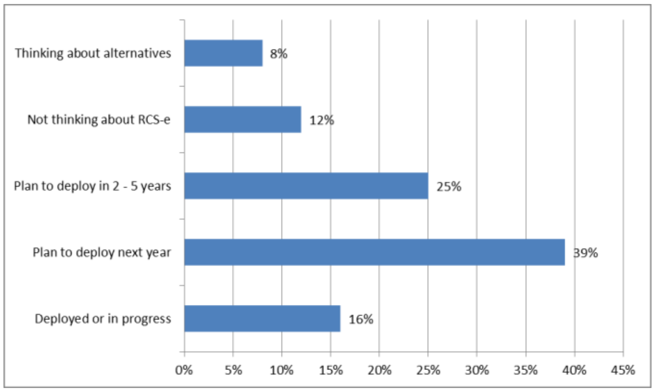


Figure 5-Operators with RCS-e Deployment (as in year 2011) [14]

But, these approaches are only addressing the outer layer of the problem, which is handling the OTT influx and hence is missing the crux of the issue. The solution lies in preparing well in advance and offering innovative services; one example is having the services on Cloud itself.

Based on the current information on GSMA website, most of the leading mobile handset manufacturers of the world seem to making heavy investments in getting their clients accredited for compliance to standards of new SIP devices across 5 leading original equipment vendors. Overall 14 companies have accredited in the February 2012 – October 2012 time period which shows the current priority for manufacturers in this segment and hence the focus needed to address some of these problems.

This business situation also opens up opportunities for software engineering service providers and telecom test vendors to create propositions and solutions around the same. Since, RCS-e technology is still in its initial stages, the approaches RCS-e proponents shall adopt in near future to resolve the implementation challenges would be a thing to observe.

V. REFERENCES

[1] http://en.wikipedia.org/wiki/Triple_play_(telecommunications)

[2] http://www.3gpp.org/Technologies/Keywords-Acronyms/article/ims

[3] http://www.3gpp.org/LTE

[4] http://www.ietf.org/rfc/rfc3261.txt

[5] rcs-e_advanced_comms_specification_v1_2_2_approved.pdf (http://www.gsma.com/rcs/specifications/rcs-e-specifications/)

[6] Traffic and Market Report from Ericsson (June 2012)

[7] http://www.gsma.com/

[8] http://openmobilealliance.org/

[9] GSMA PRD IR.90 - "RCS Interworking Guidelines" 2.1 October 2010 (http://www.gsma.com)

[10] RCS5_0_Advanced_communications_specification_version10.pdf

[11] SIP Extension for Instant Messaging IETF RFC http://tools.ietf.org/html/rfc3428

[12] RCS_e_Advanced_Comms_specification_v1_1.pdf

[13] Secure User Plane Location, Candidate Version 2.0 – 27 May 2011

[14] Mobile Squared Report, 2011

Fuzzy Redirection Algorithm for Content Delivery Network (CDN)

Thiago Queiroz de Oliveira
Universidade Estadual do Ceará (UECE)
Av. Paranjana 1700
Fortaleza - CE - Brazil
thiagoq@larces.uece.br

Marcial P. Fernandez
Universidade Estadual do Ceará (UECE)
Av. Paranjana 1700
Fortaleza - CE - Brazil
marcial@larces.uece.br

***Abstract*—Content Delivery Network (CDN) is a large distributed system of servers deployed in multiple data centers on the Internet. Its main goal is to serve requests from users providing high availability and performance. It also reduces the system failure risk providing redirection to many replica servers. It can provide load balancing between servers, avoiding network bottlenecks and, therefore, ensuring greater performance and QoE (Quality of Experience) to the end user. One of the critical issues involving CDN networks is the algorithm used to choose the replica server, because it directly influences the performance and scalability of the network. In this paper, an algorithm for choosing the best replica server is proposed. The proposal is compared in simulation against other algorithms in the literature. Finally, we show the effectiveness of the proposed algorithm to improve the choice of the best replica server in CDN.**

***Keywords*-Content Delivery Network; Fuzzy Logic; Web Server.**

I. Introduction

Content Delivery Network (CDN) is a large distributed system deployed in multiple data centers on the Internet [1]. CDN provides fast and reliable services distributing content by replica servers. The origin to the term CDN was in the 90's with the intention to provide website service with performance, scalability, replication and load balancing [1]. Websites with high-volume traffics, such as e-commerce sites, can present bottlenecks and slowdowns when it is implemented on a single server.

Research has shown that if the response time for a web request exceeds 8 sec, about 30% of users leave the request [2]. The increase in response time is directly related to performance loss, congestion and a large number of users reloading the website, making access to the website worse.

The site replication in different location's aims to: (1) reduce the response time to the nearest user, (2) eliminate a single point of failure, and (3) balance the load among multiple servers. A common approach is to redirect the user to the server closest to him, thus minimizing the bandwidth used, depending on the server's load; this way one can get a shorter response time [3].

One of the critical issues related to CDN concerns which replica server must be used. The closest server to the user is not always the best. Instead, a set of parameters could be considered during this selection process, such as distance, speed, available bandwidth and server load.

This type of algorithm, also known as request routing algorithm, can be divided into two categories: adaptive algorithms and nonadaptive algorithms [4]. In adaptive algorithms, the choice is made based on the server's status, requiring constant monitoring. In nonadaptive algorithms, the choice is based on heuristics, and then a lightweight processing by not requiring monitoring.

This paper proposes a new adaptive algorithm based on fuzzy logic to choose the best replica server. This algorithm considers the following parameters: (1) size of the service queue for each replica server; (2) time needed to answer a request from a given URL on the replica server, (3) response time of replica server (Round-Trip Time (RTT)).

The proposed algorithm can be classified as adaptive, because there is status information exchange between servers. The proposal evaluation shows the reduction of minimum and average response time for a request comparing to other models, validating the proposal.

The evaluation was done in the Network Simulator 2 (ns-2) [5], using the CDN module proposed by Cece et al. [6]. The evaluations were done in three real network topologies obtained in *Topology-zoo* [7]. The proposed algorithm is compared against three different algorithms to redirect the request: two nonadaptative, round-robin and random, and one adaptive algorithm, the *least-loaded* [8]. The metrics used to evaluate are minimum time, maximum time and average time of service request and degree of network unbalancing.

The rest of the paper is structured as follows. In section II, we present some related works, Section III introduces the Content Delivery Networks concepts and fuzzy logic. In Section IV we present our proposal, the FuzzyCDN. Section V shows the proposal evaluation, Section VI shows the results and Section VII concludes the paper.

II. Related Works

Chen Liao [9] proposes a fuzzy logic based algorithm to choose the replica server in a CDN network. However, in this proposal, the algorithm only act in case of congestion, considering the server bandwidth and the packets drop rate.

Manfredi, Oliviero and Roman [10] proposed an adaptive algorithm based on a mathematical model. In this algorithm, it is considered the request arriving rate and the requisition arriving rate variation in a certain time interval. These parameters are used in the decision-making process, where the last measured time has a weight x, and the average time for the whole period has a weight y, where x is greater than y. The results showed that this algorithm has a good performance.

One of the oldest non-adaptive algorithms is Round-Robin (RR) [11]. Each request is served by a different server, following a cyclic order. This algorithm performs well in homogeneous environment, e.g., the servers have similar capacities; they are in the same place; they share the alike network, and the requests generate a similar workload. However, when one of the conditions is not satisfied, RR algorithm gives bad results.

Another classical algorithm is the Random [12]. As its name suggests, the server will always chose at random. Generally, the system workload is much less than system capacity. Due to its stochastic characteristic, it is not possible to predict its performance.

Dahlin [13] proposes an algorithm called *Least-Loaded* that redirects the request to the server with the lowest load. Another approach would be to choose the server with the shorter response time. Least-Loaded Routing (LLR) algorithm is an algorithm that attempts to distribute the requisition to servers with the largest idle capacity, i.e., the least loaded server. The least loaded server is discovered by monitoring the current server state via protocol [14].

III. Content Delivery Networks

A CDN is a collection of network devices arranged for delivery contents to end users. A CDN network could be implemented in many architectures and topologies, which can be centralized, hierarchical, infra-structured with administrative control and decentralized [15].

The CDN provides better performance by offering content caching and server replication scattered strategically in order to address requests overloads for web content, which is called *flash crowd* [16] or *SlashDot effect* [17].

The design of a CDN network is quite complex [18]. To build a CDN network, some strategic issues must be defined [19]: (1) how many replica servers should be used and where they should be located [20]; (2) what content should be replicated and in which server should be replicated [21]; (3) what strategy should be used to keep the replica server's contents consistent; (4) how it chooses the replica server, and what mechanism should be used to redirect the client to the replica server.

The routing request process is responsible for performing customer's requisition routing to the best replica server. One solution is to redirect the request to the nearest server replica. However, not always the closest replica server is the best to serve a request [3]. The ideal is to consider another metrics to perform routing: network proximity, connection latency, distance and the replica server load.

The routing request should be divided into two mechanisms: (1) routing request algorithm, which is used to define the best replica server to answer a request; and, (2) request routing mechanism, that is responsible for performing the redirection from client to the server [18].

A. Routing Request Algorithm

The routing request algorithms are divided into two categories: adaptive algorithms and non-adaptive algorithms. The adaptive algorithms consider the current state of the replica servers before define the server to be used. The non-adaptive algorithms do not consider the server's state, i.e., no overhead is introduced to exchange status information between servers. In adaptive algorithms, the choice of replica server is based on its current state. Then, it is necessary to monitor the server's load and network congestion. As the network and server state can vary very fast, these algorithms consume many bandwidth. In non-adaptive algorithms, the replica server choice could be made using heuristics. Such solution is less complex, consumes less bandwidth, but are not efficient as the adaptive algorithms.

B. CDN Performance

To evaluate a CDN network performance, some metrics can be used [18] [22]: (1) temporal metrics, the time client expects to have his request served [23]; (2) space metrics, can be geographic distance or number of hops RTT [24]; (3) network use metrics, associated to network resources consumed, it can be internal, when communication is among servers to network management and software updates, and external, related to communication between clients and servers [18]; (4) cost metrics: they relate to server's acquisition and maintenance costs; (5) consistency metrics, related to consistency of the content accessed by users.

IV. FuzzyCDN: A Fuzzy algorithm to choose CDN replica server

This work proposes an adaptive algorithm to choose the replica server based on fuzzy logic [25]. The fuzzy logic unlike classical Boolean logic, can assign intermediate values between true and false and also, making a decision based on in-between inputs. Working with a logic that permits dealing with subjective information, imprecise and ambiguous, opens many possibilities to develop solutions to problems that classical logic is not able to solve.It considers following variables as input:

1) Queue size: queue length on replica server.
2) Service time: time to answer a request from a given URL on the replica server.

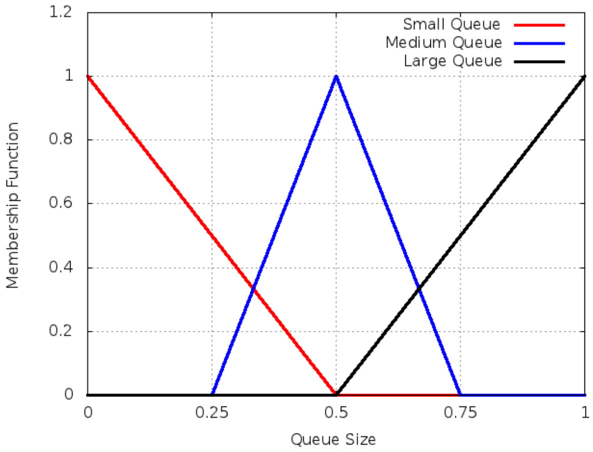


Figure 1. Membership functions to the variable queue

3) Response time: replica server response time (Round-Trip Time (RTT)).

Every server has a list of neighbor's servers. The servers periodically exchange status information with its neighbors. With this information, when a server receives a request, it performs the normalization of neighbor's status and using the inference mechanism, which will be described in the following sections, decide which server will attend the request.

A. *Fuzzification*

In this step, the mapping of input parameters, generally numerical and accurate, for fuzzy sets is realized using the membership functions. The input membership functions are triangular, and all of then have the same characteristics within the function range. The queue size variable has three linguistic values associated; they are: small queue, medium queue and large queue. The Figure 1 show graphically this membership functions, along with their respective ranges.

The service time variable also has three membership functions associated; they are: low service time, medium service time and high service time. Figure 2 show graphically this membership function. The variable response time also has three membership functions (Low, Medium and High) as shown in Figure 3.

The output consists of five membership functions using triangular functions (Very Good, Good, Normal, Bad, Very Bad), as shown in Figure 4.

B. *Rule evaluation*

The fuzzy controller inference rules definition is not a simple task, because it can produce inconsistent results. To avoid creating wrong rules, it was used the Wang-Mendel algorithm [26], in order to create consistent rules. The Wang-Mendel algorithm provides a method to create fuzzy logic rules by five phases:

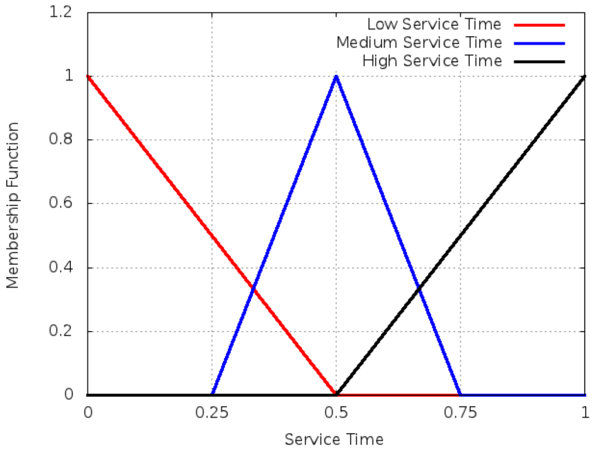


Figure 2. Membership functions to the variable queue service time

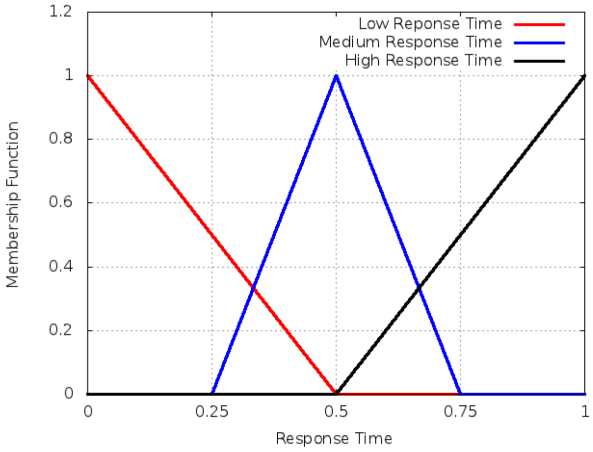


Figure 3. Membership functions to the variable queue response time

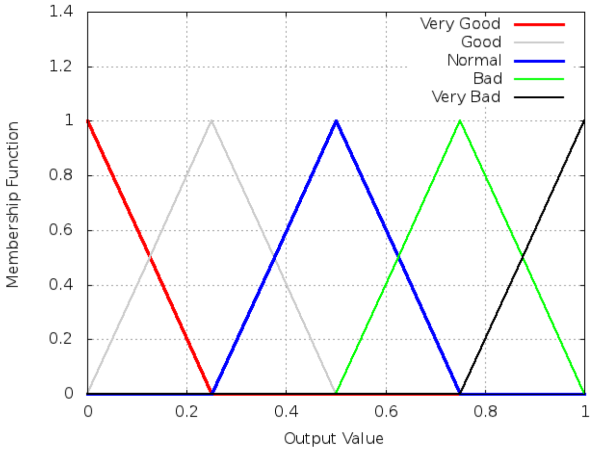


Figure 4. Membership functions to the output variable

1) Divides the input and output spaces from a given

numerical data into fuzzy regions.

2) Generates fuzzy rules based upon the given data.
3) Assigns a degree for each generated rules for resolving conflicts.
4) Creates a fuzzy rule set based on the generated rules and linguistic rules.
5) Defines a mapping from input space to output space based on the combined fuzzy rule base using the defuzzifying procedure.

The first four steps generate the knowledge base and compose the training stage. The last step generates the output values for the possible entries from the knowledge base. The inference rules obtained are shown in Table I.

Table I
FUZZY INFERENCE TABLE

Queue Size	Service Time	Response Time(RTT)	Output
Small	Low	Low	Very Good
Small	Low	Medium	Very Good
Small	Low	High	Very Good
Small	Medium	Low	Good
Small	Medium	Medium	Good
Small	Medium	High	Good
Small	High	Low	Good
Small	High	Medium	Normal
Small	High	High	Normal
Medium	Low	Low	Good
Medium	Low	Medium	Good
Medium	Low	High	Good
Medium	Medium	Low	Normal
Medium	Medium	Medium	Normal
Medium	Medium	High	Normal
Medium	High	Low	Bad
Medium	High	Medium	Bad
Medium	High	High	Bad
Large	Low	Low	Normal
Large	Low	Medium	Normal
Large	Low	High	Bad
Large	Medium	Low	Bad
Large	Medium	Medium	Bad
Large	Medium	High	Bad
Large	High	Low	Very Bad
Large	High	Medium	Very Bad
Large	High	High	Very Bad

C. Defuzzification

The Algorithm 1 shows the controller operation in the decision-making process. The algorithm gets the URL request as input and gives the server chosen as output. In line 3, it gets the servers list for a given URL. Then, from lines 4 to 13, it calculates the fuzzy value for each replica server using as parameters the server queue, service time and response time.

Then, the parameter values are mapped to fuzzy sets according to their membership functions as described above. After this step, the inference rules from Table I are applied, resulting in a fuzzy set output, showed on Table 4. Finally, it is held deffuzifaction that returns a quantitative value using the center area method. The algorithm returns the server with the lowest fuzzy value.

Algorithm 1: FuzzyCDN algorithm

```
1  $s_{min} \leftarrow NULL;$
2  $server \leftarrow NULL;$
   Input: A request $\mathcal{R}$ for a given URL
3  $n_s \leftarrow$ number of servers having the URL;
4  while $i = 1, \ldots, n_s$ do
5      $f_i \leftarrow$ queue size of replica server i;
6      $ts_i \leftarrow$ service time of server i for a given URL;
7      $tr_i \leftarrow$ response time of server i (RTT);
8      $s \leftarrow$ calculate the fuzzy value($f_i, ts_i, tr_i$);
9      if $(s < s_{min})$ or $(s_{min} = NULL)$ then
10         $s_{min} \leftarrow s;$
11         $server \leftarrow i;$
12     end
13 end
14 return the most appropriate server
```

V. PROPOSAL EVALUATION

The simulator used in the experiments was the ns-2 (*Network Simulator*network simulator), version 2.33 [5]. It was used the ns-2 CDN module developed by [6].

The client sends the request to the nearest server containing the desired content. The replica server choice mechanism is decentralized, where any server can decide whether it serves the request or for which server should forward the request.

Every server has an associated service time, which represent to the processing time required to serve a given request. Any request received is inserted into a queue, and the server will decide whether the request will be answered or redirected. The queuing model may be D/D/1 or M/M/1 for deterministic or exponential distribution, respectively.

Each server maintains a list of neighbors. The servers exchange periodically status information with their neighbors. This information is used to select the server to a given request.

A. Evaluation Topologies

For the tests, it was used three topologies of real networks, obtained at the Topology-zoo site [7], which are: Claranet [27], GridNet [28] and RNP [29].

B. Experiments Details

In the experiments, the customers' number is the same of the servers, and they vary according to the topology. The clients always send the original request to the nearest server, this request follows a Poisson process with arrival rate λ_i. Each server has a service rate μ_i.

The time that servers exchange information status is 1s. The traffic generated to the servers is CBR, the packet sent is HTTP and the simulation time for each experiment is 300 sec.

The first two are considered non-adaptive algorithms, and the last one is an adaptive algorithm. The round-robin algorithm chooses a different server every decision. The random algorithm uses a stochastic process to choose the server, and the least-loaded algorithm chooses the server with the smaller queue.

C. Traffic Model

The GridNet network is composed of nine servers, located in the southern U.S. In the experiments, nine clients generate traffic to nine servers, one server for each client. Table II shows the characteristics of each server based on parameters used by Manfredi [10].

Table II
TRAFFIC CHARACTERISTICS: GRIDNET

	1	2	3	4	5	6	7	8	9
$\lambda_i[req/s]$	12	10	7	10	13	11	11	14	17
$\mu_i[req/s]$	12	13	14	17	10	11	11	7	10

In Table II, λ_i is the requests arrival rate to the server i and μ_i is the service rate of server i. According to [10], this traffic pattern described in Table II represents a real CDN network. The *flash-crowd* effect is also simulated by increasing the arrival rate at Server 7, λ_7 of 11 requests per second to 200 requests per second, between the time $t_0 = 200sec$ and $t_1 = 250sec$. The other topologies have a similar traffic model, changing only the number of servers. Then, it will not be shown.

VI. RESULTS

The metrics used for evaluation are: minimum time, medium time and maximum time for answering a request; standard deviation of answering request and network degree of unbalancing. The network unbalancing degree is the standard deviation of the queues size in all servers along the time. Therefore, smaller network unbalancing degree is better.

The following figures show graphically the results of each metric for the four algorithms on these three topologies. Despite not having much precision in differentiating values, these graphs enable you to have an assessment of each metric.

In Figure 5, the minimum time was obtained by the fuzzy algorithm. The others' algorithms have similar results. In contrast, according to Figure 6, the proposed fuzzy algorithm presented the worst maximum time on these three topologies.

In Figure 7, it was observed that the adaptive algorithms showed better results for the average time on the GridNet and RNP. In all three topologies, the proposed algorithm showed the best results. Figure 8 shows that adaptive algorithms have the lowest standard deviation in the three topologies, especially the fuzzy algorithm which had the lowest standard deviation in all tests. This good result of the standard deviation, and with the good results of the minimum and average time to obtain a request enables the deployment of the proposed algorithm in real CDN networks.

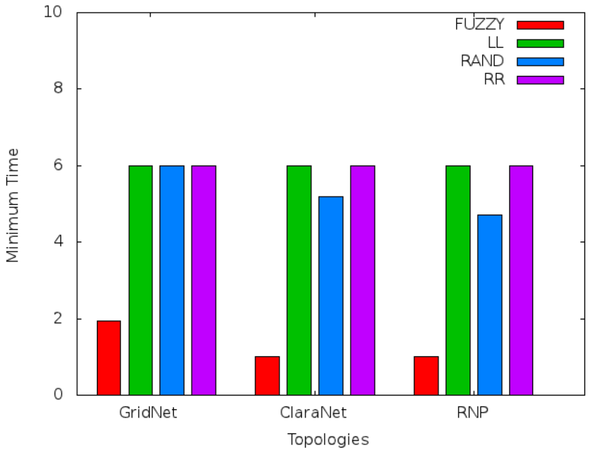


Figure 5. Minimum Response Time

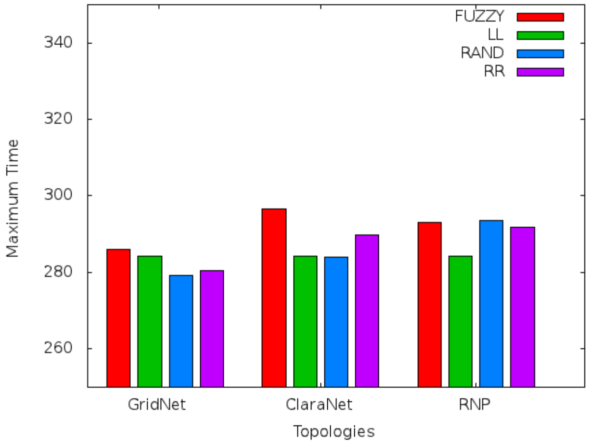


Figure 6. Maximum Response Time

VII. CONCLUSION AND FUTURE WORKS

In this paper, we proposed a new algorithm to choose the replica server in CDN networks using fuzzy logic. The fuzzy logic is feasible for this environment by simplifying the process modeling system, dispensing complex mathematical system, and leave the system closer to human thinking.

By simulations, we show that the proposed algorithm gives good results comparing with other algorithms available in the literature. The main benefit was the lowest request response time obtained in three topologies tested. Furthermore, the algorithm presented the lowest standard deviation in these topologies, showing it gives a stable solution.

However, the simulation methodology could not show the performance of the proposed algorithm in real systems.

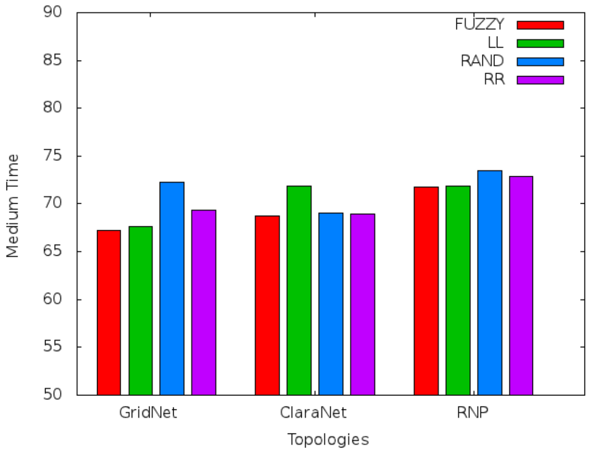


Figure 7. Average Response Time

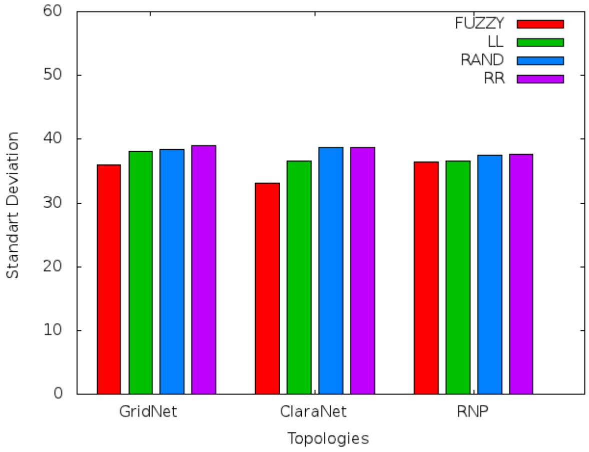


Figure 8. Std Deviation Response Time

Although we know that Fuzzy logic produces a low impact in modern processor's performance, we do not guarantee the algorithm scalability.

As future work, we can analyze another metrics algorithm, for example, the amount of available bandwidth on the link. It should be used a learning technic to choose the replica server in CDN networks, such as neural networks, Bayesian networks or Support Vector Machine (SVM).

REFERENCES

[1] G. Pallis and A. Vakali, "Insight and perspectives for content delivery networks," *Communications of the ACM*, vol. 49, no. 1, pp. 101–106, Jan. 2006.

[2] B. Davison, "A web caching primer," *Internet Computing, IEEE*, vol. 5, no. 4, pp. 38–45, jul/aug 2001.

[3] C. Chen, Y. Ling, M. Pang, W. Chen, S. Cai, Y. Suwa, and O. Altintas, "Scalable request routing with next-neighbor load sharing in multi-server environments," in *Advanced Information Networking and Applications (AINA)*. Taiwan: IEEE, Mar 2005, pp. 441–446.

[4] in *Content Delivery Networks*, ser. Lecture Notes Electrical Engineering, R. Buyya, M. Pathan, and A. Vakali, Eds., 2008, vol. 9.

[5] L. Breslau, D. Estrin, K. Fall, S. Floyd, J. Heidemann, A. Helmy, P. Huang, S. McCanne, K. Varadhan, Y. Xu *et al.*, "Advances in network simulation," *Computer*, vol. 33, no. 5, pp. 59–67, 2000.

[6] F. Cece, V. Formicola, F. Oliviero, and S. Romano, "An extended ns-2 for validation of load balancing algorithms in content delivery networks," in *Proceedings of the 3rd International ICST Conference on Simulation Tools and Techniques*. Malaga/Spain: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2010, p. 32.

[7] S. Knight, H. Nguyen, N. Falkner, R. Bowden, and M. Roughan, "The internet topology zoo," *Selected Areas in Communications, IEEE Journal on*, vol. 29, no. 9, pp. 1765–1775, 2011.

[8] M. Dahlin, "Interpreting stale load information," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 11, no. 10, pp. 1033–1047, 2000.

[9] J. Chen and S. Liao, "A fuzzy-based decision approach for supporting multimedia content request routing in cdn," in *Parallel and Distributed Processing with Applications (ISPA), 2010 International Symposium on*. IEEE, 2010, pp. 46–51.

[10] S. Manfredi, F. Oliviero, and S. Romano, "A distributed control law for load balancing in content delivery networks," *Networking, IEEE/ACM Transactions on*, vol. PP, no. 99, p. 1, 2012.

[11] Z. Xu and R. Huang, "Performance study of load balancing algorithms in distributed web server systems," *CS213 Parallel and Distributed Processing Project Report*, 2009.

[12] R. Motwani and P. Raghavan, *Randomized algorithms*. Chapman & Hall/CRC, 2010.

[13] M. Dahlin, "Interpreting stale load information," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 11, no. 10, pp. 1033–1047, 2000.

[14] V. Cardellini, M. Colajanni, and P. Yu, "Redirection algorithms for load sharing in distributed web-server systems," in *Distributed Computing Systems, 1999. Proceedings. 19th IEEE International Conference on*. IEEE, 1999, pp. 528–535.

[15] D. Verma, *Content distribution networks*. Wiley, 2002.

[16] M. Arlitt and T. Jin, "A workload characterization study of the 1998 world cup web site," *Network, IEEE*, vol. 14, no. 3, pp. 30–37, 2000.

[17] S. Adler, "The slashdot effect: an analysis of three internet publications," *Linux Gazette*, vol. 38, p. 2, 1999.

[18] S. Sivasubramanian, M. Szymaniak, G. Pierre, and M. Steen, "Replication for web hosting systems," *ACM Computing Surveys (CSUR)*, vol. 36, no. 3, pp. 291–334, 2004.

[19] J. Wang, R. Sharman, and R. Ramesh, "Shared content management in replicated web systems: A design framework using problem decomposition, controlled simulation, and feedback learning," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 38, no. 1, pp. 110–124, 2008.

[20] L. Qiu, V. Padmanabhan, and G. Voelker, "On the placement of web server replicas," in *INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 3. IEEE, 2001, pp. 1587–1596.

[21] J. Kangasharju, J. Roberts, and K. Ross, "Object replication strategies in content distribution networks," *Computer Communications*, vol. 25, no. 4, pp. 376–383, 2002.

[22] M. Harchol-Balter, B. Schroeder, N. Bansal, and M. Agrawal, "Size-based scheduling to improve web performance," *ACM Transactions on Computer Systems (TOCS)*, vol. 21, no. 2, pp. 207–233, 2003.

[23] D. Olshefski, J. Nieh, and D. Agrawal, "Using certes to infer client response time at the web server," *ACM Transactions on Computer Systems (TOCS)*, vol. 22, no. 1, pp. 49–93, 2004.

[24] B. Huffaker, M. Fomenkov, D. Plummer, D. Moore, and K. Claffy, "Distance metrics in the internet," in *Proc. of IEEE International Telecommunications Symposium (ITS)*, Natal/Brazil, 2002.

[25] L. Zadeh, "Fuzzy sets," *Information and control*, vol. 8, no. 3, pp. 338–353, 1965.

[26] L. Wang and J. Mendel, "Generating fuzzy rules by learning from examples," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 22, no. 6, pp. 1414–1427, 1992.

[27] Claranet, "Claranet Limited," Last accessed, Mar 2012. [Online]. Available: http://noc.eu.clara.net/network.jpg

[28] GridNet, "GridNet Inc Network," Last accessed, Mar 2012. [Online]. Available: http://www.nthelp.com/images/gridnet.jpg

[29] RNP, "Rede Nacional de Pesquisa," Last accessed, Mar 2012. [Online]. Available: http://www.rnp.br/backbone

Tree Construction Strategies for Survivable Overlay Multicasting in Dual Homing Networks

Wojciech Kmiecik, Krzysztof Walkowiak
Department of Systems and Computer Networks,
Wroclaw University of Technology,
Wroclaw, Poland
E-mail: wojciech.kmiecik@pwr.wroc.pl, krzysztof.walkowiak@pwr.wroc.pl

Abstract— **Due to the growing demand for high definition music and video content, an overlay multicasting providing live streaming services has been gaining popularity over last years. In this paper, we focus on applying the overlay multicasting for delivering of critical data that require to be transmitted safely, intact and with as little delay as possible, e.g., financial data, software security patches, antivirus signature database updates etc. To improve survivability of the overlay multicasting, we propose to use dual homing approach, i.e., each peer is connected to the overlay by two separate access links. We introduce several tree construction strategies and conduct simulation experiments to investigate problem of providing survivability to both static and dynamic types of network. Our studies demonstrate that the additional survivability requirements do not have a significant impact on the overlay multicasting system expressed as the streaming cost.**

Keywords-overlay multicasting; survivability; dual homing; simulation; tree construction.

I. INTRODUCTION

Nowadays, we are observing a rapid growth in the popularity of multimedia streaming in the Internet. To emphasize the growing popularity of various video streaming services, we need to quote [1], where the authors claim that Video on Demand traffic will triple and Internet TV will increase 17 times by 2015. The total share of all forms of video (already mentioned) and P2P will grow continuously to be approximately 90 percent of all global consumer traffic in the next three years. Services such as internet radio, high definition video or audio streaming are very useful for network users, but often require a lot of bandwidth, which can be costly [1]. To reduce maintenance and investment cost, the concept of overlay multicasting is applied. An overlay multicasting technology is based on a multicast delivery tree consisting of peers (end hosts). Content transmitted by the overlay multicasting can be either streaming content with additional requirements like bit rate etc. [3], or data files. In many related works, the authors assume that users of a multicast network can leave the system. That kind of overlay network is called *dynamic* or *evolving* [4]. Among examples of such a system are popular protocols like Torrent [5], eDonkey [6] or Skype [7]. Another network type is so-called a *static* network, where hosts form static structure and are not allowed to leave the system, and which well-known examples are:

- Content delivery network (CDN), e.g., Akamai Technologies,
- Set-top box (STB) technology used in IPTV,
- Critical information streaming, e.g., hurricane warnings.

In our work, we apply the overlay multicasting in a dual homing architecture to improve network survivability, defined as the ability to provide the continuous service in the presence of failures. The dual homing approach assumes that all hosts (*nodes*) have two disjoint links (*homes*) to the network. Those links provide the network protection because of redundancy. The main contribution of the paper consists of: (i) a formulation of new strategies of creating link disjoint multicast trees in survivable overlay networks with optimization of cost and number of tree levels; (ii) a new modular simulator for testing introduced strategies in both static and dynamic networks; (iii) numerical experiments based on proposed tree construction strategies showing comparative results and other characteristics of the proposed concepts. Note that the concept of overlay multicasting protection by dual homing has been introduced in our recent papers [8],[9],[10] in the context of both streaming cost and maximum delay objective functions.

The rest of this paper is organized in the following way. In Section 2, we present previous research on overlay multicasting and dual homing multicasting with a special focus on survivability and simulation. Section 3 introduces the concept of survivable overlay P multicasting based on the dual homing method. In Section 4, we formulate tree construction strategies for the survivable (link disjoint) dual homing overlay multicast. Section 5 includes description of the overlay simulator. In Section 6, we present and discuss the results of our experiments. Finally, the last section concludes this work.

II. RELATED WORKS

In this section, we present previous papers related to dual homing and overlay multicasting. Dual homing is a subject of several articles. Jianping *et al.* [11] created the multicast protection scheme based on a dual homing architecture where each destination host is connected to two edge routers. Under such an architecture, the two paths from the source of the multicast session to the two edge routers provide protection for the traffic from the source to the destination.

In [12], a novel homing scheme called *dual homing with shared backup router resources (SBRR)* is introduced. The authors claim that their approach leads to savings up to 40% of the cost of traditional dual homing architecture. New ILP model for joint optimization of dedicated working and shared backup paths of anycast and unicast demands is introduced and is followed by extensive simulation research.

Another field of application for dual homing technology is Self-healing ring networks [13][14], where new network design methods and routing algorithms are designed and developed. An integer programming formulation and the NP-completeness of the problem is presented.

In [15], Jianping *et al*. introduce a concept of partial protection for the multicast dual homing network and a new algorithm *PAS* for finding the best partial multicast protection tree is proposed. The authors claim that simulation results show that the PAS algorithm achieves performance very close to the computed lower bounds.

Wang *et al*. [16] studied IP-over-WDM network survivability with a dual homing infrastructure. The paper focuses on a problem of adding survivability to IP WDM multicasting networks for both static and dynamic traffic. The authors created and evaluated coordinated protection design.

A scalable multicast protection scheme based on the dual homing architecture was introduced in [17]. The solution proposed by Wang *et al*. can be used to choose dynamically two edge routers for a multicast host.

Thulasiraman *et al*. [18] addressed disjoint multipath routing in the dual homing network problem. An algorithm for constructing colored trees in dual-homing using colored trees is proposed.

Overlay multicast (application-layer multicast) [19], [20] is a technology, which uses the overlay network topology, that enables multicast functionality for end hosts instead of routers. The authors propose a proactive tree recovery mechanism to make the overlay multicast resilient to peer failures. Simulations are used to prove that the proactive method can recover from node failures much faster than reactive methods.

A P2P (Peer-to-Peer) simulators survey is presented in [21]. Naicken *et al*. tested the most popular P2P simulators and found that they lack some of the key functionalities and conclude that because of that the majority of P2P researchers create their own simulation environment.

III. SURVIVABILITY FOR OVERLAY MULTICASTING

Two methods are used in order to provide the network survivability: restoration and protection. The main difference between them is that restoration applies dynamic resource allocation while protection needs preallocated network resources. This results in different overall cost and restoration time. Well-known protection methods are: automatic protection switching (1+1, 1:1, etc.), p-cycles and backup paths/links. In our previous works, we proposed to use disjoint overlay multicast trees streaming the same content [8], [22]. Peers affected by the failure of one of the trees can use another tree to receive the required data in the case of a failure. This procedure guarantees very low restoration time.

In this paper, we are studying the network survivability problem for the dual homing architecture. In Fig. 1, we present a simple example to illustrate our concept.

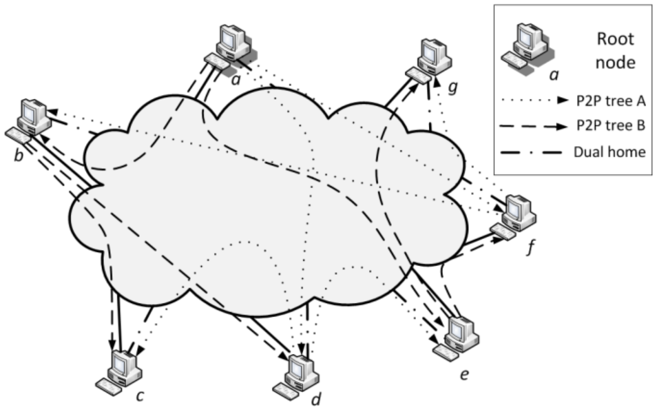


Figure 1. Simple P2P multicast scheme.

There are two disjoint multicast trees *A*, *B* that connect 7 nodes - *a*, *b*, *c*, *d*, *e*, *f*, *g*. In the case of tree *A*, nodes *a*, *d* and *f* are parents (uploading nodes), while remaining ones are leafs (nodes that are only downloading data). We use the term of *level* to describe location of nodes in the multicast tree. For example, node *a* is on level 1 of tree *B*, nodes *b* and *e* are on level 2 of tree *B* and rest of the nodes are on level 3.

The overlay multicasting is done in the application layer, i.e., end-hosts are connected using the overlay network. Connections between peers are established as unicast connection over the underlying physical layer. Each peer is connected to the overlay by an access link. We propose to use the dual homing approach to protect the system against a failure of the access link. The main idea is to create two overlay multicasting trees guaranteeing that each of access links carries traffic only of one of the trees. Since each node has two access links (dual homing), it receives the streaming data from both trees on two separate links. Thus, if one of access links is broken, the node still is connected to the stream, and moreover, it can upload the stream to subsequent peers located in the tree.

A proper configuration of the overlay multicasting with dual homing protects the network from two kinds of failure:

- Uploading node failure - failure impacts all successors of the failed peer in the tree,
- Overlay link failure - overlay link failure comprises failure of both directed links between nodes.

IV. A DISTRIBUTED APPROACH TOWARDS CREATING OVERLAY TREES

In the overlay multicasting networks we consider, hosts connect to the system by their own. A node *v* willing to join the network contacts a host with information about the network (e.g., root node) and receives a list of possible parents already connected to the multicast tree. If it successfully connects to the system, the database of feasible parents is updated. In the other case, node *v* sends a request to the root for another set of feasible parents. Another case scenario occurs when the simulated network is dynamic, meaning that nodes already connected to the system can leave the system. *CNF* message allows all the parent nodes to have updated information about their children. When node

v is disconnecting from the network, its parent informs the root node so that it can update the peers database. All the children of node v are disconnected from the network and in order to reconnect have to send the *RQT* message. It uses IP-to-Location mapping prediction method and historical connections data to gain knowledge of approximate cost of all possible connections between nodes.

In [8], in the context of streaming cost objective function, we introduced novel ILP (Integer Linear Programming) formulations of survivable overlay multicasting systems using dual homing architecture. As a natural continuation, below we introduce six tree construction strategies that include ideas related to ILP models. To formulate the problem, we use the notation as in [23]. Let indices $v,w = 1,2,\ldots,V$ denote peers – nodes of the overlay network. There are K peers (clients) indexed $k = 1,2,\ldots,K$ that are not root nodes in any tree and that want to receive the data stream. Index $t = 1,2,\ldots,T$ denotes streaming trees. We assume that $T = 2$, however the model is more general and values $T > 2$ may be used. In trees, nodes are located on levels $l = 1,2,\ldots,L$. That gives us possibility to set a limit on the maximum depth of the tree. The motivation behind this additional constraint is to improve the QoS (Quality of Service) parameters of the overlay multicasting, .e.g., network reliability and transmission delay. If node v is root of the tree t, then $r_{wt} = 1$, otherwise $r_{wt} = 0$. Constant c_{wv} denotes streaming cost on an overlay link (w,v), that can be interpreted as a network delay or a transmission cost.

We introduce constant $\tau(v)$, which denotes a virtual node associated with the node v, what follows for the dual homing Nodes v and $\tau(v)$ form a *primal node*.

Every primal node has in fact four capacity parameters – constants d_v and u_v are respectively download and upload capacity of the one access link and constants $d_{\tau(v)}$ and $u_{\tau(v)}$ are parameters of the second (dual) access link. Additionally, to be able to simulate both static and dynamic overlay multicasting networks, we introduce two constants – ts_v and te_v, which are respectively time when a host v tries to connect and to leave the network. The objective function is overall streaming cost (cost of all multicast trees). Fig. 2 depicts an example of the dual homing modeling. Dual homes are marked with a pattern of sequential lines and dots.

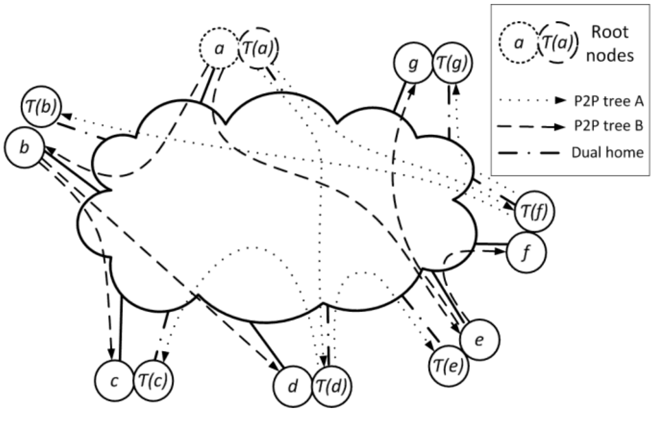


Figure 2. Modeling dual homing.

It is shown that streaming trees are using different connections to nodes.

We propose six tree construction strategies for overlay multicasting with dual homing described in details in the following section.

A. Tree construction strategies for overlay multicasting with Dual Homing

In our strategies, we use specified type of messages between hosts and root node:

- RQT – message sent from a host willing to connect to the overlay network to the root node,
- LST – possible response to the RQT; list of possible parents for the requesting host (with length of 10). In the process of selecting possible parents root node is responsible of keeping multicast trees disjoint,
- DEN – possible response to the RQT message; refusal of connection when there is no feasible parent; host can try again with RQT signal,
- ATT – message sent from a host willing to connect to the network to a feasible parent with a connection request,
- PER – possible response to *ATT*; the requesting node is connected to the network,
- REF – possible response to *ATT*; the feasible parent is refusing connection, e.g., due to lack of free upload capacity,
- CON – message from a host to the root node informing about successful connection,
- DEL – message sent from the parent of a host requesting disconnection from the network to root node,
- CNF – message sent from a child to its parent every 60 seconds. If during 120 seconds the parent does not receive CNF, it sends the DEL signal.

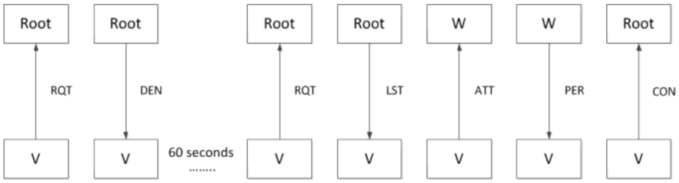


Figure 3. Example of communication process in overlay system.

In Fig. 3, we present an example of the communication in our model of the overlay system. Host v is requesting a connection to the network, receives a list of possible parents, sends *ATT* message to feasible parent w, receives positive response *PER*, connects to parent node w and sends *CON* message to the root node.

We developed the following six strategies for overlay multicasting with dual homing:

- *unprotected Cost Optimization* (*uCO*) – node v willing to connect to the network sends *RQT* to the root node, receives a *LST* message with the sorted list of possible parents selected by minimal connection cost criterion and attempts connection to the first (the cheapest) node on the list. If the connection process is successful, node v

sends a *CON* signal to the root updating information about the tree. Otherwise, the node sends *RQT* signal and receives another set of feasible parents; there is no level control; this strategy does not provide survivability because there is no requirement related to trees disjoint (nodes v and $\tau(v)$ can be in the same multicast tree),

- *unprotected Cost Optimization with Levels* (*uCOL*) the aim of this strategy is to minimize the tree depth (maximum number of tree levels). For each tree, nodes requesting connection are connected to the root until free upload capacity of the root is available. When there is not enough upload capacity on the current tree level, the next level is started. Node v willing to connect receives a *LST* message with the sorted (by cost) list of possible parents located only on the previous level. On each level the cheapest possible connection is selected; no survivability provided,
- *Cost Optimization* (*CO*) – analogous to *uCO*, but the root node in the process of selecting possible parents for the requesting node includes additional survivability constraints e.g., node v and its *virtual node* $\tau(v)$ have to be connected to disjoint multicast trees,
- *Cost Optimization with Levels* (*COL*) – analogous to *uCOL*; survivability provided,
- *Random Selection* (*RS*) and *Random Selection with Levels* (*RSL*) – those strategies perform similar to *CO* and *COL* respectively, but parent node is picked randomly from list provided by root node (also selected by random); both strategies provide survivability through disjoint of multicast trees.

All strategies are able to connect requesting hosts for both static and dynamic networks. Note that random strategies (*RS* and *RSL*) follow from real overlay systems like BitTorrent and others [5], where peers are selected by random and the transmission cost is not taken into account. For *RSL*, *uCOL* and *COL* strategies, the main goal is to limit the number of tree levels and as a result make the tree as short as possible. The motivation is to minimize the consequences of the node failure. Differently, cost optimization strategies (*uCO* and *CO*), aim in minimizing the overall streaming cost of the network, where cost can be interpreted as a network delay, transmission cost, etc.

V. OVERLAY SIMULATOR DESCRIPTION

After doing research on some well-known P2P simulators, i.e., p2pSim and PeerSim, we decided that it would be much more time efficient to create our own overlay simulator for dual homing architecture, than implementing this architecture and all the tree construction strategies into one of those simulators. The overlay simulator we developed is rather simple and concentrates on the process of creating trees. We believe that this approach is sufficient to examine the influence of survivability constraints on the overlay multicasting network. We focus on two criteria related to overlay multicasting: overall network cost and number of levels in multicast trees.

Let constant *LT* denote time of life in seconds for simulating the overlay network. Starting from $z = 0$ seconds, we can distinguish two phases of this process:

- Connection phase **-** all the hosts for which $ts_v = z$ attempt to connect to the network,
- Disconnection phase – all the hosts for which $te_v = z$ disconnect from the system (only for the dynamic networks). If disconnecting node v was a parent in multicast tree t than root is remodeling the tree by reconnecting children of node v to new parents.

After executing both phases simulator increments z by 1 (second) and repeats above actions. System is working until $z = LT$.

Pouwelse *et al*. [5] and Xiaojun *et al*. [24] study global characteristics of large P2P systems and provide measurements data useful in modeling P2P networks. They focus on phenomenon called *flashcrowd* effect, where peers join the network rapidly and after reaching peak number of host in network is decreasing gradually.

In our paper, we try to model this dependency in case of dynamic type of network. We set $LT = 10800$ seconds (3 hours) and during the first hour nodes are only connecting to the network. Between the first and the second hour, new nodes are still connecting, but there are also some that are disconnecting. In the last hour, nodes are only leaving the network.

For simulating the static networks, we set $LT = 7200$ seconds and nodes are only allowed to connect to the system, i.e., nodes do not leave the system during the simulation.

VI. RESEARCH

A. Comparing the tree construction strategies - experiment design

To compare introduced strategies, we use our overlay simulator. We randomly generated 5 different networks, where $V = 1000, 2000, 3000, 4000, 5000$ with two disjointed trees ($T = 2$). Each network consist of either symmetric nodes (100Mbps/100Mbps - 10% of all nodes) or asymmetric nodes (1Mbps/256Kbps, 2Mbps/512Kbps, 6Mbps/512Kbps, 10/1Mbps, 20/1Mbps, 50/2Mbps, 100/4Mbps). Link costs are random values from the interval [1,100]. Overall, for both static and evolving type of network, we conducted the following three experiments and each of them was executed 50 times for different sets of ts_v and te_v:

- comparing overall network cost and number of levels for different tree construction strategies and different network size (from 1000 to 5000 nodes),
- increasing the streaming rate q from 128 Kbps to 640 Kbps and verifying its impact on the overall network cost and the number of levels for different strategies,
- based on results of previous experiments, check survivability impact on overall network size.

Additional criterion of comparison for our experiments was percentage of nodes that have not been able to connect to the network, i.e., the number of rejects.

B. Comparing the tree construction strategies– results

For the purpose of the first experiment, we set streaming rate *q* to 256 Kbps and *L* value to 100. In Figs 4 and 5, we show results of tree construction strategies for different size of the network. For all strategies, the overall network cost is increasing with size of the network. In terms of the overall network cost, strategies *uCO* and *CO* prove to be most efficient.

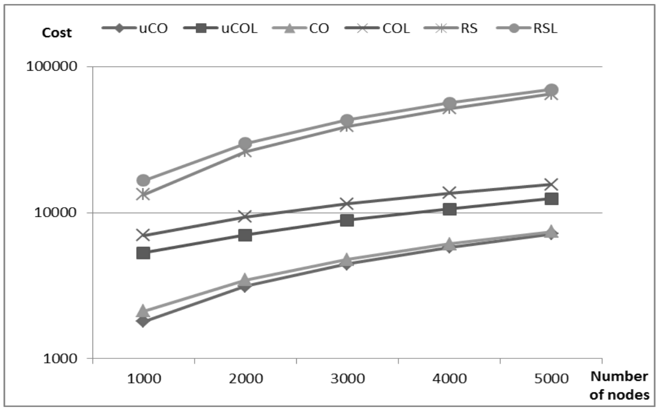


Figure 4. Streaming cost as a function of number of nodes (static).

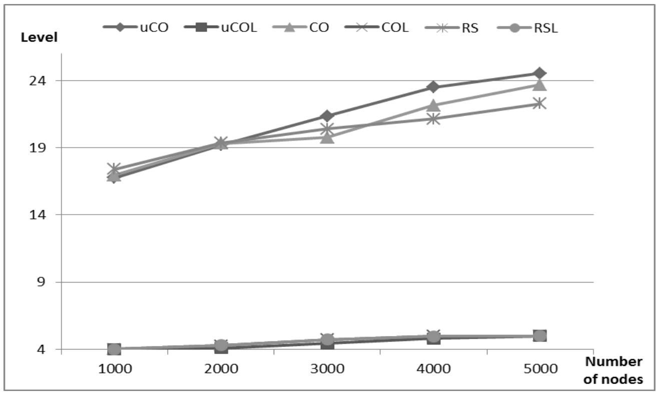


Figure 5. Number of levels as a function of size of the network (static).

uCOL strategy achieved lowest average number of levels in the multicast trees – 4,5. Results of strategies with survivability constraints (*COL* and *RSL*) were worse only by 2,2%. Results of the first experiment are in line with our expectations, since *uCO* and *CO* are minimizing the overall streaming cost, while *uCOL* and *COL* number of tree levels.

TABLE I. COMPARISION OF TREE CONSTRUCTION STRATEGIES FOR DYNAMIC TYPE OF NETWORK – SIZE OF THE NETWORK

Strategy	uCO		uCOL		CO		COL		RS		RSL	
Number of nodes	cost	level	cost	level	cost	level	cost	level	cost	level	cost	level
1000	1361.6	42.9	11643.5	4.0	1529.9	41.8	13607.3	4.1	10934.7	20.0	19819.7	4.0
2000	2472.4	56.4	14889.3	4.1	2630.5	47.9	18678.1	4.5	21692.2	22.8	33628.7	4.4
3000	3587.7	68.8	17856.0	4.5	3760.0	52.0	23568.8	4.8	32341.2	23.6	46290.6	4.8
4000	4692.3	74.8	22861.5	4.8	4859.6	57.5	28530.8	5.0	42982.9	24.7	60022.1	5.0
5000	5814.2	79.0	29218.9	5.0	5956.6	62.3	36156.5	5.1	53602.5	25.3	74523.9	5.0

TABLE II. COMPARISION OF TREE CONSTRUCTION STRATEGIES FOR STATIC TYPE OF NETWORK – STREAMING RATE *Q*

Strategy	uCO		uCOL		CO		COL		RS		RSL	
q [Kbps]	cost	level	cost	level	cost	level	cost	level	cost	level	cost	level
128	4377.4	19.0	8508.9	3.0	4648.8	18.6	11153.1	3.1	38843.5	18.1	43005.8	3.1
256	4449.4	21.4	8898.5	4.4	4780.7	19.8	11478.5	4.7	38884.9	20.4	43061.9	4.7
384	4799.7	21.8	9226.0	6.2	5335.1	21.1	12384.6	6.6	39469.4	20.0	43542.5	6.6
512	4839.1	23.9	9259.7	7.5	5430.1	21.8	12470.4	7.8	39388.4	20.1	43710.7	7.9

In Table I, we show results of analogous experiment, but for dynamic type of the network. For each strategy and network size, we report the average cost and the average number of levels obtained in the final configuration. We can observe, that results for the dynamic case are similar to those for the static approach. Random tree construction strategies (*RS* and *RSL*) are much less efficient than *CO* and *COL* in terms of overall network cost. Surprisingly, *CO* strategy is creating trees with lower number of levels than strategy with no survivability (*uCO*). This is implied by the fact that the survivability constraints cause both multicast trees to connect the same number of nodes. For the strategies with no survivability constraints there is no such requirement, so one of the multicast trees can have more nodes than the other one and in effect have more levels.

The next goal of experiments was to verify how the overall streaming cost and number of levels are influenced by the streaming rate *q*. For this experiment, we set the level limit *L* to 100 and chose network with 3000 nodes. Table II presents results related to comparison of the introduced strategies in terms of the overall streaming cost and maximum depth (level) of the multicast trees in the static environment. In line with our expectations, both the overall streaming cost and the number of levels are increasing with streaming rate *q* for all tree constructing strategies.

For the streaming rate *q* equal to 640 Kbps, strategies with the survivability constraints were unable to connect all the requesting nodes, as shown on Fig. 6:

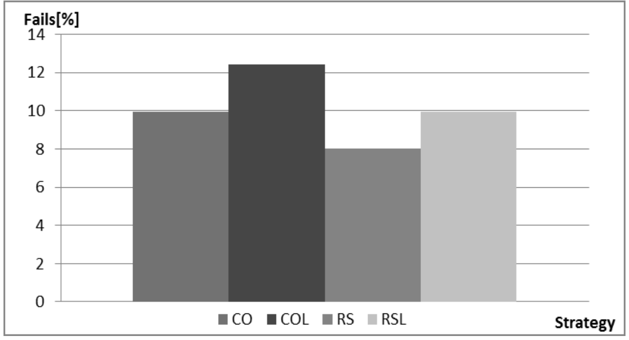


Figure 6. Percentage of not connected nodes for *q* = 640 Kbps – static network.

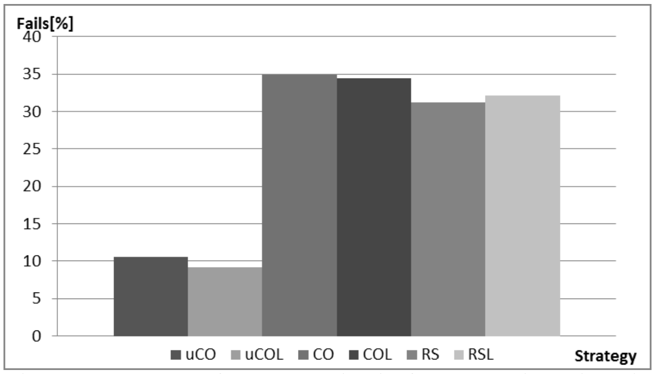


Figure 7. Percentage of not connected nodes for q = 640 Kbps – dynamic network.

Results obtained for the dynamic type of network are similar to those in Table II. Figure 7 shows percentage of nodes that failed to connect to the network for streaming rate q = 640 Kbps.

We can easily notice that all the strategies were unable to connect all of requesting nodes. Strategies with no survivability (*uCO* and *uCOL*) had average number of rejects of 9-10%. Adding the survivability constraint increases this parameter by 20-25%. The main conclusion is that for highly constrained networks (e.g. low L limit or high q value) survivability strategies may more often not be able to find any feasible solution than strategies with no survivability constraints. Those findings are in harmony with our previous conclusions [10].

In the last experiment, we tested the impact of the survivability constraints on the overall streaming cost, i.e., we compare results of *CO* against *uCO* and *COL* against *uCOL*. Results are presented in Tables III and IV. Our observations are as follows:

- The average cost (considering all experiments) of providing additional survivability was 8% for the *CO* strategy and 28% for the *COL* strategy.
- As the number of nodes grows, the gap between strategies with additional survivability requirements and normal strategies decreases. This can be explained by rapid increase in size of the solution space. As a result, it is possible to find a cost efficient solution with the survivability constraints.
- With the increase of the streaming rate q, the additional cost of providing survivability grows. Again, this is caused by the fact, that with increase of q, the solution space shrinks.

TABLE III. DIFFERENCE IN COST BETWEEN STRATEGIES WITH AND WITHOUT SURVIVABILITY [%] – SIZE OF THE NETWORK

	Static networks		**Dynamic networks**	
Number of nodes	**CO**	**COL**	**CO**	**COL**
1000	17.8%	31.8%	12.4%	16.9%
2000	10.3%	33.3%	6.4%	25.4%
3000	7.4%	29.0%	4.8%	32.0%
4000	5.5%	28.9%	3.6%	24.8%
5000	4.1%	25.3%	2.4%	23.7%

TABLE IV. DIFFERENCE IN COST BETWEEN STRATEGIES WITH AND WITHOUT SURVIVABILITY [%] – STREAMING RATE *Q*

	Static networks		**Dynamic networks**	
Streaming rate [Kbps]	**CO**	**COL**	**CO**	**COL**
128	6.2%	31.1%	4.0%	25.9%
256	7.4%	29.0%	4.8%	32.0%
384	11.2%	34.2%	7.8%	22.9%
512	12.2%	34.7%	8.9%	22.3%

VII. CONCLUSION AND FUTURE WORK

In this paper, we focused on the survivable overlay multicasting systems with the dual homing architecture in both static and dynamic networks. We introduced six different tree construction strategies (of which four provide survivability constraints), along with the overlay simulator, and compare their results for different types of networks in terms of the overall streaming cost and the number of levels in the multicast trees. According to the obtained results, we can conclude that cost optimization strategies obtain results from 4 to 8 times better than random optimization strategies in respect of the streaming cost in all multicast trees. We observe that the average cost of providing survivability in all of our experiments was 8% and 28% for *CO* and *COL* strategies, respectively. Moreover, the additional cost is decreasing with the increase of the network size. We can derive that the additional constraints that allow constructing failure-disjoint trees do not influence significantly the performance of the overlay multicasting system in terms of the streaming cost.

In future work, we plan to develop additional strategies that will provide even more survivability, like node and ISP disjoint trees, and conduct more experiments evaluating these solutions.

REFERENCES

[1] Cisco Visual Networking Index Forecast 2010–2015, http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360.pdf, 2011.

[2] Aoyama, T.: A New Generation Network: Beyond the Internet and NGN. IEEE Communications Magazine, Vol. 47, No. 5, pp. 82-87, 2009.

[3] Shen, X., Yu, H., Buford, J., and Akon, M.: Handbook of Peer-to-Peer Networking. Springer, 1st Edition, 24 November, 2009.

[4] Kwong, K-W. and Tsang, D.: Peer-to-Peer Topology Formation Using Random Walk, Handbook of Peer to Peer Networking, Springer, pp. 167-187, 2010.

[5] Pouwelse, J.A., Garbacki, P., Epema, D.H.J., and Sips, H.J.: The Bittorrent P2P File-Sharing System: Measurements and Analysis, Peer-to-Peer Systems IV, Lecture Notes in Computer Science, Volume 3640, pp. 205-216, 2005.

[6] Tutshku, K.: A Measurement-Based Traffic Profile of the eDonkey Filesharing Service, Passive and Active Network Measurement, Lecture Notes in Computer Science Volume 3015, pp. 12-21, 2004.

[7] Bonfiglio, D., Mellia, M., Meo, M., Ritacca, N., and Rossi, D.: Tracking Down Skype Traffic, INFOCOM 2008. The 27th Conference on Computer Communications. IEEE , pp. 261-265, 13-18 April 2008.

[8] Kmiecik, W. and Walkowiak, K.: Survivable P2P multicasting flow assignment in dual homing networks, 3rd

International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), pp. 1-7, Budapest, 2011.

[9] Kmiecik, W. and Walkowiak, K.: Heuristic approach for survivable P2P multicasting flow as-signment in dual homing networks, International Joint Conference CISIS'12-ICEUTE'12-SOCO'12 Special Sessions, pp. 215-224, Ostrava, 2012.

[10] Kmiecik, W. and Walkowiak, K.: Flow Assignment (FA) and Capacity and Flow Assignment (CFA) Problems for Survivable Overlay Multicasting in Dual Homing Networks, Telecommunication Systems Journal, Springer, 2013, in press.

[11] Jianping, W., Mei, Y., Xiangtong, Q., and Cook, R.P.: Dual-homing multicast protection. Global Telecommunications Conference, Vol. 2, pp. 1123-1127, USA, 2004.

[12] Palkopoulou, E., Schupke, D. A., and Bauschert, T.: Shared Backup Router Resources: Realizing Virtualized Network Resilience, Comm. Mag., IEEE, pp.140–146, May 2011.

[13] Lee, C. and Koh, S.: A design of the minimum-cost ring-chain network with dual-homing survivability: A tabu search approach, Comput. Oper. Res., vol. 24, no. 9, pp. 883–897, September 1997.

[14] Proestaki, A. and Sinclair, M.C.: Design and dimensioning of dual-homing hierarchical multi-ring networks, Communications, IEE Proc., vol. 147, pp. 96-104, April 2000.

[15] Jianping, W., Mei, Y., Xiangtong, Q., and Jiang, Y.: On finding the best partial multicast protection tree under dual-homing architecture, High Performance Switching and Routing, pp.128-132, Hong Kong, 2005.

[16] Wang, J., Vokkarane, V., Jothi, R., Qi, X., Raghavachari, B., and Jue, J.: Dual-Homing Protection in IP-Over-WDM Networks, Journal of lightwave technology, vol. 23, no. 10, pp. 3111-3124, UK, October 2005.

[17] Wang, J., Yang, M., Yang, B., and Zheng, S.Q.: Dual-Homing Based Scalable Partial Multicast Protection, IEEE Transactions on Computers, vol. 55, no. 9, pp. 1130-1141, September 2006.

[18] Thulasiraman, P., Ramasubramanian, S., and Krunz, M.: Disjoint multipath routing in dual homing networks using colored trees, Proceedings of GLOBECOM - Wireless Ad Hoc and Sensor Network Symposium, pp. 1-5, San Francisco, November/December 2006.

[19] Fei, Z. and Yang, M.: A Proactive Tree Recovery Mechanism for Resilient Overlay Multicast, IEEE/ACM Transactions on networking, vol.15, pp. 173-185, 2007.

[20] Tarkoma, S.: Overlay Networks: Toward Information Networking, Auerbach Pub., 1 edition, 9 February 2010.

[21] Naicken, S. and Wakeman, I.: Towards Yet Another Peer-to-Peer Simulator, Proc Fourth International Working Conference Performance Modelling and Evaluation of Heterogeneous Networks HETNETs' 06 , UK, 2006.

[22] Walkowiak, K. and Przewoźniczek, M.: Modeling and optimization of survivable P2P multicasting, Computer Comm., vol. 34, issue 12, pp. 1410-1424, August 2011.

[23] Pióro, M. and Medhi, D.: Routing, Flow, and Capacity Design in Communication and Computer Networks, Morgan Kaufman Publishers, ISBN:978-0-12-557189-0, July 2004.

[24] Xiaojun, H., Liang, C., Liang, J., Liu, Y., and Ross, K.: A Measurement Study of a Large-Scale P2P IPTV System, IEEE Transactions on Multimedia, Vol. 9, pp. 1672-1687, Number 8, 2007.

Evaluating OpenFlow Controller Paradigms

Marcial P. Fernandez
Universidade Estadual do Ceará (UECE)
Av. Paranjana 1700
Fortaleza - CE - Brazil
marcial@larces.uece.br

***Abstract*—The OpenFlow architecture is a proposal from the Clean Slate initiative to define a new Internet architecture where the network devices are simple, and the control and management plane is performed by a centralized controller. The simplicity and centralization architecture makes it reliable and inexpensive, but the centralization causes problems concerning controller scalability. An OpenFlow controller has two operation paradigms: reactive and proactive. The performance of both paradigms were analyzed in different known controllers. The performance evaluation was done in a real environment and emulation. Different OpenFlow controllers and distinct amount of OpenFlow devices were evaluated. The analysis has demonstrated the shortcoming of reactive approach. In conclusion, this paper indicates the effectiveness of a hybrid approach to improve the efficiency and scalability of OpenFlow architecture.**

***Keywords*-Openflow; OpenFlow Controller; Performance evaluation.**

I. INTRODUCTION

The OpenFlow [1] architecture is a proposal from the Clean Slate initiative to define an open protocol that sets up forward tables in switches. It is the basis of the Software Defined Network (SDN) architecture, where the network can be modified by the user. This proposal tries to use the most basic abstraction layer of the switch, it is the definition of forward tables, in order to achieve better performance. The OpenFlow protocol can set a tuple of condition-action assertion on switches like forward, filter and also, count the packets that matches the condition. The network management is performed by the OpenFlow Controller maintaining the switches simple, only with the packet forwarding function.

The OpenFlow architecture provides several benefits: (1) OpenFlow centralized controllers can manage all flow decisions reducing the switch complexity; (2) a central controller can see all networks and flows, giving global and optimal management of network provisioning; and (3) OpenFlow switches are relatively simple and reliable, since forward decisions are defined by a controller, rather than by a switch firmware [2]. However, OpenFlow couples two characteristics: a central controller and simple devices, that result in scalability problems.

As the number of OpenFlow switches increases, relying on a single controller for the entire network might not be feasible for several reasons: (1) the amount of control messages destined to the centralized controller grows with the number of switches; (2) with the increase of network diameter, some switches will have longer setup delay, independently where the controller is placed [2]; and, (3) since the system is bounded by the controller processor power, setup times can grow significantly when the number of switches and the size of the network grow.

Two paradigms were implemented on some OpenFlow controllers: the NOX/C++ Controller [3], the POX/Python Controller [3], the Trema/Ruby Controller [4] and the Floodlight/Java Controller [5]. In each controller and device the reactive and proactive approaches were implemented and they were evaluated.

The rest of the paper is structured as follows. In Section II, we present some related work. Section III introduces the OpenFlow architecture fundamentals; and in Section IV, we present the evaluation methodology and tests. Section V shows the results and Section VI concludes the paper.

II. RELATED WORK

Tootoonchian and Ganjali proposed the HyperFlow [6], a mechanism that changes the controller paradigm from centralized to distributed. HyperFlow tries to provide scalability, using as many controllers as necessary to reach it, but keeping the network control logically centralized. The proposal uses a publish/subscribe messaging system to propagate controller events to others, maintaining the database consistent. However, this approach can only support global visibility of rare events such as link state changes, not frequent events such as flow arrivals.

Another proposal, the DevoFlow [2], aims to deal with the scalability problem by devolving network control to switches with an aggressive use of flow wildcard and introducing new mechanisms to improve visibility. They introduced two new mechanisms to be implemented on switches: *rule cloning* and *local actions*.

The Source-Flow controller [7], proposed by Chiba, Shinohara and Shimonishi, uses a similar approach. It tries to reduce the number of flow entries using a MPLS-like tunneling approach in order do reduce the Ternary Content-Addressable Memory (TCAM) used space.

ISBN: 978-1-61208-245-5

In this paper, we evaluate some OpenFlow controller's performance working in reactive and proactive approach. Then, we propose a new controller architecture performing a hybrid approach, making better use of both paradigms.

III. OpenFlow Architecture

The OpenFlow architecture has several components: the OpenFlow controller, the OpenFlow device (switch), and the OpenFlow protocol. Figure 1 shows the components of an OpenFlow architecture. The OpenFlow approach considers a centralized controller that configures all devices. Devices should be kept simple in order to reach better forward performance and the network control is done by the controller.

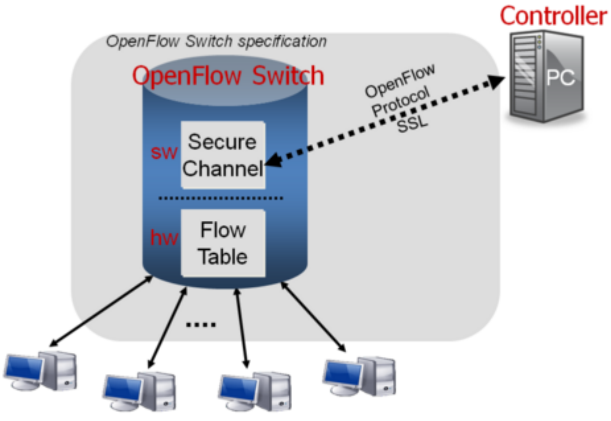


Figure 1. The OpenFlow architecture [1]

The OpenFlow Controller is the centralized controller of an OpenFlow network. It sets up all OpenFlow devices, maintains topology information, and monitors the overall status of entire network. The OpenFlow Device is any OpenFlow capable device in a network such as a switch, router or access point. Each device maintains a Flow Table that indicates the processing applied to any packet of a certain flow. The OpenFlow Protocol works as an interface among the controller and the switches setting up the Flow Table. The protocol should use a secure channel based on Transport Layer Security (TLS).

The controller updates the *Flow Table* by adding and removing Flow Entries using the OpenFlow Protocol. The Flow Table is a database that contains Flow Entries associated with actions to command the switch to apply some actions on a certain flow. Some possible actions are: forward, drop and encapsulate.

Each OpenFlow device has a Flow Table with flow entries as shown in Figure 2. A Flow Entry has three parts: Rule, Action and Statistics. The Rule field is used to define the match condition to a specific flow; Action field defines the action to be applied to this flow, and Stat field is used to count the rule occurrence for management purposes. When a packet arrives to the OpenFlow Switch, it is matched against Flow Entries in the Flow Table. The Action will be triggered if the flow Rule is matched and then, the Stat field is updated. If the packet does not match any entry in the Flow Table, it will be sent to the Controller over a secure channel to ask for an action. Packets are matched against all flow entries based on some prioritization scheme. An entry with an exact match (no wildcards) has the highest priority. Optionally, the Flow Table could have a priority field (not shown in figure) associated with each entry. Higher number indicates that the rule should be processed before.

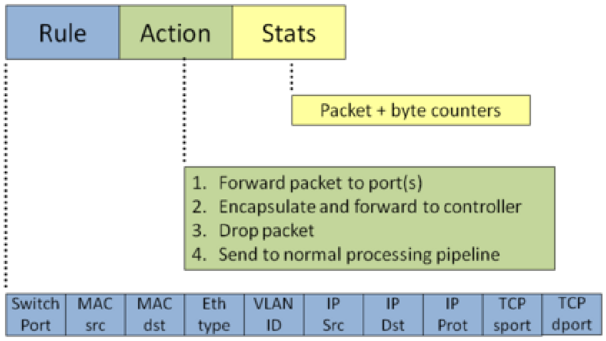


Figure 2. The OpenFlow Flow Entry [8]

The Openflow Controller presents two behaviors: reactive and proactive. In the **Reactive** approach, the first packet of flow received by switch triggers the controller to insert flow entries in each OpenFlow switch of network. This approach presents the most efficient use of existing flow table memory, but every new flow causes a small additional setup time. Finally, with hard dependency on the controller, if the switch loses the connection, it cannot forward the packet.

In the **Proactive** approach, the controller pre-populates the flow table in each switch. This approach has zero additional flow setup time because the forward rule is defined. Now, if the switch loses the connection with controller, it does not disrupt traffic. However, the network operation requires a hard management, e.g., requires to aggregate (wildcard) rules to cover all routes.

The OpenFlow Protocol uses the TCP protocol and port 6633. Optionally, the communication can use a secure channel based on TLS. The OpenFlow Protocol supports three types of messages [8]:

1) Controller-to-Switch Messages: These messages are sent only by the controller to the switches; they perform the functions of switch configuration, modifying the switch capabilities, and also manages the Flow Table.

2) Symmetric Messages: These messages are sent in both directions reporting on switch-controller connection problems.

3) Asynchronous Messages: These messages are sent by the switch to the controller to announce changes in the network and switch state. All packets received by the switch are compared against the Flow Table. If the packet matches any Flow Entry, the action for that entry is performed on the packet, e.g., forward a packet to a specified port. If there is no match, the packet is forwarded to the controller that is responsible for determining how to handle packets without valid Flow Entries [8].

It is important to note that when the OpenFlow switch receives a packet to a nonexistent destination in the Flow Table, it requires an interaction with the controller to define the treatment of this new flow. At least, the switch will need to send a message to the controller with regards to the new packet received (message Packet-In). If the path is already predefined (there is an entry in Flow Table), this procedure is not necessary, reducing the amount of messages exchanged through the network and reducing the processing at the controller.

Furthermore, the maintenance of unused Flow Entries in the switch Flow Tables requires fast TCAM memory. Therefore it is necessary to remove unused flows using a time-out mechanism. If a flow previously excluded by time-out restarts, it is necessary to reconfigure all switches on the end-to-end path.

A. OpenFlow Device

An OpenFlow device is basically an Ethernet switch supporting OpenFlow protocol. But there are different implementation approaches: OpenFlow-enabled switch and OpenFlow-compliant switch.

The OpenFlow-enabled switch uses off-the-shelf hardware, i.e., traditional switches with OpenFlow protocol that translate the rule according to the hardware chipset implementation. The OpenFlow-enabled switch re-uses existing TCAM, that in a conventional switch has no more than only few thousands of entries for IP routing and MAC table. Considering that we need at least one TCAM entry per flow, in a current hardware, it would not be enough for production environments. The Broadcom chipset switches based on Indigo Firmware [9], e.g., Netgear 73xxSO, Pronto Switch and many other, are example of this approach.

The OpenFlow-compliant switch uses specific network chipset, designed to provide better performance to OpenFlow devices. The OpenFlow philosophy relies on matching packets against multiple tables in the forwarding pipeline, where the output of one pipeline stage is able to modify the contents of the table of next stage. Some example are devices based on the EZChip NP-4 Network Processor [10]. But, nowadays, there are few commercial OpenFlow-compliant switches, one example is the NoviFlow Switch 1.1 [11].

B. OpenFlow Controller

All functions of the control and management plane are performed by the controller. It has full network topology information and the location of hosts and external paths (MAC and routing tables). When a switch receives a packet in which there is no entry in its Flow Table, it forwards the message to the controller asking for the action to take upon this new flow. The controller can define the port that the flow must be forwarded to or take other actions, such as dropping the packet. The controller must set the entire path by sending configuration messages to all switches from the source to the destination.

Scalability and redundancy are possible using a stateless OpenFlow control, allowing simple load-balancing over multiple devices [1]. Due to the OpenFlow centralized architecture, controller scalability issues have received attention by researchers. Many authors focus on distributed architectures to improve the scalability problem.

The most common OpenFlow controller operation mode is the reactive. In the reactive mode, the controller listens to switches passively and configures routes on-demand. It receives messages of connected hosts from the switches and treats ARP message from hosts in order to maintain a global MAC table. Upon receiving an ARP message, the controller looks for the destination host location and sets the path by sending OpenFlow messages to affected switches. After a time out, an unused entry is excluded from the table. The reactive behavior allows a more efficient use of memory (MAC table) at the cost of the re-establishing path later, if necessary. In the proactive mode, paths are set up in advance. This comes at the cost of lower memory space efficiency and the requirement for a priori setup of all paths.

The controller performance is a central issue of an OpenFlow architecture. A controller can only support a limited number of flow setups per second. In the former Ethane experiment [12] with similar approach, i.e., only one controller sets many switches, each controller could support 10K new flows per second. Another work, Tavakoli et al. [13] shows that one NOX controller can handle a maximum of 30K new flow installs per second maintaining a flow install time below 10 ms. From the network side, Kandula et al. [14] measure on a 1500-server cluster datacenter the creation of 100K new flows per second, implying a need for, at least, four OpenFlow controllers. As the OpenFlow philosophy relies on a single controller, we can question the feasibility of OpenFlow use in (not so big) production datacenter.

Several approaches have proposed new OpenFlow Controller architectures and implementations. One of the first approaches was the NOX Controller [3], a centralized controller that implements a reactive and proactive approach. The NOX default operation mode is the reactive, but offer an API to allow users to set flows in a proactive mode. The NOX framework is used as the basis for many controller's implementation, for example, the POX controller. POX controller is a pure Python controller, redesigned to improve the performance compared to original Python NOX. The former NOX was redesigned to provide a pure C++ controller.

The Floodlight [5] is Java-based OpenFlow Controller, forked from the Beacon controller developed at Stanford. The Floodlight controller is an open-source software Apache-licensed, supported by a community of developers. It offers a modular architecture, easy to extend and enhance.

The Trema [4] is an OpenFlow controller framework developed in Ruby and C. It is basically a framework, including basic libraries and functional modules that work as an interface to OpenFlow switches. Several sample applications are provided to permit execution of different controllers, making easy the extension to new features.

IV. Evaluating OpenFlow Controller's Paradigm

To evaluate the OpenFlow controller's performance, two scenarios were built: (1) a real network with Netgear GSM-7328SO switches and (2) a virtualized network using Mininet [15]. In each scenario, a host generates traffic to cross the entire network topology simulating a production network.

On the server, the OpenFlow controller under evaluation is installed. In the host, it runs the Cbench benchmark software to stress the controller's capacity. There is also another host to generate real traffic crossing the network using Mausezahn software [16]. The evaluation does not intend to compare the controller software; the main objective is to compare the performance and the behavior of each controller using the reactive and proactive approach.

For traffic generation, Mausezahn software was used [16]. Mausezahn is an open-source traffic generator written in C, which allows to send nearly any possible packet. In order to stress the OpenFlow controller, Mausezahn generates packets with one million different IP addresses, forcing the switch to send one million of Openlow requests to the controller. IP address was used instead of MAC address, to simulate a normal OpenFlow operation, where an ARP request starts the route request to the controller.

To evaluate the controller performance, CBench software was used [17]. CBench is a performance measurement tool designed for benchmarking OpenFlow controllers. The benchmarking measurement is the amount of flows per second that can be processed by the controller.

The controller code has been modified to implement the reactive and proactive behavior. A special configuration message was used to set the switches to work in a reactive or proactive way. The Indigo Firmware was also modified to receive this message and set the switch to work in reactive or proactive manner. The OpenVswitch code was changed to implement the same behavior on virtualized Mininet environment.

A. Evaluation Scenario

The tests were executed in two environment: real network and emulated network. The real network was made with Netgear GSM-7328SO switches running on modified Indigo Firmware release 2012.03.19, as shown in Figure 3.

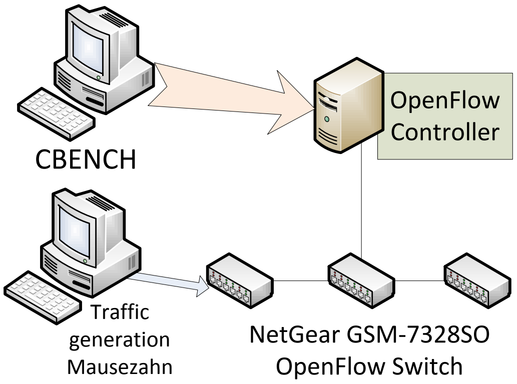

OpenFlow Controller
CBENCH
Traffic generation Mausezahn
NetGear GSM-7328SO OpenFlow Switch

Figure 3. Real switch scenario

The second test environment chosen was the OpenFlow emulation over virtual machine using Mininet [15]. This model permits an evaluation in emulation environment using only one computer. Mininet imposes a restriction on link layer configuration, e.g., we cannot specify link bandwidth or error rate, but for the validation this environment was suitable to obtain the results.

The experiment was built over VMware Workstation 8. In the virtualized environment, Mininet was used [15]. Mininet is a network emulator used to create Software Defined Networks (SDNs) scenario in Linux environment. The Mininet system permits the specification of a network interconnecting "virtualized" devices. Each network device, hosts, switches and controller are virtualized and communicate via Mininet. The traffic flow used in this test is generated by Mausezahn. A Python script is used to create the topology in Mininet and the traffic flows setup are received from a remote OpenFlow controller.

Therefore, the test environment implements and performs the real protocol stacks that communicate with each other virtually. The Mininet environment allows the execution of real protocols in a virtual network. The virtual topology created in Mininet platform is shown in Figure 4.

B. Evaluation Procedure

To define the experiment, initially it is necessary to specify the hosts and network that will be used. The OpenFlow controller has the responsibility to define the best path to connect all hosts.

To evaluate the controller performance into the test topology the *Cbench* [17] program was included, part of the OpenFlow suite that creates and sends a large amount of OpenFlow messages to the controller in order to test its performance. These messages do not represent any real

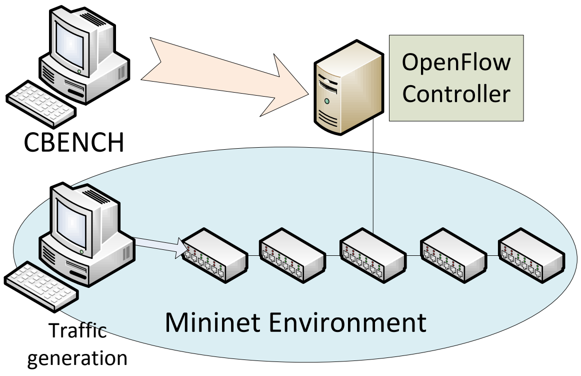


Figure 4. Virtual switch scenario

network topology, only "deceive" the controller that handles and sends messages believing it is dealing a larger network. As result, we obtain the number of OpenFlow messages the controller can support per second, besides the messages sent by real switches or virtualized switches in Mininet.

The tests with the Cbench, simulated the presence of more than 100 switches, in addition to the 50, 100 and 200 virtualized switches topology created on Mininet. In each round, 16 tests were performed, and in these tests the average and the standard deviation were calculated. Finally, the graph was plotted of average performance with 95% confidence interval. Since we are interested in studying the system in equilibrium, we do not consider the first 2 minutes of data as the warm up period.

In a normal OpenFlow network, when a host performs an ARP request to find the destination address and the switch has no such address in its Flow Table, the switch makes a query to the OpenFlow controller. The controller will decide what action would be applied to this flow, e.g., choose the path from origin to the destination to forward the packets to destination host. To optimize the memory usage in the switches, a timeout (in our experiment 10 seconds) sets the removal of this entry on its flow table, forcing the path rebuild whenever it is necessary.

In the Proactive approach, the path is already predefined for the necessary time and therefore it is not necessary to ask the controller to build the path from source to destination. The proactive approach is implemented on controller's and switch's code. The reduction of messages to discover a new path among switches allows the increase of controller performance, as shown on the results in Section V below.

V. Results

The performance test results are shown in the following graphs. They show the performance of each controller in how many OpenFlow messages could be processed according to the amount of additional switches which were created in real scenario and in Mininet environment.

Figure 5 shows the performance of each OpenFlow controller in real switch network, shown in Figure 3. We can see that every controller, independently of architecture and programming language, has better performance in the proactive approach.

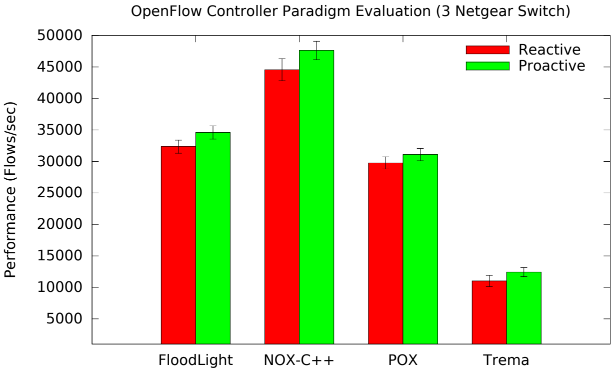


Figure 5. Netgear switch network evaluation

Figure 6 shows the performance of each OpenFlow controller in emulated Mininet environment, shown in Figure 4. We can see similar results, proactive approach gives better performance.

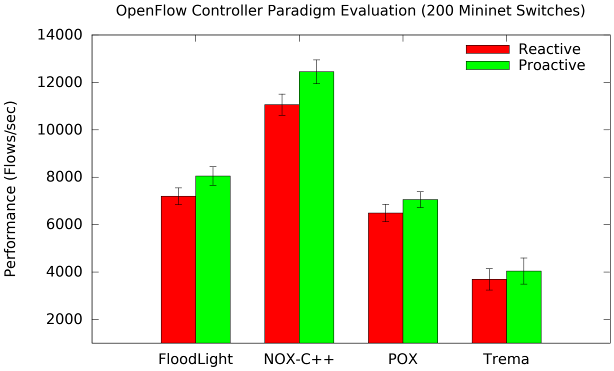


Figure 6. Mininet 200 switch network evaluation

Figure 7 shows the NOX-C++ controller performance from 3 real switches to 200 virtual switches. The performance measured by Cbench benchmark reduces according to the increase of number of switches in network.

The improvement is due to proactive controller receives less request message from switch because the path is already set. Receiving fewer messages from "real" network allows

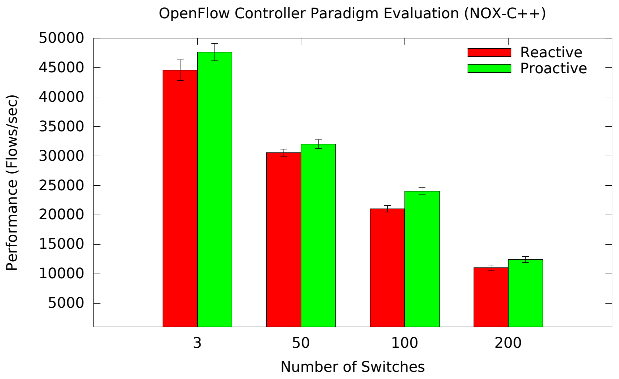


Figure 7. NOX-C++ controller evaluation

the controller to handle more “fake” messages from *Cbench* program. We can also notice that increasing the number of virtualized switches in Mininet, the performance will be reduced.

The proactive operation improves the controller performance. As the user pre-configures a path, all packets from this flow already have a Flow Entry on all switch’s Flow Table, the switches do not need to send a message to the controller. Then, the reduction of messages sent by switches reduces the amount of messages received by the controller, providing capacity to controller dealing more messages from other on-demand flows.

However, the proactive approach requires the controller know the traffic flows in advanced to configure the paths before it is used. The reactive approach reduces the controller’s performance but requires less configuration effort. In reactive approach, it is necessary only to configure an IP address to put the network in operation.

VI. Conclusion and Future Works

Although we consider that the OpenFlow architecture has a prominent future due to its simplicity and suitability to new technologies, its centralized architecture causes a scalability problems. This problem has been studied by researchers from several points of view.

This paper analyzes the performance of different OpenFlow controller operating in reactive and proactive approach. All evaluation’s results show the increase on controller performance when it used the proactive approach. However, we agree that reactive approach makes easy on the network operation and management.

Our proposal tries to indicate a possible solution to this problem, by adding a new intelligent switch that sets the paths on-demand, improving the performance of proactive approach but maintaining the facility of reactive approach. The controller acts like a learning switch without management interference but can perform a proactive set up using Reinforcement Learning. The results show an improvement in controller performance due to the fact of reduction on control messages.

But some heuristics will be able to set the paths in advance automatically, improving the performance maintaining the simplest configuration. The use of Reinforcement Learning, proposed in Boyan and Littman [18], and also, by Peshkin and Savova [19], can define the routes without the user intervention.

As future work, it will be interesting to improve the system manageability implementing the Reinforcement Learning mechanism to set the routes based on traffic behavior. Another proposal, is to adjust the OpenFlow switch Flow Table time-out based on traffic behavior.

Acknowledgment

We would like to thank Broadcom Corporation for their support.

References

[1] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, “OpenFlow: enabling innovation in campus networks,” *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 2, pp. 69–74, 2008.

[2] J. C. Mogul, J. Tourrilhes, P. Yalagandula, P. Sharma, A. R. Curtis, and S. Banerjee, “Devoflow: cost-effective flow management for high performance enterprise networks,” in *Proceedings of the Ninth ACM SIGCOMM Workshop on Hot Topics in Networks*, ser. Hotnets ’10. New York, NY, USA: ACM, 2010, pp. 1:1–1:6.

[3] N. Gude, T. Koponen, J. Pettit, B. Pfaff, M. Casado, N. McKeown, and S. Shenker, “Nox: towards an operating system for networks,” *SIGCOMM Comput. Commun. Rev.*, vol. 38, no. 3, pp. 105–110, 2008.

[4] NEC, “Trema Openflow Controller,” Last accessed, Aug 2012. [Online]. Available: http://trema.github.com/trema/

[5] D. Erickson, “Floodlight Java based OpenFlow Controller,” Last accessed, Aug 2012. [Online]. Available: http://floodlight.openflowhub.org/

[6] A. Tootoonchian and Y. Ganjali, “HyperFlow: A distributed control plane for OpenFlow,” in *Proceedings of the 2010 internet network management conference on Research on enterprise networking*. USENIX Association, 2010, p. 3.

[7] Y. Chiba, Y. Shinohara, and H. Shimonishi, “Source flow: handling millions of flows on flow-based nodes,” *SIGCOMM Comput. Commun. Rev.*, vol. 40, pp. 465–466, August 2010.

[8] B. Heller, “Openflow switch specification, version 1.0.0,” Last accessed, Dec 2011. [Online]. Available: www.openflowswitch.org/documents/openflow-spec-v1.0.0.pdf

[9] D. Talayco, “Indigo OpenFlow Switching Software Package,” Last accessed, Jun 2012. [Online]. Available: http://www.openflowswitch.org/wk/index.php/IndigoReleaseNotes

[10] O. Ferkouss, I. Snaiki, O. Mounaouar, H. Dahmouni, R. Ben Ali, Y. Lemieux, and O. Cherkaoui, "A 100gig network processor platform for openflow," in *Network and Service Management (CNSM), 2011 7th International Conference on*. IEEE, 2011, pp. 1–4.

[11] NoviFlow, "NoviFlow Switch 1.1," Last accessed, Sep 2012. [Online]. Available: http://www.noviflow.com/index.asp?node=2&lang=en

[12] M. Casado, M. J. Freedman, J. Pettit, J. Luo, N. McKeown, and S. Shenker, "Ethane: taking control of the enterprise," in *Proceedings of the 2007 conference on Applications, technologies, architectures, and protocols for computer communications*, ser. SIGCOMM '07. New York, NY, USA: ACM, 2007, pp. 1–12.

[13] A. Tavakoli, M. Casado, T. Koponen, and S. Shenker, "Applying NOX to the Datacenter," in *Proceedings of workshop on Hot Topics in Networks (HotNets-VIII)*, 2009.

[14] S. Kandula, S. Sengupta, A. Greenberg, P. Patel, and R. Chaiken, "The nature of data center traffic: measurements & analysis," in *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*. ACM, 2009, pp. 202–208.

[15] B. Lantz, B. Heller, and N. McKeown, "A network in a laptop: rapid prototyping for software-defined networks," in *Proceedings of the Ninth ACM SIGCOMM Workshop on Hot Topics in Networks*, ser. Hotnets '10. New York, NY, USA: ACM, 2010, pp. 19:1–19:6.

[16] H. Haas, "Mausezahn Traffic Generator Version 0.4," Last accessed, Jan 2012. [Online]. Available: http://www.perihel.at/sec/mz/

[17] R. Sherwood and K.-K. Yap, "Cbench (controller benchmarker)," Last accessed, Nov 2011. [Online]. Available: http://www.openflowswitch.org/wk/index.php/Oflops

[18] J. Boyan and M. Littman, "Packet routing in dynamically changing networks: A reinforcement learning approach," *Advances in neural information processing systems*, pp. 671–671, 1994.

[19] L. Peshkin and V. Savova, "Reinforcement learning for adaptive routing," in *Neural Networks, 2002. IJCNN'02. Proceedings of the 2002 International Joint Conference on*, vol. 2. IEEE, 2002, pp. 1825–1830.

Performance of an IPv6 Web Server under Congestion

A. Loukili, A. K. Tsetse, A. L. Wijesinha, R. K. Karne, and P. Appiah-Kubi
Department of Computer & Information Sciences
Towson University
Towson, MD 21252, U.S.A.
{aloukili, awijesinha, rkarne, atsetse, appiahkubi}@towson.edu

***Abstract*-We conduct experiments using an IPv6 Web server in a test LAN environment with several routers to determine the performance under congestion due to IPv6 and IPv4 traffic. The experiments use an Apache Web server and a bare PC Web server with no operating system. Requests to the servers are made using an ordinary Web browser. Different levels of congestion are created by using MGEN traffic generators. It is found that the IPv4 throughput is slightly greater than (or approximately equal to) the IPv6 throughput under the same level of congestion. When the IPv4 throughput is larger, the differences are between 4-23%. However, Apache server delays for HTTP requests over IPv6 are between 6-32 ms more than for IPv4 depending on the level of congestion. For all congestion levels, the bare PC Web server has significantly lower throughput and larger delays than Apache regardless of whether IPv6 or IPv4 is used since it does not implement any TCP optimizations. The results show that Web server throughput and delay for browser requests depend on both the congestion traffic rate and the percentage of like traffic in the congestion mix.**

Keywords-IPv6; congestion; Web server; performance; bare PC.

I. INTRODUCTION

A recent survey of 67 top ISPs in several countries showed that 97% of them have implemented, or plan to implement the next-generation IP (IPv6) by 2013 or later [1]. Yet, due to the large number of sites including home networks that currently use IPv4 and the many IPv4 address-sensitive applications that would not work over IPv6 without code modification, IPv6 and IPv4 are likely to co-exist for an extended period of time. While IPv4 performance has been researched extensively, fewer studies deal with IPv6 performance, and with performance of servers that handle requests over both IPv6 and IPv4 when the network carries a combination of IPv6 and IPv4 traffic. We evaluate the performance of Apache and bare PC Web servers under congestion resulting from different mixes of IPv6 and IPv4 traffic by measuring network throughput and delay.

The experiments are conducted in a test LAN environment consisting of several subnets connected by routers. MGEN (MultiGenerator) traffic generators are used to create IPv4 and IPv6 background traffic at moderate and higher levels of network congestion. We use primarily TCP background traffic to reflect its predominant use in the Internet and a small amount of background UDP traffic to represent applications such as VoIP, live video, or support protocols such as DNS or DHCP. The use of a bare PC Web server with no operating system enables the impact of operating system overhead, and of not using TCP optimizations or congestion control, to be determined. The throughput and delay measurements are obtained by making requests to the Web servers using an ordinary (Firefox) Web browser. The main findings are that 1) Web server throughput is not significantly different for requests over IPv6 and IPv4; 2) Delays over IPv6 can be much larger than delays over IPv4; 3) the Apache Web server performs significantly better than the bare PC Web server for all levels of congestion; and 4) throughput and delay depend on both the congestion traffic rate and the percentage of like traffic causing the congestion.

The remainder of the paper is organized as follows. In Section II, we discuss related work. In section III, we describe the experimental set up. In section IV, we present the results. In Section V, we conclude the paper.

II. RELATED WORK

A recent study on IPv6 performance [2] concludes that RTTs for IPv6 connections are less than for IPv4, although they have higher packet loss. Higher loss over IPv6 is also noted in [3], although they find that delays over IPv6 are larger, which agrees with the studies of real-time voice and video in [4]. In [5], it is claimed that native IPv6 has significantly better throughput than IPv4 due to enhanced routing capabilities. In contrast, based on measurement studies, it is found in [6] that routing inefficiencies are the cause of poor IPv6 performance although IPv4 and IPv6 performance for data are compatible. In [7], using values of throughput, delay and other metrics in a testbed, it is determined that network performance for IPv4 and IPv6 may differ depending on traffic types and the operating system. Global Internet measurements are used to compare latency and loss over IPv4 and IPv6 in [8]. While overall performance over IPv4 is often better, about 10% of the time, latency over IPv6 can be between 10-38 ms less. Internet packet traces are used in [9] to study features of IPv6 packets. It is shown that IPv6 traffic has more self-similarity than IPv4 traffic resulting in poorer performance. In addition to the above studies comparing IPv6 and IPv4 performance, extensive studies proposing a variety of approaches for TCP congestion control [10] have also been conducted.

Our study differs from the above studies since they do not specifically determine the throughput and delay associated with browser requests over IPv6 and IPv4 to a Web server under a mix of IPv6 and IPv4 congestion traffic.

It also differs from the study in [11] comparing the performance of Apache and IIS Web servers under congestion, which used IPv4 requests and IPv4 background traffic.

This study also uses an IPv6-IPv4 capable bare PC Web server with no operating system. Bare PC systems are based on the Dispersed Operating System Computing (DOSC) concept introduced in [12]. The bare PC C++ interfaces to the hardware used by applications (such as the bare PC Web server and client) are described in [13]. The implementation and performance of a bare PC Web server that runs over IPv4 are described in [14]. The IPv4-IPv6 capable bare PC Web server used in this study was built by modifying an IPv4 bare PC Web server.

III. EXPERIMENTAL SET UP

The network for our experiments is shown in Fig. 1. The test LAN consists of five Ethernets connected by routers. All the routers run Fedora 12 Kernel Linux 2.6.31.5-127.fc12.i686, and the network interface cards are 1 Gbps except for a 100 Mbps card on the client side of router R1. The Ethernet switches used are 1 Gbps except for switch S0, which is 100 Mbps. The 100 Mbps link and network act as a bottleneck to create congestion.

Two pairs of machines running MGEN [15] generate the background TCP and UDP congestion traffic. One machine (Dell OPTEPLEX GX260, CentOS Version 2.16.0) is a TCP source and UDP sink, and its peer (Dell OPTEPLEX GX260, Windows XP Professional 2002 SP3) is a TCP sink and UDP source to generate background IPv4 traffic. The other pair (same specifications as the first pair) serves to generate background IPv6 traffic in a similar manner. An Apache HTTP Server 2.2.16 running Fedora 12 (Constantine) Kernel Linux 2.6.31.5-127.fc12.i686 or a bare PC Web server is used as the Web server. A Firefox browser version 3.5.4 running Fedora Linux kernel 2.6.31.5-127.fc12.i686 (on the machine labeled as client in Fig. 1) makes individual requests to the Web server. The clients, servers, and routers run on Dell OPTEPLEX GX 520 PCs.

The background traffic consists of a mix of 10% UDP and 90% TCP v4 traffic, which reasonably represents the traffic composition for these protocols in the current Internet. Two different rates of background traffic are used: 75 Mbps representing moderate congestion and 100 Mbps representing higher congestion. Each rate was generated in three ways using different percentages of TCP/UDP traffic over IPv4 and IPv6 while maintaining the overall 90/10% TCP/UDP mix. These percentages with their respective compositions of IPv4 and IPv6 traffic are shown in Table I, where the percentages for 75 Mbps are labeled as congestion levels C1B, C2B, and C3B, and those for 100 Mbps are labeled as C4B, C5B, and C6B.

Note that percentages of IPv4 and IPv6 TCP/UDP traffic for levels C1B and C4B are equal. Likewise, the percentage of IPv6 TCP/UDP traffic for levels C2B and C5B is three times that of IPv4 traffic, and the reverse is true for levels C3B and C6B. The amounts of TCP data carried in the MGEN packets are 1440 and 1460 bytes respectively for IPv6 and IPv4, and MGEN is configured to use 10 flows at different rates to achieve the desired levels of congestion.

The throughput and delay when the Firefox Web browser requests the 320 KB file from the Apache or bare PC Web server were determined by using the network protocol analyzer Wireshark [16] (with port mirroring) at the server side (i.e., connected to switch S4). The throughput ignores retransmissions, but considers both incoming and outgoing traffic associated with a single request. From the Wireshark traces, the values of the delay were computed by using the time stamps for the HTTP Get request and the last valid ack sent by the client (before the FIN+ACK). While the throughput and delay are related, the delay is the delay for the data transfer only i.e., it is not the total delay for the request since it is not measured from the TCP SYN. Each experiment was repeated ten times (for each of IPv4 and IPv6), and the average values of the delay were used.

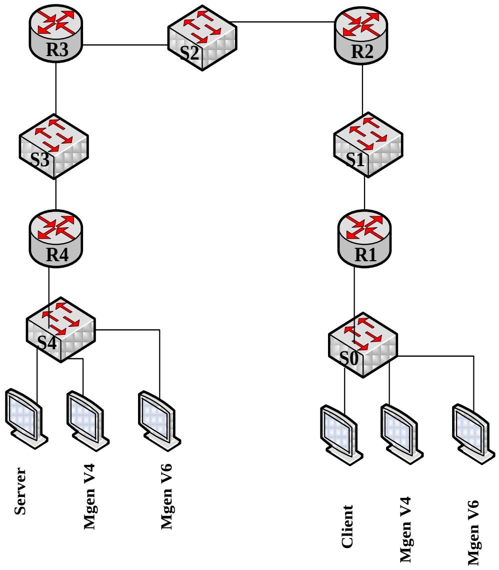


Figure 1. Test network

TABLE I. BACKGROUND TRAFFIC CONDITIONS

Congestion Level	Percentages of V4 and V6 TCP/UDP Traffic				Rate (Mbit s/sec)
	TCP V4	TCP V6	UDP V4	UDP V6	
C1B	45	45	5	5	75
C2B	22.5	67.5	2.5	7.5	75
C3B	67.5	22.5	7.5	2.5	75
C4B	45	45	5	5	100
C5B	22.5	67.5	2.5	7.5	100
C6B	67.5	22.5	7.5	2.5	100

ISBN: 978-1-61208-245-5

IV. RESULTS

A. *Apache Throughput*

The throughput associated with a single browser request to the Apache server was obtained as described above. The results are shown in Fig. 2. It is seen that the throughput for IPv4 is slightly higher than for IPv6 except for congestion level C3B when they are approximately equal (IPv6 throughput is higher, but the difference is only about 0.4 Mbps). The other differences range from approximately 1 Mbps (for congestion level C6B) to 7 Mbps (for congestion level C2B).

It can also be seen from the figure that throughput values range from about 20-27 Mbps for IPv6 and from about 22-31 Mbps for IPv4. The highest throughput is with congestion level C3B and C2B for IPv6 and IPv4 respectively, and the lowest throughput is with level C4B for both IPv6 and IPv4. So the IPv6 throughput is highest when the congestion traffic is lower (75 Mbps) and its percentage of IPv6 traffic is lower (25%), and it is lowest when the congestion traffic is higher (100 Mbps) and its percentage of IPv6 traffic is equitable (50%). Similarly, the IPv4 throughput is highest when the congestion traffic is lower and its percentage of IPv4 traffic is lower, and it is lowest when the congestion traffic is higher and its percentage of IPv4 traffic is equitable. For both IPv6 and IPv4, the throughput is highest when congestion traffic is lower and like traffic is lower, and it is lowest when congestion traffic is higher and like traffic is 50% or more. Also, the low percentage of UDP traffic appears to have a negligible impact on throughput as would be expected.

B. *Apache Delay*

In the absence of congestion traffic, the delay for the file transfer from Apache to the browser was found to be 48 milliseconds for both IPv4 and IPv6. Fig. 3 shows the transfer delay when the browser makes the request under each of the six congestion levels C1B-C6B. The delays during congestion are much larger, between 84-103 ms for IPv4 and between 105-121 ms for IPv6. Also, the delay difference between the IPv4 and IPv6 delays (for a given congestion level) varies from 6-32 ms. The difference between IPv6 and IPv4 delays are largest for congestion levels C2B and C5B, which have larger percentages of IPv6 traffic (75%), and smallest for level C3B, which has lower congestion traffic (75 Mbps) and a lower percentage of IPv6 traffic (25%).

The highest delay is with congestion level C4B and C6B for IPv6 and IPv4 respectively, and the lowest delay is with level C3B and C2B for IPv6 and IPv4 respectively. It can be seen that the IPv6 delay is highest when the congestion traffic is higher (100 Mbps) and its percentage of IPv6 traffic is equitable (50%), and it is lowest when the congestion traffic is lower (75 Mbps) and its percentage of IPv6 traffic is lower (25%). Similarly, the IPv4 delay is highest when the congestion traffic is higher and its percentage of IPv4 traffic is higher, and it is lowest when the congestion traffic is lower and its percentage of IPv4 traffic is lower. For both IPv6 and IPv4, the delay is highest when congestion traffic is higher and like traffic is 50% or more, and it is lowest when congestion traffic is lower and like traffic is lower.

The above results using the Apache server and an ordinary browser suggest that the throughput and delay for a single request over IPv6 or IPv4 in a network with mixed congestion traffic are affected by two factors. First, the rate of congestion traffic, and second, by the percentage of like traffic in the congestion mix. It should also be noted that these results are for the case when the server only receives a single request from a browser i.e., there are no additional delays at the server due to any other requests.

C. *Bare PC Web Server Throughput*

We now consider bare PC Web server performance. Fig. 4 shows the throughput for a browser request from Firefox under the six congestion levels. It can be seen that throughput is much lower than for the Apache server under the same congestion levels. The throughput for IPv4 is slightly higher than for IPv6 as seen above for the Apache server. It ranges from 3.4-4.8 Mbps for IPv6 and from 4.2-5.3 Mbps for IPv4. The difference between throughput for different congestion levels with IPv6 or IPv4 is small. For example, the throughput difference for levels C2B and C3B with IPv6, and for levels C1B and C2B with IPv4, is only 0.1 Mbps. The difference between IPv6 throughput and IPv4 throughput ranges from 0.3 Mbps (for congestion level C3B) to 1.4 Mbps (for congestion level C5B). It can be seen that the differences between IPv4 and IPv6 throughput under a given level of congestion are much smaller than for Apache. The IPv6 throughput is highest for congestion level C1B and lowest for level C5B, while the IPv4 throughput is highest for congestion level C2B and lowest for level C6B.

The likely reason that the bare PC Web server has lower throughput than Apache under congestion is the absence of TCP optimizations. In particular, the bare PC server does not use selective acks (i.e. TCP SACK) or fast retransmit/recovery, and retransmits all the data after waiting for a timeout in the event of a loss. However, the bare PC Web server implementation does not have separate IPv4 and IPv6 stacks unlike Linux and other conventional servers. Instead, it uses a single RCV (Receive) task to process an incoming IP packet regardless of whether it is an IPv6 packet or an IPv4 packet. Also, the HTTP and TCP code in the server is intertwined. This implies that there is little difference in overhead when processing the two types of IP packets. However, more studies with a bare PC server that implements TCP optimizations is needed to determine the extent of possible throughput improvement for IPv6 and IPv4 due to eliminating operating system overhead. Similarly, since the performance of the bare PC server is worse than that of Apache, the impact of not implementing any congestion control mechanisms is not known.

The IPv6 throughput is highest when the congestion traffic is lower (75 Mbps) and its percentage of IPv6 traffic is equitable (50%), and it is lowest when the congestion traffic is higher (100 Mbps) and its percentage of IPv6 traffic is higher (75%). Similarly, the IPv4 throughput is highest when the congestion traffic is lower and its percentage of IPv4 traffic is lower, and it is lowest when the congestion traffic is higher and its percentage of IPv4 traffic is higher. For both IPv6 and IPv4, the throughput is highest when congestion traffic is lower and like traffic is 50% or less, and it is lowest when congestion traffic is higher and like traffic is higher. This is similar to the results seen for Apache.

D. Bare PC Web Server Delay

The delays to transfer the 320 KB file to the browser from the bare PC server in the absence of congestion traffic are 99 milliseconds and 116 milliseconds over IPv4 and IPv6 respectively (more than double the delays for Apache). When there is congestion traffic, the delays (shown in Fig. 5) range from 727-824 ms for IPv6 and from 654-725 ms for IPv4, which are much larger than the corresponding delays for Apache. Also, it can be seen that IPv6 delay are larger than the IPv4 delays as for Apache, but the delay differences between IPv6 and IPv4 for all congestion levels are now much higher than for Apache. These differences range from 64-118 ms. As with the lowered throughput, the increased delays are likely due to the absence of any TCP optimizations.

The difference between IPv6 and IPv4 delays is largest for congestion level C5B, which has a larger percentage of IPv6 traffic (75%), and smallest for level C3B, which has lower congestion traffic (75 Mbps) and a lower percentage of IPv6 traffic (25%). The lowest delay is with congestion level C3B for IPv6 and with level C2B for IPv4, and the highest delay is with congestion level C5B for IPv6 and level C6B for IPv4. It can be seen that the IPv6 delay is highest when the congestion traffic is higher (100 Mbps) and its percentage of IPv6 traffic is higher (75%), and it is lowest when the congestion traffic is lower (75 Mbps) and its percentage of IPv6 traffic is lower (25%). Similarly, the IPv4 delay is highest when the congestion traffic is higher and its percentage of IPv4 traffic is higher, and it is lowest when the congestion traffic is lower and its percentage of IPv4 traffic is lower. For both IPv6 and IPv4, the delay is highest when congestion traffic is higher and like traffic is higher, and it is lowest when congestion traffic is lower and like traffic is lower.

These results for bare PC server throughput and delay are similar to the results for the Apache server, although the throughput is much lower and the delays are much higher. As with Apache, the throughput and delay depend on both the congestion rate of congestion traffic and the percentage of like traffic causing the congestion.

V. CONCLUSION AND FUTURE WORK

We studied the performance of IPv6 Web servers under different levels of congestion. Studies were conducted in a test LAN with several routers, and used a conventional Apache Web server and a bare PC Web server with no operating system. HTTP requests over IPv6 and IPv4 were sent to the servers using an ordinary Web browser. The results for both servers show that throughput over IPv6 throughput is slightly lower than or approximately equal to IPv4 throughput depending on the congestion level, whereas delays are much higher over IPv6 than over IPv4. For all congestion levels, bare PC server throughput and delay are significantly worse than for Apache due to not implementing any TCP optimizations. Studies using bare PC Web servers with the usual TCP optimizations and congestion control mechanisms will enable the overhead due to an operating system to be determined. The results of this study show that the performance of an IPv6 Web server for requests over IPv6 and IPv4 depend on both the congestion traffic rate and the percentage of like IP traffic.

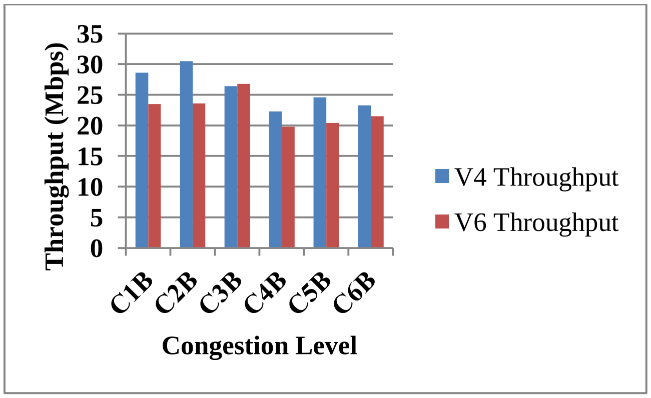


Figure 2. Apache throughput under congestion

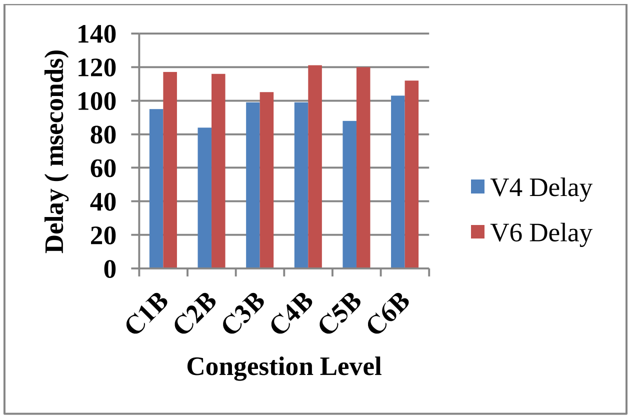


Figure 3. Apache delay under congestion

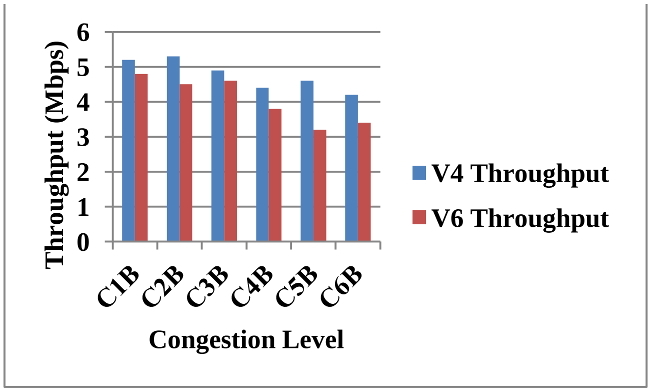


Figure 4. Bare server throughput under congestion

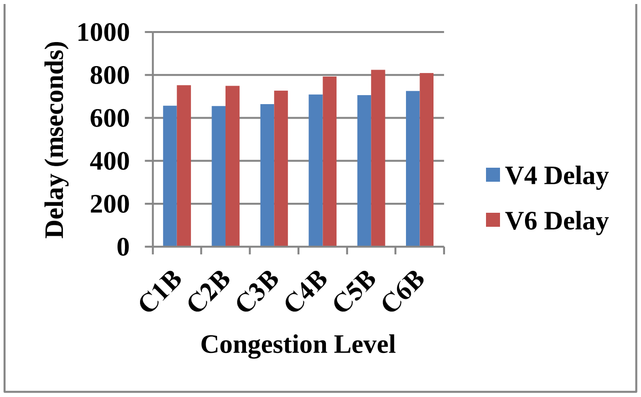


Figure 5. Bare server delay under congestion

REFERENCES

[1] Nominum Survey, http://www.nominum..com/, accessed: November 24, 2012.

[2] Y. Wang, S. Ye, and X. Li, “Understanding Current IPv6 Performance: A Measurement Study,” 10th IEEE Symposium on Computers and Communications (ISCC 2005), pp. 71-76, 2005.

[3] X. Zhou, M. Jacobsson, H. Uijterwaal, and P. Van Mieghem, “IPv6 delay and loss performance evolution,” International Journal of Communication Systems, vol. 21, no. 6, pp. 643–663, 2008.

[4] Md. T. Aziz and M. S. Islam, “Performance Evaluation of Real–Time Applications over DiffServ/MPLS in IPv4/IPv6 Networks,” Masters Thesis, Blekinge Institute of Technology, 2011.

[5] D. T. Ustundag, “Comparative routing performance analysis of IPv4 and IPv6,” Masters Thesis, Atilim University, 2009. http://acikarsiv.atilim.edu.tr/browse/100/301.pdf, accessed: July 22, 2012.

[6] M. Nikkhah, R. Guerin, Y. Lee, and R. Woundy, “Assessing IPv6 through web access a measurement study and its findings,” 7th Conference on Emerging Networking Experiments and Technologies (CoNEXT ‘11), pp. 1-12, 2011.

[7] S. Narayan, P. Shang, and N. Fan, “Performance Evaluation of IPv4 and IPv6 on Windows Vista and Linux Ubuntu,” International Conference on Networks Security, Wireless Communications and Trusted Computing, (NSWCTC '09), pp. 653-656, 2009.

[8] A. Berger, “Comparison of performance over IPv6 versus IPv4,” Intenet Statistics and Metrics Anaysis (ISMA 2012) AIMS-4 Workshop on Active Internet Measurements, pp. 1-30, 2012. http://www.caida.org/workshops/isma/1202/slides/aims1202_acox_supplement.pdf, accessed: July 22, 2012.

[9] C. Ciflikli, A. Gezer, and A. T. Ozsahin, “Packet traffic features of IPv6 and IPv4 protocol traffic,” Turk. J. Elec. Eng. & Comp. Sci., vol. 20, no. 5, pp. 727-749, 2012.

[10] A. Afanasyev, N. Tilley, P. Reiher, and L. Kleinrock, “Host-to-Host Congestion Control for TCP,” IEEE Communications Surveys and Tutorials, vol. 12, no. 3, pp. 304-342, 2010.

[11] A. Loukili, A. L. Wijesinha, R. K. Karne, and A. K. Tsetse, “Web Server Performance with Cubic and Compound TCP,” 7th International Conference on Communication, Internet, and Information Technology (CIIT), 2012.

[12] R. K. Karne, K.V. Jaganathan, T. Ahmed, and N. Rosa, “DOSC: Dispersed Operating System Computing,” 20th Annual ACM Conference on Object-Oriented Programming, Systems, Languages, and Applications (OOPSLA ’05 Onward Track), pp. 55-61, 2005.

[13] R. K. Karne, K.V. Jaganathan, and T. Ahmed. “How to run C++ applications on a bare PC,” 6th ACIS International Conference on Software Engineering, Networking, Parallel/Distributed Computing (SNPD 2005), pp. 50-55, 2005.

[14] L. He, R. K. Karne, and A. L. Wijesinha, “Design and performance of a bare PC Web Server,” International Journal of Computer Applications, pp. 100-112, vol. 15, no. 2, June 2008.

[15] Multi-Generator (MGEN), http://cs.itd.nrl.navy.mil/work/mgen, accessed: July 22, 2012.

[16] Wireshark packet analyzer, http://www.wireshark.org/, accessed: July 22, 2012.

Optimal FIR Filters for DTMF Applications

Pavel Zahradnik and Boris Šimák
Telecommunication Engineering Department
Czech Technical University in Prague
Prague, Czech Republic
{zahradni, simak}@fel.cvut.cz

Miroslav Vlček
Applied Mathematics Department
Czech Technical University in Prague
Prague, Czech Republic
vlcek@fd.cvut.cz

***Abstract*—A fast and robust analytical procedure for the design of high performance digital optimal band-pass finite impulse response filters for dual-tone multi-frequency applications is introduced. The filters exhibit equiripple behavior of the frequency response. The approximating function is based on Zolotarev polynomials. The presented closed form solution provides formulas for the filter degree and for the impulse response coefficients. Several examples are presented.**

***Keywords*-FIR filter; narrow band filter; dual-tone multi-frequency; iso-extremal approximation.**

I. Introduction

There are two basic tasks in the processing of dual-tone multi-frequency (DTMF) signals, namely the detection of DTMF frequencies and the removal of the DTMF frequencies in a broad band signal. The DTMF frequencies form two groups with four frequencies each. The lower group consists of frequencies 697, 770, 852 and 941 Hz while the higher group comprises sinusoids of 1209, 1336, 1477 and 1633 Hz. For the processing of DTMF signals, the infinite impulse response (IIR) filters are usually applied because of their lower number of coefficients. The IIR filters are usually part of the famous Goertzel procedure [1]. In the removal of DTMF frequencies in a broad band signal, the IIR filters produce substantial distortions of the output signal which appear near its flat region due to the group delay variation. This behavior is especially apparent, if pulse like components are present in the signal as demonstrated in [2]. In order to minimize these distortions in the processing of DTMF signals we propose the application of finite impulse response (FIR) filters which inherit a constant group delay. In order to maximize the discrimination of the DTMF sinusoids, the selective bands of the FIR filters should be as narrow as possible. In this paper we are focused upon the design of narrow optimal equiripple (ER) band-pass (BP) FIR filters for the DTMF decoding. They are optimal in terms of the shortest possible filter length related to the frequency specification. Note that the proposed filter design is based on formulas, i.e. no numerical procedures are involved. The presented closed form solution includes the degree equation and formulas for the robust evaluation of the impulse response coefficients of the filter.

II. Terminology

We assume a general FIR filter of type I represented by its impulse response $h(k)$ with odd length of $N = 2n+1$ coefficients and with even symmetry (1). In our further considerations we use the a-vector $a(k)$, which is related to the impulse response $h(k)$

$$a(0) = h(n) \ , \ \ a(k) = 2h(n+k) = 2h(n-k) \, , \, k = 1 \ldots n \ \ . \tag{1}$$

Further, we introduce an auxiliary real variable w

$$w = \frac{1}{2}(z+z^{-1})\big|_{z=e^{j\omega T}} = \cos(\omega T) = \cos\left(2\pi\frac{f}{f_s}\right) , \tag{2}$$

where f_s is the sampling frequency. The transfer function of the filter is

$$\begin{aligned} H(z) &= \sum_{k=0}^{2n} h(k)\, z^{-k} \\ &= z^{-n}\left[h(n) \, + 2\sum_{k=1}^{n} h(n \pm k)\, \frac{1}{2}\left(z^{k} + z^{-k}\right)\right] \\ &= z^{-n}\sum_{k=0}^{n} a(k)\, T_k(w) = z^{-n} Q(w) \end{aligned} \tag{3}$$

where $T_k(w) = \cos(k\arccos(w))$ is Chebyshev polynomial of the first kind and $Q(w)$ is the real valued zero phase transfer function which we express using the a-vector in form of the expansion of Chebyshev polynomials

$$Q(w) = \sum_{k=0}^{n} a(k)\, T_k(w) \ . \tag{4}$$

The zero phase transfer function of an ER BP FIR filter is

$$Q(w) = \frac{Z_{p,q}(w,\kappa) + 1}{y_m + 1} \ , \tag{5}$$

where $Z_{p,q}(w,\kappa)$ represents the Zolotarev polynomial [8]. For illustration, the shape of a Zolotarev polynomial is shown in Fig. 1 and the corresponding frequency response is shown in Fig. 2.

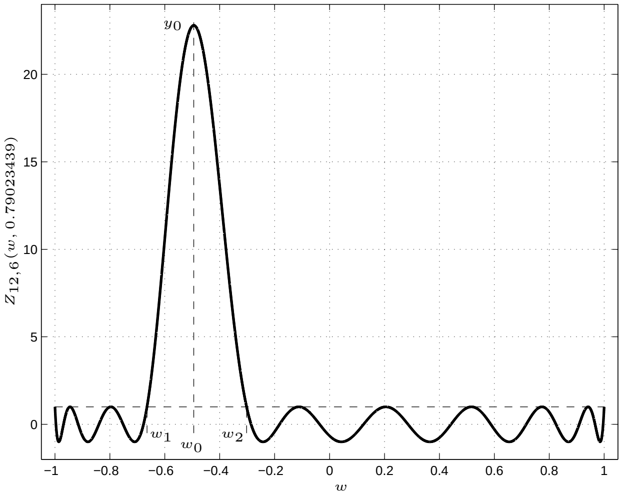


Fig. 1. Zolotarev polynomial $Z_{12,6}(w, 0.79023439)$.

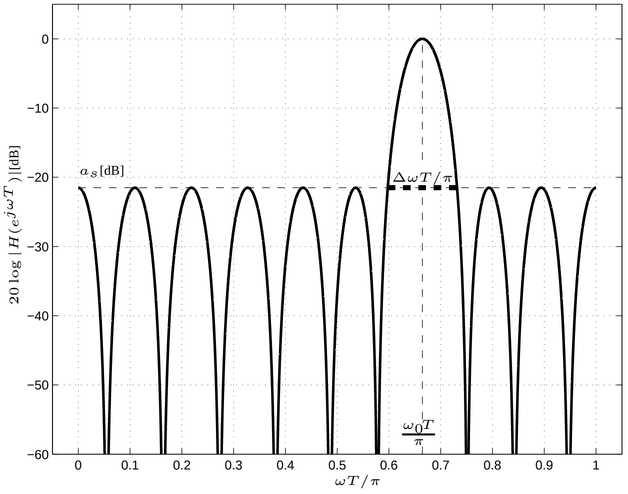


Fig. 2. Amplitude frequency response $20\log|H(e^{j\omega T})|$ [dB] corresponding to the Zolotarev polynomial from Fig. 1.

III. Optimal Band-Pass FIR Filter

An optimal ER BP FIR filter (Fig. 2) is specified by the pass band frequency $\omega_0 T$ (or $f_0 = \omega_0 T f_s/2\pi$), width of the pass band $\Delta\omega T$ (or $\Delta f = \Delta\omega T f_s/2\pi$), attenuation in the stop-bands a_s[dB] and by the sampling frequency f_s. An approximation of the frequency response of a filter is based on the generating function. The generating function of an ER BP FIR filter is the Zolotarev polynomial $Z_{p,q}(w,\kappa)$ which approximates a constant value in equiripple Chebyshev sense in two disjoint intervals $\langle -1, w_1\rangle$ and $\langle w_2, 1\rangle$ as shown in Fig. 1. The main lobe with the maximal value $y_0 = Z_{p,q}(w_0,\kappa)$ is located inside the interval (w_1, w_2). The notation $Z_{p,q}(w,\kappa)$ emphasizes the fact that the integer value p counts the number of zeros right from the maximum w_0 and the integer value q corresponds to the number of zeros left from the maximum w_0. The real value $0 \le \kappa \le 1$ is in fact the Jacobi elliptical modulus. It affects the maximum value y_0 and the width $w_2 - w_1$ of the main lobe (Fig. 1). For increasing κ the value y_0 increases and the main lobe broadens. The Zolotarev polynomial is usually expressed in terms of Jacobi elliptic functions [6]-[8]

$$Z_{p,q}(w,\kappa) = \frac{(-1)^p}{2} \tag{6}$$
$$\times \left[\left(\frac{H(u - \frac{p}{n}\ \mathbf{K}(\kappa))}{H(u + \frac{p}{n}\ \mathbf{K}(\kappa))} \right)^n + \left(\frac{H(u + \frac{p}{n}\ \mathbf{K}(\kappa))}{H(u - \frac{p}{n}\ \mathbf{K}(\kappa))} \right)^n \right] \ .$$

The factor $(-1)^p/2$ appears in (6) as the Zolotarev polynomial alternates $(p+1)-$times in the interval $(w_2, 1)$. The variable u is expressed by the incomplete elliptical integral of the first kind $F(x|\kappa)$, namely

$$u = F\left(\mathrm{sn}\left(\frac{p}{n}\ \mathbf{K}(\kappa)|\kappa\right) \sqrt{\frac{1+w}{w + 2\,\mathrm{sn}^2\left(\frac{p}{n}\mathbf{K}(\kappa)|\kappa\right) - 1}} |\kappa \right) . \tag{7}$$

The function $H\left(u \pm (p/n)\ \mathbf{K}(\kappa)\right)$ is the Jacobi Eta function, $\mathrm{sn}(u|\kappa)$, $\mathrm{cn}(u|\kappa)$, $\mathrm{dn}(u|\kappa)$ are Jacobi elliptic functions and $\mathbf{K}(\kappa)$ is the quarter-period given by the complete elliptic integral of the first kind. The degree of the Zolotarev polynomial is $n = p+q$. A comprehensive treatise of Zolotarev polynomials was published in [8]. It includes the analytical solution of the coefficients of Zolotarev polynomials, the algebraic evaluation of the Jacobi Zeta function $\mathrm{Z}(\frac{p}{n}\mathbf{K}(\kappa)|\,\kappa)$ and of the elliptic integral of the third kind $\Pi(\sigma_m, \frac{p}{n}\mathbf{K}(\kappa)|\,\kappa)$. The position w_0 of the maximum value $y_0 = Z_{p,q}(w_0,\kappa)$ is

$$w_0 = w_1 + 2\,\frac{\mathrm{sn}\left(\frac{p}{n}\mathbf{K}(\kappa)|\kappa\right)\mathrm{cn}\left(\frac{p}{n}\mathbf{K}(\kappa)|\kappa\right)}{\mathrm{dn}\left(\frac{p}{n}\mathbf{K}(\kappa)|\kappa\right)} \mathrm{Z}\left(\frac{p}{n}\mathbf{K}(\kappa)|\kappa\right) \tag{8}$$

where the edges of the main lobe are

$$w_1 = 1 - 2\,\mathrm{sn}^2\left(\frac{p}{n}\mathbf{K}(\kappa)|\kappa\right) \tag{9}$$

$$w_2 = 2\,\mathrm{sn}^2\left(\frac{q}{n}\mathbf{K}(\kappa)|\kappa\right) - 1 \ . \tag{10}$$

The relation for the maximum value y_0

$$y_0 = \cosh 2n\left(\sigma_m \mathrm{Z}(\frac{p}{n}\mathbf{K}(\kappa)|\kappa) - \Pi(\sigma_m, \frac{p}{n}\mathbf{K}(\kappa)|\kappa)\right) \tag{11}$$

is useful in the normalization of Zolotarev polynomials. The degree of the Zolotarev polynomial $Z_{p,q}(w,\kappa)$ is expressed by the degree formula

$$n \ge \frac{\ln(y_0 + \sqrt{y_0^2 - 1})}{2\sigma_m \mathrm{Z}(\frac{p}{n}\mathbf{K}(\kappa)|\kappa) - 2\Pi(\sigma_m, \frac{p}{n}\mathbf{K}(\kappa)|\kappa)} \ . \tag{12}$$

The auxiliary value σ_m in (11), (12) is given by the formula

$$\sigma_m = F\left(\arcsin\left(\frac{1}{\kappa\ \mathrm{sn}\left(\frac{p}{n}\ \mathbf{K}(\kappa)|\kappa\right)} \sqrt{\frac{w_0 - w_s}{w_0 + 1}} \right) |\kappa \right) \tag{13}$$
$$= F\left(\arcsin \frac{\sqrt{\cos\left(2\pi\frac{f_0}{f_s}\right) - \cos\left[\frac{2\pi}{f_s}\left(f_0 + \frac{\Delta f}{2}\right)\right]}}{\kappa\ \mathrm{sn}\left(\frac{p}{n}\ \mathbf{K}(\kappa)|\kappa\right)\sqrt{1 + \cos\left(2\pi\frac{f_0}{f_s}\right)}} |\kappa \right) \ .$$

The Zolotarev polynomial $Z_{p,q}(w,\kappa)$ satisfies the differential equation

$$\begin{aligned}(1-w^2)(w-w_1)(w-w_2)\left(\frac{dZ_{p,q}(w,\kappa)}{dw}\right)^2 \\ = n^2\left(1-Z_{p,q}^2(w,\kappa)\right)(w-w_0)^2 \ .\end{aligned} \tag{14}$$

Based on the differential equation (14) we have developed an algorithm for the evaluation of the a-vector and of the impulse response $h(k)$ corresponding to the Zolotarev polynomial $Z_{p,q}(w,\kappa)$ in form of its expansion into Chebyshev polynomials (4). This algorithm is summarized in Table I. There are two goals in the filter design. The first one is to obtain the minimal filter length of N coefficients satisfying the filter specification. The second one is to evaluate its impulse response $h(k)$. In the standard design of an ER BP FIR filter, which is represented by the numerical Parks-McClellan procedure (e.g. the function *firpm* in Matlab), the exact filter length is not available because of no approximating function. Consequently, the filter length is not the result of the design, it is in fact an input argument in the Parks-McClellan procedure. The filter length is either estimated or successively adjusted in order to meet the filter specification, or it is obtained from empirical approximating formulas. Moreover, a successful design is not guaranteed in the Parks-McClellan procedure. The alternative design approach that we use here is based on the analytical design which we have developed for ER notch FIR filters and introduced in [3]. The available pass band frequencies which we denote f_Q are quantized because there is always an integer number of ripples (Fig. 2). That is why we additionally tune the actual pass band frequency f_Q of the initial filter to the specified value f_0. This tuning consists in multiplying the a-vector of the initial filter by a transformation matrix, resulting in the a-vector of the tuned filter (16) which exactly meets the specified pass band frequency. For the tuning, we present an efficient algebraic procedure which is a simplified version of that one introduced in [4]. The tuning procedure results in the zero phase transfer function of the tuned filter, which is

$$Q_t(w) = \sum_{k=0}^{n} a(k)\, T_k(\lambda w \pm \lambda') = \sum_{k=0}^{n} a(k) \sum_{m=0}^{k} \alpha_k(m) T_m(w) . \tag{15}$$

Based on (15), the a-vector $a_t(k)$ of the tuned filter and the a-vector $a(k)$ of the initial filter are related by a triangular transformation matrix A

$$a_t(k) = [a_t(0)\ \ a_t(1)\ \ \cdots\ \ a_t(n)] = [a(0)\ \ a(1)\ \ \cdots\ \ a(n)]\times$$

$$\times\begin{bmatrix} \alpha_0(0) & 0 & 0 & 0 & \cdots & 0 \\ \alpha_1(0) & \alpha_1(1) & 0 & 0 & \cdots & 0 \\ \alpha_2(0) & \alpha_2(1) & \alpha_2(2) & 0 & \cdots & 0 \\ \alpha_3(0) & \alpha_3(1) & \alpha_3(2) & \alpha_3(3) & \cdots & 0 \\ \vdots & & & & & \vdots \\ \alpha_n(0) & \alpha_n(1) & \alpha_n(2) & \alpha_n(3) & \cdots & \alpha_n(n) \end{bmatrix} = a(k)A\,. \tag{16}$$

A fast algorithm for the evaluation of the coefficients $\alpha_k(m)$ is summarized in Tab. II. The presented tuning of the pass band frequency preserves the attenuation a_s[dB] in the stop bands.

IV. Design of the Optimal Band-Pass FIR Filter

Let us specify the optimal ER BP FIR filter by the pass band frequency f_0[Hz], width of the pass band Δf[Hz], sampling frequency f_s[Hz] and by the attenuation in the stop bands a_s[dB]. The design procedure reads as follows:
Calculate the Jacobi elliptic modulus κ

$$\kappa = \sqrt{1 - \frac{1}{\tan^2(\varphi_1)\tan^2(\varphi_2)}} \tag{17}$$

for the auxiliary values φ_1 and φ_2

$$\varphi_1 = \frac{\pi}{f_s}\left(f_0 + \frac{\Delta f}{2}\right) \ \ , \ \ \varphi_2 = \frac{\pi}{f_s}\left(\frac{f_s}{2} - f_0 + \frac{\Delta f}{2}\right). \tag{18}$$

Calculate the rational values $\frac{p}{n}$ and $\frac{q}{n}$

$$\frac{p}{n} = \frac{F(\varphi_1,\kappa)}{\mathbf{K}(\kappa)} \ \ , \ \ \frac{q}{n} = \frac{F(\varphi_2,\kappa)}{\mathbf{K}(\kappa)} \ , \tag{19}$$

where $\mathbf{K}(\kappa)$ is a complete elliptic integral of the first kind, which here represents the elliptic quarter-period. Determine the value y_0

$$y_0 = \frac{2}{10^{0.05 a_s[dB]}} \ . \tag{20}$$

Calculate the auxiliary value σ_m (13).
Calculate and round up the value n (12) which represents the degree $n = p + q$ of the Zolotarev polynomial $Z_{p,q}(w,\kappa)$.
Calculate the integer indices p and q of the Zolotarev polynomial $Z_{p,q}(w,\kappa)$

$$p = \left\updownarrow n\frac{F(\varphi_1,\kappa)}{\mathbf{K}(\kappa)} \right\updownarrow \ \ , \ \ q = \left\updownarrow n\frac{F(\varphi_2,\kappa)}{\mathbf{K}(\kappa)} \right\updownarrow \ . \tag{21}$$

The arrow brackets in (21) stand for rounding. For values p, q (21), κ (17) and y_0 (20) evaluate the a-vector $a(k)$ and the related impulse response $h(k)$ of the filter using the algebraical procedure summarized in Tab. I. Check the actual pass band frequency f_Q of the initial BP FIR filter

$$\begin{aligned} f_Q = \frac{f_s}{2\pi} \arccos\Bigg[1 - 2\,\mathrm{sn}^2\left(\frac{p}{n}\mathbf{K}(\kappa),\kappa\right) \\ + 2\,\frac{\mathrm{sn}\left(\frac{p}{n}\mathbf{K}(\kappa),\kappa\right)\mathrm{cn}\left(\frac{p}{n}\mathbf{K}(\kappa),\kappa\right)}{\mathrm{dn}\left(\frac{p}{n}\mathbf{K}(\kappa),\kappa\right)} \mathrm{Z}\left(\frac{p}{n}\mathbf{K}(\kappa),\kappa\right) \Bigg] \ . \end{aligned} \tag{22}$$

Because of the inherent quantization of the available pass band frequencies f_Q mentioned above, the actual pass band frequency f_Q usually slightly differs from the specified pass band frequency f_0. That is why we tune the quantized pass band frequency f_Q of the initial filter to the specified value f_0 using (16) and Tab. II. In our calculations, the Jacobi elliptic Zeta function $\mathrm{Z}(x,\kappa)$ in (8), (11), (12) and the incomplete elliptic integral of the first kind $F(x,\kappa)$ in (13) are evaluated

```
given            $p,\ q,\ \kappa,\ y_0$
initialization   $n = p + q\ ,\ w_1 = 1 - 2\,\mathrm{sn}^2\left(\frac{p}{n}\mathbf{K}(\kappa), \kappa\right)\ ,\ w_2 = 2\,\mathrm{sn}^2\left(\frac{q}{n}\mathbf{K}(\kappa), \kappa\right) - 1\ ,\ w_a = \frac{w_1 + w_2}{2}$
                 $w_m = w_1 + 2\,\frac{\mathrm{sn}\left(\frac{p}{n}\mathbf{K}(\kappa), \kappa\right)\mathrm{cn}\left(\frac{p}{n}\mathbf{K}(\kappa), \kappa\right)}{\mathrm{dn}\left(\frac{p}{n}\mathbf{K}(\kappa), \kappa\right)}\mathrm{Z}\left(\frac{p}{n}\mathbf{K}(\kappa), \kappa\right)$
                 $\alpha(n) = 1\ ,\ \alpha(n+1) = \alpha(n+2) = \alpha(n+3) = \alpha(n+4) = \alpha(n+5) = 0$
body
(for             $m = n + 2$   to   3)
                     $8c(1) = n^2 - (m+3)^2\ ,\ 4c(2) = (2m+5)(m+2)(w_m - w_a) + 3w_m[n^2 - (m+2)^2]$
                     $2c(3) = \frac{3}{4}[n^2 - (m+1)^2] + 3w_m[n^2 w_m - (m+1)^2 w_a] - (m+1)(m+2)(w_1 w_2 - w_m w_a)$
                     $c(4) = \frac{3}{2}(n^2 - m^2) + m^2(w_m - w_a) + w_m(n^2 w_m^2 - m^2 w_1 w_2)$
                     $2c(5) = \frac{3}{4}[n^2 - (m-1)^2] + 3w_m[n^2 w_m - (m-1)^2 w_a] - (m-1)(m-2)(w_1 w_2 - w_m w_a)$
                     $4c(6) = (2m-5)(m-2)(w_m - w_a) + 3w_m[n^2 - (m-2)^2]\ ,\ 8c(7) = n^2 - (m-3)^2$
                     $\alpha(m-3) = \frac{1}{c(7)}\sum_{\mu=1}^{6} c(\mu)\alpha(m+4-\mu)$
(end             loop   on   $m$)
normalization    $s(n) = \frac{\alpha(0)}{2} + \sum_{m=1}^{n} \alpha(m)$
a-vector         $a(0) = (-1)^p \frac{\alpha(0)}{2s(n)}$ ,  (for $m = 1$  to  $n$) ,  $a(m) = (-1)^p \frac{\alpha(m)}{s(n)}$ ,  (end loop  on  $m$)
impulse response $h(n) = \frac{a(0) + 1}{y_0 + 1}$ ,  (for $m = 1$  to  $n$) ,  $h(n \pm m) = \frac{a(m)}{2(y_0 + 1)}$ ,  (end loop  on  $m$)
```

TABLE I

FAST ALGORITHM FOR EVALUATING THE a-VECTOR $a(k)$ AND THE IMPULSE RESPONSE $h(k)$.

```
given                          $f_Q,\ \ f_0,\ \ f_s,\ \ k$
initialization                 if $f_Q < f_0\ :\ \lambda = \frac{\cos\left(2\pi\frac{f_Q}{f_s}\right) - 1}{\cos\left(2\pi\frac{f_0}{f_s}\right) - 1},\ s = 1,$   else : $\lambda = \frac{\cos\left(2\pi\frac{f_Q}{f_s}\right) + 1}{\cos\left(2\pi\frac{f_0}{f_s}\right) + 1},\ s = -1$
                               $\lambda^{'} = 1 - \lambda\ \ ,\ \ \alpha_k(k+1) = \alpha_k(k+2) = \alpha_k(k+3) = 0\ \ ,\ \ \alpha_k(k) = \lambda^k$
body
  (for $\mu = -3\ ...\ k-4$ )
                               $\alpha_k(k-\mu-4) =$
                               {
                                 $-2s\left[(\mu+3)(2k-\mu-3) - \frac{\lambda^{'}}{\lambda}(k-\mu-3)(2k-2\mu-7)\right]\alpha_k(k-\mu-3)$
                                                                  $+2\frac{\lambda^{'}}{\lambda}(k-\mu-2)\ \alpha_k(k-\mu-2)$
                                 $+2s\left[(\mu+1)(2k-\mu-1) - \frac{\lambda^{'}}{\lambda}(k-\mu-1)(2k-2\mu-1)\right]\alpha_k(k-\mu-1)$
                                                                  $+\mu(2k-\mu)\ \alpha_k(k-\mu)$
                               }  /  $(\mu+4)(2k-\mu-4)$
  (end loop on $\mu$)
```

TABLE II

FAST ALGORITHM FOR EVALUATING THE COEFFICIENTS $\alpha_k(m)$ OF TRANSFORMATION MATRIX A.

by the arithmetic-geometric mean [7]. The Jacobi elliptic integral of the third kind $\Pi(x, y, \kappa)$ in (12) is evaluated by a fast procedure proposed in [3].

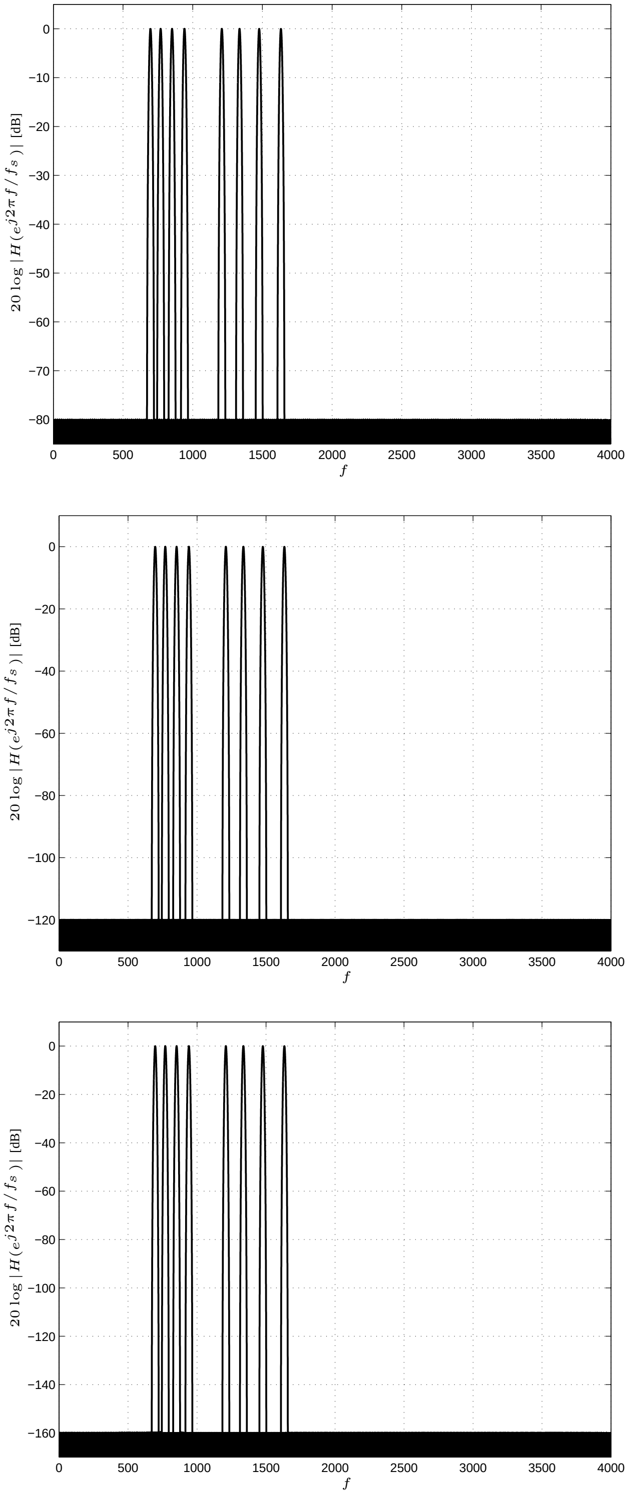


Fig. 3. Amplitude frequency responses $20\log|H(e^{j2\pi f/f_s})|$ [dB].

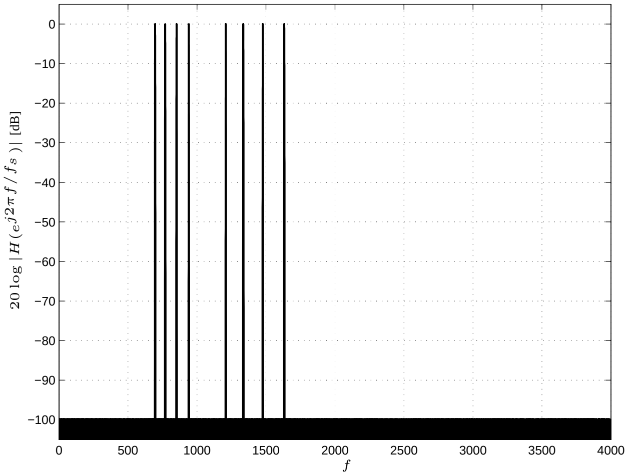


Fig. 4. Amplitude frequency responses $20\log|H(e^{j2\pi f/f_s})|$ [dB] of the filters with $N = 13141$ coefficients.

V. Examples of Design

Let us design three sets of band pass FIR filters specified by the DTMF frequencies $f_0 = 697$, 770, 852, 941, 1209, 1336, 1477, 1633 Hz, width of the pass bands $\Delta f = 50$Hz, sampling frequency $f_s = 8000$Hz and with the attenuations in the stop bands $a_s = -80$dB, -120dB and -160dB.
Using the presented design procedure, we get filter lengths $N = 1083$ coefficients for $a_s = -80$dB, 1551 coefficients for $a_s = -120$dB and 2019 coefficients for $a_s = -160$dB. The corresponding amplitude frequency responses are shown in Fig. 3. In order to demonstrate the remarkable selectivity of the ER BP FIR filters and the robustness of the presented design procedure, let us design the DTMF filters with very narrow pass band of $\Delta f = 5$Hz, sampling frequency $f_s = 8000$Hz and with the attenuation in the stop bands $a_s = -100$dB. The filter length is $N = 13141$ coefficients. The amplitude frequency responses are shown in Fig. 4.

VI. Conclusion and Future Work

We have presented a fast and robust procedure for the design of optimal equiripple narrow band-pass FIR filters for DTMF applications. In contrast to the established numerical design procedure the proposed methodology solves the approximation problem and provides a formula for the degree of the filter and formulas for the evaluation of the coefficients of the impulse response of the filter. Our future activity will include an efficient implementation of the DTMF filters using digital signal processors.

Acknowledgements

This work was thankfully supported by the grant MPO No. FR-TI2/621.

REFERENCES

[1] G. Goertzel, An Algorithm for the evaluation of finite trigonometric Series, *Amer. Math. Monthly*. Vol. 65, January 1958, pp. 34-35.

[2] M. Vlček, P. Zahradnik, Digital Multiple Notch Filters Performance, *Proceedings of the 15th European Conference on Circuit Theory and Design ECCTD'01*. Helsinky, August 2001, pp. 49-52.

[3] P. Zahradnik, M. Vlček, Fast Analytical Design Algorithms for FIR Notch Filters, *IEEE Transactions on Circuits and Systems*. March 2004 , Vol. 51, No. 3, pp. 608 - 623.

[4] P. Zahradnik, M. Vlček, An Analytical Procedure for Critical Frequency Tuning of FIR Filters. *IEEE Transactions on Circuits and Systems II*. January 2006, Vol. 53, No. 1, pp. 72-76.

[5] N. I. Achieser, Über einige Funktionen, die in gegebenen Intervallen am wenigsten von Null abweichen, *Bull. de la Soc. Phys. Math. de Kazan*, Vol. 3, pp. 1 - 69, 1928.

[6] D. F. Lawden *Elliptic Functions and Applications* Springer-Verlag, New York Inc., 1989.

[7] M. Abramowitz, I. Stegun, *Handbook of Mathematical Function*, Dover Publication, New York Inc., 1972.

[8] M. Vlček, R. Unbehauen, Zolotarev Polynomials and Optimal FIR Filters, *IEEE Transactions on Signal Processing*, Vol. 47, No. 3, pp. 717-730, March 1999.

Task Allocation Algorithms for 2D Torus Architecture

Lukasz Jakimczuk, Wojciech Kmiecik, Iwona Pozniak-Koszalka, and Andrzej Kasprzak
Department of Systems and Computer Networks
Wroclaw University of Technology
Wroclaw, Poland
e-mail: 170922@student.pwr.wroc.pl, {wojciech.kmiecik, iwona.pozniak-koszalka, andrzej.kasprzak}@pwr.wroc.pl

Abstract — **Efficient allocation of computers to incoming tasks is crucial for achieving high performance in modern networks. A good allocation algorithm should identify available computers with minimum overhead and allocate incoming tasks in as short period of time as possible. This paper concerns allocation problem for torus-structured system. The new allocation mechanism, called Improved Tree Allocation for Torus (ITAT), based on tree architecture, has been proposed. ITAT-algorithm was compared with other known allocation algorithms on the basis of simulation experiments made with the designed and implemented experimentation system. The obtained results justify a conclusion that the created allocation algorithm seems to be very promising.**

Keywords-torus; allocation; algorithm; experimentation system; efectiveness

I. INTRODUCTION

Multicomputer systems, consisting of many processing elements connected through a high speed network, have become widespread in engineering and scientific applications [1]. Such networks are intended to deal with tasks which cannot be handled by single computer. Two-dimensional (2D) torus is one of the interconnection topologies developed for mentioned system [2]. For each topology including 2D torus, predefined allocation algorithms exist. In this paper, contiguous processor allocator for torus structured network is considered (Fig. 1).

The requirement here is to allocate incoming jobs to free subtorus of appropriate size in 2D torus connected system. The allocation scheme should provide maximal resource utilization what is done by minimizing any kind of fragmentation [3]. Allocation algorithm must be fast, deliver low overhead and be able to support systems with thousands of nodes. A critical attribute of all mechanisms is ability to find available subtoruses for incoming requests, if they exist, what is called subtorus recognition ability. An allocation algorithm has complete subtorus recognition ability when it can always find a free subtorus (if one is available) for an incoming job [4].

In this paper, recognition-complete allocation scheme based on non-binary tree called Improved Tree Allocation (ITAT) is presented. It was designed with intent of maximize the utilization.

The rest of the paper is organized as follows. Section II presents definitions and notations used throughout the paper. The existing job allocation mechanisms and accessory algorithms are reviewed in Section III. Section IV describes our novel scheme in detail. In Section V experimentation system is shortly presented. Within Section VI properties of the created algorithm are analysed and compared with other well-known algorithms. Future work and conclusion are finally included in Section VII.

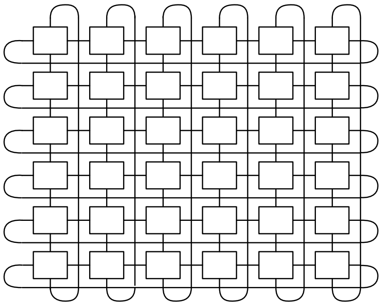


Figure 1. An example of 2D torus.

II. NOMENCLATURE

We use the classic notation presented, e.g., in [5][6][7]:

A 2D torus topology, denoted by *T(w,h)*, consists of $w \times h$ nodes arranged in a $w \times h$ 2D grid. The node in column *c* and row *r* is identified by address *<c,r>* where $0 \leq c < w$ and $0 \leq r < h$. A node *<c,r>* is connected by direct communication channel to its neighbouring nodes *<c±1,r>* and *<c,r±1>*. Thus each node has four neighbouring nodes.

A 2D subtorus *S(p,q)* in the torus *T(w,h)* is a subgrid *T(p,q)* such that $1 \leq p \leq w$ and $1 \leq q \leq h$. A job requesting a subtorus $p \times q$ is denoted by *J(p,q)*. A subtorus *S* is identified by its base (lower left node) and end (upper right end) and is denoted as $S[<x_b,y_b> <x_e,y_e>]$. In contrast to the 2D-mesh topology, in torus x_b can be greater than x_e, and y_b can be greater than y_e. However, the base still remains as lower left corner with end on the upper right node.

A busy node is a node which has been allocated to a job. A busy subtorus β is a subtorus, where all of its nodes have been allocated to jobs.

A free node is a node which is not allocated to any job. A subtorus is free when all of its nodes have not been allocated to jobs.

A busy array of a torus *T(w,h)* is a bit map *B[w,h]*, in which element *B[c,r]* has a value 1 or 0 if node *<c,r>* is busy or free, respectively.

A busy list is a set of all busy subtoruses in the system. Similarly, a free list is a set of all free subtoruses available.

Copyright (c) IARIA, 2013. ISBN: 978-1-61208-245-5

The coverage of a busy subtorus β with respect to a job J is denoted by $\zeta_{\beta,J}$ and it is a set of processors such that use of any node in $\zeta_{\beta,J}$ as the base of free subtorus for the allocation of J will cause the job J to be overlapped with β. The coverage set with respect to J is denoted by C_J and it is the set of the coverages of all busy subtoruses.

A base block with respect to a job J is a subtorus whose nodes can be used as base for free subtoruses to allocate job J. A set of disjoint base blocks is called the base set.

External fragmentation is the ratio of the number of free processors to the total number of processors in the torus, when the allocation of incoming task fails but there is sufficient number of free processors.

The given definitions are illustrated in example of a torus *T(6,6)* with respect to *J(2,3)* and *J(2,2)* (see Fig. 2).

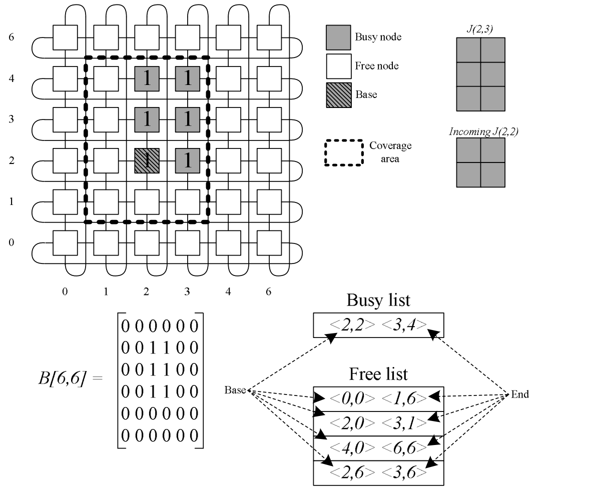


Figure 2. Busy and free nodes, coverage area, busy array and free list for a torus *T(6,6)*.

III. Known Allocation Algorithms For Torus Architecture

The algorithms, based on busy list and busy array, create coverage area set [8] in one of the first steps. For *k-array 2-cube* four possible cases of task allocation can be distinguish and they are presented in Fig. 3.

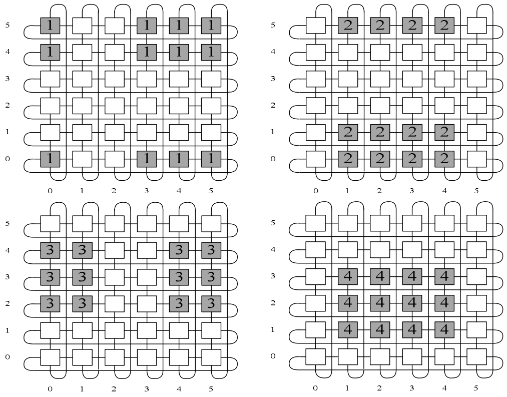


Figure 3. Four different cases of job *J(4,3)* allocation.

These cases are characterized by: 1) $x_b > x_e$ and $y_b > y_e$,; 2) $x_b \leq x_e$ and $y_b > y_e$,; 3) $x_b > x_e$ and $y_b \leq y_e$; 4) regular case known from 2D-mesh networks.

For every presented instance, coverage needs to be determined in a different way. For a given $\beta=[<x_b,y_b> <x_e,y_e>]$, its coverage with respect to *J(p,q)* is $\zeta_{\beta,J}=[<x_1,y_1> <x_2,y_2>]$, where x_1, y_1, x_2, y_2 are determined according to the Construction of Coverage algorithm described in [1].

IBMAT Algorithm. First existing allocation algorithm is Improved Bit Map Allocation for Torus [9]. The general idea of the IBMAT is based on the approach used in IFF algorithm [4]. With respect to an incoming job, the busy array is scanned to create a coverage array C_T in the form of bit map. Each coverage $\zeta_{\beta,J}$ is divided into three regions: job coverage, left coverage and bottom coverage, presented in Fig. 4. In the worst case, two inspections through a C_T are required:

All rows from right to left, each row two times (determining the left coverage of a job).

All columns from top to bottom, each column two times (creating bottom coverage of a job).

The IBMAT is recognition complete by manipulating the job orientation. If for a given *J(p,q)* the allocation fails and $p \neq q$, the scheme will change the orientation of the job and then *J(q,p)* possibility is checked. When both attempts fail, the allocation of the job fails.

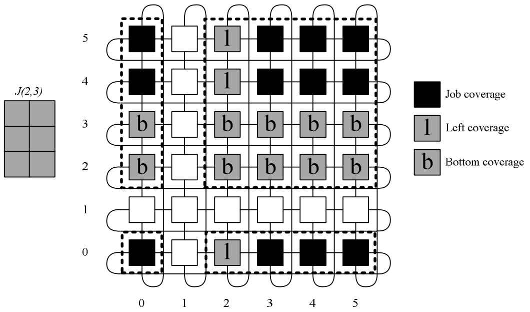


Figure 4. Coverage of job *J(4,3)* with respect to incoming job *J(2,3)*.

IBLAT Algorithm. Second existing allocation algorithm is Improved Busy List Allocation for Torus [10]. The IBLAT is based on the strategy employed in the IAS scheme [11]. For an incoming job *J(p,q)*, the IBLAT scans a busy list and creates coverage set C_T which is also in a list form. Both busy and coverage lists contains coordinates of each β and $\zeta_{\beta,J}$ respectively. When C_J is created, each node is tested for membership in C_J , what is done by inspecting the whole C_J for every node. Node which is not in C_J can be a base for given job, in the other case, the algorithm checks another node. The IBLAT scheme is recognition complete.

IRAT Algorithm. Third existing mechanism, based on randomness, is called Improved Random Algorithm. It can only pick random node from system and check if it can become base for an incoming request. If node is available as base job is allocated, otherwise scheme will change orientation of job J(p,q) and J(q,p) possibility is checked. When both attempts fail, the allocation of the job fails [12].

IV. Improved Tree Allocation For Torus Algorithm

The task allocation algorithm proposed in this paper has complete subtorus recognition ability. The allocation scheme is particularly attractive for large systems, what is confirmed in experimentation section of this paper.

ITAT achieves recognition completeness by manipulating the orientation of the subtorus request. In allocation a job *J*(*p*,*q*), the scheme first tries to allocate the task using the given orientation *p* × *q*. If allocation fails, the algorithm creates a new request *J(q,p)* by rotating the original orientation and tries to allocate rotated request. If this attempt also fails, the allocation of the job also fails.

The following definitions are introduced:

A busy tree, denoted by *T(n),* incorporates *n* nodes including root, free and busy leaves and base nodes.

A root of the tree is the node with address <0,0>.

Subtree, denoted by $S_T(x)$, is a tree incorporates *x* nodes including only base nodes of tree *T(n)* such that $0 < x < n$.

Free leave, denoted by $L_f(c,r)$, is the node with address *<c,r>* such that it is not base of any existing subtoruses and it is not allocated to a job.

Busy leave, denoted by $L_b(c,r)$, is the node with address *<c,r>* such that it is not base of any existing subtoruses but it is allocated to a job.

In Fig. 5(a), the introduced notions are illustrated for an exemplary torus *T(6,6).*

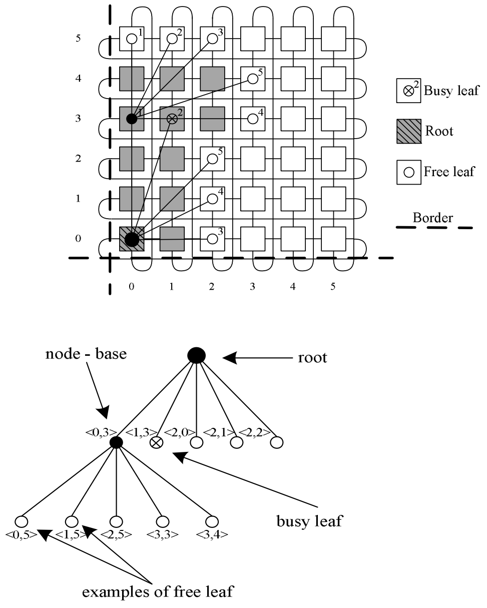


Figure 5(a). Illustration of the introduced notions for torus *T(6,6)*.

Allocation scheme always starts from node with address *<0,0>* which is a root of busy tree. The root is a constant element of tree and exists even if it is not allocated as base to any job. With first allocated job, tree evolution process starts, that lasts until there are no more requests to allocate. Tree evolution process is divided into two main sections.

Base of every allocated job generates children nodes (leaves, shown in Fig. 5(a)) as follows:

Add adjacent nodes, from left to right, along top of the busy subtorus.

Add adjacent nodes, from bottom to top, along the right side of the busy subtorus.

After that, busy tree is recursively updated and every leaf receives its status – free or busy. Free leaves are potentially base nodes and hence are directly under consideration. Tree is searched level by level from left to right, start with root level. First free leaf that meets the requirement is returned as base for given job. Search scheme is presented in Fig. 5(b).

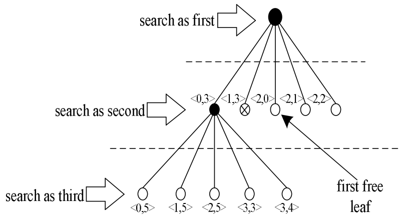


Figure 5(b). Search scheme for busy tree *T(11)*.

The search scheme, with respect to children creation mechanism, provides good fit of the incoming task to the existing configuration of torus, marked as *border* in Fig. 6(b).

Every allocated job, after elapse of its duration time is deallocated. It means that base node and its leaves are deleted from busy tree. Children that represent base are reattached to other base from current level or level up as follows:

If any base exists on the left side of deleted base descendants are attached to the base on the left side.

If none base does not exist on the left side of deleted base but does exist on the right side descendants are attached to the base on the right side.

If none base does not exist in current level descendants are attached to the parent base of deleted one.

V. Experimentation System

Experimentation system was developed in C++.

Input parameters:

The following task allocation problem parameters are taken into consideration:

P_1: the number of jobs in the queue (important data for static case of allocation).

P_2: the range of uniformly distributed pseudorandom sizes of each job in the queue (range of *p* and *q*).

P_3: the range of uniformly distributed pseudorandom numbers for execution time of each job in the queue.

P_4: the size of torus *T(w,h)* i.e., the values of *w* and *h*.

Output parameters:

The following indices of performance (measures of efficiency, criteria) are treated as system outputs:

Simulation time t_s [ms]: defined as total time of simulation. It is the time needed to allocate and process all given jobs. This criterion is being analysed during static experiments.

Effectiveness E [%]: defined as percentage of process jobs with respect to all given jobs. This criterion is being analysed during static experiments. Effectiveness is also measured in specific period of time for dynamic experiments.

Unreliability U [%]: defined as complement of effectiveness in specific period of time. Unreliability can be calculated using eq. 1.

$$U = 1 - E \tag{1}$$

This criterion is not being analysed itself due to the fact it is calculated based on knowledge of effectiveness. It is important during dynamic experiments.

Remark: The series of simulation experiments were carried out on the Intel Pentium i5 machine with 4 GB of RAM memory.

VI. INVESTIGATION

A. Static Allocation – Experiment Design

First experiment was focused on comparing total simulation times t_s and effectiveness *E* for the considered algorithms. List of tasks was generated according to range of P_2 and P_3 presented in Table I.

TABLE I. RANGE OF INPUT PARAMETERS

Parameter	Range
Job size	2÷7
Job time [s]	5÷20

Parameter P_4 was equal: 25×25, 50×50, 75×75, and 100×100. For each system, ten measurements were made on the basis of which the average result for each system was calculated. Each algorithm had to allocate respectively 25, 50, 75, 100, 150, 200, 300, and 400 tasks.

B. Static Allocation – Results

The averaged results, concerning the total simulation time t_s and the effectiveness *E* are presented in Fig. 6 and Fig. 7.

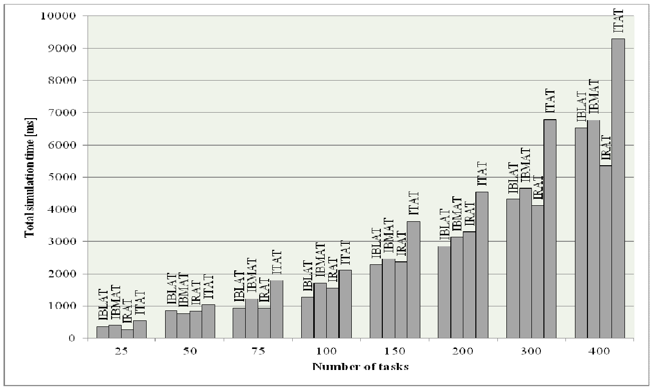


Figure 6. Average total simulation time t_s.

It may be observed, that the best algorithm, with respect to t_s is *IRAT* which is not reliable because it is based on randomness. Thus it is able to get through the whole list of tasks in short period of time but a substantial part of them will not be allocated and processed (Fig. 7). Due to the fact, that the best results were achieved for *IBLAT* and *ITAT* algorithms.

More complex structure of system does have noticeable impact on allocation time. Although differences are not significant and can be consider as neglected due to simulation error – experiments were done in the multitasking operating system that can cause measurement errors.

Testimony that the Improved Random Allocation Algorithm is not reliable is shown in Fig. 7. The *IRAT* works faster than the other mechanisms but a substantial part of tasks is not allocated and processed. Due to this fact the effective results were achieved for *IBLAT, ITAT* and *IBMAT* algorithms.

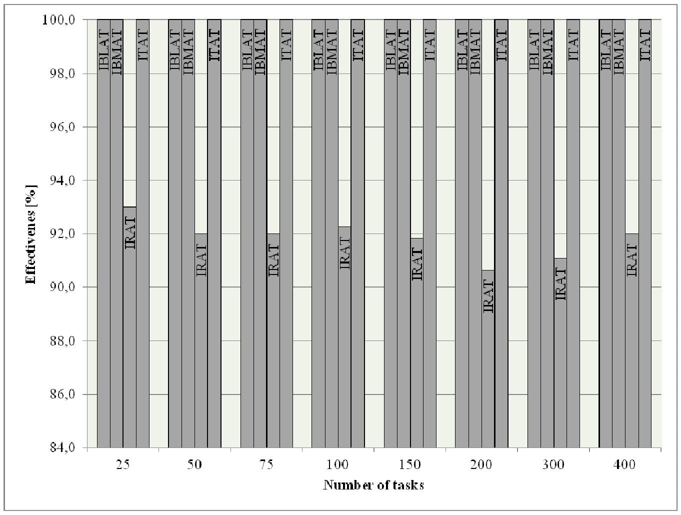


Figure 7. Average effectiveness *E[%]*.

C. Dynamic Allocation – Experiment Design

The second complex experiment was focused on comparing the total number of processed jobs *N* and unreliability *U*. Two cases were considered depending on the two sets of input parameters.

In the Case 1, incoming tasks were created according to the ranges of P_2 and P_3 parameters presented in Table II. Input parameter P_4 was equal: 25×25, 50×50, 75×75, and 100×100.

TABLE II. CASE 1: RANGE OF INPUT PARAMETERS P_2 AND P_3

Parameter	Range
Job size	2÷7
Job time [s]	10÷30

In the Case 2, incoming tasks were created according to range of P_2 and P_3 parameters presented in Table III. Input parameter P_4 was equal: 100×100, 200×200, and 300×300.

TABLE III. CASE 2: RANGE OF INPUT PARAMETERS P_2 AND P_3

Parameter	Range
Job size	20÷40
Job time [s]	50÷100

For the both sets of parameters and for each considered torus 10 measurements were performed on the basis of which the average result for each system was calculated. Each algorithm had to allocate incoming tasks during respectively 10, 20, 30, 40, 50, and 60 seconds.

D. Dynamic Allocation - Results

The averaged results of the experiments for both cases are presented in Fig. 8 and 9.

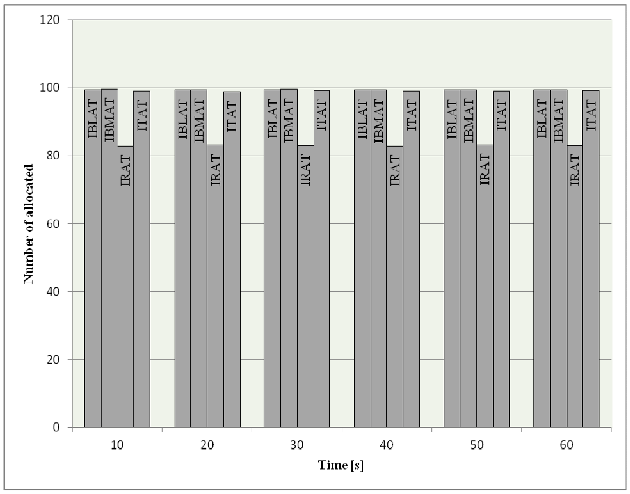


Figure 8. Average effectiveness *E* for Case 1

As expected, randomness has noticeable impact on the effectiveness and thus also on number of allocated tasks and unreliability. It allows processing large number of jobs (tasks) but at the cost of high unreliability what is a result of this, that system gets more tasks than it can be processed. It is important to know, that quality of algorithms cannot be determined solely by the number of allocated and processed tasks.

It can be said that effectiveness of algorithms for case 1 is comparable, or even the same - differences are barely noticeable. The effectiveness is not an ideal parameter because it is based on the speed of processing tasks within the system. Simplify algorithm, i.e. IRAT are faster, thus whole process of allocation for one task does not take a lot time. Because of that the number of processed tasks can be higher when compared with other algorithms, however in comparison with the all given jobs results are worse.

The probability of allocation success decreases with every processed job and so unreliability decreases.

It does not mean that every algorithm with ability to allocate large number of jobs, even at the cost of low effectiveness, is efficient. The efficient allocation technique needs to have balanced parameters.

Every incoming task can be described by the probability of its allocation. As the size of torus grows, the probability of allocation grows as well. The same can be said about effectiveness and total allocation time parameters that achieve higher values for larger systems. The reverse situation can be observed for duration time and size of incoming tasks. As the size and duration time of tasks falls, probability of allocation, effectiveness and total allocation time grows.

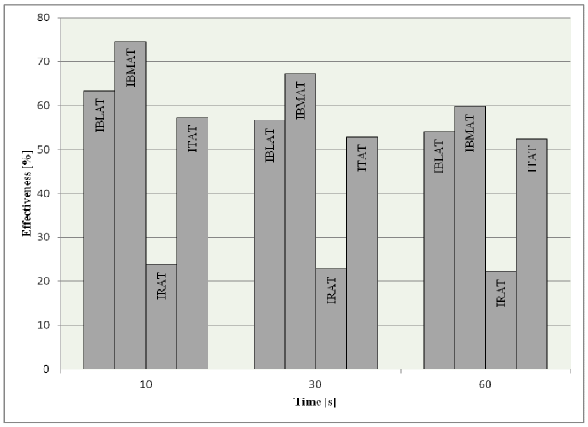


Figure 9. Average effectiveness *E* for Case 2.

VII. CONCLUSION AND FUTURE WORK

In this paper the Improved Tree Allocation for Torus (*ITAT*) was proposed. Based on experiments it may be observed that each allocation algorithm has its advantages and disadvantages. The proposed algorithm is particularly attractive for large toruses and large tasks. *ITAT* uses non-binary tree to identify free subtoruses which can be allocated to an incoming request.

Busy tree as the algorithm is a very complex structure that is why for smaller systems *ITAT* work worse than other schemes. It is important to know that modern networks with torus topology do have hundreds of nodes. Thus *ITAT*'s quality is comparable with other existing schemes.

The future work includes plans to extend *ITAT* by implementing more intelligent scheme for leaves creation process. It is necessary in order to maximize utilization of nodes. Other process that needs to be reviewed is recursively update of busy tree that has negative impact on total simulation time and allocation time. We also intend to combine leaves update scheme with leaves creation process.

REFERENCES

[1] T. Srinivasan,, J. Seshadri, A. Chandrasekhar, and J. B. Siddhart, “A minimal fragmentation algorithm for task allocation in mesh-connected multicomputers,” Report in Department of Computer Science, College of Engineering, Sriperumbudur, India, 2004.

[2] W. J. Dally and C. L. Seitz, “The torus routing chip,” Journal of Distributed Computing, vol. 1, no. 4, 1986, pp. 187-196.

[3] W. Kmiecik, L. Koszalka, I. Pozniak-Koszalka, and A. Kasprzak, “Evaluation scheme of tasks allocation with metaheuristic algorithms in mesh connected processors,” Proceedings of 21st International Conference on Systems Engineering (ICSEng), IEEE CPS, 2011, pp. 241-246.

[4] S. Yoo and R. Das, “An efficient task allocation scheme for 2D mesh architectures,” IEEE Transaction on Computers, vol. 8, no. 9, 2002, pp. 934-938.

[5] T. Liu, W. Huang, F. Lombardi, and L. N. Bhuyan, “A submesh allocation scheme for mesh connected multiprocessor systems,” Proceedings of International Conference on Parallel Processing, vol. II, 1995, pp. 193-200.

[6] I. Pozniak-Koszalka, L. Koszalka, and M. Kubiak, „Allocation algorithm for mesh structured networks” Proceedings of IARIA International Conference on Systems, 2006, pp. 24-30.

[7] D. M. Zydek and H. Selvaraj, “Implementation of processor allocation schemes for mesh-based chip multiprocessors,” Journal of Microprocessors and Microsystems, ISSN 0141-9331, vol. 34, no. 1, 2011, pp. 39-48.

[8] J. Ding, and L. N. Bhuyan, L. N., “An adaptive submesh allocation strategy for two-dimensional mesh connected systems" Proceedings of International Conference on Parallel Processing, vol. II, 1993, pp. 193-200.

[9] Y. Zhu, “Efficient processor allocation strategies for mesh-connected parallel computers,” Journal of Parallel and Distributed Computing, vol. 16, no. 4, 1992, pp. 328-337.

[10] D. M. Zydek, “Processor allocator for chip multiprocessors,” PhD. Dissertation, University of Nevada, Las Vegas, USA, 2010.

[11] D. M. Zydek and H. Selvaraj, “Fast and efficient processors allocation algorithm for torus-based chip multiprocessors,” Journal of Computers & Electrical Engineering, ISSN 0045-7906, October 2010.

[12] T. Baba, Y. Iwamoto, and T. Yoshinaga, “A network-topology independent task allocation strategy for parallel computers,” Proceedings of ACM/IEEE conference on Supercomputing, IEEE Computer Society Press Los Alamitos, CA, USA, 1990, pp. 878-887.

A Real Time Synchronous V - C System with the Extracted Data from Buffering Function

Hyung-Se Kim[1], Chan-Seok Jeong[2] ,
Moon-Hwan Lee[3]
CAAS (Center for Army Analysis & Simulation),
R.O.K.A
Gyeryong, Korea
e-mail: kjk9311@gmail.com[1], csjeong7@paran.com[2],
doleok445@gmail.com[3]

Baek-Yeol Seong[4], Sung-Woo Choi[5],
Mi-Seon Choi[6], Young-Kuk Kim[7]
Department of Computer Science & Engineering,
Chungnam National University
Daejeon, Korea
e-mail: {sby86[4], oneloveyou08[5], mschoi27[6],
ykim[7]}@cnu.ac.kr

Abstract—**This study introduces the state-of-art of the Real-time LVC (Live-Virtual-Constructive) interoperation systems that are being used for the purpose of warfare training and presents the first Real-time V-C (Virtual-Constructive) interoperation system building project in Korea, recently performed by ROKA (Republic of Korea Army). In this project, we have developed a new UAV (Unmanned Aerial Vehicle) image simulation system, which corresponds to V (Virtual) system and made it interoperate with the existing C (Constructive) system, ChangJo21, held by ROKA. Through this V-C interoperation, it became possible for the UAV image simulation system to simulate surveilance of a large group of troops like a real battlefield. But this V-C interoperation system can be suffered from severe overload caused by lots of data transmission between two systems. To solve this problem, we apply buffering function on data extraction procedure and use different data transmission strategies on the types of object. As a result, we have decreased more than 53% of amount of data transmission needed for V-C interoperation.**

Keywords-simulator; simulation; Interoperation system; Live; Virtual; Constructive; LVC; data extraction

I. INTRODUCTION

Due to rapid progress in IT technology over the decades it is anticipated that the theatre of the future is dependent to the network-based operational environment [1].

For now, it is said that combat training based on LVC training environment is the most economical and effective way for combat training because LVC virtually presents the real battlefield using computer and network technology.

In LVC training environment, “L” means a “Live” training system where real people operate real systems such as a pilot flying a jet. “V” means a “Virtual” training system where actual players use simulated systems in a synthetic environment. “C” means a “Constructive” training system where simulated players use simulated systems in a synthetic environment. Constructive training system is often referred to as “wargame”.

Any of L, V, C systems, which is inadequate to simulate complex battlefield situations by oneself, can complement each other by interoperating with one another, leading to a more realistic combat simulation. In general, LVC requires that at least two of three types, which are Live (L), Virtual (V), and Constructive (C), are involved.

To make L, V, C systems interoperable, assets, models, and effects from one training environment should be seen, affect, and be affected within the rest of the training environment. This implies that huge amount of data transmission between several training systems is needed. Therefore, it is important to effectively exchange and process data in real time which is served as the foundation for strategizing tactics and allocating commander and staff in punctuality [1][2].

Recently, we have developed a new UAV (Unmanned Aerial Vehicle) image simulation system, which corresponds to V system. As the next step, we have constructed a real time synchronous V-C interoperating system by making the UAV simulator interoperate with the existing C system, ChangJo21 [5], held by ROKA. In order to enable interoperation between the UAV image simulation system and ChangJo21 with different resolution(UAV simulator: entity-level, ChangJo21: unit-level), unit information of ChangJo21 is disaggregated into individual objects (entities) of UAV image simulator and individual objects of UAV image simulator are aggregated into units on real-time basis (within a second) [5]. We also reduced the amount of data transmission by applying buffering function on data extraction procedure for information exchange and use different data transmission strategies on the types of object.

The rest of the paper is organized as follows. In Section 2, we will introduce some restricted L, V, C training examples that are being operated in developed countries [3][4]. In Section 3, we will introduce a new V-C interoperating system consisting of a UAV image simulation system (V) and ChangJo21 (C). And then, we are going to describe the problems and our solutions for interoperating two heterogeneous simulation systems. Some modifications in simulation logic for supporting V-C system interoperation will be described in Section 4. Finally, we conclude our work in Section 5.

II. RELATED RESEARCH

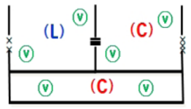
A. LVC examples in developed countries

1) U.S. Army

U.S. Army has tried to construct STOW (Synthetic Theater of War) as in Table I since 1994. However, they recognized some problems such as difficulty in sharing battlefield situation between Live (L) and Virtual (V), Constructive (C) training systems. It was very difficult to achieve different training goals for each system at the same time. A large amount of budget has to be injected to solve these problems. Instead, COP (Common Operation Picture) Based LVC (not to interoperate systems but gather unit identifications on COP to utilize in commander / staff procedure training) integration has been adopted in training.

In U.S. Army, Live training is mainly applied to an individual or a team and Virtual training is applied to middle echelon such as a brigade or a division. Constructive training is used for joint military exercises [1].

TABLE I. TRANSITION OF LVC SYSTEM

STOW : '94～'01 (Synthetic Theater of War)	**L.V.C : '02 ～** (Live Virtual Constructive)
INTEROPERABLE LIVE CONSTRUCTIVE VIRTUAL	
● Concept - Synthesis of L.V.C in the same theater ● Restriction - Restriction on situational description for L, V and C - Difficult to achieve all the purposes of various training - Need to secure network stability reliance	● Concept - Divide L and C, coexist with V ● Restriction - Restriction on situational description for L, V and C

2) German Army

The German Army thought that verification is needed to figure out if LVC really boosts effects in training. They are also worried that LVC might create situations that cannot happen in reality and lead to discontinuities in situation [4]. Table II shows the German army's evaluation results of the effects that LVC interoperation gives to combat training.

TABLE II. THE GERMAN ARMY'S EVALUATION RESULTS OF THE EFFECTS OF LVC INTEROPERATION IN COMBAT TRAINING

Sorting	Virtual (A)	Live (A)	Constructive (A)
Virtual (B)	+	-	◦
Live (B)	-	+	-
Constructive (B)	◦	-	-
+ (Expected effect), ◦ (Restricted effect), - (Insufficient effect)			

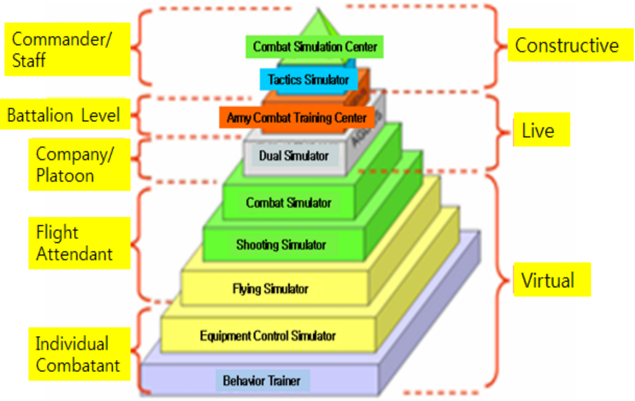


Figure 1. Diagram of scientific training system for the German Army.

Combat training systems are classified by the size of training into an individual - a crew - a platoon ～ battalion - a commander or a staff officer training. They are also divided into L, V, C according to application system.

The German army's LVC concept is similar to STOW (Synthetic Theater of War) of the U.S. Army. However, it is different from that of U.S. Army because each system operates solely by stages as in Fig. 1.

LVC training exercises continues step by step as follows. At first, individuals and crew are trained by virtual simulators (Virtual). Next, teams and platoons are trained interactively in real situation with dual simulator (similar to MILES: Multiple Integrated Laser Engagement Simulation), which are considered as Live training. Finally, commanders and staff officers are trained in virtual situation with war-game model (Constructive) [4].

B. Characteristics and development of virtual training

The purpose of virtual training is to make teams and individuals experience battle fields indirectly and execute tactical training by utilizing developed simulator in case of restriction in armament because of high price and danger.

However, most of simulators that ROK (Republic Of Korea) Army possess or currently under development are just for acquiring simple skills [3].

There is no simulator for tactical training linked to other systems in ROK Army. Therefore, a simulator which can facilitate the training environment in accordance to the training characteristics and interests has to be taken in place in the near future.

1) System Operations

When we develop simulators, there are many things that should be considered such as acquiring cost, developing range according to purposes of training, adjusting to environment, evolving concepts of operating battle, and developing new technologies.

There are little data that can approve the effect of PC game-based simulator training and it needs to be checked by many types of experiments. Especially, PC game-based simulator training has to be used as an auxiliary tool for real maneuver in an echelon from a soldier to a battalion, not just simulation or simple game method.

Simulator for instrument training in the past was just for acquiring skills. However, it is being developed to be used in tactical training for a platoon or a company, and synthetic theater in the same environment of the crew.

Therefore, it should be used to train operating instrument, controlling and managing in various environments as preceding training instruments.

2) System construction / development

Simulators need to be made according to the purposes such as acquiring skills and tactical training. Simulators for acquiring skills should be settled in institutions and simulators for tactical training should be settled in field units to practice tactical training.

However, all of the simulators in ROK Army (24 types 175 items) are put in practice for acquiring skills. Therefore, all the simulators should be used in training connected with the same system.

Ongoing simulator development project is developing simulators for tactical training. However, connection with other systems is not considered at this moment. Therefore, simulators in the future should be connected to not only the same system, but also the other systems for integrated training.

Simulators should be constructed for enabling not only improvement in controlling and firing ability under various virtual reality training environment but also training for a team, a platoon, or a company.

In the future, simulators should be developed in the way of satisfying requirements such as interoperation property, ease in improving capacity, ease in setting up and supporting military demands, ease in movement and operation, ease in various kinds of training, adjustment to environment, and cost effectiveness.

Like the above, virtual training is the training that real forces participate with simulation equipments. It can give opportunities for acquiring equipment control skills / proficient training method and HITL (Human-in-the-Loop), such as human reaction process, decision process, human-machine interface, etc. [1].

However, it can only be used in training of a person or a small force for simultaneous training.

C. Characteristic and development of a war-game model

War-game (Constructive) models that ROK Army possesses, ChangJo21 and Combat21 [13], do not reflect the characteristic of terrain of Korea and brand-new weapon systems.

Future development of war-game model has to move toward improving the performance of training model currently being used and making possible to interoperate with newly developed simulation models for auxiliary functions. It is also moved in the direction of developing a war-game model which is able to train upper and lower echelon in collective way and to train and test various echelons, and to train multi-function echelons [6].

1) Development of simulation capabilities according to battle field function

A war-game model should be developed to provide Command and Control/Communication capability similar to reality in a network environment on the way that directions deliver, and this enables Command and Control/Communication training. It also should simulate the procedure to collect information by analyzing, processing, merging secret information gathered by a number of secret information collecting system. Especially, it should be able to provide collected data with images and photos with same form to real system.

Since mechanized corps has high mobility and strong combat capability, a close combat must be differentiated according to simulation level. Also, a war-game model should provide detection-decision-fire procedure for operating integrated fire-power and simulation of a precise strike by a high precision projectile weapons [12].

In addition, the model must be able to simulate air defense, automation system, real time air observation and radio system as well as details about a combat service support system of a whole army.

2) Development of image information field

When simulating a field of information, although it is possible to operate equipments such as a UAV, TOD (Thermal Observation Device) and a RASIT (Radar Surveillance Intercept), the fact is that training of a field of information is restricted because detection results are offered dissimilar report form to actual fighting.

On the other hand, the U.S army practices using methods providing virtual image information produced by collected data from UAV interlocked with CBS (Corps Battle Simulation) model for direction command center in every 10 minutes. Therefore, ROK Army needs a system that can give information analysis training opportunities in ASIC (All Sources Intelligence Center) by providing image information that is similar to real system to command center.

As stated above, as a method for scientific corps practice using computer simulation in order to enhance combat direction abilities, war-game models are economical and offers us large scale integrated practice environment, but, have defects that they cannot embody actual fighting combat situations (equipment controlling exercise and the rest) [11].

D. Development of LVC integrated simulation system

The purpose of constructing LVC integrated simulation system is to provide a virtual integrated battle field environment similar to a real combat circumstance, by interlocking Live (L), Virtual (V) and Constructive (C) system with inter-working interface.

If providing a virtual integrated battlefield environment by using available M&S (Modeling and Simulation) system, it is possible to do integrated training; a training from an individual to theater level, the staff and the director training by function, and an individual soldier training.

1) *Working system*

Most LVC working systems of U.S army have the main purpose of providing a natural integration of LVC training assets and a large scale distributed environment by integrating weapon system simulators, and constructive simulations which depict a large scale battle-ground with a huge military strength in order to conduct training which considers operation situation of theater level.

Live Assets should be selectively connected to provide highest level of tough-to-feel considering the restricted usefulness of real system integration and various safety precautions. In addition, it should provide integrated simulation that scenarios like actual fighting, real-world protocols and interfaces are integrated with other systems to support pertinent training of live asset managers.

A virtual simulator asset should provide flexibility of training with a low cost by establishing an integrated synthetic training environment and training of strategies and tactics under dangerous areas or virtual situations. It should make it possible to conduct flexible training through simulating sensors and tough-to-feel three-dimensional terrain and pause/resume function not in reality.

A war-game (Constructive) asset encourages to provide construction synthetic environment for sufficing a established training goal, production of a number of friendly forces, enemies and neutral forces, and a perfect training atmosphere complementing a platform produced by combatants in real and a virtual simulator, as well as dynamic training environment by creating unexpected conditions and enemies.

2) *Establishing and development of System*

Steps of establishing and development plan for a LVC integrated simulation system of ROK Army is like below [3]; Table III, Table IV and Table V.

- Step1: Establishing each model separately considering interoperation in the future (’14 ~ ‘20).
- Step2: Establishing interoperating system between the other systems after between the same systems
- Step3: Establishing division level/ brigade level LVC interoperating system (’19～’25).

TABLE III. ESTABLISHING AND DEVELOPMENT OF EACH SYSTEM

Live (L)	Constructing a foundation system interlocked with other systems when promoting project with a brigade level (～'14)
Virtual (V)	Developing a simulator which allows integrated simulation for tactical training (～'24)
Constructive (C)	Developing a brand-new army’s combat direction training model which enables to be interlocked with other systems and models (～’17)

TABLE IV. INTERLOCK BETWEEN THE SAME/OTHER SYSTEMS

V-V	Support L, C system by establishing connection between V systems
C-C	Support a war-game training by establishing connection between C systems
V-C	Firstly construct an interlock system (V-C) considering effect of training and technical level

TABLE V. ESTABLISHMENT/DEVELOPMENT OF LVC SYSTEM DEPENDING ON ECHELON

Division level LVC	Establish LVC system considering training demand and technology after training interlocked between same systems.
Brigade level LVC	Establish with minimizing errors by analyzing division level LVC cases

If we establish the LVC integrated simulation system as above, we attain vertical integration by echelons and horizontal integration by function. The system allows conducting unified combat training based on network by integrating various training capabilities such as detection asset, decision of direction, and means of attack.

E. Restriction on establishing LVC integrated system

Provided that we utilize LVC integrated simulation system, we can achieve the goal, “A training likes a combat, a combat likes a training.” for directors, chiefs and combatants by attaining upmost effects similar to real combats by implementing virtual simulation environments. However, there are some restrictions for building LVC integrated system as follows.

First, a standard architecture which is able to satisfy all interoperability demand or interlock system (middle-ware) is absent.

Since each L, V, C system is developed based on field-specific requirements of various users they differ so much in communication protocol, object model, and so on. It implies that data converting and mapping process are required when interlocking between systems, which not only increases complexity of the system but also causes a bottleneck due to the tremendous amount of data it has to deal with

Second, when interlocked between simulation-systems which have different simulation resolution (degree of depiction: object unit, corps unit), the systems need disaggregation and aggregation process due to difference of resolution of information. While ChangJo21 training model of ROK army like Fig. 2 depicts battles in a corps unit, the UAV simulator depicts battles by individual weapon systems.

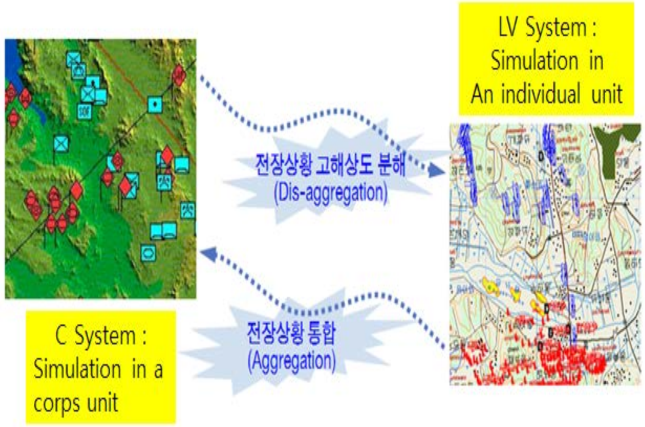


Figure 2. Aggregation/Disaggregation process between LVC systems.

Because the simulator is unable to depict such a direct combat between corps and individual weapons, it is required that corps is disaggregated into individual equipment in necessity or numerous individual weapon systems aggregated into a unit. After several rounds of disaggregation and aggregation process, it is difficult to maintain accuracy and consistency of information due to loss of information.

Third is a technical problem concerning about how to transmit status information of corps/objects depicted in each L, V, C training system to the counterpart systems. We have a technical problem about transmitting various status information including position information in V training system (UAV) to C training system (ChangJo21) and how to apply to V training system when damages occur in combat in C training system [3].

Another technical problem is due to the network overload caused by transmission of about 49,000,000 byte per second (participated corps * combat equipment possessed by corps, the number of supporting equipment * fixed objects). This data transmission quantity of 49,000,000 byte/sec is hard to operate and apply to different systems (V-C) in short amount of time. Thus, it is hard to guarantee actual combat condition.

III. DEVELOPMENT OF NEW V-C INTEROPERATING SYSTEM

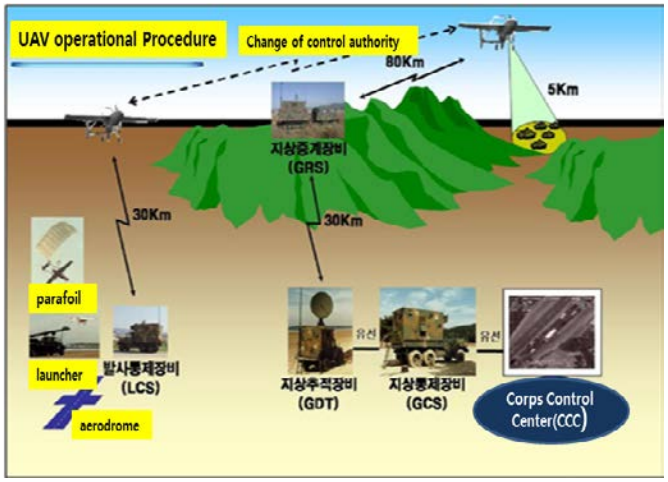
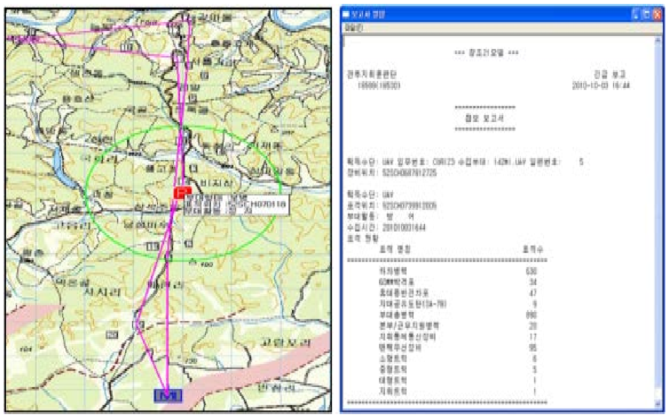


Figure 3. Operations of Real UAV System.

A UAV system is operated as in Fig. 3. UAV flies without pilot by remote control and performs a reconnaissance mission. Images acquired during reconnaissance flight are transmitted to GRS (Ground Repeater System) by the radio, then transmitted to CCC (Corps Command Center) [4][11].

ChangJo21, one of the constructive training models that are owned by ROK Army, can simulate several information assets (UAV, TOD, RASIT, etc.). Among these, UAV automatically detects units / armaments of enemy located within radius of 2km by applying certain probability when inputting flight-related information (flight route, prior-target decision, etc.) as in Fig. 4.



*** 창조21모델 ***

2010-10-03 16:44

첩보 보고서

UAV

[illegible]

[illegible]	[illegible]
[illegible]	630
60MM[illegible]	34
[illegible]	47
[illegible](SA-7B)	9
[illegible]	890
[illegible]	20
[illegible]	17
[illegible]	95
[illegible]	6
[illegible]	5
[illegible]	1
[illegible]	1

Figure 4. ChangJo21 UAV Simulation (Result of target detection).

However, ChangJo21 provides the reconnaissance result of UAV simulation in text-based report without analysis procedure of target detection result. The result is that realistic training is restricted [5].

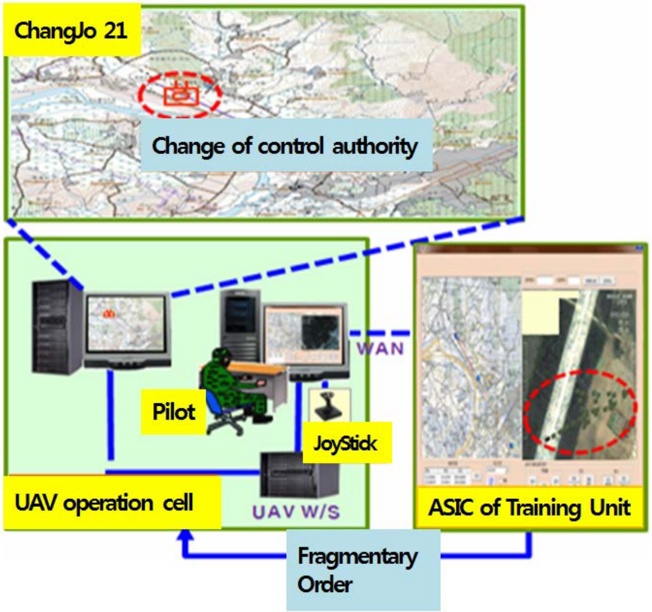


Figure 5. Newly developed V-C interoperating simulation system.

In order to overcome this restricted training condition, we have developed a new UAV image simulator and made it interwork with ChangJo21 constructive training model as in Fig. 5. The UAV simulator can produce reconnaissance images close to the level of images produced by real UAV. This V-C interoperating simulation system provides similar effectiveness to real training through virtual simulation environment and enables training for commands and staffs in intelligence field by simulating target detection/analysis /processing functions and mastering UAV image photographing at the same time [9][10].

A. Problem of interworking between ChangJo21 training model and the UAV image simulator

1) Object data transmission reduction strategy

There is a problem in supporting real-time (1 second) training using our V-C system due to the network overload caused by 49,000,000 byte of object data transmission.

To solve this, we have applied a dispersion method that only considers objects in the inner section of UAV detection capacity. First, the required information (approximately 18,000,000byte) within UAV detection capacity (40km) is transmitted. Then, we applied the strategy that the updated information within detection territory is transmitted differently on the type of object such as fixed object or moving object.

Different transmission strategies of object information are shown in Table VI.

TABLE VI. TRANSMISSION STRATEGY OF OBJECT INFORMATION

Item	Transmission Strategy	Remarks
1. Transmission of initial data	Transmit data only within detection territory considering UAV detection capacity (30Km)	- Reduction of 80% of load - Increase in complexity of related logic
2. Transmission of fixed object	Need to register or delete of object along the UAV flight	- Increase of load - Increase in complexity of related logic
3. Transmission of moving object	Register or delete objects only within detection territory considering UAV detection capacity (30Km)	- Reduction of load - Increase in complexity of related logic

a) Designing Concept

As in Fig. 6, the UAV image simulator transmits objects within detection territory (40km) which is determined by applying buffering function (10km) to UAV detection capacity (30km) to prevent frequent addition, deletion of objects. Reconfiguration also has been made to have average network load at the moment of large capacity.

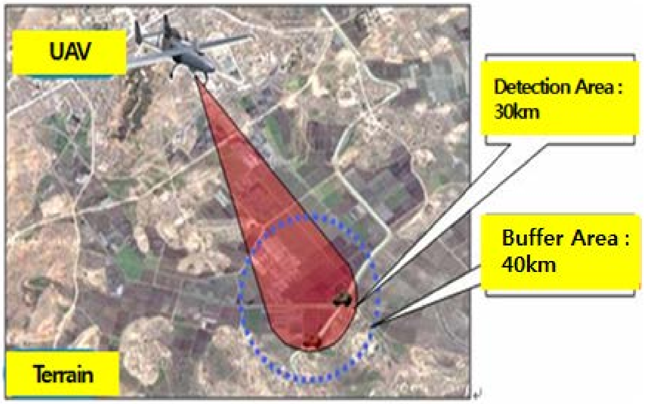


Figure 6. Domain of object processing that is capable of buffer function.

b) Comparison of reduced load caused by reconfiguration

Table VII is a table of results that measures and compares quantity of data that has transmitted to the UAV system. This is based on the assumption that the number of participated military units is about 6,000 in ChangJo21 constructive training system, and the average number of combat equipments and supporting equipments is 30.

TABLE VII. RESULTS OF MEASURING DATA QUANTITY SENT TO THE UAV SYSTEM

Classification		Transmission (byte/sec)	Remarks
Quantity of transmission before reconfiguration		49,000,000	
After applying detector capacity area	Initial (40Km)	18,000,000	First Reconfiguration
	Fixed object	30	
	Moving object	78,000	
After applying buffer function	Initial (30Km) +Buffer	900,000	Final Reconfiguration
	Fixed object	30,000	
	Moving object	15,600	

If the image diameter (width) of image sensor is 40km, radius of area that object information should be exchanged between our V-C interoperation system is more than 40Km, which is too large to continuously transmit in real-time.

We have accomplished more than 53% of data reduction by reconfiguring V-C interoperation system in order that only object information inside detection territory including buffered area is transferred with different transmission strategies depending on the types of object, while satisfying requirement of interoperation system that has to be provided by V system, which requires real-time information exchange.

2) Information resolution conversion between systems

ChangJo21 training system describes battlefield in unit (battalion) while UAV simulator describes battlefield in individual entity (tank, armored vehicle, etc.). This difference of information resolution between two systems implies that information resolution conversion function is needed to interconnect them. A Unit should be disaggregated to individual entities, or many of weapon systems should be aggregated into one unit as in Fig. 7.

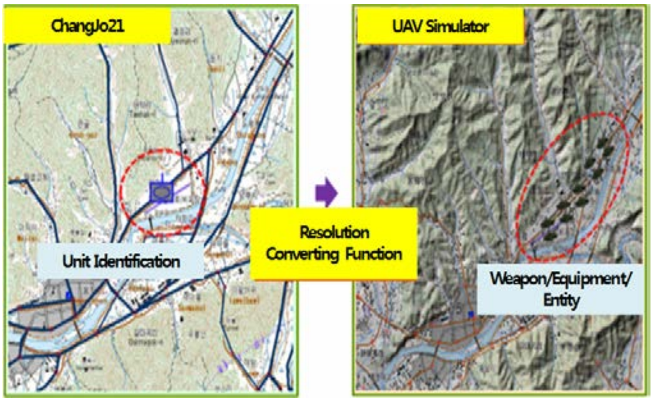


Figure 7. Information resolution conversion / exchange.

In addition, state information (location, state of equipment, etc.) for the UAV is transmitted from the UAV simulator to ChangJo21 model in real-time as in Fig. 8. It makes operating UAV simulator impossible when it is damaged by air defense weapons.

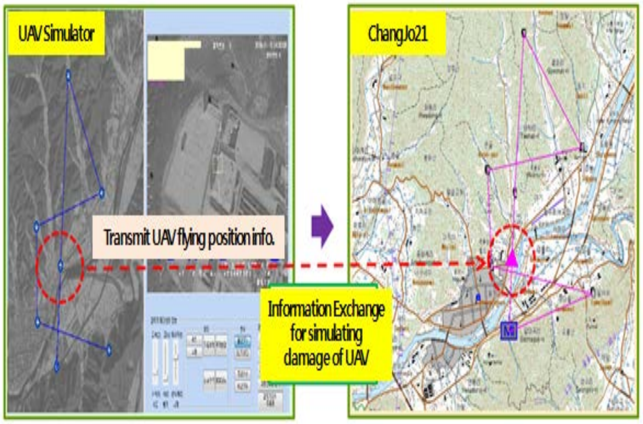

UAV Simulator
ChangJo21
Transmit UAV flying position info.
Information Exchange for simulating damage of UAV

Figure 8. Transmission of UAV state information.

3) Correspondence of topographical information between ChangJo21 and the UAV simulator

Due to differences of topographical information between ChangJo21 and the UAV simulation system, unrealistic situations may occur. For example, tanks may move along rivers or mountains as in Fig. 9. In this case, it is required to match topographical information between two systems. Through this work, objects (tanks, etc.) can move along the roads.

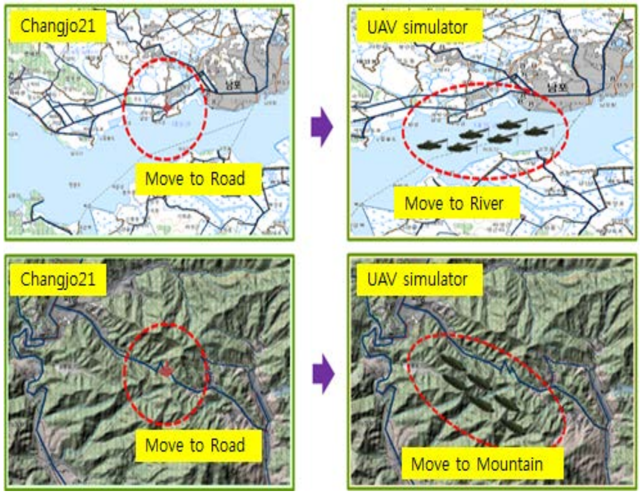


Figure 9. Inconsistency of geographical information between systems.

4) Optimization of simultaneous target processing for UAV simulator

More than 7,000 military units are operated in corps combat-command training using ChangJo21, and approximately 350,000 objects (entities) are operated when units are transformed to objects (weapons) in order to be expressed in the UAV simulation system. It is impossible to operate systems normally due to overload caused by processing such a large amount of objects simultaneously.

We have reduced system loads by processing only updated objects located within feasible photographing regions by UAV sensor (with diameter 00km) as in Fig. 10.

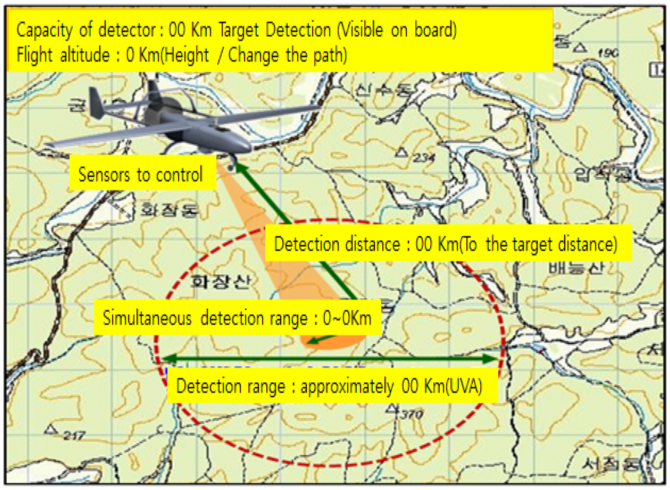

Capacity of detector : 00 Km Target Detection (Visible on board)
Flight altitude : 0 Km(Height / Change the path)
Sensors to control
Detection distance : 00 Km(To the target distance)
Simultaneous detection range : 0~0Km
Detection range : approximately 00 Km(UVA)

Figure 10. Target detection / Discerned territory.

B. Construction of a real-time UAV image sharing system

Gamers who are responsible for controlling combat units in combat command training by corps are located in BSC (Battle Simulation Center) and CCCs (Corps Command Centers) are located in that corps. That is, BSC and CCCs are placed and operated in different locations. The UAV image simulation system is operated in BSC's information simulation department.

We have constructed a real-time UAV image sharing system by transmitting acquired secret information of UAV image to ASIC and CCCs as in Fig. 11. It enables commanders and staff officers to have exercise for information analyzing /handling process.

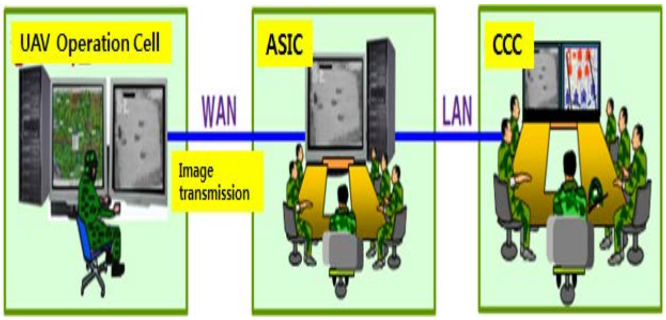


Figure 11. Architecture of real-time UAV image sharing system.

C. Comparison of UAV image systems held by Korea and U.S

Fig. 12 shows the differences in a visual aspect between our UAV image simulation system and MUSE, UAV virtual reality system that has been used to real training since August 2003 by U.S. Army.

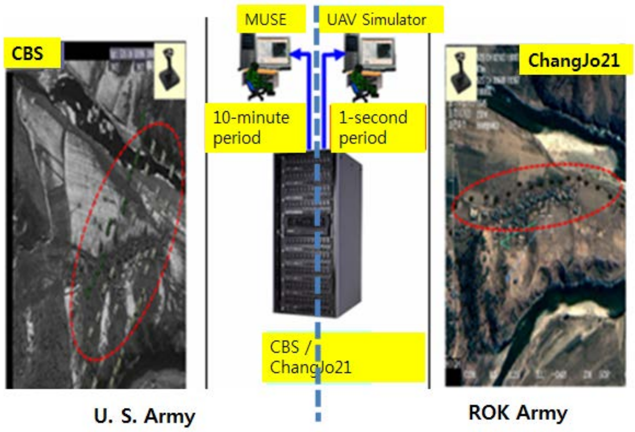


Figure 12. Comparison of UAV image systems in Korea and U.S.

Table VIII also describes system configuration for each system in detail.

TABLE VIII. COMPARISON OF KOREA-U.S. UAV IMAGE SYSTEM

Item	Contents
MUSE (U.S. Army)	• MUSE operator: located in ASIC • Map image: Black / White • Sensor capacity: Diameter 0Km • System overloads due to the transmitting information of all units within operating area (by minute)
UAV (ROK Army)	• UAV system operator: Located in simulated Information team ⇒ transmitting UAV image to ASIC • Map image: Color • Sensor capacity: Diameter 00Km • Minimize system overloads by selected transmission of data which is in UAV screen (by second)

IV. MODIFICATION OF SIMULATION LOGIC FOR V-C INTEROPERATION

A. Visual factor

In order to reflect shielding effect to the objects (weapons) by applying altitude to plateau/ridge considering 3-dimensional topographical characteristics, we have to match topographical information in training model with those in the UAV simulator. In addition, topographical information should not be applied to simulator in artificially edited way.

Unfortunately, 3-dimensional topographical map of our UAV simulator does not support forests (plants), civilization, and topographical shape (mountains, clouds), so that it is necessary to consider the effect of detection of hidden and concealed objects in real theater as in Fig. 13.

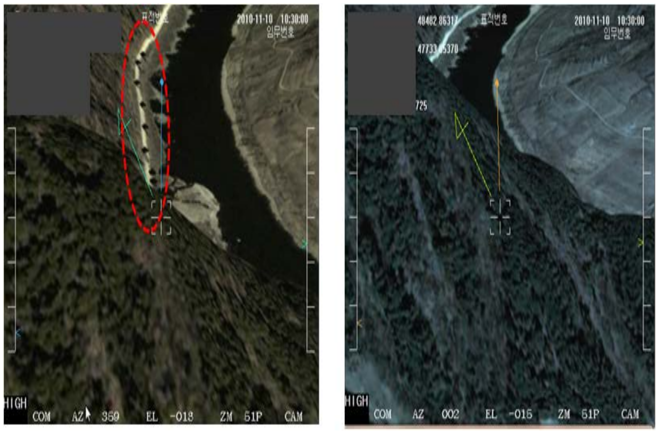


Figure 13. Shielding effect considering three-dimensional topography.

We have solved these problems by applying different detection rate on topographical factors in the training model. Specifically, we have applied 60% of detection rate on objects of stopped units considering various kinds of hiding and concealment in combat situation, and have applied 100% of detection rate on objects of moving units using roads except for infantry outfits. We decided the detection rates based on the experiences in real combat training.

B. Weather factor

In operation of real UAV, it is general to set up flying plans considering rainfall, snowfall, the direction of the wind/the velocity of the wind. Therefore, we have made UAV flight restricted considering weather condition in the UAV image simulation system as well. The flight is restricted when rainfall more than midway (0mm /H above) or snowfall more than midway (00mm /H above) occurs or when right wind 00m/s / side wind 0～0m/s occurs in respect of the direction and the speed of wind.

On the other side, it is necessary to carry out consistent research on special effects such as blurred image occurred by weather condition on sensor image photographing / diminished visible area.

C. Application of moving distance depending on unit formation / armament

We have applied different movement intervals depending on weapons (tanks, armored vehicles, etc.) and vehicles to provide each object location to simulator applying 6 formations (column, row, tripod, inverse tripod, right echelon, left echelon) and tactical intervals between objects.

Simulation logic for unit movement of training model has been modified, which made it possible for marching columns to follow along the bending roads instead of keeping straight disposition as in Fig. 14. This logic modification has brought a positive effect on not only damage estimation on moving units but also interoperation between simulator and training model.

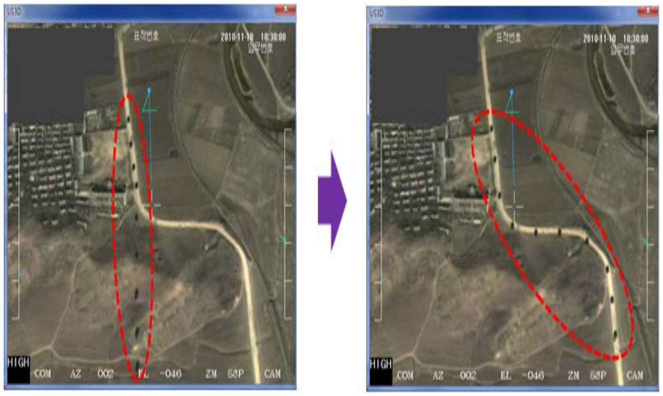


Figure 14. Objects moving along the roads.

D. Further developments

Newly developed UAV image simulation system is the first system that supports V-C interoperation in ROK Army but is not perfect yet. There are several points to be complemented and further developed.

First, in terms of personnel / portable equipment, the reflection of three-dimensional image of objects, is restricted if the number of objects (military forces) is excessive. In case of considering all personnel, the system loads will be aggravated.

The reflection is also restricted due to limitation of three-dimensional image processing capability of portable equipments. Thus, it is necessary to estimate appropriate number of persons and reflect it to the system after system overload test for further personnel. Displaying of three-dimensional object images on portable equipment should be considered after developing three-dimensional image processing capability.

Second, it is necessary to complement reality such as blurred image effect and reduction of visibility range in target detection / discernment considering image photographing effect which includes weather condition like rainfall, snowfall, mist, cloud and the rest.

Third, real UAV operation has been executed mainly on providing target information associated with tactical plan and secret information for BDA (Battle Damage Assessment) evaluation. Consequently, newly developed UAV image simulation system supports 3D image photographing and represents objects considering combat damage as that objects destroyed by attack are deleted in real-time. However, the flame and smoke from objects (equipments) when damage occurs are not included in the image. Therefore, this weak point should be complemented.

V. CONCLUSION AND FUTURE WORK

Current simulation models (system) held by ROK Army have limitations on vicarious execution of mission and action of combat personnel. Moreover, there are restrictions on providing comprehensive training / exercise opportunity because simulator training is also restricted for upper level unit's commanders / staff training. Therefore, LVC training system is the most suitable means for commanders, staffs, and combat personnel to train together in ordered ways.

The purpose of LVC training system is to embody the system for accomplishing similar effect with real combat execution through providing virtual simulation environment similar to real world. Still, the embodiment of LVC integrated simulation system is currently restricted in ROK Army because LVC training system construction is planned from year 2019 to 2025 in successive manner.

We described about a V-C interoperation system first developed in Korea that connects UAV simulator (V) and ChangJo21 model (C). We have enhanced the performance of the V-C interoperation system by applying buffer function only to changed object information that is within the diameter of image sensor, and by constructing interoperation system.

As this study suggests, when V system has self-image and interoperates with C system in real-time, and when V system interoperates with war-game model (C) for a large troop, it is possible to do real-time data synchronization between two models.

We have developed the foundation of L-V-C system without trial and error that many of the M&S developed countries experienced in interoperation of different systems.

ACKNOWLEDGMENT

This work was supported by Defense Acquisition Program Administration and Agency for Defense Development under the contract UD110006MD, Korea

REFERENCES

[1] ROK Joint Chiefs of Staff, National Defense M&S System Improvement Points (12~26 Integrated Concept Guide), Appendix Ⅲ, Annex 2, 2011.

[2] ROK Army Headquarter, M&S Policy Guide (2009 ~ 2025 Fundamental Policy Guide of Army), 2009.

[3] ROK Army Headquarter, Scientific Training System (LVC) Mid-long Development Planning, 2010.

[4] ROK Air Force Headquarter, Air Force LVC System Interoperation Technique Concept Research, 2010.

[5] ROK Army Training & Doctrine Command, ChangJo21 Model '11 Simulation Logic Analysis Guide (Education Reference 25-14), ROK Army Press, 2011.

[6] S. Y. Choi, National Defense Modeling and Simulation, National Defense University, 2007.

[7] K. T. Kim and Y. S. Moon, Data Structure Theory, Jeong-Ik Press, 1990.

[8] G. R Ash, Traffic Engineering and QoS Optimization of Integrated Voice & Data Networks, MORGAN KAUFMANN Publishers, 2007.

[9] D. J. Hatley and I. A. Pirbhai, Strategies for Real-Time System Specification, Dorset House Publishers Co., 1987.

[10] M. Fowler, Refactoring, Dae-Chung Media, 2003.

[11] K. S. Choi, H .G. Jung, and T. Y. Park, "Mission Analysis Research on UAV System Operation Effectiveness", Korea National Defense Management Analysis Society Conference Division 3: Management Science/ OR, 2011.

[12] ROK Army, Principle of Ground Weapon System(I), Chapter 4, ROK Army Press, 2010.

[13] ROK Army Training & Doctrine Command, Combat21 User Manual(I), ROK Army Press, 2003.

Avoiding Border Effect in Mobile Network Simulation

Raid Alghamdi, John DeDourek, Przemyslaw Pochec
Faculty of Computer Science
University of New Brunswick
Fredericton, Canada
Email: r.alghamdi, dedourek, pochec@unb.ca

***Abstract*—Simulation of mobile networks requires reliable movement generation. Random movement pattern is frequently used in simulators. Standard movement generator setdest in ns2 suffers from border effect, i.e., shows bias towards placing the nodes in the center of the simulated area. We propose and implement a different method for random movement generation in ns2 simulator based on the boundless movement mobility model. By using the quadrats count statistical testing, we show that our movement generator improves the randomness of the node distribution during the simulation.**

Keywords- movement generator; network simulation; setdest utility; MANET; VMR; Quadrats Count; ns2.

I. Introduction

Communication networks are divide into two main categories: wired and wireless. Wired networks exist between a number of devices connected to each other using connecting media, such as cables and routers. Wired networks can be applied within an area limited by the cables and routers that allow for sending and receiving of data. Wireless networks, on the other hand, are free of such space limitations, and are more easily able to connect different devices to each other. Wireless nodes can play the roles of both hosts and routers, which forward the packets to neighboring nodes.

The mobile ad hoc network (MANET) [1] is a subcategory of ad hoc networks. With the advent of newer technologies, mobile ad hoc networks are becoming an integral part of next-generation networks because of their flexibility, autoconguration capability, lack of infrastructure, ease of maintenance, self-administration capabilities, and cost-effectiveness [2]. A MANET contains mobile nodes that can be connected wirelessly to each other, for example through either Wi-Fi or Bluetooth. Nodes can be connected over wireless links in ad hoc fashion without central control; this is one of the main advantages of MANETs. In addition, MANET is dynamic and does not rely on fixed or static structure. Consequently, the frequent changes that occur in network topology impact mobile ad hoc network protocols' performance [2]. Because of this very dynamic structure, designing a new MANET often relies on a simulation modeling. In turn, simulation requires a reliable movement generation. In this paper we propose an improvement to setdest, the existing and popular ns2 utility.

The structure of this paper is as follows. In Section II, we discus different movement models for MANETs, including the setdest utility used in ns2. In Section III, we introduce performance measures for evaluating a movement generator. In Section IV, we present a new movement generator. The performance of the new generator is discussed in Section V.

II. MOVEMENT TYPES USED IN MANETS

The mobility models for MANETs can be grouped into two categories: random movement models (which will be discussed late in this paper) and uniform movement models. The uniform movement models include four well-known models: Boundless Simulation Area Mobility Model, Gauss-Markov Mobility Model, A Probabilistic Version of the Random Walk Mobility Model, and City Section Mobility Model [9]. First, a boundless simulation area model is based on the velocity of the mobile node of the current direction and the previous direction [4]. A Gauss-Markov Mobility Model resembles a random model but in fact it is not because it follows a pattern that could be calculated in advance [9]. The Gauss-Markov model is calculating two main parameters of each mobile node which are speed and direction at a certain time instance based on last instances update [9][18][19]. The Probabilistic Version of the Random Walk Mobility Model is a model that uses the probability to determine the next position by using the node state at each position [9][20]. The probability of the Probabilistic model is going higher when the mobile node keeps following the previous direction and is lower if the direction is to be changed [9]. Finally, the City Section Mobility Model is a realistic movement model where the movement of the nodes is still random but the paths are constrained to the grid representing streets in a city [8][9].

A. Random Movement

The Random mobility models used in MANETs differ in the way the nodes move. These principal movement types are: Random Walk, Random Waypoint and Random Direction. Each mobile node in Random Walk has a randomly generated starting position. The nodes travel from their starting position to a randomly generated new location by generating random direction and velocity [9]. A node changes direction and speed either at the end of a time

ISBN: 978-1-61208-245-5

interval t or if it traveled a distance d. In Random Waypoint, the movement is not constant whereas pause times are introduced. The nodes start at randomly chosen positions, then ”pause” for some time and then start moving at a random velocity towards a chosen destination. The nodes in this model have to ”pause” for some time before they change direction or speed [9]. The drawback in Random Waypoint is clustering near the center (i.e. having the nodes near each other near the center of the experimental area). Random Direction Mobility model was introduced to overcome this drawback in the Random Waypoint model. In Random Direction model, a node travels at a chosen velocity in a chosen direction until it reaches the boundaries of the area rather that until it reaches a randomly chosen location. Once a node hits the boundaries it pauses for a time t and chooses a new direction and starts moving in this new direction again, and so on [7]. The direction is changed only when a node hits a boundary.

B. ns2 Setdest Utility

Setdest is a tool used to generate nodes movements for the mobile nodes in the network simulation ns2 by positioning network nodes in a bounded area and setting the movement in a random direction [10][11]. Setdest tool (version v2) uses the random waypoint mobility model algorithm to create the random movements for the mobile nodes [10][11][12][21] and accept the following parameters: number of nodes, maximum speed, minimum speed, speed type, pause time type, simulation time, x coordinate, and y coordinate) [15][22].

III. INVESTIGATING MOVEMENT GENERATOR PERFORMANCE

We tested the Setdest utility and our new movement generator in two ways: by evaluating the randomness of the position of mobile nodes at different times in the course of simulation, and by comparing the delivery ratio in a simulated MANET in different regions of the experimental area.

A. Randomness Performance

For testing the randomness performance we used the Quadrats Count methodology and the Variance to Mean Ratio (VMR). The Quadrats Counts methodology is an established technique used for analyzing spatial point patterns present in an area by dividing this area into a certain number of sub-areas and then counting the number of points in each sub-area independently [16]. The Variance to Mean Ratio is a statistical test that describes a spatial distribution. We calculate the mean $\bar{x}$, and the variance s^2, of the number of points (network nodes) in each subarea in the Quadrats Count method. The closer the ratio

$$VMR = \frac{s^2}{\bar{x}} \tag{1}$$

is to 1, the more the points are randomly distributed.

B. Delivery Ratio

We tested the boundary effect (i.e., changes in nodes density along the edges of experimental area) by using a random movement generator to run the MANET simulation and then transmitting the packets through the center and also transmitting the packets along the edges. We used a CBR (constant bit rate) [23] traffic over UDP [24], using AODV routing [25], and counted the total number of bytes delivered at the destination node. The higher the delivery ratio for transmission along the edges, the better the movement generator.

IV. CUSTOM JAVA MOVEMENT GENERATOR

The setdest utility uses the random waypoint mobility model algorithm. The waypoint algorithm is known to have the border effect, which creates a kind of clustering at the center of the simulation area [7]. The border effect problem can be avoided by following a different algorithm called the boundless simulation area mobility model. The main idea of the boundless model is to allow going over the edges thus avoiding the influence of the simulation areas edges. Moreover, going over the edges in the boundless simulation means that once the mobile node reaches the boundary from any side of the simulation area, it does not bounce the same as in the other models, but it disappears from the side and reappears from the other side continuing moving with the same direction, which makes the simulation area look more like a tube than a plane. Figure 1 shows one interpretation of what the simulation area might look like.

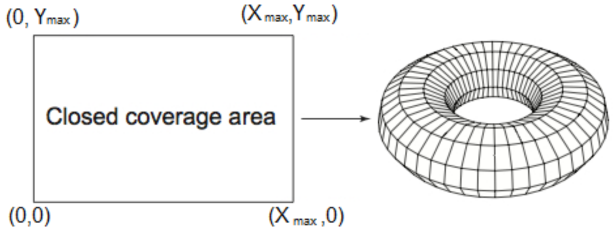


Figure 1. Simulation area of Boundless Simulation Area Mobility Model shown as a tube [18]

In our random generator, we used this method to avoid the border effect without affecting the randomness of node movement. Since the ns2 tool is based on TCL [26] object-oriented programming language, we decided to use Java programming language to build the new random generator. The new random generator generates a file readable by ns2 that has the definitions of all the movements of all the nodes during the simulation time. The Java code firstly generates the initial random positions of the nodes. The x,y and z coordinates are given in ns2 TCL format, for example:

```
$node_(0) set X_ 323.81544267544473
$node_(0) set Y_ 576.9394231828528
$node_(0) set Z_ 0.000000000000
```

The movements' commands can be for either a node that does not cross the border(s) or a node that does cross the border. The nodes that do not go across the border(s) will have just one movement command as in the following example:

```
  $ns_ at 63.904477062 "$node_(0)
        setdest 118.025456608
        271.953011717 9.048003190"
```

On the other hand, the nodes that go across the border(s) will have three movement commands. The following is an example of the generated three commands by the custom random generator which represent the movement that has a jump from one border to the other through the movement.

```
  $ns_ at 102.779352050 "$node_(0)
setdest
        247.578795505 0.000000001
        12.273796854"
$ns_ at 113.299385328 "$node_(0) setpos
        247.578795505 999.999999999"
$ns_ at 113.299385328 "$node_(0) setdest
        247.984526589 998.357412892
        12.273796854"
```

The first line represents the movement from the current position to the border, the second line represents the jump from one border to the opposite border and the last line represents the movement from the border after the jumping to the destination.

The speed of movement is the same in the first and the third line which are making the movement to be exactly the same before the jump and after it. Figure 2 shows the three movements that happen once the node across the border(s) generated by the custom random generator. Algorithm 1 describes in details the algorithm used in the new custom movement generator.

We modified ns2 adding a new setpos command (analogous to setdest) that allows for placing a node at specified location during the simulation.

V. RESULTS

We investigated the randomness performance and the packet transmissions of both the setdest utility and the custom generator. The randomness performance of both generators is shown in Figure 3.

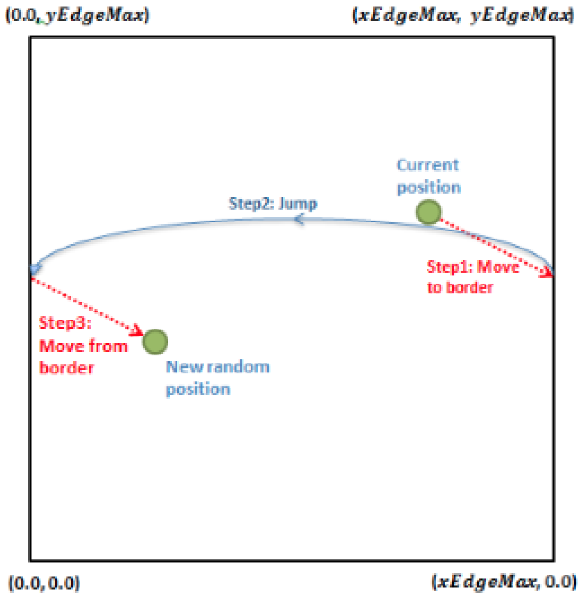


Figure 2. Three steps those nodes have to follow once they reach boundary

In this figure, the VMR values represent the randomness in the distribution of mobile nodes at different times in the simulation. The results obtained with the custom generator are closer to the VMR value of one than the values obtained from using the setdest utility. Taking in to account the main deficiency of the setdest utility, which is the tendency of placing the mobile nodes in the center of the simulation area we also investigated the packet transmissions in a simulated mobile networks controlled by both generators in two ways: we measured transmitting the packets through the center and along the edges of the simulated area. Once the packets were transmitted through the center, there was no significant difference between the movement generators while using the 60 nodes, but with the 30 and 20 nodes there were significant differences with the setdest utility showing more packets delivered throught the center, suggesting some clustering of nodes in the center (Table I). Transmitting the packets along the edges (results are shown in Table II), we observed a significant difference between the two generators in the experiments with 30 and 60 nodes.

For example, when transmitting through the centre in the experiments with 30 nodes we observed on average 7453 packets delivered in the scenarios generated with setdest utility vs 5600 packets in the case when our custom generator was used. For the same scenarios with 30 nodes, transmissions along the edges gave 716 vs 2031 packets received. This shows a strong bias for packet delivery in the centre of the experimental area for the setdest utility: 7453 in the center vs 716 along the edges, almost 10 to 1 ratio. Similar comparison for the new generator gives only 2 to 1 ratio (5600 vs 2031), which is close to what would be expected based on the model described in [27].

Overall the simulation with the custom generator delivered

Algorithm 1 Pseudocode (Custom Generator)

```
(maxX, maxY) = simulation area dimensions
maxTime = max duration of node's movement
maxMovement = max distance of node's movement
currentTime = 0
(currentX, currentY) = random position

while Simulating do
    angle = rand(0..1) * 360°
    destX = rand(0..1) * cos(angle) * maxMovement //set the movement destination
    destY = rand(0..1) * sin(angle) * maxMovement
    MovementTime = rand(0..1) * maxTime
    speed = geometric distance (currentX, currentY) to (destX, destY)/movementTime

    if path does not cross the boundary then
        writeMovement ($ns_ at current Time "$nodeX setdest destX, destY, speed")
    else
        if path crosses the boundary then
            (interceptX, interceptY) = boundary intercept
            distToBound = geometric distance to the boundary in the direction of the movement path
            writeMovement ($ns_ at currentTime "$nodeX setdest interceptX, interceptY, speed")
            timeAtBoundary = currentTime + distToBound/speed
            writeMovement ($ns_ at timeAtBoundary "$nodeX setpos interceptX, interceptY")
            destX = destX % maxX
            destY = destX % maxY
            writeMovement ($ns_ at timeAtBoundary "$nodeX setdest destX, destY, speed")
            (currentX, currentY) = (destX, destY)
            currentTime = currentTime + movementTime
        end if
    end if
end while
```

packets better and more uniformly than the setdest utility.

Table I
AVERAGE PACKET DELIVERY FOR TRANSMITTING THROUGH THE CENTER (WITH 95% CONFIDENCE INTERVAL)

	60 Nodes	30 Nodes	20 Nodes
Setdest Utility	9595 (9459 - 9731)	7453 (6724 - 8182)	3878 (2945 - 4811)
Custom Generator	9490 (9371 - 9610)	5600 (5150 - 6050)	1886 (1325 - 2448)

Table II
AVERAGE PACKET DELIVERY FOR TRANSMITTING ALONG THE EDGES (WITH 95% CONFIDENCE INTERVAL)

	60 Nodes	30 Nodes	20 Nodes
Setdest Utility	4233 (3664 - 4802)	716 (478 - 955)	384 (193 - 576)
Custom Generator	8085 (7933 - 8238)	2031 (1712 - 2351)	332 (197 - 468)

VI. CONCLUSION

We investigated the performance of the popular setdest utility used in the ns2 network simulator. The movement generated with setdest utility tends to cluster the mobile nodes in the center of the experimental area. This has an effect on the VMR coefficient used to measure randomness of the positions of points in the simulation. We proposed and demonstrated the advantage of new random motion generator for use in ns2 simulator. Testing the new generator shows a marked advantage over the standard setdest utility and should improve the quality of MANET simulation models when randomness of node movement is required.

ACKNOWLEDGMENT

This work is sponsored and funded by the Ministry of Higher Education of Saudi Arabia through the Saudi Arabian Cultural Bureau in Canada.

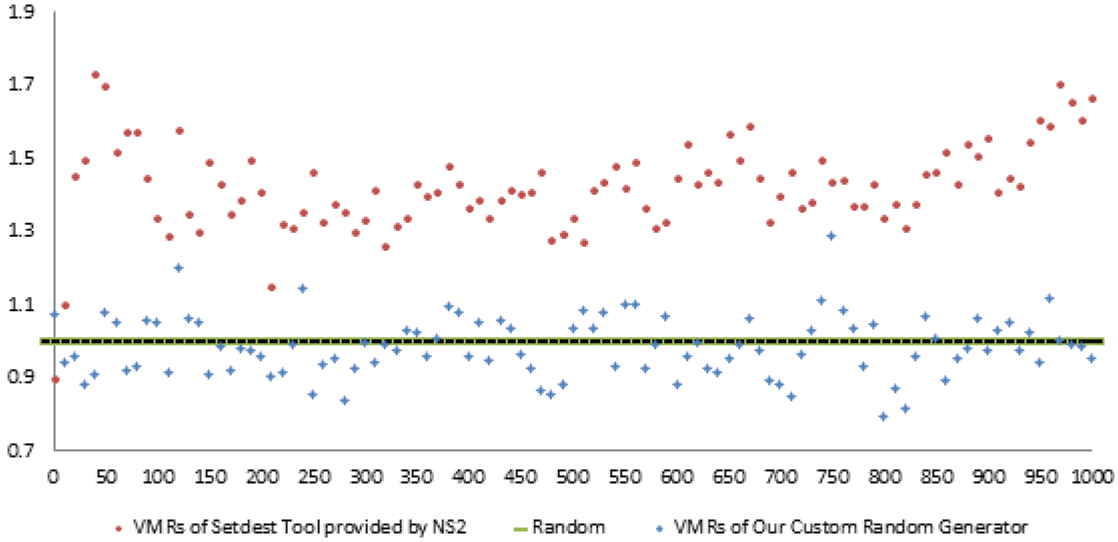


Figure 3. Comparison of the 101 means (of five VMR runs) of the setdest utility and the custom generator

References

[1] Wikipedia, the free encyclopedia (January 2012), Mobile ad hoc network, http://en.wikipedia.org/wiki/Mobile_ad_hoc_networks (accessed November 2012)

[2] R. Suprio, Realistic mobility for MANET simulation, Masters thesis, The University of British Columbia, 2003.

[3] Wikipedia, the free encyclopedia (November 2009), Basic Station, http://en.wikipedia.org/wiki/Base_station (accessed October 2012)

[4] Z. J. Haas, "A new routing protocol for the reconfigurable wireless networks," in 1997 IEEE 6th International Conference on Universal Personal Communications Record, San Diego, CA, USA, 12-16 Oct. 1997, vol.2, pp. 562-566.

[5] Wikipedia, the free encyclopedia (November 2010), Wireless Ad-hoc Network, http://en.wikipedia.org/wiki/Wireless_ad-hoc_network (accessed December 2012)

[6] N. S. Dalton and R. C. Dalton, "The theory of natural movement and its application to the simulation of mobile ad hoc networks (MANET)", in Proc. the Fifth Annual Conference on Communication Networks and Services Research 2007, Fredericton, N.B., pp. 359-363.

[7] Y. Zhang and W. Li, "An integrated environment for testing mobile ad-hoc networks", in Proc. the Third ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '02), Lausanne, Switzerland, June 2002, pp. 104-111.

[8] V. Davies, Evaluating Mobility Models Within an Ad Hoc Network, Masters thesis, Colorado School of Mines, Colorado, 2000.

[9] T. Camp, J. Boleng, and V. Davies, "A survey of mobility models for ad hoc network research", Wireless Communications & Mobile Computing, Special Issue on Mobile Ad Hoc Networking: Research, Trends and Applications, vol. 2, no. 5, pp. 483-502, 2002.

[10] B. J. Culpepper and H. C. Tseng, "Sinkhole intrusion indicators in DSR MANETs, in Proc. First International Conference on Broadband Networks 2004, San Jose, CA, USA, 25-29 October 2004, pp. 681-688.

[11] N. Frangiadakis, M. Kyriakakos, S. Hadjiefthymiades, and L. Merakos, "Realistic mobility pattern generator: design and application in path prediction algorithm evaluation," in Proc. The 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications 2002, Lisboa, Portugal, 15-18 Sept. 2002, vol.2, pp. 765-769.

[12] P. Caballero-Gil, C. Caballero-Gil, J. Molina-Gil, and C. Hernandez-Goya, "A simulation study of new security schemes in mobile Ad-hoc NETworks", in EUROCAST 2007, LNCS, vol. 4739, R. Moreno Daz, F. Pichler, A. Quesada Arencibia, Ed., Heidelberg: Springer, 2007, pp. 73-81.

[13] J. Hu and R. Marculescu, "DyAD: smart routing for networks-on-chip", in Proc. the 41st annual conference on Design automation, San Diego, CA, USA, June 07-11, 2004, pp. 260-263.

[14] A. L. Cavilla, MANET extensions to ns2. Available: http://www.cs.toronto.edu/andreslc/software/MANET extensions.tgz (accessed November 2012)

[15] B. Carbone, Routing Protocols for Interconnecting Cellular and Ad Hoc Networks, Masters thesis, Universite Libre De Bruxeles, Faculte des Sciences, Department d'Informatique, 2006.

[16] J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan, Statistical analysis and modelling of spatial point patterns. West Sussex, England: John Wiley & Sons Ltd, 2011.

[17] R. R. Roy, Handbook of mobile ad hoc networks for mobility models (1st Ed.), New York, NY: Springer Science Business Media, 2011.

[18] J. Ariyakhajorn, P. Wannawilai, and C. Sathitwiriyawong, ”A Comparative Study of Random Waypoint and Gauss-Markov Mobility Models in the Performance Evaluation of MANET,” International Symposium on Communications and Information Technologies, 2006, ISCIT '06, Bangkok, Thailand, , Oct. 18 2006 - Sept. 20 2006, pp. 894-899.

[19] B. Liang and Z. J. Haas, ”Predictive distance-based mobility management for PCS networks,” Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies INFOCOM '99, New York, NY, USA, 21-25 Mar 1999, vol.3, pp. 1377-1384.

[20] C. Chiang, Wireless Network Multicasting. Ph.D. thesis, University of California, Los Angeles, 1998.

[21] F. Bai and A. Helmy, ”A Survey of Mobility Models in Wireless Ad hoc Networks”, in Wireless Ad Hoc and Sensor Networks, Chapter 1, Kluwer Academic Publishers, June 2004, pp. 1-29.

[22] M. Pascoe, J. Gomez, V. Rangel, and M. Lopez-Guerrero, ”Route duration modeling for mobile ad-hoc networks”, ACM Wireless Networks Journal (WiNet), 16(3), 2010, pp. 743-757.

[23] Wikipedia, the free encyclopedia (June 2012), Constant bitrate, http://en.wikipedia.org/wiki/Constant_bitrate (accessed December 2012)

[24] Wikipedia, the free encyclopedia (May 2012), User Datagram Protocol, http://en.wikipedia.org/wiki/User_Datagram_Protocol (accessed December 2012)

[25] Wikipedia, the free encyclopedia (December 2011), Ad hoc On-Demand Distance Vector Routing, http://en.wikipedia.org/wiki/Ad_hoc_On-Demand_Distance_Vector_Routing (accessed December 2012)

[26] Wikipedia, the free encyclopedia (December 2011), Tcl, http://en.wikipedia.org/wiki/Tcl (accessed October 2012)

[27] C. Bettstetter and J. Zangl, ”How to achieve a connected ad hoc network with homogeneous range assignment: an analytical study with consideration of border effects”, 4th International Workshop on Mobile and Wireless Communications Network, Stockholm, Sweden, Sept. 2002, pp. 125-129.

A Numerical Approach for Performance Evaluation of Cellular Mobile Networks with Channels Breakdowns

Nawel Gharbi
Computer Science Department
University of Sciences and Technology, USTHB
Algiers, Algeria
Email: ngharbi@wissal.dz

***Abstract*—In this paper, we propose a numerical approach to study performance and reliability of cellular mobile networks, taking into account the repeated attempts of users whose call was refused due to the lack of available resources and random breakdowns of the base station channels, using the Generalized Stochastic Petri Nets (GSPNs) model as a support. In fact, one of the major drawbacks of this high-level formalism in performance evaluation of large networks is the state space explosion problem. Hence, the novelty of this investigation is the presentation of an approach which allows a direct computing of the infinitesimal generator describing the users behavior and channels allocation, without the generation of the reachability graph nor its reduction. In addition, we develop the formulas of the main stationary reliability and performance indices as a function of the network parameters, the stationary probabilities and independently of the reachability set markings.**

Keywords-Cellular Mobile Networks; Repeated calls; Channels breakdowns; Infinitesimal generator; Performance Evaluation.

I. Introduction

Modeling and performance evaluation are essential for design of cellular mobile networks, where, the number of users and the need for higher data rates and multimedia services increase more and more. Hence, the study of users behavior and in particular, the repeated attempts (called repeated calls or retrials) of users whose call was refused due to the lack of available resources and the consideration of random breakdowns of the base station channels, are crucial to determine the network performance because they can have quite a negative impact on the quality of service offered to users and should therefore not be neglected in network design and planning.

The modeling of repeated attempts has been a subject of numerous investigations dealing with the performance analysis of switching systems, communication networks, cellular mobile networks [1], [2], [3] and wireless sensor networks [4].

In modern cellular networks, micro cells are under consideration. These small cells operate in licensed and unlicensed spectrum that have a range of 10 to 200 meters, compared to macrocells which might have a range of a few kilometers. Hence, the cell size gets smaller, and thus the number of users served in a cell will be relatively smaller, such that traffic models with a finite source of users should be considered.

This paper aims at presenting a numerical approach for performance and reliability evaluation of cellular mobile networks, where the supported area is divided into micro cells, each of them contains a finite number of users and is served by a base station having a limited number of channels which could be subject to random breakdowns. We focus specially on the effect of repeated calls of blocked users and channels breakdowns on the network performances.

Although the reliability study is of great importance, there are only few works that take into consideration repeated calls involving the unreliability of the servers and the finite source of users, as it can be seen in the recent classified bibliography of Artalejo [5]. Moreover, most studies deal with single unreliable server retrial queues or an infinite customers source [6], [7], [8], [9], [10]. However, papers treating finite-source retrial systems with multiple unreliable servers are fewer. The unreliable heterogeneous servers case was considered by Sztrik [11] using a retrial queueing model. On the other hand, retrial mobile networks with several homogeneous servers subject to breakdowns were modeled and analyzed by means of Generalized Stochastic Petri nets (GSPNs) in the recent paper of Gharbi [3].

From a modeling point of view, and compared to retrial queues, the generalized stochastic Petri nets (GSPNs) high level formalism allows an easier description of the behavior of complex systems. This is particularly true for mobile networks. However, the model analysis requires the generation of the reachability graph and then its reduction to obtain the corresponding Markov chain. These two steps require a large storage space and a long execution time. Moreover, the state space increases exponentially as function of the users source size and the base station channels number. So, for practical cellular networks, the models may have a huge state space. Hence, the novelty of this investigation is the presentation of an approach to deal with this problem. To this end, we develop an algorithm for automatically calculating the Markov chain infinitesimal generator, without the generation of the reachability graph nor its reduction. In that way, the storing of the entire state space is avoided. In addition, we develop the formulas of the main stationary performance and reliability indices, as a function of the base station channels number, the users source size, the stationary probabilities and independently of the reachability set markings.

The paper is organized as follows: First, we give the syntax and semantics of GSPNs formalism. Next, the GSPN model describing a network cell with repeated calls and base station channels breakdowns is developed in Section 3. Then, the proposed analysis approach is detailed. In Section 5, the computational formulas for evaluating exact performance indices of these networks are derived. Next, based on some experimental examples, we validate our approach in the reliable case and we illustrate the effect of retrial rate and base station channels number on the mean response time. Finally, we give a conclusion.

II. Syntax and semantics of Generalized Stochastic Petri nets

Generalized stochastic Petri nets (GSPNs) [12], [13] are a powerful mathematical and graphical formalism, well suited for modeling and evaluating the performances of stochastic systems involving concurrency, nondeterminism and synchronization. In the past decade, GSPNs have received much attention from researchers, and have been extensively used for analytical modeling of performance and performability of computer, communication, manufacturing and aerospace systems.

Formally, a GSPN can be defined as an eight-tuple $(P, T, \alpha, I, O, H, W, M_0)$ where P is the set of places, T is the set of transitions such that $T \cap P = \emptyset$, it consists of timed and immediate transitions, $\alpha : T \rightarrow \{0, 1\}$ is the priority function which associates the priority $\alpha(t) = 1$ to immediate transitions and $\alpha(t) = 0$ to timed transitions, I, O, $H : T \rightarrow Bag(P)$ are the input, output and inhibition functions, respectively, where $Bag(P)$ is the multiset on P, $W : T \rightarrow R^+$ is a function that assigns a rate of negative exponential distribution to each timed transition and a weight to each immediate transition, $M_0 : P \rightarrow IN$ is the initial marking, a function that assigns a nonnegative integer value to each place, and describes the initial state of the system.

In the graphical representation, places are represented by circles, timed transitions by boxes (or rectangles) and immediate transitions by thin bars. Arcs, leading from places to transitions (from transitions to places resp.) describe the input (the output resp.) function and the arcs, denoting the inhibition function are circle-headed. Arcs are labeled with an integer $d \geq 1$ called the multiplicity of the arc, a value of 1 is usually omitted for readability.

The system state is described by means of markings. The marking of a place is the number of tokens which the place contains. A marking of a Petri net is a mapping $M : P \rightarrow IN$, which specifies the number of tokens in each place of the net. The dynamic behavior of a GSPN results from the firing of transitions yielding other markings than M_0.

A transition t is enabled in a marking M iff each of its ordinary input places contains at least as many tokens as the multiplicity of the input arc, and each of its inhibitor input places contains fewer tokens than the multiplicity of the corresponding inhibitor arc. One more condition for timed transitions is that no immediate transition is enabled simultaneously in M because immediate transitions have priority over timed transitions. Moreover, an enabled timed transition t fires after a delay which is exponentially distributed with rate $W(t)$ while an enabled immediate transition t fires in zero time. In case of conflicts between immediate transitions in a marking M, a given transition t fires with probability $W(t)/\sum_{t':M[t'>} W(t')$. On the other hand, a timed transition has a single-server, n-servers or ∞-servers semantics. For the single-server semantics, the firing rate of a transition t equals its rate $W(t)$, however, for the infinite-servers semantics, the firing rate of transition t in marking M is marking dependent and so equals $W(t) \cdot ED(t, M)$, where $ED(t, M)$ is the enabling degree of t in the marking M. The condition of marking dependent firing is represented by the symbol # placed next to transition.

The firing of any enabled timed or immediate transition t from a marking M, produces a new marking $M' = M - I(t) + O(t)$. All markings created due to the firing of transitions are called reachable and the *reachability graph* is obtained by representing each marking by a vertex and placing a directed edge from vertex M_i to vertex M_j, if marking M_j can be obtained by the firing of some transition enabled in marking M_i. In the reachability graph, markings enabling no immediate transitions are called *tangible markings*. In this case, one of the enabled timed transitions can fire next (application of race policy commonly). Markings in which at least one immediate transition is enabled, are called *vanishing markings* and are passed through in zero time. Since the process spends zero time in the vanishing markings, they don't contribute to the dynamic behavior of the system, so, they are eliminated from the reachability graph by merging them with their successor tangible markings. This reduction process which corresponds to the elimination of vanishing markings results in a *tangible reachability graph*, which is isomorphic to a continuous time Markov chain (CTMC). The states of the CTMC are the markings in the tangible reachability graph, and the state transition rates are the exponential firing rates of timed transitions in the GSPN.

The solution of this CTMC at steady-state is the stationary probability vector π which can be expressed as the solution of the linear system of equilibrium equations $\pi.Q = 0$ with the normalization condition $\sum_i \pi_i = 1$, where π_i denotes the steady-state probability that the process is in state M_i and Q is the infinitesimal generator. Having the probabilities vector π, we can compute several stationary performance indices of the system.

III. GSPN MODEL OF CELLULAR MOBILE NETWORKS WITH REPEATED CALLS AND CHANNELS BREAKDOWNS

We observe a cellular mobile network where a supported area is divided into small cells, with a finite source of users (mobiles) of size N in each cell and a base station that consists of c $(c \geq 1)$ identical and parallel channels subject to breakdowns and repairs. Each user is either free, under service or in orbit at any time. Each channel can be in operational (up) or non-operational (down) state, and it can be idle or busy (on service). User requests are assigned to operational idle channels randomly and without any priority order. If one of the channels is *up and idle* at the moment of the arrival of a call, then the user starts being served immediately. Service times are independent identically-distributed random variables, whose distribution is exponential. After service completion, the channel becomes idle. Otherwise, if all channels are busy or down at the arrival of a request, the user joins the orbit.

In Fig. 1, we present the GSPN model describing the users behavior and the channels allocation. In this model, the place Cus_Free contains the free users, place $Choice$ represents the arrival of a primary or a repeated call for service and place Ser_Idle represents the operational idle channels. Initially, it contains c tokens because all channels are up and available. Place Cus_Serv contains the users in service. Place $Orbit$ represents the orbit and place Ser_Down contains the failed channels. Hence, the initial marking of the net is given by:

$$\begin{cases} M_0(Cus\_Free) = N \\ M_0(Ser\_Idle) = c \\ M_0(p) = 0 \quad \forall p \in P,\ p \notin \{Cus\_Free, Ser\_Idle\}. \end{cases}$$

The firing of transition $Arrival$ indicates the arrival of a primary request. The service semantics of this transition is ∞-servers (represented by symbol $\#$) because free users can independently generate primary calls. Hence, the firing rate depends on the marking of place Cus_Free and is equal to $\lambda.M(Cus\_Free)$.

At the arrival of a primary or repeated call to place $Choice$, if place Ser_Idle contains at least one token, i.e., if there is at least one idle operational channel, immediate transition $Begin\_Serv$ fires. Hence, the user starts being served and the channel moves into busy state. Otherwise, if place Ser_Idle is empty, immediate transition Go_Orbit fires and the user immediately joins place $Orbit$ and starts generating a flow of repeated calls with rate ν, until it finds an operational idle channel. In fact, users in orbit behave independently of each other and are persistent in the sense that they keep making retrials until they receive their requested service, after which they have no further effect on the network. The firing of transition $Retrial$ represents the arrival of a repeated call. As users independently generate repeated calls, this transition has an ∞-server semantics.

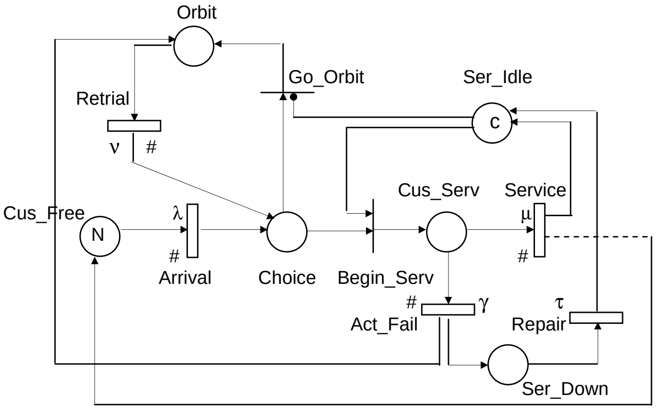


Figure 1. GSPN model of small cell networks with retrials and channels breakdowns

At the end of a service period, timed transition $Service$ fires. The users under service returns to free state (to place Cus_Free) and the channel becomes idle and ready to serve another user. As services take place in parallel, transition $Service$ has an ∞-servers semantics.

If a channel fails during a service period, which is represented by the firing of timed transition Act_Fail, the interrupted user joins the orbit and will restart service later, while the failed channel joins place Ser_Down, where it will be repaired. The firing of transition $Repair$ represents the end of the repair time which is exponentially distributed with rate τ, and the fact that the repaired channel returns to the operational idle state (to the place Ser_Idle). The repairman repairs one channel at a time. Thus, the service semantics of transition $Repair$ is single-server semantics. This means that the firing rate is constant.

IV. STOCHASTIC ANALYSIS

When modeling real cellular mobile networks, generating the GSPN reachability graph and then the reduced underlying Markov chain, may require a huge storage space and a long execution time, since the state space increases exponentially as a function of the users source size and channels number. To overcome this problem, this paper aims to avoid these two steps by designing an algorithm that compute directly the Markov chain infinitesimal generator as a function of system parameters and without generating neither the reachability graph nor the reduced Markov chain. In that way, the complete storing of the reachability set is avoided.

In the following, we describe in detail, the applied steps to derive the corresponding algorithm.

Whatever the values of N and c (with $c < N$), the conservation of users and channels gives the following equations:

$$\left\{\begin{array}{l} M(Cus\_Free) + M(Cus\_Serv) + M(Orbit) = N \\ M(Ser\_Idle) + M(Cus\_Serv) + M(Ser\_Down) = c \end{array}\right. \quad (1)$$

Observing these two equations, we note that the system state at steady-state can be described by means of three variables (i, j, k), where:

- i represents the number of users in service (in place Cus_Serv);
- j is the number of users in orbit (in place $Orbit$);
- and k is the number of failed channels (in place Ser_Down).

Hence, having (i, j, k), the markings of all places can be obtained. On the other hand, applying (1), we can deduce:

$$\left\{\begin{array}{l} 0 \leq i + j \leq N \\ 0 \leq i + k \leq c \end{array}\right. \quad (2)$$

The behavior of the system can be described by a CTMC, whose infinitesimal generator is an $R \times R$ matrix Q. When there are i users in service, the remaining $N - i$ users must be dispatched between places Cus_Free and $Orbit$, and the remaining $c - i$ channels are idle or down. However, when active breakdowns are considered, state $(0, 0, c)$ where all users are free and all channels are down is not reachable, because channels can fail only in busy state. But the model with (in)dependent breakdowns includes this state. Hence, the number R of accessible tangible markings equals: $R = [\sum_{i=0}^{c}(N - i + 1).(c - i + 1)] - 1$, which can be rewritten as: $R = [\sum_{i=1}^{c+1}(N - c + i) * i] - 1$.

The infinitesimal generator Q is constructed as follows:

$$Q[(i,j,k),(x,y,z)] = \left\{\begin{array}{l} \theta[(i,j,k),(x,y,z)] \\ \text{if } (i,j,k) \neq (x,y,z), \\ -\sum_{(x,y,z)\neq(i,j,k)} \theta[(i,j,k),(x,y,z)] \\ \text{if } (i,j,k) = (x,y,z). \end{array}\right.$$

where $\theta[(i, j, k), (x, y, z)]$ is the transition rate from state (i, j, k) to state (x, y, z). By analyzing the firings of the GSPN transitions, we obtain the following rates:

- $[k > 0] : (i, j, k) \xrightarrow{\tau} (i, j, k - 1)$
- $[i > 0] : (i, j, k) \xrightarrow{i\mu} (i - 1, j, k)$ and $(i, j, k) \xrightarrow{i\gamma} (i - 1, j + 1, k + 1)$
- $[j > 0$ and $i + k < c] : (i, j, k) \xrightarrow{j\nu} (i + 1, j - 1, k)$
- $[i + j < N$ and $i + k < c] : (i, j, k) \xrightarrow{(N-i-j).\lambda} (i + 1, j, k)$
- $[i + j < N$ and $i + k = c] : (i, j, k) \xrightarrow{(N-i-j).\lambda} (i, j + 1, k)$

As a consequence, the infinitesimal generator can be automatically calculated by means of Algorithm 1. In this case, when dealing with line 7, the case where $i + j = 0$ should not be considered, as the state where all users are free and all channels are down does not exist. The same holds for line 29 when $i = j = 0$ and $k = c$.

Algorithm 1 Computation of the infinitesimal generator

```
                                          ▷ Primary arrivals : $i + j < N$
 1: for $i \leftarrow 0, c - 1$ do
 2:     for $j \leftarrow 0, N - i - 1$ do
 3:         for $k \leftarrow 0, c - i - 1$ do
 4:                                          ▷ admission in service
 5:             $\theta[(i,j,k),(i+1,j,k)] \leftarrow (N - i - j).\lambda$
 6:         end for
 7:                                          ▷ admission in orbit
 8:         $\theta[(i,j,c-i),(i,j+1,c-i)] \leftarrow (N - i - j)\lambda$
 9:     end for
10: end for

                           ▷ Successful retrials : $j > 0$ and $i + k < c$
11: for $i \leftarrow 0, c - 1$ do
12:     for $j \leftarrow 1, N - i$ do
13:         $\theta[(i,j,k),(i+1,j-1,k)] \leftarrow j.\nu$
14:     end for
15: end for

              ▷ End of service and channel breakdown : $i > 0$
16: for $i \leftarrow 1, c$ do
17:     for $j \leftarrow 0, N - i$ do
18:         for $k \leftarrow 0, c - i$ do
19:                                          ▷ end of service
20:             $\theta[(i,j,k),(i-1,j,k)] \leftarrow i.\mu$
21:                                          ▷ channel breakdown
22:             $\theta[(i,j,k),(i-1,j+1,k+1)] \leftarrow i.\gamma$
23:         end for
24:     end for
25: end for
                                          ▷ Repairs : $k > 0$
26: for $i \leftarrow 0, c - 1$ do
27:     for $j \leftarrow 0, N - i$ do
28:         for $k \leftarrow 1, c - i$ do
29:             $\theta[(i,j,k),(i,j,k-1)] \leftarrow \tau$
30:         end for
31:     end for
32: end for
```

V. Performance and reliability indices

The aim of this section is to derive the formulas of the most important stationary performance and reliability indices. As, the proposed models are bounded and the initial marking is a home state, the underlying process is ergodic. Hence, the steady-state solution exists and is unique.

The infinitesimal generator Q can be obtained automatically by applying the above algorithms. Then, the steady-state probability vector π can be computed by solving the linear equation system:

$$\left\{ \begin{array}{l} \pi.Q = 0 \\ \sum_{(i,j,k)} \pi_{i,j,k} = 1, \\ \text{where } (i,j,k) \text{ satisfy the conditions given in (2).} \end{array} \right.$$

Having the probability distribution π, we can derive several exact performance and reliability measures. Although state $(0, 0, c)$ is not reachable, we consider it in order to have an homogeneous presentation of formulas. In this case, we assign it a null probability.

- Mean number of busy channels (n_s): This corresponds to the mean number of tokens in place Cus_Serv which is also the mean number of customers under service.

$$n_s = \sum_{i=0}^{c} \sum_{j=0}^{N-i} \sum_{k=0}^{c-i} i.\pi_{i,j,k} \qquad (3)$$

- Mean number of users in orbit (n_o): This corresponds to the mean number of tokens in place $Orbit$.

$$n_o = \sum_{i=0}^{c} \sum_{j=0}^{N-i} \sum_{k=0}^{c-i} j.\pi_{i,j,k} \qquad (4)$$

- Mean number of users in the system (n):

$$n = n_s + n_o = \sum_{i=0}^{c} \sum_{j=0}^{N-i} \sum_{k=0}^{c-i} (i + j).\pi_{i,j,k} \qquad (5)$$

- Mean number of failed channels (n_f): This represents the mean number of tokens in place Ser_Down.

$$n_f = \sum_{i=0}^{c} \sum_{j=0}^{N-i} \sum_{k=0}^{c-i} k.\pi_{i,j,k} \qquad (6)$$

- Mean number of operational idle channels (n_i): This represents the average number of tokens in place Ser_Idle.

$$n_i = c - (n_s + n_f) = c - \sum_{i=0}^{c} \sum_{j=0}^{N-i} \sum_{k=0}^{c-i} (i + k).\pi_{i,j,k} \qquad (7)$$

- Mean rate of generation of primary calls $(\overline{\lambda})$: This represents the throughput of transition $Arrival$, which equals the throughput of transition $Service$.

$$\overline{\lambda} = (N - n).\lambda = \sum_{i=0}^{c} \sum_{j=0}^{N-i} \sum_{k=0}^{c-i} (N - i - j).\lambda.\pi_{i,j,k} \qquad (8)$$

- Mean rate of service $(\overline{\mu})$: This represents the throughput of transition $Service$.

$$\overline{\mu} = \mu.n_s = \overline{\lambda}$$

- Mean rate of generation of repeated calls $(\overline{\nu})$: This represents the retrial frequency of customers in orbit. It corresponds to the throughput of transition $Retrial$.

$$\overline{\nu} = \sum_{i=0}^{c} \sum_{j=1}^{N-i} \sum_{k=0}^{c-i} j.\nu.\pi_{i,j,k} = \nu.n_o$$

- Failure frequency of busy channels $(\overline{\gamma})$: This represents the throughput of transition Act_Fail.

$$\overline{\gamma} = \sum_{i=1}^{c} \sum_{j=0}^{N-i} \sum_{k=0}^{c-i} i.\gamma.\pi_{i,j,k} = \gamma.n_s$$

- Failure frequency of idle channels $(\overline{\delta})$: This represents the throughput of transition $Idle\_Fail$.

$$\overline{\delta} = \sum_{i=0}^{c-1} \sum_{j=0}^{N-i} \sum_{k=0}^{c-i-1} (c - i - k).\delta.\pi_{i,j,k} = \delta.n_i$$

- Blocking probability of a primary call (B_p):

$$B_p = \frac{\sum_{i=0}^{c} \sum_{j=0}^{N-i-1} (N - i - j).\lambda.\pi_{i,j,c-i}}{\overline{\lambda}}$$

- Blocking probability of a repeated call (B_r):

$$B_r = \frac{\sum_{i=0}^{c} \sum_{j=1}^{N-i} j.\nu.\pi_{i,j,c-i}}{\overline{\nu}}$$

- Blocking probability (B):

$$B = \frac{\overline{\lambda}}{\overline{\lambda} + \overline{\nu}} \ . \ B_p + \frac{\overline{\nu}}{\overline{\lambda} + \overline{\nu}} \ . \ B_r$$

- Mean rate of repair $(\overline{\tau})$: This represents the throughput of transition $Repair$.

$$\begin{array}{rcl} \overline{\tau} & = & \tau. \sum_{i=0}^{c-1} \sum_{j=0}^{N-i} \sum_{k=1}^{c-i} \pi_{i,j,k} \\ & = & \left\{ \begin{array}{ll} \overline{\gamma}, & \text{in active breakdowns,} \\ \overline{\gamma} + \overline{\delta}, & \text{in dependent breakdowns.} \end{array} \right. \end{array}$$

- Utilization of s channels (U_s): $(0 \leq s \leq c)$ This corresponds to the probability that s channels are busy :

$$U_s = \sum_{j=0}^{N-s} \sum_{k=0}^{c-s} \pi_{s,j,k}$$

- Availability of s channels (A_s): $(0 \leq s \leq c)$ This corresponds to the probability that s channels are operational and idle.

$$A_s = \sum_{i=0}^{c-s} \sum_{j=0}^{N-i} \pi_{i,j,c-s-i}$$

Table I
VALIDATION IN THE RELIABLE CASE

	Reliable [14]	Non-reliable
Number of channels	4	4
Number of users	20	20
Primary call generation rate	0.1	0.1
Service rate	1	1
Retrial rate	1.2	1.2
channel's failure rate	-	1e-25
channel's repair rate	-	1e+25
Mean number of busy channels	1.800748	1.800764
Mean number of sources of repeated calls	0.191771	0.191786
Mean rate of generation of primary calls	1.800748	1.800745
Mean waiting time	0.106495	0.1065036

- `Failure probability of` s `channels (`F_s`):` $(0 \leq s \leq c)$ This corresponds to the probability that s channels are failed:

$$F_s = \sum_{i=0}^{c-s} \sum_{j=0}^{N-i} \pi_{i,j,s}$$

- `Utilization of the repairman (`U_r`):` This corresponds to the probability that at least one channel is failed:

$$U_r = \overline{\tau}/\tau$$

- `Mean response time (`$\overline{R}$`):` The mean response time is defined as the mean time from the instant a customer generates a primary request until it is served. In the steady state, it can be obtained using Little's formula:

$$\overline{R} = \frac{n_o + n_s}{\overline{\lambda}}$$

- `Mean waiting time (`$\overline{W}$`):`

$$\overline{W} = \frac{n_o}{\overline{\lambda}} = \overline{R} - \frac{1}{\mu}$$

VI. EXPERIMENTAL EXAMPLES

In order to test the feasibility of our approach, we developed a tool to implement the above algorithm and the performance indices formulas. Hence, we tested it for a large number of examples. In particular, the results obtained in the reliable case were compared to those generated by the program given in the book of Falin and Templeton [14] for analysis of finite-source retrial queues with reliable servers, since if the failure rate in non-reliable models is very low and repair rate is very high, the performance parameters should approach the corresponding ones in reliable models. From table I, we can see that the results of the proposed approach are very close to those obtained in the reliable case.

Next, we illustrate the effect of retrial rate and base station channels number on the mean response time. The results are presented in Figure 2 and Figure 3, respectively, where the two curves correspond to the reliable case and non-reliable one. From these two figures, we see that the mean response time is a decreasing function of retrial rate and channels number. Moreover, the reliable model gives the best mean response times in the two cases.

Figure 2 also shows that the retrial rate has a significant influence on the mean response time for low retrial rate values. However, when more and more repeated requests arrive, the decrease is not considerable in the case of channels breakdowns.

Figure 3 shows that a small change in the number of base station channels, particularly from 1 to 3 channels, produces a big difference in the mean response time ($\approx -61\%$ for the unreliable channels). However, after a certain value ($c = 4$), the decrease is not considerable.

VII. CONCLUSION AND FUTURE WORK

This paper aims at presenting an approach that allows performance evaluation of cellular mobile networks, taking into account the repeated calls of blocked customers, the finite number of customers served in a cell and the breakdowns of base station channels. The flexibility of GSPNs modeling approach allowed us a simple construction of detailed and compact models for these systems. The models are used as a support to derive the balance equations of the networks, so that the infinitesimal generator can be obtained without building the reachability graph of the model nor reducing it. Exact stationary performance and reliability parameters can then be computed.

In conclusion, the GSPNs method holds promise for the solution of several systems with repeated attempts. Hence, it is worth noting that our approach can be further extended to more complex systems with different breakdowns disciplines.

REFERENCES

[1] Artalejo, J.R. and Lopez-Herrero, M.J.: Cellular mobile networks with repeated calls operating in random environment. Computers & operations research 37, no7, pp. 1158-1166 (2010)

[2] Tien Van Do: A new computational algorithm for retrial queues to cellular mobile systems with guard channels. Computers & Indus. Engin. 59, pp. 865-872 (2010)

[3] Gharbi, N.: Modeling and Performance Evaluation of Small Cell Wireless Networks with Base Station Channels Breakdowns. In Proceedings of The Eighth Inter. Conf. on Wireless and Mobile Communications, pp. 42-48, Italy (2012)

[4] Wuechner, P., Sztrik, J., and De Meer, H.: Modeling Wireless Sensor Networks Using Finite-Source Retrial Queues with Unreliable Orbit. Proc. of the Workshop on Perf. Eval. of Computer and Communication Systems (PERFORM'2010), vol. 6821 of LNCS Publisher: Springer-Verlag (2011)

[5] Artalejo, J.R.: Accessible bibliography on retrial queues: Progress in 2000-2009. Mathematical and Computer Modelling 51, pp. 1071-1081 (2010)

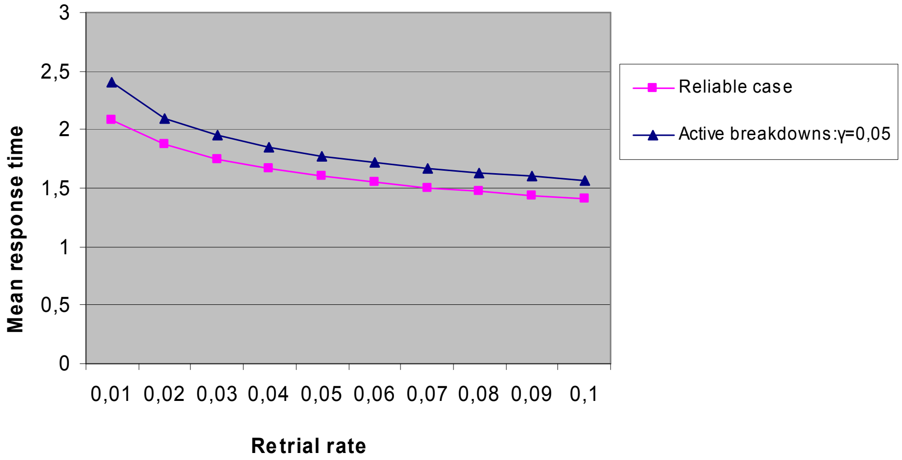


Figure 2. Mean response time versus retrial rate

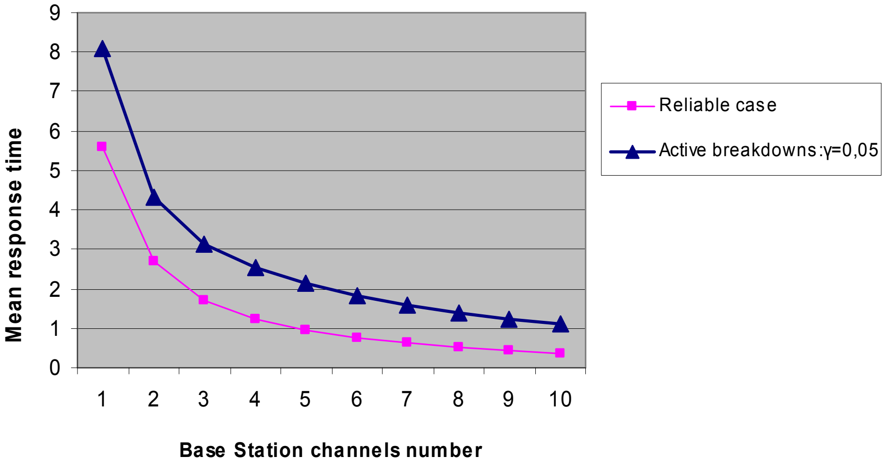


Figure 3. Mean response time versus base station channels number

[6] Almasi, B., Roszik, J., and Sztrik, J.: Homogeneous finite-source retrial queues with server subject to breakdowns and repairs. Mathematical and Computer Modelling 42, pp. 673-682 (2005)

[7] Almasi, B., Roszik, J., and Sztrik, J.: Heterogeneous finite-source retrial queues with server subject to breakdowns and repairs. Journal of Mathematical Sciences 132, pp. 677-685 (2006)

[8] Sztrik, J. and Efrosinin, D.: Tool supported reliability analysis of finite-source retrial queues. Automation and Remote Control 71, pp. 1388-1393 (2010)

[9] Sztrik, J. and Kim, C.S.: Tool supported performability investigations of heterogeneous finite-source retrial queues. Annales Univ. Sci. Budapest., Sect. Comp. 32, pp. 201-220 (2010)

[10] Wang, J., Cao, J., and Li, Q.: Reliability analysis of the retrial queue with server breakdowns and repairs. Queueing Systems 38, pp. 363-380 (2001)

[11] Roszik, J. and Sztrik, J.: Performance analysis of finite-source retrial queues with nonreliable heterogenous servers. Journal of Mathematical Sciences 146, pp. 6033-6038 (2007)

[12] Ajmone Marsan, M., Balbo, G., Conte, G., Donatelli, S., and Franceschinis, G.: Modelling with Generalized Stochastic Petri Nets. John Wiley & Sons, New York (1995)

[13] Diaz, M.: Les réseaux de Petri - Modèles Fondamentaux. Paris, Hermès Science Publications (2001)

[14] Falin, G.I. and Templeton, J.G.C.: Retrial Queues. Chapman and Hall, London (1997)

MTRP: Multi-Topology Recovery Protocol

Paulo V. A. Pinheiro
Universidade Estadual do Ceará (UECE)
Av. Paranjana 1700
Fortaleza - CE - Brazil
paulovap@larces.uece.br

Marcial P. Fernandez
Universidade Estadual do Ceará (UECE)
Av. Paranjana 1700
Fortaleza - CE - Brazil
marcial@larces.uece.br

***Abstract*—Link and node failure recovery is critical in any production network and recover the failures in times below 50 milliseconds is desired to maintain the quality of real time application. In this paper, we propose the Multi-Topology Recovery Protocol (MTRP) that provides network protection using pre-calculate routes and a multi-topology approach. The protocol was based on the state-of-the-art recovery and protection techniques. MTRP is based on Resilient Routing Layers (RRL) algorithm, used to generate the sub-topology. The MTRP prototype was implemented and tested in a virtualized environment, providing real IP stack in an actual operating system. The tests show that the MTRP provides a quick convergence, below few milliseconds, similar to ring protection protocol for partial mesh topology. MTRP evaluation shows that it also produces recovery paths with costs (distance) almost as low as the primary ones.**

Keywords-network recovery; protocol; resilience.

I. Introduction

Since the inception of the Internet, the problem of protection and failure's recovery on networks has aroused interest of researchers. This area has received enough attention because it keeps the quality of services provided by communication networks as regards the stability and availability. This is done by using mechanisms that aim to ensure protection of the network. Not limited to this, these mechanisms also restore the network to its normal operating condition, since there is a failure situation. This ability that the network has to keep itself alive, thus in an operational state, is called, in literature, survivability or resilience [1]. The most networks are designed to take advantage of that assumption, exploring the use of two topologies types: mesh and ring topology.

Ring recovery protocols are simple and provide recovery in a shorter time, restoring the network to its fully functional state in a time close to 50 milliseconds. However, this topology has limited recoverability: only a single failure can be recovered (only one spare path). These are some examples of protocols for this type of architecture: Ethernet Automatic Protection Switching (EAPS) [2], Ethernet Ring Protection Switching (ERPS) [3] and SONET/SDH. A mesh recovery protocol permits broader recovery, limited only by the amount of multiple paths available, but it has a greater recovery time. Since the protocol does not know the network topology to perform its protection, it always takes longer time to calculate a new viable path, typically in the order of seconds. Despite this recoverability, it provides a slow return to the natural state of the network because, when there is an error, a signaling process to notify the topology change to all nodes starts. These are examples of this type of protocol: Spanning Tree Protocol (STP) [4], Open Shortest Path First (OSPF) [5] and Intermediate System to Intermediate System Routing Exchange Protocol (IS-IS) [6].

This paper proposes a new protocol to ensure recovery of a partially mesh network quickly and efficiently. This proposal takes advantage of better recoverability from a mesh network, but it offers a recovery time close to the ring topology networks, such as SONET/SDH. The protocol will be based on the pre-computation of paths and the use of multi-topologies for network recovery, based on a technique known as Resilient Routing Layers (RRL) [7]. Our contribution is to propose and validate a new recovery protocol based on a small variation of this algorithm. The proposal validation was made by implementing a prototype in a virtualized network, created with Mininet tool [8].

This article is organized as follows. In Section II, we present related works. In Section III, the proposal of the Multi-Topology Recovery Protocol (MTRP) protocol is described. Then, in Section IV, the proposal evaluation is described. In the Section V, results are shown. Finally, in Section VI, we present the conclusion and future works.

II. Related Works

These are some related works that propose failure recovery protocol in mesh topologies in a short convergence time.

Barreto [9] presents a proactive approach that is added to the OSPF protocol. Emergency paths are computed in advance in each node, adding a secondary route for each available neighbor. This proposal is based on IP Fast reroute scheme.

Psenak et al. [10] propose an RFC that describes an extension of the OSPF protocol, called OSPF-MT. This extension suggests the use of multi topologies for using of the routing protocol. Among the possible uses, there are: new route's creation, isolation of classes of service and the management. Przygienda et al. [11], proposed an extension to the IS-IS Protocol, in which they suggest the use of multiple topologies for general purpose.

The literature presents many proposals of IP Fast Reroute (IPFRR) technologies [12]. IPFRR refers to the set of mechanisms aiming to provide fast rerouting using pure IP protocol forwarding and routing. Several proposals have been made to IETF IPFRR, such as, Release Point, Downstream Routes, Loop-Free Alternates, U-Turns and Not-Via Adresses [13]. The goal of the IPFRR mechanism is to set alternate routing paths, which avoid micro loops under node or link failures. However, IPFRR recalculates new routes after failure detection (reactive), and it requires to work only over IP protocol.

III. Multi-Topology Recovery Protocol (MTRP)

In this section, a new protocol for network recovery will be presented; whose primary goal is to ensure the recovery of failure in links or nodes in times near to the telecommunications networks based on rings (typically below 50 ms), using the least of redundant resources.

The protocol presented in this work was called Multi-Topology Recovery Protocol (MTRP), due to its most striking characteristic: using a set of sub topologies created from the actual topology of the network to keep the packets routing. MTRP has features to optimize the recovery process and provide new features to the routing protocol, making it more flexible, efficient and looking forward to becoming a good option for network's deployment.

A. Characteristics

This section aims to present several of these features which were incorporated to MTRP.

1) Local Recovery: When a failure occurs, the affected traffic is redirected by passing through the recovery mechanism, following a new path. This path, or more specifically the track used in the path, can be built through the principles of local and global recovery. The local recovery is done as close as possible to the point of failure, so in general, detection and recovery are performed much faster than the global recovery, and in most cases, requires fewer states. In order to achieve times shorter than $50ms$, MTRP protocol uses this local recovery scheme, where the traffic is rerouted to an alternative route to join a node next to the failure.

2) Hardware Failures Detection: The main mechanism to detect a failure is the hardware mechanisms in equipment physical layers. A mechanism on the equipment operating systems detects the link down by the loss of optical signal at an interface, starting the recovery procedure. So it can achieve a shorter failure detection time, and it can change to a new topology. In these there is not a hardware detection engine, it is necessary to send `HELLO` messages to supply this lack, consuming a little more network bandwidth and delaying the failure detection time.

3) Precalculated Recovery Paths: Using reactive schema to perform network nodes and link's recovery has been shown to be inefficient when the recovery time is important. Much time is spent in signaling; the dissemination of the new topology is slow and, finally, it is necessary to run an algorithm to generate the shortest path tree. A way to optimize the recovery time is the pre calculation of some steps that can be in advance for quick recovery. Using pre-computed paths to ensure recovery of all possible failures is impractical due to the large number of states that would be necessary. However, it is likely to cover 100% of single failures and ($\geq 80\%$) of multiple failures with few layers [7].

4) Multiple Topologies: The use of Multiple Topologies (MT) is a consequence of pre-calculating paths. Each possible path takes part of one (or more) viable topology. It is a relatively new approach, which many authors have been proposing to incorporate this feature to the traditional protocols such [10] [11]. Its basic principle is to generate virtual topologies based on its real network topology, where resources belonging to the network may not be present in all topologies.

5) Centralized Processing: The best set of routes to the network recoverability can be obtained when you have all the information about the topology and define the configuration in the same place. The Central Processor unit is the equipment responsible for the generation and distribution of routes to the routers, giving a huge management power to network operator.

B. MTRP Architecture

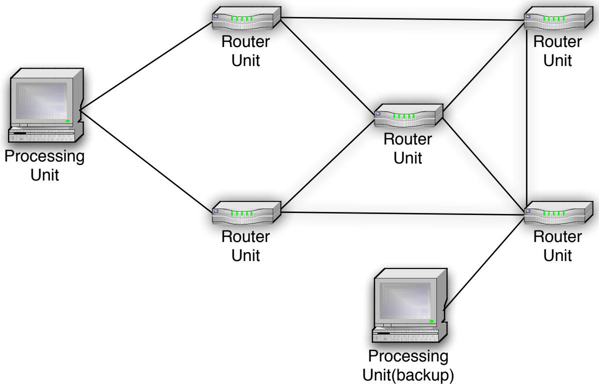


Figure 1. MTRP Architecture

The MTRP routing process involves two entities: the central node, called **Processing Unit (PU)**, which will be responsible for the generation of multiple network topologies and the routes' calculation, as well as some minor managerial activities, and the router, called **Router Unit**

(RU), which will manage the traffic that passes through the network using route tables generated by PU.

During the network initializing, each RU recognizes its neighbors and discovers the link metrics through their *HELLO* message. At the same time, the PU sends a message identifying itself as the control node. This message, called **Processing Unit Advertisement (PUA)**, is sent via broadcast to all the network RUs, which in their turn, learn by which interface it should communicate to the PU. If the PU receives this message from more than one interface, it will consider only the interface from which the PUA message came first.

Once the information about the neighbors and the path to the PU are established, each RU sends its routing table to the PU, using the message **Neighbor Advertisement (NA))** through the newly discovered PUA message path. PU receives all NAs and creates the actual topology, called **Real Topology (RTOP)**. With the RTOP defined the PU starts the process to create n virtual topologies, called **Virtual Topology (VTOP)**. The VTOP will be used to define the predefined routing paths (**Routing Table (RT)**).

C. Protocol Messages

The MTRP protocol uses the following messages: (1)Forwarding Packet, When a data packet enters on the network, it is encapsulated within a MTRP and it can be forwarded through the network; (2)Hello packets, used to maintain the neighborhood relationship between nodes, and detect failures that were not detected by the hardware; (3) Processing Unit Advertisement, Message sent just by PU to all RUs in the network to inform the active PU; (4) Neighbor Advertisement, message sent only by the RU to inform PU the list of its neighbor's nodes and the metrics of each link; (5)Route Database Update, after the routes are created, the PU will send this packet to update each RU routing tables, according to the number of sub topologies; (6)Acknowledgement, message used to confirm the messages received.

D. Processing Unit (PU)

The Processing Unit (PU) is the entity responsible for characterization, topologies generation, route's computation, topologies distribution and, finally, network monitoring.

1) Topologies Generation: The main feature of PU is the route generation, including pre-calculated primary and backup routes. In order to have it, it is necessary that PU has a set of VTOPs which will ensure these alternative routes, no matter how they are created. There are several ways to create VTOPs according to different network routing characteristics. VTOPs can be created prioritizing the size of backup paths, increasing coverage to multiple failures, minimizing the amount of VTOPs to simple failure's protection and many other possible customizations.

As a standard tool for VTOPs generation, the MTRP uses a variation of the Resilient Routing Layers (RRL) algorithm proposed in [14], showed at Algorithm 1. Our algorithm aims to ensure greater network redundancy, expanding coverage to multiple failures (*k-fault*) and avoid the back hauling effect [1]. The Algorithm 1 shows the layers generation.

Algorithm 1: MTRP: VTOP Generation for k-failures

```
Input: $G(V, E)$: Bidirectional Graph of Real Network
       Topology.
Input: $nTopo$: Minimum Number of Layers to be
       Generate
Input: $k$: Number of Simultaneous Failures Supported
Input: $artPoints(G)$: List of articulated nodes, i.e.,
       nodes that if removed would split the network
Result: $L[i]$: List of Graphs of Virtual Topologies
        VTOP
$S = artPoints(G)$
foreach $n \in V$ do
    $c(n) = 0$
    $cl(E) = 0$
end
$i = 0$
while $(i < nTopo)$ or $(|S| < |V|)$ do
    $L_i(V_i, E_i) = G$
    $P = \{\}$
    while $|P| < |V|$ do
        $n = min(c(n))$ such that $n \notin P$
        if $n \notin artPoints(L_i)$ then
            $\{l_1, \ldots, l_k\} = links(n, E_i)$
            $E_i = E_i - \{l_1, \ldots, l_k\}$
            $e = min(cl(\{l_1, \ldots, l_k\}))$
            $E_i = E_i + e$
            $cl(e) = cl(e) + 1$
            $S = S \cup \{n\}$
            $c(n) = c(n) + 1$
        end
        $P = P \cup \{n\}$
    end
    $i = i + 1$
end
```

The algorithm starts by creating two lists ($c()$ and $cl()$) containing integers indicating how many times a node ($c(n)$) and a link ($cl(E)$) were used for layer's generation. Then, a loop starts in each sub topology created, which will be added to the list L. P stores the list of nodes already used for creation of the L_i sub topology. In the innermost algorithm loop, there is a link removal to ensure they will be saved by other layers. For each layer, the node less used is selected $min(c(n))$, so that, their edges are removed $(E_i = E_i - l_1, l_k, \ldots)$ excepting the not as much used edge ($E_i = E_i + e$)). After all nodes have been saved for at least one layer ($S = V$), the algorithm ends and returns the list $L(i)$ of sub topologies.

If the network is a connected graph, the path is guaranteed by the routing algorithm, e.g., in our experiment Dijsktra's was used. The only constraints we assume is that the graph should have, at least, one path from any node to any other node in the graph.

E. Router Unit (RU)

The Router Unit (RU) is the entity responsible for the packets' forwarding on the network. Unlike to what happens with traditional routing protocols, such as OSPF and IS-IS, where the network control is done in a distributed way, the RU takes care only for the packet forwarding, report changes and errors. This ensures that the RU may become a simpler equipment with less processing power and thus, lower cost. Another advantage is the reduction of signaling overhead that occurs in distributed routing protocols, e.g., the tables' synchronization among routers and designated router election in OSPF. The following sections describe the RU state machine, the messages between RU-PU and the algorithms to recovery mechanism.

1) Signaling: There are two possible signals that start from RU: topology change and packet drop alert. The topology change may occur due to a failure, resulting in the loss or addition of links or nodes. The change notification is sent only to the PU and occurs in parallel to other RU activities, there is no dependency in that signals nor other critical router functionalities.

2) Recovery Mechanism: Since its route tables are completed, the RU is able to perform the packet forwarding and repair paths down. As described previously, the protocol used the table referring to RTOP during its regular functioning. Once there is the failure, the neighbor, that would use it as the next hop to deliver the packet, runs the Algorithm 2 to select a new routing table that does not present the failure element.

Algorithm 2: MTRP: RU Packet Forward Algorithm

```
Input: $T_n$: Route table
Input: $pkt$: Packet to be forward

sent = False;
if sendNextHop(pkt) not True then
    foreach $t \in T_n$ do
        if canSendVia(t,pkt) then
            mark(pkt,t);
            sendNextHop(pkt);
            sent = True;
        end
    end
    if sent not True then
        drop(pkt);
    end
end
```

At the beginning of the Algorithm 2, the RU tries to send the packet using the RTOP routing table, or the VTOP table in use marked on the packet. The packet is sent using the `sendNextHop()` function that receives the packet to be sent as a parameter. This function checks which table should be used reading $TOPOIDX$ field on packet and checks if the next hop in this table is possible (if the link has no failures). If it is not possible to send, this function returns the value `False`, which causes the algorithm to find a new route table to be used for this packet. Once the destination is found on the table, the function `canSendVia` returns the value `True` indicating that the table t can be used. The packet is marked by the `mark` function to indicate the next RU which table should be used. Then it is sent by `sendNextHop` function, but now using another table.

IV. Prototype Validation

The computational virtualization enables various operating systems to run concurrently on shared hardware equipment, so it is possible to have multiple operating systems logically isolated between them running on a single hardware. The use of virtual machines connected through network interfaces has been shown to be quite efficient and presents a higher degree of realism and interactivity than the use of network simulators like ns-2 [15].

MTRP evaluation shall be carried out in a scenario created for the rapid prototyping network tools Mininet [8]. This tool creates a network topology through the use of *Lightweight Virtualization*, i.e., a virtualization scheme where an isolated environment is created to group processes in containers, where logical devices of each container are not shared among themselves.

Scenarios were created based on real-world topologies: GEANT, IINET and SPRINT. Database containing information about these and other actual topologies can be found on [16]. For worst case evaluation; it was used the bigG topology, an artificial network generated from the algorithm proposed in [17], containing 123 nodes and 243 links.

For each topology, the metrics, network initialization time, recovery path distance, memory used to store routing table and time to recover a simple failure will be reviewed. With such data, we can have an overview of the performance of the protocol on these networks.

V. Results

In this section, the results of the tests will be presented and analyzed, demonstrating the effectiveness of the protocol.

A. Network Initialization

Tests to establish the network startup time were made to define how the use of a centralized point for processing and distribution routes would affect the network startup. The startup time is the time interval between the first `HELLO` message sent by PU until the last `ACK` message received by it, indicating that all routes were sent successfully. Table I shows the startup times for each topology [18].

Table I
NETWORK INITIALIZATION TIME

Topology	Network Initialization Time (sec)
Sprint$(11, 18)$	0.89882
Geant$(27, 38)$	0.999489
Iinet$(31, 35)$	1.03966
medG$(42, 81)$	1.51030
bigG$(123, 243)$	5.1478

According to the Table I, the network boot time is around one second, gradually increasing with the rise of network complexity. The MTRP protocol presented low boot times compared to others such as Routing Information Protocol (RIP) and OSPF. According to [18], in a network with seven routers, the OSPF protocol takes, in average, ten seconds to carry out the convergence of its nodes, which is about the time of a network initialization. RIP takes, in average, more than 100 seconds.

B. Average Route Length

Table II shows the average of all possible paths between two different nodes in the whole topology in all topologies generated by MTRP algorithm. To perform the calculation the Equation 1 was used:

$$\bar{x} = \sum_{s,t \in V} \frac{d(s,t)}{n(n-1))} \tag{1}$$

where s and t are nodes belonging to the network, n represents the total number of nodes and $d(s,t)$ is the smallest path which goes from s to t. Table II shows the results of average of route's length.

In a failure situation, the protocol will use one of the sub-topologies created to perform the packets' forwarding, choosing the shortest path between source and destination. The failure was simulating by removing a link or a node. However, we only considered the situations where the graph still connected, i.e., there is at least one backup path. Non-connected graphs cannot be fully restored, so it is impossible to infer an average route length.

On the Sprint network, it was noted that the paths generated by sub-topologies are on average $1.19x$ to $1.49x$ larger than the original path without failures. Geant presented a variation from $1.10x$ to $1.50x$ and, finally, the algorithm obtained from $1.10x$ to $1.24x$ in Iinet. BigG, in its turn, presented sub-topologies with the smallest paths (Topo3 - $1,002x$) as well as the largest ones (Topo1 - $3.6x$) in the tests.

The VTOP generation algorithm used does not consider the topology metrics to split the network into layers. So, it was created the "bad" Topo 1 topology on bigG. However, the table shows the worst path in this topology, so there are other topologies that give the optimal paths, thus more eligible to be used as recovery topology. This is not really an issue once you use a fair number of virtual topologies (5, 6..), and it would not be difficult to generate the VTOP that the "good" paths would be fairly distributed by the VTOPs.

Table II
MTRP - AVERAGE PATH LENGTH (NR. OF HOPS)

Topology	Real	Topo 1	Topo 2	Topo 3
Sprint$(11, 18)$	1.89091	2.2545	2.8363	2.3818
Geant$(27, 38)$	2.9373	4.5641	4.2222	3.2421
Iinet$(31, 35)$	2.8215	3.1053	3.5053	3.3677
bigG$(123, 243)$	2.9648	10.6743	3.3145	2.9722

It is noticed that the algorithm generates good recovery paths, with sizes close to the original one. Even with good results, there are some ways to improve the algorithm so that it generates shorter paths. The use of more topologies is necessary, and each topology may have a larger set of bindings. It is important to note that as a consequence of the number of topologies rises, there is also the increase use of memory in the router.

C. Memory Use

The amount of memory used for tables storing is an important parameter for the protocol, because the creation of new sub topologies implies the increase of memory usage for storing the routing tables. For the scenarios described earlier, Table III shows the amount of memory use for each topology, when the protocol is written in Python [19]; and the amount of the memory use in an equivalent protocol written in C.

Table III
MEMORY USED TO STORE ROUTING TABLE

Topology	Topo number	Mem(Py)	Mem(C)
Sprint$(11, 18)$	3+1	7.5923 kb	352 b
Geant$(27, 38)$	3+1	21.1996 kb	864 b
Iinet$(31, 35)$	3+1	22.2615 kb	992 b
bigG$(123, 243)$	3+1	84.1110 kb	3.84375 kb

The Python implementation of the protocol implies in high memory usage. It is caused by the excessive use of the object orientation. To store the value of the link cost or router ID in Python, we need to use an *integer* type, which is 24 bytes long. In a more efficient implementation in C, these elements would be stored in an *unsigned int* of 4 bytes. In Table III we can see that the memory usage in an implementation in C is small, that permit to store more layers in the router. For the memory estimation in C, the we use the formula $mem = |V| * (t+1) * obj$, where $|V|$ is the number of vertex of the graph, t is the number of sub topologies created and obj is the memory size required to store the RouterID and the link cost (8 bytes). In summary, the use of memory space for storage of routing tables grows linearly ($O(t)$) with the number of created sub topologies.

D. Network Recovery Time

Recovery time of MTRP protocol basically consists of three steps: failure detection, next-hop table lookup and set the packet tag. In our proposal, any failure detection protocol can be used, e.g., the Bidirecional Forwarding Detection (BFD) protocol [20]. BFD protocol works with IP protocol and guarantee link failure detection in milliseconds (usually below $50ms$).

The Table IV shows the times found in tests to perform the table lookup of the next hop, returning the network address of the next hop of the new path, without failures. The values showed in Table IV did not consider the time of failure detection, in order of tens of milliseconds. The packet should follow the next hop and the time of packet tag, indicating which recovery table will be used. Then, to perform the recovery time, the tasks were performed one thousand times, to define the table lookup execution time with greater accuracy. To get the recovery time, only the worst case was considered. The tests were conducted by generating different amounts of sub topologies.

Table IV
RECOVERY TIME OF SINGLE FAILURE

Topology	t(topo=4)	t(topo=8)	t(topo=12)
Sprint$(11, 18)$	2.740 μs	5.279 μs	7.463 μs
Geant$(27, 38)$	2.995 μs	5.524 μs	7.204 μs
Iinet$(31, 35)$	2.790 μs	5.031 μs	7.422 μs
bigG$(123, 243)$	3.119 μs	5.139 μs	7.049 μs

The table lookup time grows linearly considering the quantity of topologies used by network, as it was expected due to its linear complexity algorithm. Packet tagging time is constant and was included in the test's execution time. Considering one thousand executions, we can realize that the tasks of lookup and packet tagging are performed on the microseconds' scale (10^{-6}), becoming insignificant compared to the hardware detection time. Therefore, the recovery time may be considered, only the failure detection time; it shows a big performance gain compared to the reactive protocols that take more than a second to perform the recovery, such as OSPF.

VI. CONCLUSION AND FUTURE WORKS

Through the results presented, it is possible to verify the viability of the MTRP protocol for network recovery in mesh topology. The MTRP shows recovery times in the milliseconds range, equivalent to SDH/SONET networks. In addition, you can see that the use of a central authority reduces the amount of processing required for signaling, compared to distributed protocols. Reducing the amount of messages relieves the individual processing of each router, ensuring the possibility of the use of equipment with less processing power to perform the transfer of packets. The subtopologies generation is done only on network initialization, 5 sec in BigG network (bigger than tradicional operator backbones), we may guarantee the proposal scalability for real networks. As the recovery is done using pre computed paths, there is no impact of network length in recovery time.

The MTRP promotes a quick startup, because link state distribution throughout the network is not necessary, only to the control node, PU. Our work also confirms the validation of RRL algorithm done in [7] and [14].

The MTRP protocol presents a range of possibilities for future works. To continue the development of the MTRP protocol, a better definition of message authentication system and use of header fields are important. Several improvements can be made to increase the protocol performance, such as implementation in C and integration of routing packets with the kernel of the operating system. Another possibility is the addition of Quality of Service (QoS) extension in protocol. A network can provide different classes of service to clients and divide its network into sub topologies with distinct QoS classes, where their packets will be routed exclusively for each sub topology. When a packet from a client joins the network edge, this packet would be already marked for the topology to which the class belongs.

Finally, a study on the integration of the protocol with the *OpenFlow* technology can be made. The centralizing feature of the MTRP protocol is similar to the OpenFlow control architecture, so the adaptation should not be difficult.

REFERENCES

[1] J.-P. Vasseur, M. Pickavet, and P. Demeester, *Network Recovery: Protection and Restoration of Optical, SONET-SDH, IP, and MPLS*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2004.

[2] S. Shah and M. Yip, “Extreme Networks’ Ethernet Automatic Protection Switching (EAPS) Version 1,” RFC 3619 (Informational), Internet Engineering Task Force, Oct. 2003. [Online]. Available: http://www.ietf.org/rfc/rfc3619.txt

[3] J. Ryoo, H. Long, Y. Yang, M. Holness, Z. Ahmad, and J. Rhee, “Ethernet ring protection for carrier ethernet networks,” *Communications Magazine, IEEE*, vol. 46, no. 9, pp. 136–143, 2008.

[4] R. Perlman, *Interconnections: bridges and routers*. Redwood City, CA, USA: Addison Wesley Longman Publishing Co., Inc., 1992.

[5] J. Moy, “OSPF Version 2,” RFC 2328 (Standard), Internet Engineering Task Force, Apr. 1998, updated by RFCs 5709, 6549. [Online]. Available: http://www.ietf.org/rfc/rfc2328.txt

[6] D. Oran, “OSI IS-IS Intra-domain Routing Protocol,” RFC 1142 (Informational), Internet Engineering Task Force, Feb. 1990. [Online]. Available: http://www.ietf.org/rfc/rfc1142.txt

[7] A. Kvalbein and A. F. Hansen, “Fast recovery from link failures using resilient routing layers,” in *ISCC ’05: Proceedings of the 10th IEEE Symposium on Computers and Communications*. Washington, DC, USA: IEEE Computer Society, 2005, pp. 554–560.

[8] B. Lantz, B. Heller, and N. McKeown, “A network in a laptop: rapid prototyping for software-defined networks,” in *HotNets*, G. G. Xie, R. Beverly, R. Morris, and B. Davie, Eds. ACM, 2010, p. 19.

[9] F. Barreto, “Esquema de caminhos emergenciais rápidos para amenizar perdas de pacotes,” Ph.D. dissertation, Universidade Técnologica Federal do Paraná, 2010.

[10] P. Psenak, S. Mirtorabi, A. Roy, L. Nguyen, and P. Pillay-Esnault, “Multi-Topology (MT) Routing in OSPF,” RFC 4915 (Proposed Standard), Internet Engineering Task Force, Jun. 2007. [Online]. Available: http://www.ietf.org/rfc/rfc4915.txt

[11] T. Przygienda, N. Shen, and N. Sheth, “M-ISIS: Multi Topology (MT) Routing in Intermediate System to Intermediate Systems (IS-ISs),” RFC 5120 (Proposed Standard), Internet Engineering Task Force, Feb. 2008. [Online]. Available: http://www.ietf.org/rfc/rfc5120.txt

[12] M. Shand and S. Bryant, “IP Fast Reroute Framework,” RFC 5714 (Informational), Internet Engineering Task Force, Jan. 2010. [Online]. Available: http://www.ietf.org/rfc/rfc5714.txt

[13] M. Gjoka, V. Ram, and X. Yang, “Evaluation of ip fast reroute proposals,” in *Communication Systems Software and Middleware, 2007. COMSWARE 2007. 2nd International Conference on*. IEEE, 2007, pp. 1–8.

[14] T. Čičić, A. Hansen, S. Gjessing, and O. Lysne, “Applicability of resilient routing layers for k-fault network recovery,” *Networking-ICN 2005*, pp. 173–183, 2005.

[15] L. Breslau, D. Estrin, K. Fall, S. Floyd, J. Heidemann, A. Helmy, P. Huang, S. McCanne, K. Varadhan, Y. Xu *et al.*, “Advances in network simulation,” *Computer*, vol. 33, no. 5, pp. 59–67, 2000.

[16] S. Knight, H. Nguyen, N. Falkner, R. Bowden, and M. Roughan, “The internet topology zoo,” *Selected Areas in Communications, IEEE Journal on*, vol. 29, no. 9, pp. 1765–1775, october 2011.

[17] S. Dorogovtsev, A. Goltsev, and J. Mendes, “Pseudofractal scale-free web,” *Physical Review E*, vol. 65, p. 066122, Jun 2002.

[18] H. Pun, “Convergence behavior of rip and ospf network protocols,” Ph.D. dissertation, University of British Columbia, 1998.

[19] M. Lutz, *Programming python*. O’Reilly Media, Inc., 2011.

[20] D. Katz and D. Ward, “Bidirectional Forwarding Detection (BFD),” RFC 5880 (Proposed Standard), Internet Engineering Task Force, Jun. 2010. [Online]. Available: http://www.ietf.org/rfc/rfc5880.txt

Bandwidth on Demand over Carrier Grade Ethernet Equipment

Christos Bouras[1,2], Theoni-Katerina Spuropoulou[2] and Kostas Stamos[1,2,3]
[1] Computer Technology Institute and Press “Diophantus”, N. Kazantzaki Str, University Campus 26504, Rio Greece
[2]Computer Engineering and Informatics Dept., University of Patras
[3]Technological Educational Institute of Patra, Greece
bouras@cti.gr, spuropoulo@ceid.upatras.gr, stamos@cti.gr

***Abstract*—This paper presents the prototype implementation of a Bandwidth on Demand (BoD) service over equipment using Carrier Grade Ethernet. The BoD multi-domain service is based on AutoBAHN (Automated Bandwidth Allocation across Heterogeneous Networks) software. The paper describes the steps that have taken place for designing and implementing a prototype technology proxy that is able to match the Carrier Grade Ethernet with AutoBAHN, based on the implementations of the relevant standards and technologies by Extreme Networks. The equipment in particular is comprised of BlackDiamond 12804 switches, running ExtremeXOS version 12. The paper demonstrates how a suitable testbed can be created and utilized and how a new technology at the data plane can be supported by a Bandwidth on Demand tool.**

Keywords-Carrier Grade Ethernet; BlackDiamond switches; Bandwidth on Demand; AutoBAHN

I. INTRODUCTION

The GN3 European project [1] is a research project funded by the European Union and Europe's National Research and Education Networks (NRENs). It is a continuation of the previous GN2 project and aims at building and supporting the next generation of the pan-European research and education network, which connects universities, institutions and other research and educational organizations around Europe and interconnects them to the rest of the Internet using high-speed backbone connections.

In the context of this project, a BoD service is being developed and the service is supported by the AutoBAHN tool.

The AutoBAHN system [2] is capable of provisioning circuits in heterogeneous, multi-domain environments that constitute the European academic and research space and allows for both immediate and advanced circuit reservations. The overall architecture of the AutoBAHN system, its goal and the network mechanisms it employs are thoroughly presented in [3].

This paper presents the prototype implementation of a Technology Proxy (TP) for Carrier Grade Ethernet (CGE) equipment. AutoBAHN may support multiple underlying technologies, and Carrier Grade Ethernet features a promising set of characteristics for network carriers. This is one of the first attempts at using a CGE network within a multi-domain Bandwidth on Demand service, and the conclusions of this work are therefore useful in order to determine the effectiveness and required effort of using CGE as the underlying technology for such purposes. It also enables us to compare the degree to which CGE implementations lend themselves to participation in a multi-domain automated reservation service.

The rest of the paper is structured as follows:

Section II presents the Carrier Grade Ethernet technology and related standards, while Section III introduces the general architecture of the AutoBAHN system with an emphasis on its lower levels that interact with the underlying equipment. Section IV describes the TP framework that has been used for handling the creation of the interface towards the underlying network technology. Section V focuses on the implementation that took place including the testing to verify our work. Finally, Section VI concludes the paper and presents future fields of related study.

II. CARRIER GRADE ETHERNET

Carrier Grade Ethernet (CGE) in general refers to the enhancements to Ethernet standards in order to be suitable as a carrier-grade and transport technology [4].

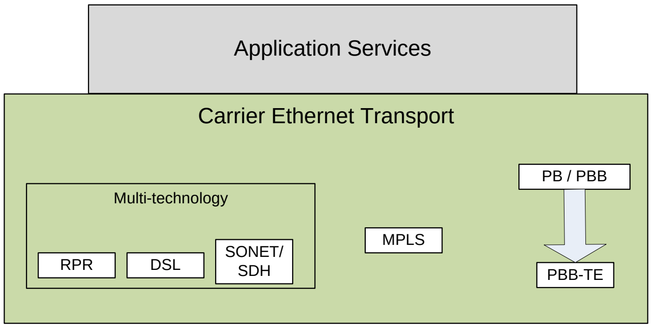


Figure 1. Carrier Ethernet technologies

This section presents briefly the main standards and how these relate to the Carrier Grade Ethernet concepts. There are several aspects of Ethernet that need to be modified or enhanced in order for a technology that was initially designed for local area networks to be suitable for carrier grade deployments ([5][6][7]). In general, several technologies can be considered Carrier Ethernet, including

ISBN: 978-1-61208-245-5

solutions based on carrying Ethernet over Resilient Packet Ring (RPR), SDH or DSL technologies, or over MPLS. All these approaches can be used to carry applications over Ethernet backhaul. However we are focused on so-called pure Ethernet centric solutions (Figure 1). Below we give a brief overview of the main related protocols that comprise a CGE deployment based on pure Ethernet centric solutions.

The Link Layer Discovery Protocol (LLDP) is a vendor-neutral Link Layer protocol used by network devices for advertising their identity, capabilities, and neighbours on an IEEE 802 local area network. The protocol is formally referred to by the IEEE as Station and Media Access Control Connectivity Discovery specified in standards document IEEE 802.1AB [8]. There are several proprietary protocols that perform functions similar to LLDP, such as Cisco Discovery Protocol, Extreme Discovery Protocol, Nortel Discovery Protocol (also known as SONMP), and Microsoft's Link Layer Topology Discovery (LLTD).

IEEE 802.1ad (Provider Bridges) [9] is an amendment to IEEE standard IEEE 802.1Q-1998 (aka QinQ or Stacked VLANs) [10], intended to develop an architecture and bridge protocols to provide separate instances of the MAC services to multiple independent users of a Bridged Local Area Network, in a manner that does not require cooperation among the users, and requires a minimum of cooperation between the users and the provider of the MAC service. The idea is to provide, for example, the possibility for customers to run their own VLANs inside service provider's provided VLAN. This way the service provider can just configure one VLAN for the customer and customer can then treat that VLAN as if it was a trunk.

Provider Backbone Bridges (PBB) or IEEE 802.1ah-2008 is a set of architecture and protocols for routing of a customer network over a provider's network allowing interconnection of multiple Provider Bridge Networks without losing each customer's individually defined VLANs. It was initially created as a proprietary extension by Nortel before being submitted to the IEEE 802.1 committee for standardization. The final standard was approved by the IEEE in June 2008. [11]

Provider Backbone Bridge Traffic Engineering (PBB-TE) is an approved networking standard, IEEE 802.1Qay-2009 [12]. PBB-TE adapts Ethernet technology to carrier class transport networks. It is based on the layered VLAN tags and MAC-in-MAC encapsulation defined in IEEE 802.1ah (Provider Backbone Bridges (PBB), but it differs from PBB in eliminating flooding, dynamically created forwarding tables, and spanning tree protocols. Compared to PBB and its predecessors, PBB-TE behaves more predictably and its behavior can be more easily controlled by the network operator, at the expense of requiring up-front connection configuration at each bridge along a forwarding path. PBB-TE operational administration and maintenance (OAM) is usually based on IEEE 802.1ag. It was initially based on Nortel's Provider Backbone Transport (PBT).

PBB-TE's connection-oriented features and behaviors, as well as its OAM approach, are inspired by SDH/SONET. PBB-TE can also provide path protection levels similar to the UPSR (Unidirectional Path Switched Ring) protection in SDH/SONET networks.

IEEE 802.1ag IEEE Standard for Local and Metropolitan Area Networks Virtual Bridged Local Area Networks Amendment 5: Connectivity Fault Management [13] is a standard defined by IEEE. It defines protocols and practices for OAM (Operations, Administration, and Maintenance) for paths through 802.1 bridges and local area networks (LANs). It is an amendment to IEEE 802.1Q-2005 and was approved in 2007. IEEE 802.1ag is largely identical with ITU-T Recommendation Y.1731, which additionally addresses performance management.

III. Architecture Description

The aim of the BoD service, as defined in GEANT, is to provide dedicated channels for data transport, which are necessary for demanding applications and research fields with strict demands for the provisioning of guaranteed and dedicated capacity and high security level in the sense that the carried traffic is isolated from other traffic. This service is offered collaboratively by GÉANT and a set of adjacent domains (NRENs or external partners) that adhere to its requirements. These joint networks form a multi-domain area where the service is provided between two end points, which may belong to the same or different domains.

The AutoBAHN system operates in a multi-domain environment and consists of several modules that take over the resource negotiation, pathfinding, topology abstraction and exchange, admission control and other necessary operations in order for a multi-domain circuit reservation to be processed. At the final stage of processing, a circuit has to be realized by sending the appropriate configuration commands to the network devices. The AutoBAHN software stack is shown in Figure 2 and it is repeated at each domain that participates in the BoD service. Domain instances communicate with each other via the Inter-Domain Manager (IDM) module.

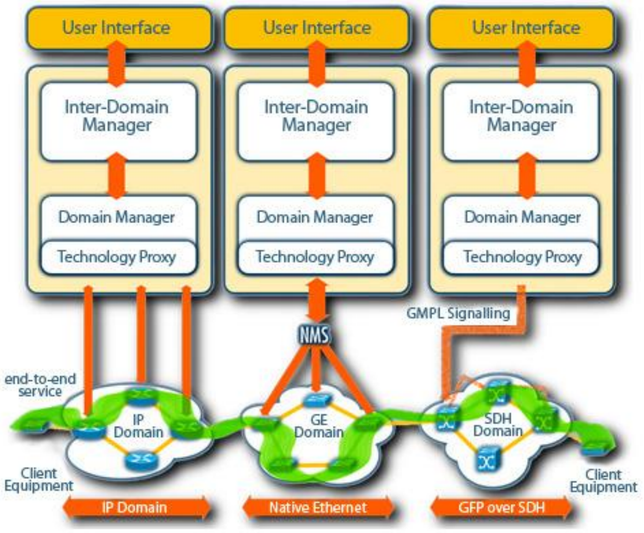


Figure 2. AutoBAHN architecture [2]

Network device configuration may take place via a Network Management System (NMS) or with direct application of configuration commands on devices. In both cases, the AutoBAHN module that is responsible for bridging the generic AutoBAHN software stack with the technology-specific NMS or network devices is called Technology Proxy (TP). The TP translates the reservation requests, received from the DM, to the appropriate commands to be sent (usually via SSH) to a network or an NMS. Each domain participating in the BoD service, depending on the underlying technology, uses a customized version of the TP module.

IV. TP FRAMEWORK

Developing a customized Technology Proxy module for each domain participating in the BoD service is time-consuming and may be difficult to accomplish as the necessary development manpower may not be available. Therefore, AutoBAHN provides the TP framework [2], which enables the creation of a customized TP without the need to develop software. Instead, all that has to be done is to edit the TP configuration with the appropriate network configuration commands for setting up and tearing down circuits. The TP framework then takes care of generating a TP that is capable of bridging the AutoBAHN software with the underlying data plane equipment. The TP framework configuration is based on XML, and allows the definition of several communications methods (such as Telnet or SSH), which can also be customized or extended. It can also accommodate various vendors, equipment models and Operating System versions, using an extensible loader architecture. Furthermore, it provides scripting functionality for more complicated command structures. Finally, the TP framework provides detailed logging so that the network administrator may troubleshoot erroneous behavior or failed requests.

V. IMPLEMENTATION

The Carrier Grade Ethernet TP was prototyped for the case of a network implemented with ExtremeNetworks equipment and especially with BlackDiamond 12804 switches. The BlackDiamond 12800 switches allow a single Ethernet network to deliver both residential and business services. They are chassis-based, Ethernet service core switches designed for core applications. Their features include hot-swappable I/O modules, Management Switch Fabric Modules (MSMs) that provide the active switching fabric and CPU control subsystem, auto-negotiation for half-duplex or full-duplex operation on 10/100/1000 Mbps ports and load sharing on multiple ports. They are running ExtremeXOS operating system, which supports PBB and PBB-TE.

The PBB-TE technology in particular is the one chosen for implementing the dynamic circuit service. PBB-TE allows the creation of a traffic engineered service instance path that behaves much like a dedicated service line and operates over an Ethernet network. A PBB-TE path is a static path through a Provider Backbone Bridge Network (PBBN), which is a 2 network that supports 802.1ad frames (also called Q-in-Q or vMAN in Exterme Networks terminology). If a vMAN connects to a PBB-TE path, vMAN frames always follow the same path to the egress PBB. If a vMAN connects to a PBB, vMAN frames are switched based on the configuration of the PBBN switches. PBB-TE changes the existing forwarding behavior as defined in 802.1Q to a new forwarding behavior by introducing a new port state: forwarding with address learning disabled. On a PBB-TE link, all broadcast, multicast, and unicast packets with an unknown destination MAC address are discarded. PBB-TE relies on Ethernet OAM (CFM) for fault detection and restoration of provisioned paths. CFM actively monitors all provisioned (working and protection) paths. A protection path is selected by mapping a Service VLAN (SVLAN) to a different BVLAN, which defines all the switch ports that link the tunnel endpoints.

Each switch offers a command line interface which can be used by the CGE TP to independently configure each switch in the network. This means that the CGE TP needs to have knowledge of the underlying network and make complicated decisions. Access to the command line interface is provided via Telnet, from a strict set of machines within the internal testbed network (Figure 3). The testbed, as shown on figure 2, consists of three Extreme BlackDiamond 12804 Carrier Grade Ethernet switches and virtual machines as traffic source and destination. MAC addresses of all switches and of the virtual machine are either already known or discovered through the LLDP protocol supported in the switches. Connectivity to the outside world is provided over JANET, the research and education network of the UK.

The Domain Manager (DM) of AutoBAHN submits its requests for reservation or deletions of reservations and the CGE TP sends its responses using a pre-defined Web Services (WS) interface for the DM-TP communication. The reservation request as sent by the DM fully describes the desired route and request parameters (such as required VLAN, capacity, start and end times), leaving to the TP only the configuration of the relevant switches.

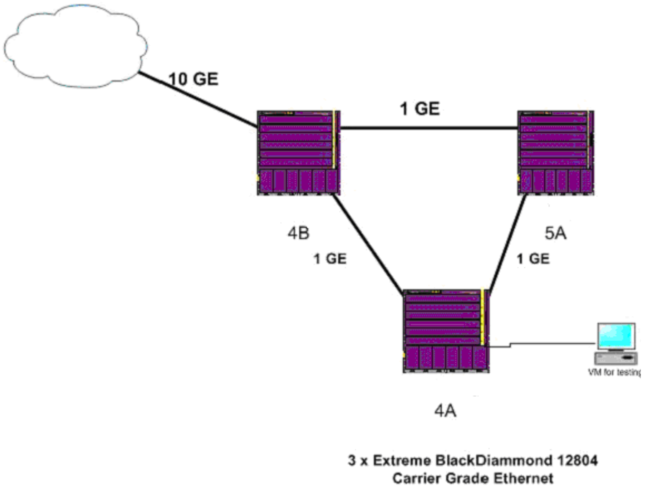


Figure 3. The testbed network, provided by University of Essex

For the purpose of the PBB-TE tunnel function, we initially defined two BVLANs, creating them statically on the interface, so that the TP does not have to recreate them every time it processes a reservation request from AutoBAHN. The first BVLAN consists of edge switches 4A and 4B, while the second includes core switch 5A. Edge switches receive customer VLAN traffic from virtual machines and transform it into SVLAN traffic, to be inserted into the PBBN. By disabling address learning on the switches, we gain complete control over the PBBN path, since each path is a static route. On a PBB-TE link, all broadcast, multicast, and unicast packets with an unknown destination MAC address are discarded. The PBB-TE trade-off is that it takes away the Ethernet self-configure and self-healing mechanisms. We rely on AutoBAHN for the selection of the desired route. By disabling flooding we ensure that all path traffic is limited to the configured path. Finally, by configuring FDB entries on the egress port of each switch along the route, we define the possible paths.

The implementation described above was tested using a simplified client application that provided the TP with incoming requests. The TP was able to successfully configure the testbed switches as described in the relevant sections, setting up a PBB-TE tunnel and enabling layer 2 connectivity between the desired end points. The implemented architecture allows the network administrator to define pre-determined paths (using the BVLAN configuration described above), which leads to predictable traffic management and load balancing. AutoBAHN is then used for the creation of the circuits on-demand, making dynamic use of the pre-determined paths. It is also possible for the administrator to devote a subset of the available capacity for the Bandwidth on Demand service, reserving the rest for manual configuration or other purposes.

VI. Conclusion and Future Work

Our work focused on integrating a testbed network based on equipment supporting Carrier Grade Ethernet standards with a multi-domain BoD service. The outcome of our work demonstrates that Carrier Grade Ethernet is a viable technology for such purposes, even in cases where the equipment does not provide a proprietary network management solution. Furthermore, we verified that the TP framework component that has been developed within the GN3 project greatly eases the necessary work for developing a bandwidth on demand module for a novel technology. This means that different technologies or different vendor implementations can be accommodated faster and with fewer resources because emphasis needs only to be put on properly aligning the technology with the service requirements rather than on low level programming tasks.

Our future work will focus on further testing and performance evaluation of Carrier Grade Ethernet operation in a broader Bandwidth on Demand context. There are a number of possible technology stitching requirements that may arise from the interoperation of Carrier Grade Ethernet with other technologies supporting Bandwidth on Demand in a multi-domain environment, and we plan to investigate them in both testbed and production settings. Another aspect that we were unable to cover was the one of bandwidth limiting or shaping techniques, since VLAN tag was the only information available to identify and separate network traffic, but BlackDiamond 12804 series switches did not support VLAN based traffic groups. We intend to investigate solutions to this issue in the future either with the equipment under consideration or different equipment.

Acknowledgments

The authors would like to acknowledge the contribution of their GN3 partners and especially Ramanujam Jayakumar and Mayur Channegowda from University of Essex who provided access and support for usage of the CGE equipment.

References

[1] "GN3 European Project," [Online]. Available: http://www.geant.net/pages/home.aspx.

[2] AutoBAHN, [Online]. Available: autobahn,geant.net (Accessed September 2012)

[3] M. Campanella, R. Krzywania, V. Reijs, A. Sevasti, K. Stamos, C. Tziouvaras, and D. Wilson, "Bandwidth on Demand Services for European Research and Education Networks", 1st IEEE International Workshop on Bandwidth on Demand, 27 Nov 2006, San Francisco (USA).

[4] R. Sanchez, L. Raptis, and K. Vaxevanakis, "Ethernet as a carrier grade technology: developments and innovations", Communications Magazine, IEEE, Vol. 46, Issue 9, pp. 88-94, September 2008

[5] Marwan Batayneh, Dominic A. Schupke, Marco Hoffmann, Andreas Kirstaedter, Biswanath Mukherjee, and Biswanath Mukherjee, "Lightpath-Level Protection versus Connection-Level Protection for Carrier-Grade Ethernet in a Mixed-Line-Rate Telecom Network", GLOBECOM 2009, pp. 2178-2182

[6] K. Ogaki, and T. Otani, "GMPLS Ethernet and PBB-TE (A carrier's view)", Conference on Optical Fiber Communication (OFC) 2009, pp. 1-4

[7] Wonkyoung Lee, Chang-Ho Choi, Taesik Cheung, Sun-Me Kim, and Ho-Young Song, "Implementation of hierarchical QoS mechanism on PBB-TE system", 9th International Conference on Optical Internet (COIN), 2010, pp. 1-3

[8] The Institute of Electrical and Electronics Engineers, "802.1ab: Station and Media Access Control Connectivity Discovery" IEEE Standard for Local and metropolitan area networks, 6 May 2005, IEEE

[9] The Institute of Electrical and Electronics Engineers, "802.1ad: Virtual Bridged Local Area Networks - Amendment 4: Provider Bridges" IEEE Standard for Local and metropolitan area networks, 26 May 2006, IEEE

[10] The Institute of Electrical and Electronics Engineers, "802.1Q: Virtual Bridged Local Area Networks" IEEE Standard for Local and metropolitan area networks, 19 May 2006, IEEE

[11] The Institute of Electrical and Electronics Engineers, "802.1ah: Virtual Bridged Local Area Networks - Amendment 7: Provider Backbone Bridges" IEEE Standard for Local and metropolitan area networks, 14 August 2008, IEEE

[12] The Institute of Electrical and Electronics Engineers, "802.1Qay: Virtual Bridged Local Area Networks - Amendment 10: Provider Backbone Bridge Traffic Engineering" IEEE Standard for Local and metropolitan area networks, 5 August 2009, IEEE

[13] The Institute of Electrical and Electronics Engineers, “802.1ag: Virtual Bridged Local Area Networks - Amendment 5: Connectivity Fault Management” IEEE Standard for Local and metropolitan area networks, 17 December 2007, IEEE

A Reliability and Survivability Analysis of Local Telecommunication Switches Suffering Frequent Outages

Andrew P. Snow[1]
School of Information & Telecommunication Systems[1]
Ohio University
Athens, Ohio, USA
e-mail: asnow@ohio.edu

Julio Arauz[1], Gary Weckman[2], Aimee Shyirambere[1]
Department of Industrial & Systems Engineering[2]
Ohio University
Athens, Ohio, USA
e-mail: arauz@ohio.edu, weckmang@ohio.edu

Abstract— **This paper presents a reliability analysis of local telecommunication switches experiencing frequent outages in the United States, based upon empirical data. Almost 13,000 switch outages are examined and over 2,500 are found to originate with just 156 switches experiencing eight or more outages each over a 14-year period. Telecommunication switch outage statistics are analyzed for this multiyear period, allowing examination into switch failure frequency, causes, trends, and impacts. Failure categories are created by reported outage cause codes, including human error, design error, hardware failure, and external factor causality categories. Principal findings are that there are significant differences in the switch and outage characteristics for switches experiencing more frequent outages/failures. Additionally, time series analysis indicates significant reliability/survivability deterioration in switches experiencing more frequent outages.**

Keywords- telecommunication; reliability; local switches; mobile switching centers; public switched telephone network; wireless systems.

I. INTRODUCTION

Historically, the Public Switched Telephone Network (PSTN) in the U.S. has been used predominantly for landline voice services. However, with the exponential growth of mobile voice services, the PSTN has been integrated with wireless systems. In fact, for calls outside of regional areas, wireless serves as radio interface technology, taking place of local loop connectivity. A call over hundreds or thousands of miles travels a very small percentage of the total distance over wireless infrastructure, as the wireless system connects to the PSTN for long haul transport. Both the PSTN and wireless systems use circuit switches manufactured by the same equipment suppliers. In fact, the switches are very similar. As such, we expect local PSTN telecommunication systems to be very reliable and survivable, as they are the access nodes to transport services in voice networks, for both landline and wireless calls. As there are many thousands of these switches in the PSTN, monitoring and improving local switch reliability is of great importance. Also, wireline switch outage characteristics serve as a good proxy for wireless mobile switching centers.

Continuous improvement of any communications device, such as local telecommunication switches, requires documenting today's performance, and measuring against that baseline. It is important to know reliability trends, not merely to predict, but to influence the future in a proactive way. The key to managing highly reliable systems is the recognition of an important precept – a reliability trend does not have to be accepted and actions may in fact be taken to alter the trend. However, all failures cannot be prevented, as products are put into environments with hazards of all types. But understanding failure modes and how to avoid certain failures is important. Additionally, management must endeavor to decrease the chance of human induced errors throughout he lifecycle. This can be done by training, tools, and other support, but management must make reliability a priority and fund reliability programs, effectively managing reliability engineering [1]. In order to change a trend, we look for approaches that will offer insights into why failures are occurring. Telecommunication switch reliability is determined by the complex interaction between software, hardware, operators, traffic load, and a variety of environmental factors. By knowing failure causes, designers (switch vendors) and operators of telecommunication switches (service providers) may take corrective action to alter future trends. Likewise, Barnard argues that modern reliability engineering must embrace the principal of continually improving products throughout the lifecycle, to include FRACAS (Failure Reporting, Analysis and Corrective Action System) before and during the operational phase of products [2].

There has been a paucity of published empirical research concerning the reliability of operational telecommunication switches in the US. Snow investigated local switch outage from 1992 through 1995, and documented reliability growth [3]. Later, Snow also noticed that some switches failed many times, while others failed infrequently [4]. This paper extends that early research over the years 1996 through 2009.

II. RESEARCH QUESTIONS

This Research will address the following questions regarding telecommunication local switch reliability and survivability over a 14 year period:

1. Are the characteristics of switches failing more frequently different from those that are not?
   a. Switch size (lines)
   b. Rural or Urban location
2. Are the event characteristics for switches failing more frequently different from those that are not?
   a. Causes
   b. Duration of outages and impact
   d. Time of day (TOD), Day of Week (DOW), and Month of Year (MOY)

ISBN: 978-1-61208-245-5

3. Are the failure trends for switches experiencing failures more frequently different from those that are not? Has there been reliability/survivability:
   a. Growth,
   b. Constancy, or
   c. Deterioration?

Reliability is the probability a system will perform its intended function, in the intended environment and at a particular level of performance. Thresholds are very commonly used to declare a system as in either an "operational" or "degraded" mode [5]. Others define reliability as "conformance to specifications over time"[6].

If the system is in a degraded mode, there is a failure event. Survivability is "The capability of a system to fulfill its mission, in a timely manner, in the presence of attacks, failures, or accidents." and is also a resiliency characteristic [7]. Outage frequency and impact, resulting from failures and accidents, are survivability measures. Therefore, for this paper we will principally analyze failure events to assess reliability, and scheduled/unscheduled outages to assess survivability. The data consists of local switches experiencing outages 2 minutes or more in duration. If a switch experiences a failure, it results in an outage, which has an impact until the failure is mitigated and service restored.

III. CONTEXT AND IMPORTANCE

The PSTN is a complex, distributed system, and its functions are executed by the close cooperation between switching, signaling and transmission entities. These entities cooperate in order to provide circuit switching, or the establishment, maintenance and termination of temporary end-to-end connections between subscribers through a network, as shown in Figure 1. The switching entities are responsible for concatenating individual transmission links into an end-to-end circuit, while the signaling entities coordinate the establishment, maintenance and termination of the end-to-end circuit. Lastly, transmission entities provide links between switches. Local switches are defined as those having local loop access lines, including standalone, host, or remote local switches. Tandem switches that also have access lines, or access tandems, are also included in this study, but represent a small number of the total population. Tandem switch outages are beyond the scope of this research.

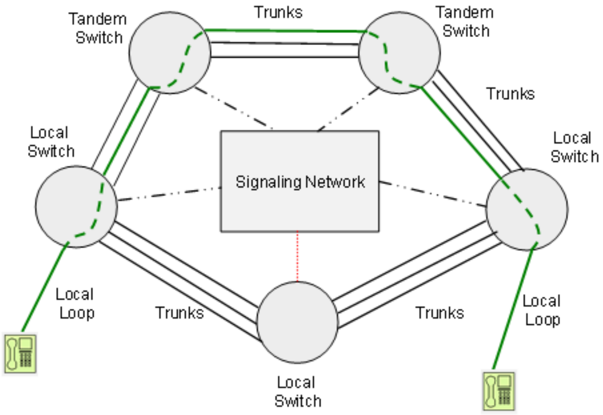


Figure 1. PSTN Infrastructure

Wireless voice infrastructure also includes switches, as seen in Figure 2. These switches, called Mobile Switching Centers (MSC) or Mobile Switching Telephone Office (MTSO) , are very similar to the wire line switches studied in this research:

> "The mobile telephone office (MTSO) is the switch that serves a cellular system. It is similar in function to a class 5 end office switch….Prominent makers of MTSOs include Northern Telcom, Ericsson, Motorola, DSC, and Lucent Technologies", the same manufacturers of local telecommunication switches."[8]

MSCs switch mobile calls in the wireless coverage area, and also interface to the PSTN if the call goes outside the wireless area being served.

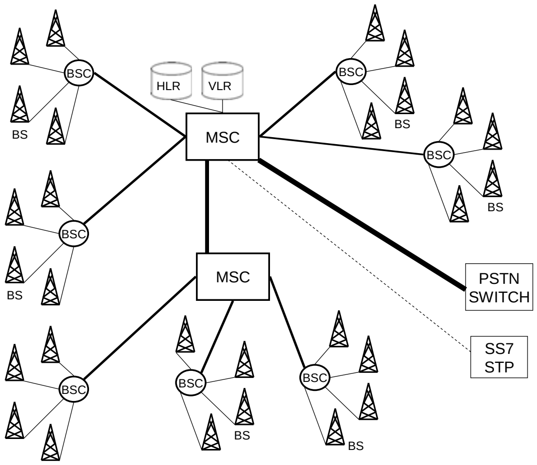


Figure 2. Wireless Infrastructure

In this work, the PSTN is viewed as a single system, made up of switching, signaling and transmission segments. The switching segment is made up of the tandem and local switch subsystems. The purpose of this paper is to investigate the reliability of the local exchange switching subsystem as a whole, by investigating the pooled failures of all individual local switches in the PSTN. There are a large number of different manufacturers and models of local switches in this infrastructure. Even the same model switch varies substantially from serial to serial because of differences in customers served and features offered. By pooling failures from different switches, we may assess the reliability of local switching as a whole, rather than the reliability of a single switch.

IV. EMPIRICAL DATA

Individual switch outage incidents of at least two minutes in duration have been reported to the Federal Communications Commission (FCC) by all price cap regulated local exchange carriers, accounting for over 90% of the wire line telephone access lines in the U.S. This data is part of the quality-of-service statistics required by the FCC's Automated Reporting and Management Information System (ARMIS) reports made to the FCC by the carriers. For each reported switch incident the date, time, duration, and outage cause are included, along with the number of access lines connected to the switch experiencing the outage. Very importantly, the reporting Carrier classifies each incident using one of fifteen different cause codes. This research presents a comprehensive reliability analysis of LEC local telecommunication switches in the United States using this public data over an extended period, January 1996 through December 2009 [9].

Local switches are repairable systems in that they are repaired by means other than replacing the entire unit. Renewal processes are often used for modeling such phenonoma, as it is hoped the system is made "good as new" through modular replacement. However, switches involve software, wherein some repairs result in a slightly different switch as the software changes. This means failures are not independent or identically distributed. This is the case of the non-homogeneous Poisson process (NHPP), the most common model used for repairable systems. For such systems the failure rate changes over time, and is nonstationary [10].The two-minute reporting threshold is recognized as a reliability threshold in this study. An outage is different from a failure event, as the outage also has a duration (how long the switch failed) and a size (how many subscriber lines connect to the switch). An impact metric, called "Lost Line Hours" or LLH is a survivability metric, and is used here to assess outage and reliability deficit impact. If a 10,000 line switch experiences a 2 hour outage, that is equivalent to a 20,000 line switch down for 1 hour, or 20,000 LLH. The results presented below make an important distinction between an outage and a failure event. Lastly, availability is another important aspect of switch quality-of-service too, but not in the scope of this study.

V. SUMMARY ANALYSIS

As mentioned, the reporting carrier attributes an individual switch outage incident to one of fifteen different cause codes, as required by the ARMIS reporting instructions. It is important to note that only total switch outages are reported. A partially failed switch is not a reportable outage, irrespective of the size of the partial switch outage. Neither are outages less than two minutes. An abbreviated definition for each cause code and the number of outages reported for each category is shown in Table 1. From here on, a distinction is made between a failure and an outage. Cause code one is recognized as a planned maintenance outage, while cause codes two through fifteen are treated as failures resulting in outages. From Table 1, note that the largest cause of outages was scheduled outages (about 30%) while the next largest was random hardware failure (about 23%), followed by roughly equal percentages of 8% for software design and acts of god.

A. Causal Analysis

Another way to summarize the failure data is to combine some of the codes into categories that might offer more insight into the reliability performance of local switches. The following categories are created by combining cause codes:

- Human error: Procedural errors made in installation, maintenance or other activities by Telco employees, contractors, switch vendors, or other vendors.
- Design error: Software or hardware design errors made by the switch vendor prior to installation.
- Hardware error: A random hardware failure, which causes the switch to fail.
- External circumstances: An event not directly associated with the switch, which causes it to fail or be isolated from the PSTN.
- Other/unknown: A failure for which the cause was not ascertained by the carrier.

These categories, their composition, and the distribution of failures to each category are shown in Table 2, where scheduled outages are left out. Note that the largest categories causing failures in about equal proportions are hardware failure and external causes. The next largest is procedural error and design error, each with about half the failures as either hardware failures or external causes.

B. Time Series Analysis of All Switch Outages

A time series analysis of outages is shown in Figure 3. From this figure there appears to be a period of reliability growth followed by a period of reliability deterioration. However, during this study period, the number of local exchange switches decreased somewhat, as shown in Figure 4. From these results a time series of outage rate can be determined, as seen in Figure 5 (dividing the outage count per year by the number of switches per year). Here it is seen that the initial reliability growth is not as pronounced, and that the reliability deterioration is slightly more pronounced than that indicated by Figure 3.

TABLE I. LOCAL SWITCH OUTAGE AND OUTAGE CAUSE DISTRIBUTION

Code	Description	Number	%
1	Scheduled	3,885	30.2%
2	Procedural error (Telco install./maintenance)	446	3.5%
3	Procedural error (Telco non-install./maintain.)	376	2.9%
4	Procedural error (System vendor procedural error)	315	2.4%
5	Procedural error (Other vendor procedural error)	257	2.0%
6	Software design	1,078	8.4%
7	Hardware design	136	1.1%
8	Hardware failure	2,951	22.9%
9	Acts of god	935	7.3%
10	Traffic Overload	17	0.1%
11	Environmental	83	0.6%
12	External power failure	896	7.0%
13	Massive line outage, cable cut, other	660	5.1%
14	Remote - loss of facilities between host/remote	309	2.4%
15	Other/unknown	516	4.0%
	Total	12,860	100%

TABLE II. LOCAL SWITCH FAILURE CAUSE CATEGORY DISTRIBUTION

Codes	Failure Category	Numb.	%
2,3,4,5	Human Proc. Error	1,394	15.5%
6,7	Design Error	1,214	13.5%
8	Hardware Failure	2,951	32.9%
9 thru 14	External Circumstances	2,900	32.3%
15	Other/unknown	516	5.7%
2 thru 15	Total	8,975	100%

VI. SWITCHES WITH MORE FREQUENT OUTAGES/FAILURES

Do some switches experience more outages than others? The logarithmic plot in Figure 6 indicates this is in fact the case. Here we see that 156 unique switches experienced 8 or more outages during the study period, while 5,976 switches experienced 7 or less outages. The selection of 8 or more outages is somewhat arbitrary, but partly selected because the data points at 8 or more outages deviate from the smooth curve formed by the data points for 7 or less outages per switch.

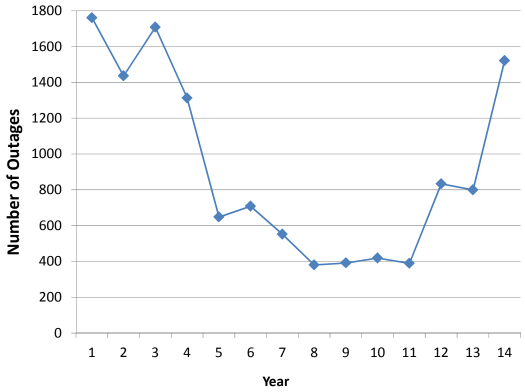


Figure 3. Time Series of Switch Outages Over the Study Period

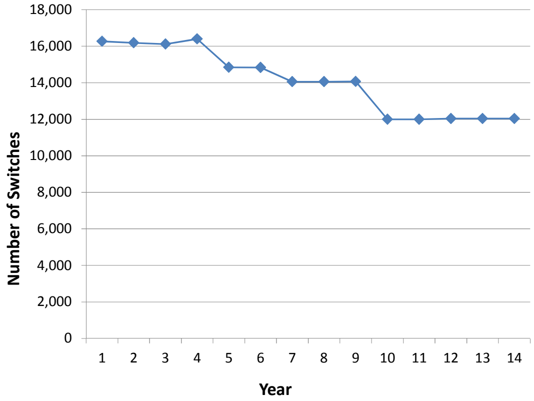


Figure 4. U.S. Local Switches Over the Study Period

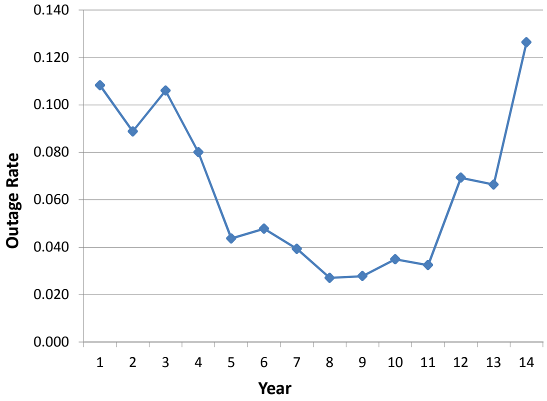
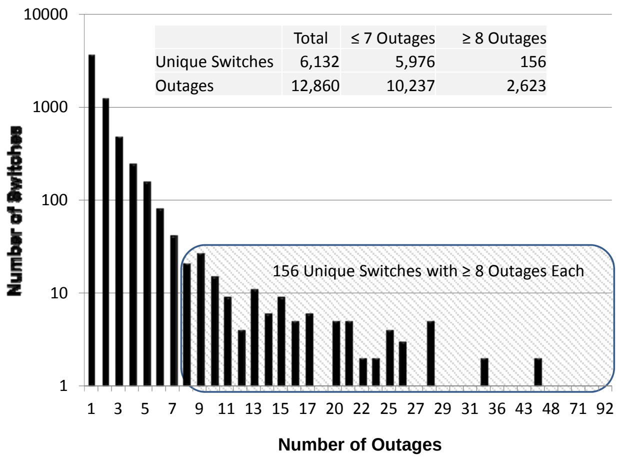


Figure 5. Switch Outage Rate Over the Study Period



	Total	≤ 7 Outages	≥ 8 Outages
Unique Switches	6,132	5,976	156
Outages	12,860	10,237	2,623

Figure 6. Number of Unique Switches One or More Outages

A. Causal Comparison of Switches with More Frequent Outages/Failures

The percentages of outages by cause code and category are shown in Tables 3 and 4, respectively. For a cause code comparison, note the following regarding switches with more failures compared to switches with less failures:

- One-half the percent of scheduled outages
- Double the percent for
  1. Acts of god,
  2. Line outage,
  3. Loss of connection to host switch, and
  4. Other/unknown

For major cause categories, note the following regarding switches with more failures compared to switches with less failures:

- One-third the human error
- Double external circumstances

B. Summary Analysis of Switches with Frequent Outages/Failures

A summary comparison of switches is shown in Table 5. First note that although less frequently failing switches account for but 3% of unique failed switches, they represent 25% of the switch failures. Also note that the less frequent failing total switch lines represent 10% of the total, they represent 22% of the total duration. On a better note, the more frequently failing switches represent only 7% of the survivability deficit due to the outages induced by these failures (lost line hours, or LLH). Note that the more frequently failed switches are about one-third the size of the less frequently failing switches. However, note little differences in the average duration of outages (4.61 versus 4.18 hours). Also note that the average LLH for the more frequently failed switches is about one-fourth of the less frequently failed switches. Also note that the median location for all failed switches is rural rather than urban (MSA stands for Metropolitan Statistical Area).

Lastly, refer to Tables 5 and 6 for the following temporal comparisons:

- Time of Day (TOD), where the day was divided into 6 timeslots
- Day of Week (DOW), where 1 is Monday
- Month of Year (MOY). Where 1 is January

Note that although the differences in average TOD and DOW week shown in Table 5 are small, Table 6 results indicates statistically significant differences. No differences are indicated for month of year.

TABLE III. COMPARISON OF CAUSE CODE DISTRIBUTION

Cause Code	≥ 8 Outages	≤ 7 Outages
1	15.7%	33.9%
2	1.9%	3.9%
3	0.5%	3.6%
4	0.5%	2.9%
5	1.1%	2.2%
6	10.8%	7.8%
7	0.7%	1.1%
8	27.4%	21.8%
9	14.0%	5.5%
10	0.2%	0.1%
11	0.5%	0.7%
12	8.5%	6.6%
13	8.2%	4.3%
14	4.0%	2.0%
15	5.9%	3.5%
Total	100.0%	100.0%

TABLE IV. COMPARISON OF CAUSE CATEGORY DISTRIBUTION

Cause Category	≥ 8 Outages	≤ 7 Outages
Scheduled (1)	15.7%	33.9%
Human Proced. Error (2-5)	4.0%	12.6%
Design Error (6-7)	11.6%	8.9%
Hardware (8)	27.4%	21.8%
External Circumst. (9-14)	35.4%	19.3%
Other/Unknown (15)	5.9%	3.5%
Total	100.0%	100.0%

TABLE V. RELIABILITY AND SURVIVABILITY COMPARISON

Cause Codes 2-15	Outages	≥8 Outages	≤7 Outages	% ≥8	% ≤7
Number Outages	8,975	2,210	6,765	25%	75%
No. of Switches	4,517	154	4,363	3%	97%
Total Lines	65.6 M	6.6 M	59.0 M	10%	90%
Total Dur. (Hrs)	41.3 K	9.2 K	32.1 K	22%	78%
Total LLH	307.8 M	20.5 M	287.3 M	7%	93%
Avg Sw. Lines	7,313	3,000	8,723		
Avg Dur. (Hours)	4.61	4.18	4.75		
Average LLH	34,295	9,257	42,475		
Median TOD	10:57 AM	12:00 PM	10:34 AM		
Mean TOD	11:02 AM	11:54 AM	10:45 AM		
Median DOW	4.09	4.29	4.06		
Mean DOW	4.22	4.32	4.19		
Median MOY	6.89	7.08	6.82		
Mean MOY	6.87	6.90	6.86		
Median MSA	Rural	Rural	Rural		

VII. TIME SERIES ANALYSIS OF SWITCHES WITH FREQUENT OUTAGES

Here the outage data is investigated for trends and arrival process assessment. The perspective is that the PSTN is viewed as a single repairable system, and that we are investigating the local switch subsystem as a whole. The first method in assessing a trend is visual, using the cumulative failures versus time plot. A linear plot means constant arrival process. This is a homogeneous-Poisson-process (HPP) if the time-to-failures are i.i.d. and exponentially distributed. If the events are i.i.d. and the distribution is other than exponential, then we may classify the process as renewal [10]. However, if the cumulative plot bends downward or upward, the failures are not from a common distribution, and we have either reliability growth or reliability deterioration, respectively. In this instance, the most common classification is expected to be the nonhomogeneous Poisson process (NHPP), where subsequent failures come from a different distribution [10]. This should be expected, as switches commonly receive new software versions and feature upgrades.

The Cox-Lewis trend test (Laplace test) can be used to tease out whether subtle upward or downward bending of cumulative outage/failure plots are statistically significant cases of reliability deterioration or growth, respectively. The Laplace test looks for trends where the homogeneous Poisson process (HPP) is the null hypotheses. The resulting test statistic rapidly converges to a normal score with very few data points [10]. However, the sample results presented here are visually convincing, with no need for formal trend testing to detect periods of reliability growth, constancy, and deterioration.

TABLE VI. TEMPORAL COMPARISON FOR OUTAGES AND FAILURES

T-TEST (Results Summary)	≥ 8 Outages	≤ 7 Outages	Result
TOD (All Cause Codes)	11:41:23 AM	10:43:48 AM	Difference
TOD (Cause Code 1)	10:32:08 AM	10:39:36 AM	No Difference
TOD (Cause Codes 2-15)	11:54:19 AM	10:45:57 AM	Difference
DOW (All Cause Codes)	4.37	4.25	Difference
DOW (Cause Code 1)	4.68	4.36	Difference
DOW (Cause Codes 2-15)	4.32	4.19	Difference
MOY (All Cause Codes)	6.90	6.93	No Difference
MOY (Cause Code 1)	6.89	7.06	No Difference
MOY (Cause Codes 2-15)	6.90	6.86	No Difference

Reliability growth, constancy, and deterioration can be examined through cumulative outage plots. A cumulative plot of all outages reported during the study period is seen on Figure 7. Note three general regions of the curve :

- Region I: Reliability constancy – outage rate constant for years 1 to 4
- Region II: Slight reliability growth for years 4 to 11
- Region III: Slight reliability deterioration for years 11 to 14

Region I constant outage rate is about 1500 outages per year, indicating process stability. For Region II, the outage rate is demonstrably lower than Region I, and slowly decreases nonlinearly over years 4 to 12. The average outage rate in Region II is about 540 per year. Reliability growth is indicated from Region I to II. However, the reliability starts to deteriorate again in Region III, albeit not as bad as Region I reliability, with an average rate of about 1000 per year. Overall however, the Laplace trend test indicates very strong evidence of reliability growth over the entire 14 year period.

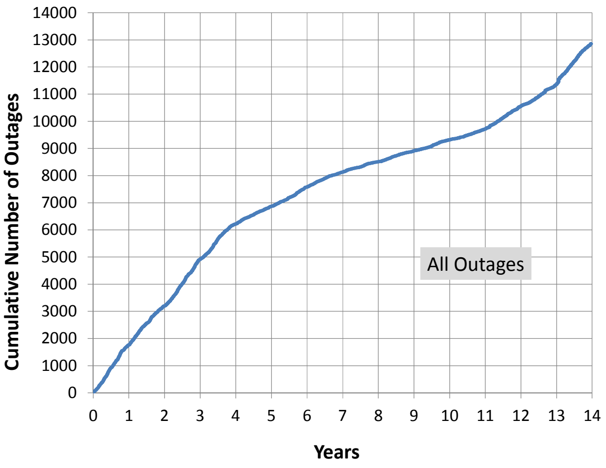


Figure 7. Cumulative Outage Plot: All Outages

To compare switches with more frequent outages to those with less, refer to Figures 8 and 9. The switches with less outages exhibit a very strong reliability growth, with improvement starting about year 4. If we linearize into two regions, we see very significant reliability growth about 1400 outages per year (years 1 to 4 years) to about 460 per year (years 4 to 14). There is a slight tailing up of outages in year 14. However, for the switches with more frequent outages, from Figure 9 we see that there are three well defined regions of constancy, reliability growth, followed by a very strong region of reliability deterioration.

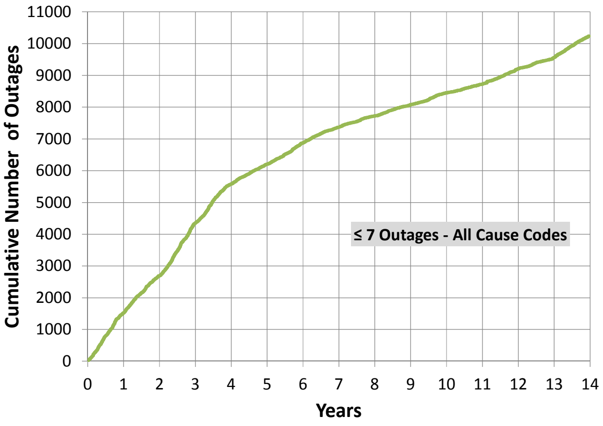


Figure 8. Cumulative Outage Plot: All Outages
(Seven or Less per Switch)

Further insights into the switches with more frequent outages are shown in Figures 10 and 11. First, note very significant decreases in scheduled outages (Figure 10), and a very rapid increase in outages due to unplanned failures during the last three years of the study period (Figure 11). In Figure 11 the failure rate the first 11 years is about 55 per year, while for the last three years it is about 530. This is a tenfold increase and represents very severe reliability deterioration.

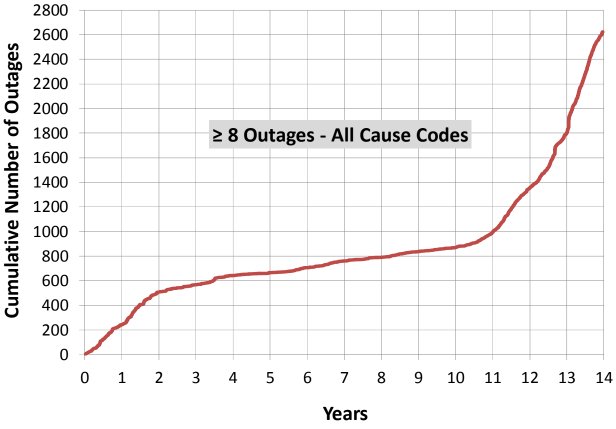


Figure 9. Cumulative Outage Plot: Outages
(Eight or More per Switch)

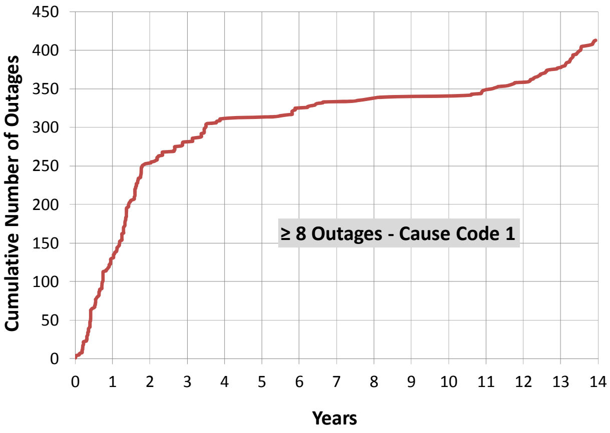


Figure 10. Cumulative Outage Plot: Scheduled Outages
(Eight or More per Switch)

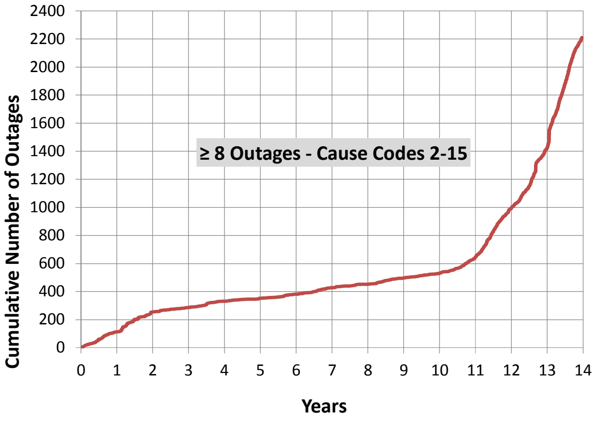


Figure 11. Cumulative Outage Plot:
Outages Due to Failures
(Eight or More per Switch)

VIII. SUMMARY OF FINDINGS AND CONCLUSIONS

Over the study period there was significant reliability growth for all local switches. However, the outage rate (number of switches with outages divided by number of switches) accelerated over the last four years of the study period.

A summary of the more interesting and significant findings for switches that experience more frequent outages are:

- There is very severe reliability deterioration in switches that go down more frequently while there is good reliability growth for switches that fail less frequently.
- Switches that go down more frequently are decidedly smaller, more rural, and receive much less scheduled maintenance than those failing less frequently.
- Human error induced failures are much less of a problem for the more frequently out switches. Perhaps this is due to their rural nature, where there are less visits by technicians. Or perhaps the rural switches are hosted to larger switches more often, and require less visits by technicians.
- Scheduled outages occur less frequently for the more frequently out switches, perhaps indicating less frequency preventive maintenance. On the other hand, this could be an artifact of smaller switches hosted to large switches.
- Acts of god, massive line outage, loss of connection to host switch, and other/unknown causes are much more of a problem for switches with more frequent outages. This could be suggestive of weaknesses in host-remote switch architectures and/or susceptibility of physical plant to natural disasters.
- There are significant differences in the times-of-day and days-of-week between the more and less frequently out switches. This suggests a different maintenance/disaster-recovery approaches for large vs. small and/or rural vs. urban switches.
- The more frequently out switches represent 2.5% of the switches with outages, but account for 7% of the lost line hours. This means that frequently out switches are about three times less survivable than switches that are out less frequently.

In conclusion, there are demonstrable differences in (1) the causality of outages and (2) the characteristics of switches suffering outages, and (3) switch resiliency when it comes to switches out more and less frequently. Also, there is a slight uptrend in outages in the last several years of the 14 year study period. Unfortunately, the FCC stopped collecting this data from carriers in 2009, masking the trends since then and in the future. This research demonstrates that very pronounced reliability and survivability trends are identifiable, some of which are troublesome. This is a good example of retrospective quantitative research analysis, yielding important trends that can be investigated further and corrective action taken.

REFERENCES

[1] O'Conner, P.D.T., *Practical Reliability Engineering*, 4th edition, John Wiley & Sons, England, 2001.

[2] R.W.A. Barnard, "Reliability Engineering : Futility and Error", Second Annual Chapter Conference, South African Chapter, International Council on Systems Engineering (INCOSE), 31 August, 1 September 2004.

[3] Snow, Andrew P., "The Reliability of Telecommunication Switches, Six International Conference on Telecommunications Systems: Modeling and Analysis (March 1997): 288-295.

[4] Snow, Andrew P., "Internet Implications of Telephone Access", IEEE Computer, Volume 32, Number 9, (September 1999): 108-110.

[5] Leemis, L., Reliability: Probabilistic Models and Statistical Methods, Prentice-Hall, Englewood Cliffs, NJ (1995).

[6] Levin, M.A. and Kalal, T.T., *Improving Product Reliability, Strategies and Implementation*. John Wiley & Sons, England, 2003.

[7] R. J. Ellison, D. A. Fisher, R. C. Linger, H. F. Lipson, T. Longstaff, N. R. Mead, Survivable Network Systems: An Emerging Discipline, Carnegie-Mellon Software Engineering Institute Technical Report CMU/SEI-97-TR-013, 1997 revised 1999.

[8] Beddel, Paul, *Cellular/PCS Management*, McGraw-Hill, ISBN 0071346457, 1999.

[9] FCC Report 43-05, ARMIS Service Quality Report Table Iva, downloaded from http://transition.fcc.gov/wcb/armis/ September 2012.

[10] Louit, D.M., Pascual, R., and Jardine, A.K.S, "A procatical procedure for the selection of time-to-failure models based on the assessment of trends in maintenance data", *Reliability Engineering and System Safety 94* (2009) 1618-1628.

ISBN: 978-1-61208-245-5